Some Impliciations of Constraints in Phasefield Models

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation
von

Dipl.-Ing. Andreas Reiter

Tag der mündlichen Prüfung:	04.12.2020
Erstgutachterin:	Prof. Dr. rer. nat. Britta Nestler
Zweitgutachterin:	Prof. Dr. Katharina Schratz

Kurzfassung

In dieser Arbeit werden verschiedene Aspekte, die sich durch Zwangsbedingungen in der Phasenfeldmodellierung ergeben, untersucht.
Zum einen wird, im Rahmen eines reinen Phasenfeldmodells, der Einfluss des häufig verwendeten Hindernispotentials in Bezug auf die Diskretisierung und algorithmische Gesichtspunkte der Verwendung von Projektions-basierten Algorithmen in nicht-gewichteten und gewichteten Mobilitätsformulierungen betrachtet.
Zum anderen werden "Grandchem"-artige Modelle in einem chemischen, mechanischen und chemomechanischem Kontext diskutiert, in denen eine gegebene phasenunabhängige Größe innerhalb von Mehrphasenbereichen als gewichtetes Mittel der entsprechenden Größen innerhalb der Einzelphasen aufgefasst wird. Die so eingeführten zusätzlichen Freiheitsgrade ermöglichen durch eine geschickte Festlegung der phasenspezifischen Werte in Abhängigkeit der restlichen Parameter eine verbesserte Modellbildung, durch welche sich der Einfluss der Breite der Übergangsbereiche auf die Ergebnisse deutlich reduzieren lässt. In vielen Fällen lässt sich die meistens direkt physikalisch motivierte Festlegung der phasenspezifischen Größen zugleich als die Lösung eines parametrisierten Minimierungs- oder Maximierungsproblems unter der Nebenbedingung des vorgegebenen Mittelwerts interpretieren. Hier wird untersucht, welche Konsequenzen sich aus dieser Interpretation ergeben und weshalb das Zusammenspiel dieses lokalen Extremalproblems mit dem globalen variationellen Ansatz des Phasenfeldmodells von entscheidender Bedeutung ist.

Abstract

In this work, several aspects resulting from constraints in phasefield models are investigated.
On the one hand, within the framework of a pure phasefield model, the influence of the commonly used obstacle potential is considered, both with respect to the discretization and to some algorithmic implications for projection-based algorithms using nonweighted and weighted mobility formulations.
On the other hand, "grandchem"-type models, which are based upon interpreting a given phaseindependent quantity within the multiphase regions as a weighted average of the corresponding quantity within the individual phases, are disussed within a chemical, mechanical and chemomechanical context. The additional degrees of freedom introduced by these phasespecific quantities lead, if fixed in an clever manner in terms of the remaining parameters, to improved models with a significantly reduced dependence upon the width of the diffuse interface. In many cases, this typically physically motivated specification of the phasespecific quantities can also be interpreted as the solution of a paramterized minimization or maximization problem under the constraint of maintaining the correct average. Here, the investigation focuses on consequences of this interpretation and the importance of the interplay of this local extremum problem with the global variational Ansatz of the phasefield model.

Contents

Part I

Introduction and Outline

Chapter 1

Introduction

Over the course of the last decades, the phasefield method has evolved from a popular model in material science problems, in particular for solidification problems in which surface tensions play a major role, to a more widely used computational method for a large class of multi-physics interface and free discontinuity problems. Its primary advantage in problems involving moving interfaces is that, whereas sharp interface approaches typically require some relatively expensive interface tracking, the phasefield method is in many cases capable of reproducing the essential physical effects associated with the interfaces without requiring an explicit identification of their location. This is primarily achieved by approximating the sharp interface with a diffuse transition region with a small but finite width. Within this region, the underlying physics are then approximated through appropriate interpolation procedures, which, if done properly, leads to an approximating the solution of the original sharp interface problem. As the interpolation procedures do not require explicitly identifying the precise location of any surfaces, the phasefield method is often referred to as an interface capturing rather than an interface tracking approach.

There is a price to be paid though, as the diffuse interface approximation introduces an additional lengthscale in the form of the width of the interfaces. In order to provide accurate approximations to the solution of sharp interface problems, it is usually necessary to choose an interface width which is significantly smaller than the other lengthscales of the problem. This in turn requires a discretization which is, at least locally, sufficiently fine to faithfully solve the underlying equations and can thus give rise to relatively large and potentially expensive problems. Whether or not the phasefield approach is competitive is thus clearly a question of both how well the physics are approximated in the diffuse setting and of the efficiency with which the numerical problems can be solved.

Two major developments in this context have been the increasingly widespread use of obstacle-based potentials and an improved modeling approach for the physics within the transition region. In contrast to well-based potentials, obstacle-based potentials in principle allow to significantly decrease the computational cost as they ensure that the computationally expensive interfaces are strictly contained within a small region of width $\mathcal{O}(\epsilon)$. Their only disadvantage is that this requires introducing inequality constraints on the phasefield values, which leads to some additional complications in both the equations and various questions related to their solution. Some primarily practical aspects related to this will be discussed in Chapter 6.

A somewhat similar observation can be made for the improved modeling approach first proposed in the chemically driven context by [42]. Its basic idea is that, considering the interface region as a mixture of different phases, the values of any further physically relevant quantity such as e.g. the concentration, temperature or mechanical strains, can also be considered as the average of the respective quantity over all phases present at this point, i.e. to introduce phase-specific versions of these quantities which need to average to the total ones. As there are of course many ways to redistribute a given quantity onto several phase-specific ones, their precise specification

introduces an additional degree of freedom into the model, which can be used to include more of the underlying physics. Even though this is an intuitively appealing idea, it is also clear that this introduces additional complexities into the model as there is now a "submodel" for deriving the phase-specific quantities from the remaining parameters of the problem. Besides a higher computational difficulty of such models, the additional unknowns further leads to some questions regarding the interpretation and consistency with respect to the standard variational approach underlying the phasefield method. These questions, in particular in relation with various models extending the one in [42] to a mechanical setting, will be the focus of Chapter 7.

Chapter 2

A Quick Sketch of the Phasefield Method

In its most basic form, the phasefield method provides a means of approximating the surface, or, more generally, a surface energy associated with subsets of $\boldsymbol{R}^n$. If Ω is such a (sufficiently smooth) subset, its surface area is given by $|\partial\Omega| = \int_{\partial\Omega} \mathrm{d}\boldsymbol{s}$ with $\mathrm{d}\boldsymbol{s}$ denoting the surface measure on $\partial\Omega$. In addition, if one assumes that to the surface one can associate a surface energy density γ, $[\gamma] = \frac{\mathrm{J}}{\mathrm{m}^2}$, this surface can, in the simplest case, be associated with a surface energy of $\mathcal{E}_s(\Omega) = \gamma|\partial\Omega|$. More generally, γ may depend on $\boldsymbol{x}$ and other parameters such as the temperature, leading to the expression

$$\mathcal{E}(\Omega) = \int_{\partial\Omega} \gamma(\boldsymbol{x}, ...)\, \mathrm{d}\boldsymbol{s}. \tag{2.1}$$

An alternative formulation for the surface energy (2.1) may be obtained by introducing the characteristic function

$$\chi_\Omega(\boldsymbol{x}) = \begin{cases} 1 & , \boldsymbol{x} \in \Omega, \\ 0 & , \text{else} \end{cases}$$

of Ω. Even though χ is discontinuous on $\partial\Omega$, one can define its distributional derivative $\mathrm{D}\chi_\Omega$, whose total variation can be identified with the surface measure on $\partial\Omega$. Formally, one thus has $|\partial\Omega| = \int_{\boldsymbol{R}^n} |\mathrm{D}\chi_\Omega|$ for the surface area, and, by again using the surface energy density γ,

$$\mathcal{E}(\Omega) = \int \gamma(\boldsymbol{x}, ...)|\mathrm{D}\chi_\Omega|. \tag{2.2}$$

While the rigorous form of this this approach in terms of functions of bounded variation turns out to be very fruitful from a theoretical point of view (see e.g. [2] and [6] for an introduction to this topic), it is clear that the use of Equation (2.2) in numerical approximations is somewhat difficult due to the lack of smoothness.
A first remedy would be to replace the discontinuous indicator function by a smooth approximation obtained e.g. by convolution with an appropriate smoothing kernel, $\chi^\epsilon_\Omega = \psi_\epsilon * \chi_\Omega$. As the approximation χ^ϵ_Ω is smooth, its distributional derivative reduces to its classical gradient, and thus (2.2) could be approximated by

$$\mathcal{E}_\epsilon(\Omega) = \int \gamma(\boldsymbol{x}, ...)|\nabla\chi^\epsilon_\Omega|\, \mathrm{d}\boldsymbol{x}. \tag{2.3}$$

While this provides a useful approximation of the surface energy **given** an appropriate subset Ω, it is still not particularly convenient for the purpose of actually finding subsets with minimal surfaces. On the one hand, it is not obvious how one can integrate the "constraint" of corresponding

to a regularized characteristic function into the problem of minimizing the approximated surface energy (2.3). On the other hand, the integrand in (2.3) is still not differentiable due to the kink in the mapping $\boldsymbol{\zeta} \mapsto |\boldsymbol{\zeta}|$.

The solution to this dilemma in the phasefield approach is also based on using an approximation ϕ to an indicator function, but by approximating the surface energy through a combination of two competing terms. The first contribution is given by the **gradient energy density** $\epsilon a(\nabla\phi)$ acting as a (smooth) penalization of gradients. The second contribution is by the **bulk potential** $\frac{1}{\epsilon}w(\phi)$ penalizing values different from 0 and 1. By a proper choice of a and w, the approximation to the minimal surface energy problem in the phasefield setting can then be put as that of minimizing the energy functional

$$\mathcal{E}_\epsilon(\phi) := \int_\Omega \epsilon a(\nabla\phi) + \frac{1}{\epsilon}w(\phi)\,\mathrm{d}\boldsymbol{x} \tag{2.4}$$

over an appropriate set of admissible functions.

It is clear that any minimizer of the w-term alone would take only the values 0 and 1, and that this term thus forces ϕ towards some indicator function. In contrast, the contribution from the a-term enforces some spatial regularity of ϕ, and has the purpose of "smoothing" the minimizers that would result from w alone, leading to the desired smooth approximation of χ_Ω. The different scalings in ϵ in Equation (2.4) provide a means of regulating the balance between the two terms, with decreasing values of ϵ increasingly favoring the w-contribution, therefore leading to a stronger similarity of the minimizers with actual characteristic functions (i.e. smaller transition regions between the two "phases"). As ϵ tends to zero, this approach can further be shown to converge to the desired sharp-interface limit in the appropriate sense[1].

As one now disposes of a formulation for (approximately) describing minimization problems involving surface energies, it is natural to extend the approach to more general problems involving additional competing energy contributions as these can easily be integrated by modifying the energy functional (2.4). A simple such example is obtained by considering melting/solidification processes involving a liquid and a solid phase where, depending on the temperature, it is energetically favorable for the substance to be either in its solid or liquid form. This leads, besides the surface energy associated with liquid-solid interfaces, to an additional energy contribution. By identifying for example ϕ with an approximation of the indicator function for the liquid region, this effect can be included into the approach above by considering the modified energy functional

$$\tilde{\mathcal{E}}_\epsilon(\phi, T) := \int_\Omega \epsilon a(\nabla\phi) + \frac{1}{\epsilon}w(\phi) + f(\phi, T)\,\mathrm{d}\boldsymbol{x}. \tag{2.5}$$

Here T is the temperature (for the moment assumed to be **fixed**) and f represents a suitable approximation of the energy associated with being liquid instead of solid for a given temperature, for example expressed in the form $\Delta f^{ls}(T)h(\phi)$, where $\Delta f^{ls}(T)$ is the energy required for changing from the solid to the liquid state at the given temperature and h is a suitable interpolation function.

It is clear that similar ideas can in principle be applied to a relatively large class of problems[2]. The primary use of the phasefield method within the material sciences is for the description

[1] This is usually done in terms of de Giorgi's Γ-convergence, see [50] and [45] for some early works in this direction within the phasefield setting and e.g. [15] or [47] for a general introduction to Γ-convergence. That this limit procedure requires some slightly more complex definition of convergence is of course to be expected, since the desired limit is not part of the original function space and the functional (more specifically the quadratic gradient energy density) is ill-defined on this limit.

[2] If any additional parameters such as T in the functionals are not fixed, the situation becomes somewhat more delicate, a point which will discussed in more detail in Section 3.2.

of dynamic processes involving phase-transformations and/or moving interfaces. In these, the process itself is typically considerered to play an important role, and one can not content oneself with a "steady-state" analysis based solely on any actual minimizers of some functionals such as in (2.4) or (2.5). While an appropriately chosen thermodynamic potential (e.g. the entropy or free energy) can serve as a guideline and motivation for the description, these potentials are inherently related to equilibrium thermodynamics, i.e. purely spatial. Therefore, by themselves, they do not provide any direct information on the dynamic behavior of the system.
Nevertheless, the standard approach within the phasefield setting is a variationally motivated one, i.e. one chooses a suitable thermodynamic potential and then postulates that the dynamics of the independent variables are proportional to the gradient of that functional. For example, when considering the pure surface minimization problem corresponding to (2.4) without any additional constraint, one postulates that the phasefield evolution is governed by

$$\begin{cases} \frac{\partial \phi}{\partial t} \sim \epsilon \nabla \cdot \frac{\partial a}{\partial \nabla \phi} - \frac{1}{\epsilon} w'(\phi) & \text{in } \Omega, \\ \frac{\partial \phi}{\partial \boldsymbol{n}} = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.6}$$

i.e. a steepest-descent-type flow. When considering problems where the physically problem is that of maximizing an entropy $\mathcal{S}(\phi)$ instead of minimizing an energy, one would similarly postulate a gradient-ascent-type flow based on $\mathcal{S}$.

This approach has a number of advantages. Firstly, for a reasonable choice of the proportionality, once any potential steady-state is reached, this state will also be a (local) minimizer of the chosen potential, and will therefore be compatible with the equilibrium formulation. Secondly, it still leaves a large flexibility in the actual description - and therefore the parameterization - of the dynamics, which can be adjusted to fit a large class of problems. Thirdly, the approach is conceptually easy, and, due to the presence of an underlying functional, allows for a relatively simple analysis of the resulting equations. Finally, again due to its close relation with minimization/maximation problems, there is a number of potentially attractive algorithms which, possibly after minor modifications, suggest themselves for the numerical solution at the discrete level.

Despite these advantages, there are also a number of good reasons for deviating from the purely variational setting.
The most obvious is that many physical effects can simply not be properly described based on equilibrium constructs such as a free energy or entropy functional alone. This includes in particular inertial effects, i.e. a resistance to change in time, and therefore a concept which is meaningless for "steady-state" functionals[3].
A different (and less obvious) reason within the phasefield approach is that adding "artificial" dynamic effects into the evolution equations can help to better match some experimental or sharp-interface results. A particularly popular example of such a modification is given by **anti-trapping** currents in the context of solidification problems.

After this short summary of the basic idea underlying the phasefield method, the following chapter will outline a number of selected applications in order to clarify some of the typical settings (and the unavoidable challenges) associated with them.

[3]Even though the dynamics of many of the classical physical equations can be related to other variational principles based on e.g. space-time integral over some action (see e.g. [44]), this need not be the case and in addition need not be well-suited for what one is trying to achieve. Others, such as e.g. the Navier-Stokes equations, do not derive from a classical variational principle.

Chapter 3

An Outline of the Topics Considered

3.1 Pure Phasefield Problems

Even though the ability to capture the effect of surface energies through a smooth description as in Equation (2.4) is theoretically interesting and does already have a number of useful applications, in particular in its multiphase generalization, this by itself is not the primary reason for the increasing popularity of the phasefield method. Instead, it is, as indicated above, the relative ease with which additional effects intervening in combination with surface energy (or entropy) effects can be included into the basic model.

In the simplest cases, this can be achieved by adding e.g. an additional energy contribution $f(\phi, ...)$ depending only, besides the phasefield values, on a number of **fixed** external parameters. One such example is given by the energy minimization given a fixed undercooling as outlined above in the discussion preceding Equation (2.5). Physically, this corresponds to the case when the solidification is slow enough such that one is able to maintain the material at a constant undercooling, e.g. when one assumes that the material is in contact with an outer thermal reservoir and the heat-conduction within the material happens at a much faster rate than the phase-transformation itself. In this case, an initial solid nucleus will grow, provided that the curvature of its surface is sufficiently low, such that the solidification of the material can liberate enough latent heat to counterbalance the associated increase in the surface energy[1]. There are a large number of other physical effects which can be included in much the same manner. These range from in principle relatively simple ones, as for example the influence of gravity for studying the deformations this induces in droplets on a surface, to significantly more complex ones such as the effect of (measured) stored elastic energy contributions on recrystallization processes [78].

Remark 1. Without this ability to include additional "driving forces", the phasefield method would essentially be reduced to a numerical tool for treating minimal surface problems. This can, even for two phases only, be of some practical interest when considering such problems within geometrically complex regions. There are also a number of more complex questions which can be adressed when one is able to model the interaction of a larger number of phases, as this allows studying e.g. energetically favorable arrangements of bubbles such as in the honeycomb conjecture. Nevertheless, it is clear that this type of problem only accounts for a fraction of the large number of questions being investigated by the phasefield method. ◇

A big advantage of this type of energy contributions through some interpolation of **fixed** bulk

[1] The increase in volume upon an outwards movement, and therefore the rate of energy transformed from a latent to a thermal form during the solidificatiton, is proportional to the surface area $\sim R^2$, whereas the growth in the surface and the associated energy is proportional to the curvature $\sim \frac{1}{R}$. Due to this difference in scalings, the volume increase dominates for a large radius whereas the surface energy dominates when R is small.

energy contributions is that the phasefield is the only actual unknown, and one does therefore not have to consider any of the additional complexities associated with coupled problems, some of which will be outlined in Sections 3.2.1 and 3.2.2. Despite the seeming simplicity, these problems already contain, as far as the phasefield itself is concerned, essentially all the "prototypical ingredients" and challenges also arising in the more complex coupled case.
This will be made use of in Chapter 6 for discussing some of the generic practical issues associated with the numerical treatment of phasefield problems, which are primarily due to the interplay of a and w alone and do not really depend on the precise structure of f. A particular focus there will be on some practical consequenes of the use of the obstacle potential for w and the associated constraint of lying within the Gibbs-simplex this imposes on the phasefield values.

This should certainly not be misunderstood in the sense that the choice of f is not important. In fact, one of the major challenges for phasefield models is that large driving forces as compared to the contributions by the basic phasefield term a and w can lead to significant deviations with respect to the expected sharp interface limits, and the majority of Chapter 7 is concerned with more accurate models for the additional energetic contributions to concentration- and/or elasticity-based effects.
This is essentially a question of how one can better model the physics of these additional fields in the presence of diffuse interfaces, and **not** on how the phasefield should react to them. From the point of view of the evolution equation for the phasefield, this is still a problem of precisely the same form, i.e. two contributions due to the bulk and gradient energy (or entropy) densities and some additional driving force - regardless of a potentially increased "internal" complexity - and the primary challenges from a numerical point of view are due to the surface energy contributions and not f.

Remark 2. A different but also very interesting class of problems is obtained when the energy functional (2.4) is not (or not exclusively) supplemented by an additional volumetric contribution but by an additional energy contribution associated with the boundary $\partial\Omega$. In this case, one can consider a part Γ_S of $\partial\Omega$, corresponding to a substrate S, on which wetting phenomena take place. These can be included into the model by adding a contribution of the form ([49] and [9])

$$\int_{\Gamma_S} \Delta\gamma_S h(\phi)\,\mathrm{d}\boldsymbol{s}, \tag{3.1}$$

where $\Delta\gamma_S$ corresponds to the difference of the surface energies between the two phases and the substrate, and h is a suitable interpolation function.
This, together with a gravity term, for example allows studying equilibrium shapes obtained through capillary rise problems within complex geometrical settings but will not be discussed in detail here. ◇

3.2 Coupled Problems

Assuming a fixed undercooling for the energy functional in Equation (2.5) represents a major simplification. Even though the functional does then depend on the temperature, this temperature is assumed to be given and uniform und is thus essentially just an additional parameter for controling the energy differences through the amount of undercooling without "actively" participating in the minimization problem. In many cases though, this simplifying assumption cannot be made. In this example, the solidification of the material will lead to a local increase in temperature. If the heat conduction takes place on a time-scale which is comparable with the phase transformation itself or if the material is well-isolated from its surroundings, this cannot (or at least not immediately) be counteracted by giving off heat to the environment, thus invalidating

the assumption of a constant temperature. In order to obtain a physically meaningful model, it therefore becomes necessary to change the role of the temperature to that of an additional unknown.

More generally, a phase transformation process may, besides the temperature, depend on a large number of additional parameters. For more complex solidification problems, it is for example necessary to include additional effects depending on the concentration, which, in most cases, leads to the introduction of an additional unknown. In other processes such as solid-solid transformations, even if performed sufficiently slow to allow for the approximation of a constant temperature, the phase transformation will induce additional stresses within the material, which in turn may significantly influence the transformation process itself (see e.g. [35], [66], [4] and [5] for some recent application-oriented works within the phasefield context). From a modeling point of view, it is again often possible to incorporate these additional effects by complementing the basic phasefield functional with additional mechanical contributions such as the strain energy and an appropriate set of equations for determining the displacement and strains.

3.2.1 Solidification Problems

The Basic Model

As the simplest extension of the situation considered in Section 2, one can consider a setting in which the solidification process is driven solely by an undercooling below the melting temperature of the material under consideration, but where the heat-conduction itself is not fast enough to allow for the approximation of a fixed (external) temperature. In order to capture the dynamics of the temperature field T in a physically meaningful manner, it is necessary to introduce an evolution equation based on the conservation of energy. In the simplest case, this energy can be expressed as a function on ϕ and T alone as $e = e(\phi, T)$, and one postulates

$$\frac{\partial e}{\partial t} = \nabla \cdot \boldsymbol{q}, \tag{3.2}$$

where the heat-flux $\boldsymbol{q}$ is usually specified in terms of the material parameters and the temperature.

Alternatively, and in a generalized setting, one may also consider solidification problems which, besides the temperature, also depend on the concentration of one or several chemical components. One such model, with the additional ability to describe an arbitrary number of phases, is proposed in [52] (see also [71]). In this model, the underlying functional is given in terms of the entropy functional

$$\mathcal{S}(\boldsymbol{\phi}, \boldsymbol{c}, e) = \int_\Omega s(\boldsymbol{\phi}, \boldsymbol{c}, e) - \left(\epsilon a(\boldsymbol{\phi}, \nabla \boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi}) \right) \mathrm{d}\boldsymbol{x} \tag{3.3}$$

where the vectorial functions $\boldsymbol{\phi} : \Omega \to \boldsymbol{R}^N$ and $\boldsymbol{c} : \Omega \to \boldsymbol{R}^K$ describe the different phases and components and a and w represent suitable entropy-based generalizations of the surface and bulk energy densities to the multiphase setting. In addition, $\boldsymbol{\phi}$ and $\boldsymbol{c}$ are locally subject to the constraint $\sum_{\alpha=1}^N \phi^\alpha = 1$ and $\sum_{i=1}^K c_i = 1$, and it is postulated that, based on the conservation of energy and mass and with an appropriate choice of the interaction matrix $\boldsymbol{L} = (L_{ij})_{0 \le i \le K, 0 \le j \le K}$,

the energy and concentration follow the evolution equations

$$\frac{\partial e}{\partial t} = -\nabla \cdot \left(L_{00} \nabla \Big(\frac{1}{T} \Big) + \sum_{j=1}^{K} L_{0j} \nabla \Big(-\frac{\mu_j}{T} \Big) \right),$$
$$\frac{\partial c_i}{\partial t} = -\nabla \cdot \left(L_{i0} \nabla \Big(\frac{1}{T} \Big) + \sum_{j=1}^{K} L_{ij} \nabla \Big(-\frac{\mu_j}{T} \Big) \right),$$

where the **chemical potential** $\boldsymbol{\mu}$ is given by $\boldsymbol{\mu} = \frac{\partial f}{\partial \mathbf{c}}$. In order to be consistent with the maximization of the entropy functional (3.3), the phasefield is subject to the evolution equation[2]

$$\tau \epsilon \frac{\partial \phi^\alpha}{\partial t} = \epsilon \Big(\nabla \cdot \frac{\partial a}{\partial \nabla \phi^\alpha} - \frac{\partial a}{\partial \phi^\alpha} \Big) - \frac{1}{\epsilon} \frac{\partial w}{\partial \phi^\alpha} - \frac{1}{T} \frac{\partial f}{\partial \phi^\alpha} - \Lambda,$$

where Λ is a suitable Lagrange multiplier for the sum-constraint $\sum_{\alpha=1}^{N} \phi^\alpha = 1$.

Quantitative Phasefield Models

While early phasefield models for purely temperature-driven problems have been quite successfull in matching experimental results, it has been observed that the results based on an influence of a concentration vector $\boldsymbol{c}$ - while qualitatively correct - are plagued by a number of artefacts if the interface width is, for numerical reasons, chosen much larger than the physical one ([41], [56] and [19]).

In particular for isothermal solidification problems, a more advanced modeling approach through so-called **quantitative phasefield models** has had, besides a simple increase in computational power, a very significant impact on the ability of phasefield models to more accurately describe the physics of phase transformations. The central idea, initiated by [42], is that the conditions determining the steady-state are based on an equilibration of the temperature and chemical potentials. Whereas models based on simple interpolation procedures of the phase-specific free energy contributions such as $f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \sum_{\alpha=1}^{N} f^\alpha(\boldsymbol{c}, T) h^\alpha(\boldsymbol{\phi})$ are, with respect to the temperature, therefore expressed in terms of a quantity which is "naturally" expected to be equal between different phases, the same is not true for the additional dependence on the common concentration and leads to excess energy contributions within the interface regions.

A way to significantly reduce these artefacts is, instead of using a single "averaged" concentration field, to introduce an additional internal degree of freedom into the model by "splitting" the given concentration into contributions from each of the coexistent phases, i.e. by introducing **phase-specific concentrations** $\boldsymbol{c}^\alpha$ for each phase α, which, weighted by a suitable interpolation function h^α, satisfy $\sum_{\alpha=1}^{N} \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c}$. This clearly increases the number of unknowns, and therefore requires introducing additional equations for their specification. As shown in [42], postulating that these phase-specifc concentrations $\boldsymbol{c}^\alpha$ are such that the associated chemical potentials corresponding to the f^α are equal allows for a siginificant improvement of the model. This model has further been analyzed and extended to the multiphase and multicomponent setting by various authors.

In particular, it was realized relatively early by Eiken et al. [25] that this model can also be elegantly interpreted as one where the free energy density is defined as the minimal averaged one $\sum_{\alpha=1}^{N} f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi})$ that can be achieved subject to the constraint of the $\boldsymbol{c}^\alpha$ averaging to the total one $\boldsymbol{c}$. Later works by e.g. [56] and [19] in contrast have primarily focused on a direct description in terms of the chemical potential itself and a very efficient but slightly misleading derivation based on the use of a grand chemical potential approach instead of a free-energy based one. In addition, they replace the evolution equation for $\boldsymbol{c}$ through an in principle equivalent one for $\boldsymbol{\mu}$, which is constructed such that the corresponding evolution of the average concentration

[2] See the discussion of Equation (3.5) below for an explanation of the term $\frac{1}{T} \frac{\partial f}{\partial \phi^\alpha}$.

as a function of the phasefield and chemical potential matches the standard one through Fick's law.
The focus in Section 7.1 will be on some primarily practical implications of this type of model. As will be seen there, even though the description in terms of $\boldsymbol{\mu}$ can be a very efficient one for some simple (but practically quite relevant) cases, it is unfortunately much less so in the general case, a point which will again be seen in Section 7.1. In particular, after outlining the model in the spirit of [25] and discussing its relation with the approach in [56] and [19], some practical aspects involved in its actual numerical evaluation - and potential pitfalls - will be discussed. Furthermore, it will be seen that the model generalizes in a straightforward manner to the non-isothermal setting based on a similar definition in terms of an entropy-functional of the form in Equation 3.3, and that these different definitions are consistently linked through the "standard" thermodynamic transforms.
Remark 3. Note that the coupled problems with an increased number of unknown fields and different ways to parameterize the problem give rise to a somewhat delicate question.

As pointed out above, the evolution equation for the phasefield is usually obtained by postulating a gradient-type flow with respect to some functional. This functional is written down in terms of a number of "natural" parameters such as e.g. the phasefield, concentrations, energy or temperature and elastic strains. In order to determine these additional parameters, the evolution equation for $\boldsymbol{\phi}$ needs to be supplemented with physically appropriate evolution laws for the relevant remaining unknowns (i.e. those which are considered as the independent variables) of the problem. Almost always, there will be interdependencies between the various parameters of the phasefield-functional and the unknowns, either directly by construction (the simpler case), or implicitly. In fact, as noted in e.g. [41] in the solidification context, the steady-state solutions are not truly independent. The question therefore arises how the approach using independent unknowns differs from a **reduced formulation**, where e.g., such as in [41], $\boldsymbol{c}$ is treated as a function of $\boldsymbol{\phi}$ (and possibly other parameters) in terms of the steady-state equation it satifies and whether or not the final result is actually related to a local maximizer or minimizer of the functional.

In the thermodynamic setting, this issue of interrelations between various functions, thermodynamic potentials and parameters is of course well known and in principle also well understood. In particular when adhering to the standard physicist notation of identifying functions and their values, one can easily forget about any possible - then hidden - dependencies. That this notation is nevertheless very useful and does not, at least in many of the more classical situations, cause any serious issues relies fundamentally on an underlying variational structure between the various thermodynamic potentials. This entails that, despite the ubiquitious changes between various interrelated variables, any hidden dependencies have a tendency of simply "dropping out".
As a simple illustration, consider the example of determining the derivative of an entropy density $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ such as in [52] with respect to $\boldsymbol{\phi}$. As working directly with the entropy and its (natural) dependence on e can be cumbersome, it is often advantageous to instead rely on a description in terms of the free energy $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$, in terms of which one has $\frac{\partial s}{\partial \boldsymbol{\phi}} = -\frac{1}{T}\frac{\partial f}{\partial \boldsymbol{\phi}}$.
Formally, this is easily obtained from a partial derivative (i.e. while keeping e constant) of the well-known equality $s = \frac{e-f}{T}$ with respect to $\boldsymbol{\phi}$. While this seems logical at first sight, at a closer look this calculation actually raises some questions. Firstly, if f is a function of $\boldsymbol{\phi}$, $\boldsymbol{c}$ and T, then defining s through $\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}$ would make the entropy a function of $\boldsymbol{\phi}$, $\boldsymbol{c}$, e **and** T. That this is of course not the case is due to the fact that T and e are not independent in this relation. Instead, given $(\boldsymbol{\phi}, \boldsymbol{c}, e)$, there is, under natural assumptions on f, a single $T > 0$ for which the values of $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ and $\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}$ coincide. Based on this, one could define T - now as a function of $(\boldsymbol{\phi}, \boldsymbol{c}, e)$ - such that this relation always holds, i.e. such that one has

$$s(\boldsymbol{\phi}, \boldsymbol{c}, e) = \frac{e - f\big(\boldsymbol{\phi}, \boldsymbol{c}, T(\boldsymbol{\phi}, \boldsymbol{c}, e)\big)}{T(\boldsymbol{\phi}, \boldsymbol{c}, e)}. \tag{3.4}$$

This returns the parameters of s to the original ones, but would seem to invalidate the simple formal differentiation above, as, again under natural assumptions, one a priori actually has

$$\frac{\partial s}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi},\boldsymbol{c},e) = -\frac{1}{T(\boldsymbol{\phi},\boldsymbol{c},e)}\frac{\partial f}{\partial \boldsymbol{\phi}}\big(\boldsymbol{\phi},\boldsymbol{c},T(\boldsymbol{\phi},\boldsymbol{c},e)\big) + \frac{\partial}{\partial T}\Big(\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}\Big)\big(\boldsymbol{\phi},\boldsymbol{c},T(\boldsymbol{\phi},\boldsymbol{c},e)\big)\frac{\partial T}{\partial \boldsymbol{\phi}}. \quad (3.5)$$

That the second term actually drops out is due to the fact that s and f are not arbitrary functions but are actually related to one another through a variational property, namely given $\boldsymbol{\phi}$, $\boldsymbol{c}$ and e, s and $T(\boldsymbol{\phi},\boldsymbol{c},e)$ satisfy[3]

$$s(\boldsymbol{\phi},\boldsymbol{c},e) = \inf_T\Big\{\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}\Big\} \quad \text{and} \quad T(\boldsymbol{\phi},\boldsymbol{c},e) = \operatorname{argmin}_\theta\Big\{\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},\theta)}{\theta}\Big\},$$

i.e. $T(\boldsymbol{\phi},\boldsymbol{c},e)$ is precisely the T minimizing $\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}$. It is then clear (at least formally) that T is characterized by the Euler-Lagrange equation $\frac{\partial}{\partial T}\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T} = 0$, and therefore the relation $\frac{\partial s}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi},\boldsymbol{c},e) = -\frac{1}{T}\frac{\partial f}{\partial \phi}\big(\boldsymbol{\phi},\boldsymbol{c},T\big)$ does indeed hold where T is the temperature corresponding to $(\boldsymbol{\phi},\boldsymbol{c},e)$.

This has of course been known in a similar form for more than a century (that the derivative is with respect to $\boldsymbol{\phi}$ instead of e.g. the more classical case of the concentration plays no real role here), but it highlights the fundamental importance that variational interrelations can play for treating potentially interdependent variables as independent[4].
The same basic principle also provides the answer to the two questions posed above: The formulations in terms of independent variables and in a reduced form are compatible with each other[5], and the final state will be a local maximizer/minimizer if (and essentially only then!) the steady-state equations allow for an appropriate variational interpretation in terms of the functional. Otherwise, while one can impose independent evolution or steady-state equations on some of the variables, one then either has to explicitly take their dependence on the other parameters into account - a potentially arduous and expensive task - or one looses the relation of the final solution with any constrained (local) maximizer or minimizer of the given functional. ◇

3.2.2 Solid-Solid Phase Transformation Problems

In recent years, often based on a similar reasoning, a number of quantitative phasefield models have also been proposed for more complex physical models involving - instead of (or in addition to) the concentration fields - the influence of additional elastic energy contributions due to mechanical interactions on phase transformation processes with several solid phases. These additional energy contributions may for example arise from either a prestress of the material, or stresses induced at interfaces due to different crystal structures and/or orientations between different phases and can have a significant influence on the phase-transformation process.
Whereas elastic effects within the bulk-phases are relatively well understood, their modeling within an - often artificially large - diffuse interface region in terms of the phasefield approach

[3] This is similar to the more standard Legendre-Fenchel transform and will be discussed in more detail in Section 5.

[4] In this context, it is important to stress that this is **not** a consequence of (3.4) by itself, which is simply a definition of T for two **given** functions s and f. If s and f are "well-behaved", one can e.g. invoke an implicit function theorem to obtain information on the derivatives of T based on those of s and f. However, Equation (3.4) does not a priori imply anything particular about $\frac{\partial}{\partial T}\Big(\frac{e-f(\boldsymbol{\phi},\boldsymbol{c},T)}{T}\Big)$ and could in principle be used for relatively arbitrary functions f and s, provided one can find a function $T(\boldsymbol{\phi},\boldsymbol{c},e)$ such that the equality holds. In particular, it is (obviously) not sufficient to write down an equation similar to (3.4) which "looks like" a Legendre-transformation and then to proclaim independence.

[5] In the sense that, while both approaches might lead to different dynamics and therefore potentially different local minimizers, each will accept the steady-state solution of the other.

has faced similar challenges as the models underlying solidification processes.

Remaining within the small-strain setting, one of the earlier models was based on the use of a single average strain-field $\boldsymbol{\epsilon}$ and a "suitable" interpolation of the stiffness-tensor $\boldsymbol{C}(\boldsymbol{\phi})$ as e.g. $\boldsymbol{C}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{n} \boldsymbol{C}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$ of the indivual stiffness-tensors $\boldsymbol{C}^{\alpha}$ with some interpolation function h^{α}. Using these two quantities, one can naturally define (assuming for simplicity the absence of any prestresses or eigenstrains) a volumetric free energy contribution $f_{el}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \frac{1}{2}\boldsymbol{\epsilon} : \boldsymbol{C}(\boldsymbol{\phi}) : \boldsymbol{\epsilon}$ and the resulting stress tensor $\boldsymbol{\sigma}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \boldsymbol{C}(\boldsymbol{\phi}) : \boldsymbol{\epsilon} = \frac{\partial f_{el}(\boldsymbol{\phi}, \boldsymbol{\epsilon})}{\partial \boldsymbol{\epsilon}}$.
While this provides a seemingly reasonable interpolation scheme, it was soon realized that its use led to similar artefacts as the ones encountered in the earlier, simpler solidification models. Given the success of the models based on the use of a common chemical potential of all phases instead of the common concentration and based on the analogous roles of $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ with respect to the underlying energies, Seinbach and Apel proposed a different scheme in [69] relying on the equality of the stresses instead of that of the strains.

Unfortunately, it has been observed that (see [23]), that, depending on the mechnical setting, both models suffer from excess energy effects within the interface. Durga et al. further observed that these deviations for both models can primarily be attributed to the fact that one does in general neither expect a full equilibration of the strains nor of the stresses. Instead, the corresponding continuity conditions within a sharp interface setting are given by an equality of the normal stresses and an equality of the tangential strains. Based on this, they proposed a model using a "mixed" interpolation approach mimicking this expected behavior within the diffuse interface region, showing a significant improvement in the modeling results. The model was then, within a two-phase setting, further developed and extended in [51], [24] and [64].
In constrast to the simpler mechanical models and the solidification models from Section 3.2.1, these mechanical models, while very elegant in the two-phase case, are unfortunately very difficult to generalize to a multiphase setting. Even though different extensions to the multiphase case have been proposed in e.g. [61], [74], [63] and [62], none of them can be considered fully satisfactory since they either have to rely on a physically undesirable "geometric" simpliciation through a common normal vector between all phase-pairings, or will suffer from a violation of the jump conditions for at least parts of the phases. These difficulties are not unexpected since the multiphase regions in the mechanical model correspond to the intersection of several interfaces in the sharp interface setting - and thus points which are generally associated with singularities in the behavior of the mechanical fields - but it is important to be aware of the limitations and relative advantages and disadvantages associated with each of them. Since the differences in the models are primarily inherited from the particular description chosen in the simpler two-phase case (where, except for the model of [23] and [24], the models coincide) and the chosen description in addition has a significant influence on the computational cost and implementation effort, Section 7.2 will start by summarizing and comparing the various descriptions, both in terms of the formulation itself and in terms of some computational aspects. Subsection 7.2.4 then contains an outline of the different extensions to the multiphase case and some of the issues associated with those. The remainder of Section 7.2 is then devoted to a discussion of some aspects concerning coupled mechanical and chemical calculations, in particular in combination with the jump-condition based mechanical model and the more advanced free energy model from Section 7.1.

Part II

Background

The outline in Chapter 3 clearly indicates the importance of the variational structure and the constraints underlying the phasefield approach in general as well as for the description of the more quantitative models for both the solidification and solid-solid transformations. For this reason, before returning to a more detailed discussion of the applications introduced above, this section will provide some mathematical background on such problems and the closely associated topic of Lagrange multipliers. As this is a topic of both high practical importance and independent mathematical interest, it has been studied in great detail and the literature on the subject is vast. The purpose here is thus not to provide an in-depth discussion, but just to outline some basics which will be used in the sequel. This will be complemented by numerous literature references where a more detailed presentation and related but more advanced results may be found.

The central difference between constrained minimization problems and their unconstrained counterparts is that both the characterization of and the search for (local) minima is made more difficult by the fact that, in the former case, it is not sufficient to focus just on some differential information of the objective function alone. Instead, it is obvious that the constraints must also be a part of the formulation of any necessary and/or sufficient condition describing minima and of any algorithm aiming at their determination. As a first step, the following Chapter 4 contains a discussion of equality and inequality-constrained problems. In both cases, the focus is on the description of first-order necessary conditions for local minimizers, which naturally leads to the notion of Lagrange-multipliers.
The results obtained for both the equality- and inequality constrained settings can be conveniently summarized using the concept of Lagrange-functionals, which will be discussed in Section 5.1. These on the one hand allow for a simple formal derivation of first-order necessary and second-order necessary resp. sufficient conditions for constrained local minimizers. On the other hand, they are also the basis for both the design and the analysis of a number of algorithms for the numerical treatment of such problems. Section 5.2 then gives some background on an a priori different topic, namely the Legendre-Fenchel transform, which has a fundamental role in thermodynamics.

Chapter 4

Equality and Inequality Constrained Problems

4.1 Linear and Nonlinear Equality Constraints - the Finite-Dimensional Case

Linear Equality Constraints The simplest setting for constrained minimization problems is that of minimizing a smooth (e.g. C^2) function f defined on $\mathbb{R}^n$ subject to $m \leq n$ linear equality constraints

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{b}_i \cdot \boldsymbol{x} = c_i,\ i = 1, ..., m. \end{cases} \tag{4.1}$$

The question to be answered is then under what conditions one can assert that a point $\bar{\boldsymbol{x}}$ is a (local) minimizer for the problem (4.1).

The first obvious point is that $\bar{\boldsymbol{x}}$ needs to satisfy $\boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} = \boldsymbol{b}_i^T \bar{\boldsymbol{x}} = c_i$ for $i = 1, ..., m$ (and thus in particular the constraints need to be consistent). In order for such a point $\bar{\boldsymbol{x}}$ to be local minimizer, the value $f(\bar{\boldsymbol{x}})$ has to be no larger than those of $f(\boldsymbol{x})$ with $\boldsymbol{x}$ in a neighborhood of $\bar{\boldsymbol{x}}$ which, in addition, also satisfy $\boldsymbol{b}_i \cdot \boldsymbol{x} = c_i, i = 1, ..., m$. Due to this restriction and the linearity of the constraints, any admissible variation $\delta\boldsymbol{x}$ of $\bar{\boldsymbol{x}}$ has to satisfy $\boldsymbol{b}_i \cdot \delta\boldsymbol{x} = 0$, $i = 1, ..., m$. Combining the equality-constraints into a vector-equation by inserting the vectors $\boldsymbol{b}_i$ row-wise into the matrix

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1^T \\ \boldsymbol{b}_2^T \\ ... \\ \boldsymbol{b}_m^T \end{pmatrix},$$

this can be written in a more compact form as $\boldsymbol{B} \cdot \delta\boldsymbol{x} = \boldsymbol{0}$ or $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$. As all admissible points are of the form $\bar{\boldsymbol{x}} + \delta\boldsymbol{x}$ with $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$, the optimality condition is thus given by $f(\bar{\boldsymbol{x}} + \delta\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}})$ for all sufficiently small $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$.

It remains to convert this into a differential characterization. Given any $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$, the vector $\boldsymbol{x} = \bar{\boldsymbol{x}} + t\delta\boldsymbol{x}$ clearly satisfies $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{c}$ provided $\bar{\boldsymbol{x}}$ does so, i.e. is admissible, and, since $\bar{\boldsymbol{x}}$ is assumed to be a local minimizer, will satisfy $f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}})$ or

$$0 \leq \frac{f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}})}{t} = \frac{f(\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}) - f(\bar{\boldsymbol{x}})}{t}$$

for sufficiently small t. Letting $t \to 0$ and since f is assumed smooth[1], it then follows that any minimizer $\bar{\boldsymbol{x}}$ has to satisfy the **first-order necessary condition (FONC)**

$$\nabla f(\bar{\boldsymbol{x}}) \cdot \delta \boldsymbol{x} \geq 0 \quad \forall \delta \boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B}). \tag{4.2}$$

As $\mathrm{Ker}(\boldsymbol{B})$ is a linear subspace of $\mathbb{R}^n$, this condition can be further simplified to the equality [2]

$$\nabla f(\bar{\boldsymbol{x}}) \cdot \delta \boldsymbol{x} = 0 \quad \forall \delta \boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B}). \tag{4.3}$$

In contrast to the unconstrained case where $\delta \boldsymbol{x}$ can be chosen arbitrarily, this does not imply $\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{0}$. Instead, condition (4.3) only states that $\nabla f(\bar{\boldsymbol{x}})$ lies in the orthogonal complement of the kernel of $\boldsymbol{B}$,

$$\nabla f(\bar{\boldsymbol{x}}) \in \mathrm{Ker}(\boldsymbol{B})^{\perp}.$$

In order to obtain a more explicit condition on $\nabla f(\bar{\boldsymbol{x}})$, it is thus sufficient to give a characterization of $\mathrm{Ker}(\boldsymbol{B})^{\perp}$. Each of the equations $\boldsymbol{b}_i \cdot \delta \boldsymbol{x} = 0$ characterizing the kernel of $\boldsymbol{B}$ is nothing but the description of a hyperplane with the vector $\boldsymbol{b}_i$ acting as the normal vector. In the case of a single constraint, it is geometrically obvious that $\mathrm{Ker}(\boldsymbol{b}_i^T)^{\perp} = \mathrm{Span}(\boldsymbol{b}_i)$ (see Figure 4.1), or, equivalently, if $\boldsymbol{b}_i$ is considered as the linear mapping $\boldsymbol{b}_i : \mathbb{R} \ni \lambda \to \lambda \boldsymbol{b}_i \in \mathbb{R}^n$, that $\mathrm{Ker}(\boldsymbol{b}_i^T)^{\perp} = \mathrm{Range}(\boldsymbol{b}_i)$.
More generally, the kernel of $\boldsymbol{B}$ is just the intersection of such hyperplanes, $\mathrm{Ker}(\boldsymbol{B}) = \cap_i \mathrm{Ker}(\boldsymbol{b}_i^T)$. As $\cap_{i=1}^m \mathrm{Ker}(\boldsymbol{b}_i^T) \subset \mathrm{Ker}(\boldsymbol{b}_i^T)$ and $A \subset B$ implies $A^{\perp} \supset B^{\perp}$ for any subsets A, B of $\mathbb{R}^n$, it is clear that $\mathrm{Ker}(\boldsymbol{B})^{\perp} \supset \mathrm{Ker}(\boldsymbol{b}_i^T)^{\perp}$, $i = 1, ..., m$, and thus also[3]

$$\mathrm{Ker}(\boldsymbol{B})^{\perp} \supset \mathrm{Span}\left(\{\mathrm{Ker}(\boldsymbol{b}_i^T)^{\perp}\}_{1 \leq i \leq m}\right) = \mathrm{Span}(\{\boldsymbol{b}_i\}_{1 \leq i \leq m}) = \mathrm{Range}(\boldsymbol{B}^T).$$

A restatement of the converse conclusion $\mathrm{Ker}(\boldsymbol{B})^{\perp} \subset \mathrm{Range}(\boldsymbol{B}^T)$ is simply that any vector which is not in $\mathrm{Range}(\boldsymbol{B}^T)$ is not orthogonal to $\mathrm{Ker}(\boldsymbol{B})$. In fact, any vector $\boldsymbol{x} \notin \mathrm{Range}(\boldsymbol{B}^T)$ can be written in the form $\boldsymbol{x} = \boldsymbol{y} + \sum_{i=1}^m \alpha_i \boldsymbol{b}_i$ where[4] $\boldsymbol{y} \neq \boldsymbol{0}$ is orthogonal to $\mathrm{Range}(\boldsymbol{B}^T)$ (apply e.g. a Gram-Schmidt orthogonalization procedure using $\boldsymbol{x}$ and the $\{\boldsymbol{b}_i\}_{1 \leq i \leq m}$). This means precisely that $\boldsymbol{y} \perp \boldsymbol{b}_i$, $i = 1, ..., m$ or $\boldsymbol{y} \in \mathrm{Ker}(\boldsymbol{B})$, and thus $\boldsymbol{y} \cdot \boldsymbol{x} = \|\boldsymbol{y}\|^2 \neq 0$ shows that $\boldsymbol{x} \notin \mathrm{Ker}(\boldsymbol{B})^{\perp}$.
In combination with the same argument applied to $\mathrm{Ker}(\boldsymbol{B}^T)$ and $\mathrm{Range}(\boldsymbol{B})$, this implies the following fundamental theorem:

Theorem 1. *(Fundamental theorem of linear algebra)*
Let $\boldsymbol{B} : \mathbb{R}^n \to \mathbb{R}^m$, $m, n \geq 1$ be any real matrix. Then the orthogonal decompositions

$$\mathbb{R}^n = \mathrm{Ker}(\boldsymbol{B}) \oplus \mathrm{Range}(\boldsymbol{B}^T) \tag{4.4}$$

and

$$\mathbb{R}^m = \mathrm{Ker}(\boldsymbol{B}^T) \oplus \mathrm{Range}(\boldsymbol{B})$$

hold.

Using this result, $\nabla f(\bar{\boldsymbol{x}}) \in \mathrm{Ker}(\boldsymbol{B})^{\perp}$ is thus equivalent to $\nabla f(\bar{\boldsymbol{x}}) \in \mathrm{Range}(\boldsymbol{B}^T)$, i.e. there exists a vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ of **Lagrange-multipliers** such that[5]

$$-\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{B}^T \boldsymbol{\lambda}. \tag{4.5}$$

[1] Clearly, Gâteaux-differentiability is sufficient here.

[2] In fact, otherwise assuming that $\delta \boldsymbol{x}$ is a variation such that $\nabla f(\bar{\boldsymbol{x}}) \cdot \delta \boldsymbol{x} > 0$, $-\delta \boldsymbol{x}$ is necessarily also a legitimate variation, but, with $\nabla f(\bar{\boldsymbol{x}}) \cdot (-\delta \boldsymbol{x}) = -\nabla f(\bar{\boldsymbol{x}}) \cdot \delta \boldsymbol{x} < 0$, contradicting the optimality of $\bar{\boldsymbol{x}}$.

[3] This is again intuitively clear in $\mathbb{R}^n$, as this simply expresses that any linear combination of the vectors $\{\boldsymbol{b}_i\}_{1 \leq i \leq m}$ is orthogonal to all vectors which are orthogonal to all of the $\{\boldsymbol{b}_i\}_{1 \leq i \leq m}$.

[4] As will be seen below, the analogue of this seemingly simple statement in $\mathbb{R}^n$ in terms of a separation theorem is the major source of difficulty for an analogous result in the infinite-dimensional setting. The basic difficulty in the latter case is that, while the range of any continuous linear operator is always a subspace, this subspace (contrary to the situation in $\mathbb{R}^n$) need not be closed. An element x not in the range of the operator might thus lie in its closure and can therefore be approximated arbitrarily close by elements in the range.

[5] Note that using $-\nabla f$ instead of ∇f is simply a matter of convenience here as one could just as well change the sign of $\boldsymbol{\lambda}$.

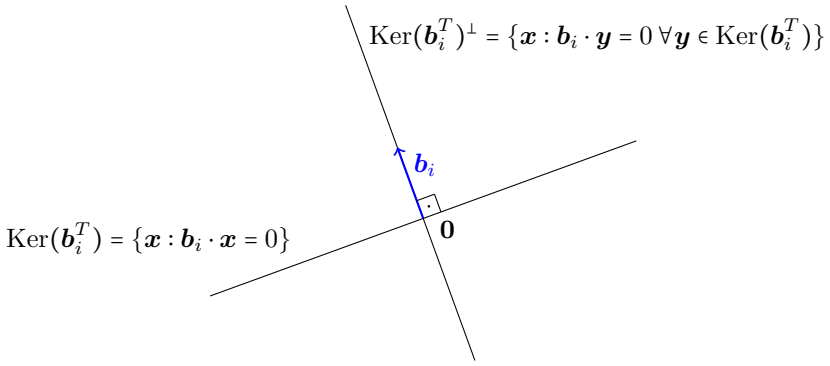


Figure 4.1: Relationship between $\mathrm{Ker}(\boldsymbol{b}_i^T)$ and $\mathrm{Ker}(\boldsymbol{b}_i^T)^\perp$

Remark 4. Despite its simplicity, the result above is, as the name suggests, fundamental for finite-dimensional linear problems (constrained or not) as it gives a very convenient decomposition of the domain and image spaces of any linear operator.
Among the many consequences, one should mention the well-known **Fredholm alternative**: The equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable iff $\boldsymbol{b}$ is orhogonal to all solutions $\boldsymbol{y}$ of the homogeneous transposed equation $\boldsymbol{A}^T\boldsymbol{y} = \mathbf{0}$, which is just a reformulation of the second statement $\mathrm{Range}(\boldsymbol{A}) = \mathrm{Ker}(\boldsymbol{A}^T)^\perp$ above.
In particular, the operator $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^m$ is surjective iff $\mathrm{Ker}(\boldsymbol{A}^T) = \{\mathbf{0}\}$, i.e. $\boldsymbol{A}^T$ is injective. As all (if any) solutions of an equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ only differ by elements lying in $\mathrm{Ker}(\boldsymbol{A})$, a similar conclusion is that a linear operator $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$ is injective (or $\mathrm{Ker}(\boldsymbol{A}) = \{\mathbf{0}\}$) iff $\boldsymbol{A}^T$ is surjective (i.e. $\mathrm{Range}(\boldsymbol{A}^T) = \mathbb{R}^m$) as the first statement then shows that $\mathbb{R}^m = \mathrm{Ker}(\boldsymbol{A}) \oplus \mathbb{R}^m$. ◇

Remark 5. The preceding arguments are prototypical for the more difficult situations considered below.
First of all, under mild assumptions, they can be applied essentially without change in the nonlinear case considered in the next paragraph after a simple linearization.
In addition, these result can in many cases of practical interest be extended to the infinite-dimensional setting where the matrix $\boldsymbol{B}$ above is replaced by a suitable linear operator acting between two Banach spaces $\boldsymbol{X} \to \boldsymbol{Y}$. The only fundamental change here (at least for smooth functionals $\mathcal{F} : X \to \mathbb{R}$) is that this requires some additional topological assumptions, under which the decisive Theorem 1 can be generalized in the form of the **closed range theorem** (see Thm. 2 below).
Finally, the treatment of inequality constrained problems of the form $\boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{c}$ for some given matrix $\boldsymbol{B}$ and a vector $\boldsymbol{c}$ also runs much along the same lines. A major difference in this case is that, while the Euler-Lagrange equation characterizing any local minimizers could be simplified to an equality in the setting above, this is no longer possible when inequalities are present. This leads to the requirement of replacing the subspace defining the kernel of $\boldsymbol{B}$ through an intersection of the hyperplanes by a cone defined through a suitable intersection of half-spaces. ◇

Nonlinear Equality Constraints While most of the equality constraints arising in the applications considered in this thesis are indeed linear, this is of course not always the case. Fortunately, under relatively mild additional hypothesis on the equality constraints, the FONC for the more general case where $\boldsymbol{x}$ is subject to $m \leq n$ nonlinear equality constraints

$$h_i(\boldsymbol{x}) = c_i, \quad i = 1, ..., m \tag{4.6}$$

can be deduced from the previous consideration after a simple linearization of the h_i, $i = 1, .., m$. More precisely, one would like to replace the admissible variations $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$ at a feasible

point $\bar{\boldsymbol{x}}$ from the FONC of the previous section with the vectors $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{h}(\bar{\boldsymbol{x}})')$, where $\boldsymbol{h}$ is the (column) vector composed of the $h_i(\boldsymbol{x})$, $\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} h_1(\boldsymbol{x}) & h_2(\boldsymbol{x}) & \dots & h_m(\boldsymbol{x}) \end{pmatrix}^T$. Provided this is legitimate, one can apply the theory from the linear case to deduce the existence of a (not necessarily unique) Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that any local minimizer satisfies

$$\nabla f(\bar{\boldsymbol{x}}) = (\boldsymbol{h}'(\bar{\boldsymbol{x}}))^T \boldsymbol{\lambda} = \nabla \boldsymbol{h}(\bar{\boldsymbol{x}})\boldsymbol{\lambda}.$$

Before entering into a more detailed discussion, it is instructive to consider a simple counterexample in $\mathbb{R}^2$ which shows that this is not always possible. Consider a smooth function $f(\boldsymbol{x})$ with $\boldsymbol{x}$ constrained to lie in the admissible set consisting of all points on the vertical axis, but with this (in its most natural form) linear constraint artificially made nonlinear by rewriting this set using the constraint $h(\boldsymbol{x}) = x^2 = 0$. It is clear from the discussion in the previous section that at any minimizer, there will be a $\lambda \in \mathbb{R}$ such that $\nabla f = \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, i.e. the only non-vanishing component of the gradient is along the x-direction.
Linearizing h though, any admissible point with $x = 0$ will satisfy

$$\nabla h(\boldsymbol{x}) = h'(\boldsymbol{x})^T = \begin{pmatrix} x \\ 0 \end{pmatrix}^T = \mathbf{0}.$$

It is obvious that this vector is useless for expressing the FONC above.

Remark 6. Roughly speaking, one can see that the difficulty in the above example arises as there are two different "linearizations" involved, which in this case do not lead to the same conclusions. One linearization is the more geometric one based on the admissible set $\mathcal{A}$ (in this case the y-axis) itself and the translated tangent space $T_{\mathcal{A}}(\boldsymbol{x})$ to this set obtained by considering the tangents to all smooth curves lying in $\mathcal{A}$ and passing through $\boldsymbol{x}$ (in this case again the y-axis). The other one is a more algebraic one based on the linearization of the function $\boldsymbol{h}$ describing the admissible set and the set of directions lying in $\mathrm{Ker}(\boldsymbol{h}'(\boldsymbol{x}))$. As seen in the example above, this set corresponds to the y-axis for the choice $h(\boldsymbol{x}) = x$ and to the whole of $\mathbb{R}^2$ (i.e. a significantly larger set) for the choice $h(\boldsymbol{x}) = x^2$. In constrast to $T_{\mathcal{A}}(\boldsymbol{x})$, this set thus depends on the particular expression chosen for $\boldsymbol{h}$ and may or may not coincide with the former.
While it is intuitively to be expected that one is interested in the first set for the constrained problems considered here, it is also clear that the second one is much easier to deal with.
The principal question to be answered is therefore under which conditions both subsets coincide. These so-called **constraint qualificiation conditions** appear generally in constrained optimization problems where the definition of the admissible set $\mathcal{A}$ involves systems of equalities and inequalities. The difficulty is always due to the above dichotomy, where one has, on the one hand, a purely geometrically defined tangent set to $\mathcal{A}$ which appears naturally in the Euler-Lagrange equation characterizing the minimizer, and, on the other hand, an appropriate linearization of the constraint equations which one would like to combine with Lagrange multipliers in order to obtain a more explicit representation of this set. ◇

As for the linearly constrained case, the FONCs at any purported local minimizer $\bar{\boldsymbol{x}}$ are again based on considering nearby points also satisfying the inequality constraints. For all such points sufficiently close to $\bar{\boldsymbol{x}}$, one has, by the assumption on $\bar{\boldsymbol{x}}$, that $f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}) \geq 0$. In order to obtain a first-order necessary condition based on directional derivatives, one would now like to divide by t and take the limit of $t \to 0$ for points of the form $\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}$ with an admissible direction $\delta\boldsymbol{x}$. There is a slight difficulty with this approach, as, due to the potential curvature of the admissible set, it is in general not possible to approach $\bar{\boldsymbol{x}}$ along a straight line while remaining within the admissible set $\mathcal{A}$, i.e. it is not exactly legitimate to compare the values of $f(\boldsymbol{x})$ along a line of the type above with those of $\bar{\boldsymbol{x}}$. Instead, one has to compare the value of $\bar{\boldsymbol{x}}$ with points lying "almost" on such a straight line, i.e. points $\boldsymbol{x} \in \mathcal{A}$ which can be written in the form $\boldsymbol{x} = \bar{\boldsymbol{x}} + t\delta\boldsymbol{x} + o(t)$ for some direction $\delta\boldsymbol{x}$. In fact, provided f is sufficiently smooth and there is some sequence t_n

tending to 0 such that an associated sequence $\boldsymbol{x}_n = \bar{\boldsymbol{x}} + t_n\delta\boldsymbol{x} + o(t_n)$ lying in $\mathcal{A}$ exists, one then has

$$0 \leq \frac{f(\boldsymbol{x}_n) - f(\bar{\boldsymbol{x}})}{t_n} = \frac{f(\bar{\boldsymbol{x}}) + t\nabla f(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} + o(t_n) - f(\bar{\boldsymbol{x}})}{t_n} = \nabla f(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} + o(1)$$

and thus, in the limit, $\nabla f(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} \geq 0$.
Denoting the "convergence in $\mathcal{A}$" with $\boldsymbol{x}_n \underset{\in\mathcal{A}}{\to} \bar{\boldsymbol{x}}$, i.e. all points $\boldsymbol{x}_n$ (and $\bar{\boldsymbol{x}}$) lie in $\mathcal{A}$, the set of directions for which this construction is possible is given by the tangential set

$$T_{\mathcal{A}}(\bar{\boldsymbol{x}}) = \{\delta\boldsymbol{x} \in \mathbb{R}^n : \exists t_n \to 0, \boldsymbol{x}_n \underset{\in\mathcal{A}}{\to} \bar{\boldsymbol{x}} \text{ such that } \delta\boldsymbol{x} = \lim_{t_n \to 0} \frac{\boldsymbol{x}_n - \bar{\boldsymbol{x}}}{t_n}\} \tag{4.7}$$

called the **sequential/Bouligand cone**. With this set, the FONC in this more general case is given by

$$\nabla f(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} \geq 0 \quad \forall \delta\boldsymbol{x} \in T_{\mathcal{A}}(\bar{\boldsymbol{x}}). \tag{4.8}$$

According to the preceding discussion, in order to obtain a description in terms of Lagrange multipliers as outlined above, one now needs to ensure that $T_{\mathcal{A}}(\bar{\boldsymbol{x}})$ is in fact a subspace (allowing to simplify the inequality into an equality) and that this subspace is given by the kernel of $\boldsymbol{h}'(\bar{\boldsymbol{x}})$ in order to conclude that $\nabla f(\bar{\boldsymbol{x}}) \in \mathrm{Ker}(\boldsymbol{h}')^{\perp} = \mathrm{Range}\left((\boldsymbol{h}'(\bar{\boldsymbol{x}}))^T\right)$ and thus that

$$-\nabla f(\bar{\boldsymbol{x}}) = \left(\boldsymbol{h}'(\bar{\boldsymbol{x}})\right)^T \boldsymbol{\lambda} = \nabla \boldsymbol{h}(\bar{\boldsymbol{x}})\boldsymbol{\lambda} \tag{4.9}$$

for some $\boldsymbol{\lambda} \in \mathbb{R}^m$.

A first observation in this sense (also recall the example above) is that, if $\boldsymbol{h}$ is smooth, e.g. $\boldsymbol{h} \in C^1(\mathbb{R}^n)$, then $T_{\mathcal{A}}(\bar{\boldsymbol{x}}) \subset \mathrm{Ker}\left(\boldsymbol{h}'(\bar{\boldsymbol{x}})\right)$ as, for any direction $\delta\boldsymbol{x} \in T_{\mathcal{A}}(\bar{\boldsymbol{x}})$ and the associated sequences $\boldsymbol{x}_n$ and t_n, one has

$$\boldsymbol{0} = \frac{\boldsymbol{h}(\boldsymbol{x}_n) - \boldsymbol{c}}{t_n} = \frac{\boldsymbol{h}\big(\bar{\boldsymbol{x}} + t_n\delta\boldsymbol{x} + o(t_n)\big) - \boldsymbol{c}}{t_n} = \frac{\boldsymbol{h}(\bar{\boldsymbol{x}}) + t_n\nabla\boldsymbol{h}'(\bar{\boldsymbol{x}})\delta\boldsymbol{x} + o(t_n) - \boldsymbol{c}}{t_n} = \boldsymbol{h}'(\boldsymbol{x})\delta\boldsymbol{x} + o(1)$$

and thus in the limit $\delta\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{h}'(\bar{\boldsymbol{x}}))$.
All that can a priori be said about vectors $\delta\boldsymbol{x} \in \mathrm{Ker}\left(\boldsymbol{h}'(\bar{\boldsymbol{x}})\right)$ is that

$$\boldsymbol{h}\big(\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}\big) = \boldsymbol{h}\big(\bar{\boldsymbol{x}}\big) + t\boldsymbol{h}'(\bar{\boldsymbol{x}})\delta\boldsymbol{x} + o(t) = \boldsymbol{c} + o(t), \tag{4.10}$$

i.e. they satisfy the equality constraint up to an error of $o(t)$. What is required in order to show that $\delta\boldsymbol{x}$ lies in $T_{\mathcal{A}}$ though is not that the equation defining $\mathcal{A}$ is satisfied up to an order of $o(t)$, but that there is a point $\boldsymbol{x}$ within a distance of $o(t)$ to $\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}$ actually satisfying the equation. This conclusion based on Equation (4.10) requires a stability result ensuring that the $o(t)$-error in the defining equation for $\mathcal{A}$ can be compensated by an equal-order correction to the point $\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}$. In other words, one has to ensure that $\boldsymbol{h}$ is a local homeomorphism at $\bar{\boldsymbol{x}}$, i.e. a continuous mapping with a continuous inverse. This is a classical question though, for which it is known (see e.g. [27]) that this is the case iff $\boldsymbol{h}'(\bar{\boldsymbol{x}} + t_n\delta\boldsymbol{x})$ is surjective, which is in turn guaranteed by $\boldsymbol{h}'(\bar{\boldsymbol{x}})$ being surjective.

4.2 Linear and Nonlinear Equality Constraints - the Infinite-Dimensional Case

Even though the setting considered in the previous section already contains the essential ideas underlying the Lagrange multiplier approach for equality-constrained problems, it is not strictly speaking applicable to the applications outlined above. Instead, one has to consider a more general situation in which the domain of the function(al) to be minimized is given by an appropriate subset of a function space and thus generally infinite-dimensional. In addition, while it may happen - as in the case of the minimization with a prescribed volume - that there is only a finite number of equality constraints, there are other situations in which this need not be the case anymore[6].

Both aspects entail some additional difficulties absent from the considerations in the previous section. Even though the discretization of any of the examples considered in this section ultimately reduces the situation to the previous finite-dimensional one, the underlying continuous description should not be completely forgotten.

Remark 7. This is not just a matter of theoretical interest but also has implications from a "practical" point of view.
On the one hand, despite the ultimately discrete nature of the computational problems, a continuous description almost always leads to significantly shorter calculations as compared to an analogous one at the discrete level, and is in particular also heavily made use of when deriving e.g. the phasefield equation and other related ones from the underlying functional. Even though the engineering community enjoys some additional liberties in this respect as compared to the mathematical one (which will also be taken here), it can still be helpful to have an idea of which calculations are potentially problematic and which are not. In particular, it is clear that the continuous description will become increasingly important as one increases the resolution in the discrete case, and an ill-defined description at the continuous level is then likely to cause issues at the discrete one.
On the other hand, a basic understanding of some important relations at the continuous level - such as e.g. the gradient and negative divergence being related in a "tranpose-like" fashion - allows for a relatively simple intuitive interpretation of some a priori quite complex questions at the discrete level. ◇

The remainder of this section will therefore provide a quick sketch of how and to what degree the arguments from Section 4.1 can be carried over to the function space setting. The central result replacing the use of the fundamental theorem of linear algebra in the previous section is a very similar characterization of the decomposition of the domain and range space through the closed range theorem 2 below, with the role of the matrix $\boldsymbol{B}$ and its transpose $\boldsymbol{B}^T$ in Theorem 1 being replaced by the opertator $\mathcal{B}$ defining the constraints and its adjoint $\mathcal{B}^*$. Provided the conditions of the theorem are satisfied, one can use this characterization to derive, based on essentially identical arguments as in the finite-dimensional case, the existence of an appropriate Lagrange multiplier for the constraint and the way it enters into the first-order necessary codition through $\mathcal{B}^*$.

Remark 8. The closed range theorem is a fundamental theorem of functional analysis and can therefore be found in essentially any introductory text on this topic. The outline below mostly borrows from [16], [48],[80] and [20], to which the reader is referred for more details. More in-depth descriptions and generalizations can be found in e.g. [14], [37] and [57]. ◇

[6] A particularly important example is given by the incompressibility condition $\nabla \cdot \boldsymbol{u} = 0$ encountered when dealing with fluid flow problems.

Linear equality constraints Consider first the case when $\mathcal{F} : X \to \mathbb{R}$ is a (smooth) functional defined on some Banach or Hilbert space[7] X and the equality constraints are specified through a (possibly unbounded[8]) linear operator $\mathcal{B} : \mathrm{Dom}(\mathcal{B}) \subset X \to Y$,

$$\mathcal{B}x = y,$$

where Y is a second Banach space.

Remark 9. The simplest case is the one where Y is finite-dimensional, corresponding to a finite number of (linear) equality constraints, specified through m continuous linear functionals $b_i^* \in X^*$,

$$\langle b_i^*, x\rangle_{X^*,X} = c_i, \quad i = 1, ..., m.$$

For the phasefield, one may for example have $X = H^1(\Omega) \cap L^\infty(\Omega)$, $\mathcal{F} = \mathcal{E}_\epsilon$ and $\int_\Omega \phi \, \mathrm{d}\boldsymbol{x} = V$. For the (Navier-)Stokes equation one instead has to consider the more general case where X may e.g. be a subspace of $H^1(\Omega)$ and Y equal to $L^2(\Omega)$. ◇

As in the previous section, standard differential calculus implies that the derivative $\mathcal{F}'(\bar{x})$ of $\mathcal{F}$ (as an element of X^*) at any minimizer $\bar{x}$ must satisfy

$$\langle \mathcal{F}'(\bar{x}), \delta x\rangle_{X^*,X} \geq 0 \quad \forall \delta x \in \mathrm{Ker}(\mathcal{B}),$$

and then again, as $\mathrm{Ker}(\mathcal{B})$ is a linear subspace of X,

$$\langle \mathcal{F}'(\bar{x}), \delta x\rangle_{X^*,X} = 0 \quad \forall \delta x \in \mathrm{Ker}(\mathcal{B}). \tag{4.11}$$

Defining [9]the **annihilator** M^a of a subset of M as the set of all $x^* \in X^*$ such that $\langle x^*, x\rangle_{X^*,X}$ for all $x \in M$, one thus has $\mathcal{F}'(\bar{x}) \in \mathrm{Ker}(\mathcal{B})^a$. As in the case of $\mathbb{R}^n$, one would now like to, whenever this is possible, provide a more explicit representation of $\mathrm{Ker}(\mathcal{B})^a$. This requires introducing, in analogy to the transposed matrix, the **adjoint operator** $\mathcal{B}^* : \mathrm{Dom}(\mathcal{B}^*) \subset Y^* \to X^*$:

Definition 1. (**Adjoint operator**, see e.g. [16], p. 43f)
Let $\mathcal{B} : \mathrm{Dom}(\mathcal{B}) \subset X \to Y$ be a linear operator which is densely defined[10]. The domain of $\mathcal{B}^*$ is defined to be the set of $y^* \in Y^*$ such that there exists a constant c such that

$$|\langle y^*, \mathcal{B}(x)\rangle_{Y^*,Y}| \leq c\|x\|_X \ \forall x \in \mathrm{Dom}(\mathcal{B}),$$

i.e. $\mathrm{Dom}(\mathcal{B}^*)$ is the (linear) subspace of Y^* for which the linear mapping $D(\mathcal{B}) \ni x \mapsto g(x) = \langle y^*, \mathcal{B}x\rangle$ is uniformly bounded. As $\mathrm{Dom}(\mathcal{B})$ is by assumption dense in X, g can be extended by continuity to a unique bounded linear operator on all of X, allowing to identify g with an element of X^*. This association in turn defines a linear mapping associating with each $y^* \in \mathrm{Dom}(\mathcal{B}^*)$ an element of X^*, which will be denoted by $\mathcal{B}^* y^*$ and by definition satisfies

$$\langle y, \mathcal{B}x\rangle_{Y^*,Y} = \langle \mathcal{B}^* y^*, x\rangle_{X^*,X} \quad \forall y^* \in \mathrm{Dom}(\mathcal{B}^*), x \in \mathrm{Dom}(\mathcal{B}). \tag{4.12}$$

[7]I.e. a complete normed vector space, resp. one where the norm can additionally be derived from an inner product.

[8]Meaning that $\mathcal{B}$ need not be continuous on the whole space. Typical examples for such operators are differentiation operators between various spaces, which need not be bounded on the whole space but are so on a dense subset of sufficiently smooth functions.

[9]The reason for introducing a slightly "modified version" of the orthogonal complement here is that in the more general Banach space setting, not every linear subspace admits an orthogonal complement. When X is a Hilbert space, this is the case though for every **closed** subspace. As the kernel of every linear continuous operator is always closed, in this case one can thus always replace $\mathrm{Ker}(\mathcal{B})^a$ with $\mathrm{Ker}(\mathcal{B})^\perp$ since both notions coincide then (see e.g. [16]).

[10]Meaning its definition is such that it "makes sense" on a dense subset of functions in X.

The adjoint operator is thus constructed such as to extend the defining property of the transposed matrix $\boldsymbol{B}^T : \mathbb{R}^m \to \mathbb{R}^n$ with respect to the Euclidian scalar product in $\mathbb{R}^m$ resp. $\mathbb{R}^n$,

$$\boldsymbol{y} \cdot (\boldsymbol{B}\boldsymbol{x}) = (\boldsymbol{B}^T \boldsymbol{y}) \cdot \boldsymbol{x},$$

to the more general setting given by the duality product between the spaces above. The hope is of course to obtain an analogue of the fundamental theorem of linear algebra for this case. Unfortunately, this is not possible without some additional topological assumptions, which hold in many cases of practical interest though.

Their statement requires another definition (see e.g. [16], p. 43):

Definition 2. (Graphs and closed maps)
The **graph** $\mathrm{graph}(\mathcal{B})$ of a map (linear or not) $\mathcal{B} : X \to Y$ is the subset of $X \times Y$ given by

$$\mathrm{graph}(\mathcal{B}) = \{(x, y) \in X \times Y : y = \mathcal{B}(x)\}.$$

$\mathcal{B}$ is **closed** if $\mathrm{graph}(\mathcal{B})$ is a closed subset of $X \times Y$ (equipped with the standard product topolgy)[11].

With this definition, the fundamental result for the characterization of $\mathrm{Ker}(\mathcal{B})^a$ is contained in the following (see e.g. thm. 2.16 [16], thm. 2.13 [48], §7.5 [79])

Theorem 2. *(Closed range theorem)*
Let $\mathcal{B} : \mathrm{Dom}(\mathcal{B}) \subset X \to Y$ be an unbounded linear operator which is densely defined and closed. Then the following are equivalent:

1. $\mathrm{Range}(\mathcal{B})$ *is closed.*
2. $\mathrm{Range}(\mathcal{B}^*)$ *is closed.*
3. $\mathrm{Range}(\mathcal{B}) = \mathrm{Ker}(\mathcal{B}^*)^a$.
4. $\mathrm{Range}(\mathcal{B}^*) = \mathrm{Ker}(\mathcal{B})^a$.

Remark 10. Note that there are two different notions of "closedness" involved in this theorem. The first one - that of being a closed map - depends on the simultaneous convergence of a sequence of points (x_n) in X and its image sequence $(y_n) = \mathcal{B}x_n$ in Y and requires that, if both converge, the limit in Y is the image of the limit in X. The second one - that of having a closed range - is more concerned with the image space, i.e. given an arbitrary sequence (y_n) in Y which converges in Y, the limit has to lie in $\mathrm{Im}(B)$, meaning there only has to be some point in X such that $y = \mathcal{B}x$ (but which, a priori, need not be related to any particular sequence in X). ◇

Assuming that the conditions of this theorem are satisfied for $\mathcal{B}$, one can finally conclude from (4.11) that at a local minimizer $\bar{x}$,

$$-\mathcal{F}'(\bar{x}) \in \mathrm{Ker}(\mathcal{B})^a = \mathrm{Range}(\mathcal{B}^*),$$

or, more explicitly, that there exists a Lagrange multiplier $\lambda \in Y^*$ such that

$$-\mathcal{F}'(\bar{x}) = \mathcal{B}^* \lambda. \tag{4.13}$$

Remark 11. An important example for the use of this theorem is e.g. the role of the pressure gradient in relation with the divergence-free constraint in fluid dynamics. If one considers a sufficiently smooth domain Ω and the (negative) divergence operator $-\mathrm{div}$ as an operator from $\boldsymbol{H}^1_0(\Omega) = \left(H^1_0(\Omega)\right)^n$ into $L^2(\Omega)$, then it is actually a continuous operator between these

[11] Here this reduces to: For any sequence $x_n \in D(\mathcal{B})$ such that $x_n \to x \in X$ and $y_n := \mathcal{B}x_n \to y \in Y$, it must hold that $x \in D(\mathcal{B})$ and $y = \mathcal{B}x$

spaces (since $|\sum_{i=1}^N \frac{\partial u}{\partial x_i}|^2 \leq n\sum_{i=1}^N |\frac{\partial u}{\partial x_i}|^2$ and thus $\|-\operatorname{div}(\boldsymbol{u})\|_{L^2(\Omega)} \leq \sqrt{n}\|\boldsymbol{u}\|_{H_0^1(\Omega)}$) and therefore defined on all of $\boldsymbol{H}_0^1(\Omega)$. By Schwarz's inequality, one futher has $\|-\operatorname{div}(u)q\|_{L^2(\Omega)} \leq \|-\operatorname{div}(u)\|_{L^2(\Omega)}\|q\|_{L^2(\Omega)}$ and thus the mapping $\boldsymbol{u} \mapsto \int_\Omega -\operatorname{div}(\boldsymbol{u})q\,\mathrm{d}\boldsymbol{x}$ is uniformly bounded for all q in $L^2(\Omega)$ with constant $c = \sqrt{n}\|q\|_{L^2(\Omega)}$. The domain of its adjoint $-\operatorname{grad} : L^2(\Omega) \to \boldsymbol{H}^{-1}(\Omega) = \boldsymbol{H}^{-1}(\Omega)$, where $\boldsymbol{H}^{-1}(\Omega) = \left(H^{-1}(\Omega)\right)^n$ is the dual of $\boldsymbol{H}_0^1(\Omega)$, will therefore be all of $L^2(\Omega)$. For smooth functions, Green's formula shows that

$$\int_\Omega -\operatorname{div}(\boldsymbol{u})q\,\mathrm{d}\boldsymbol{x} = \int_\Omega \boldsymbol{u}\cdot\nabla q\,\mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} q\boldsymbol{u}\cdot\boldsymbol{n}\,\mathrm{d}\boldsymbol{s} = \int_\Omega \boldsymbol{u}\cdot\nabla q\,\mathrm{d}\boldsymbol{x}$$

since $\boldsymbol{u}$ vanishes on the boundary and the adjoint is therefore an extension of the standard gradient operator[12]. It further follows that the "Green-type" formula

$$\langle \operatorname{grad} q, \boldsymbol{u}\rangle_{\boldsymbol{H}^{-1}(\Omega),H_0^1(\Omega)} = \int_\Omega (-\operatorname{div}\boldsymbol{u})q\,\mathrm{d}\boldsymbol{x}$$

holds, regardless of whether the duality pairing on the left may be written as an integral and whether an integration by parts may be justified or not.
The assumption for Theorem 2 can be shown to hold (see e.g. [73]) and thus $\operatorname{Ker}(\operatorname{div})^a = \operatorname{Range}(-\operatorname{grad}) = \operatorname{Range}(\operatorname{grad})$. The crucial implication in this case is that any functional $\boldsymbol{f} \in \boldsymbol{H}^{-1}(\Omega)$ which vanishes on all divergence-free functions in $H_0^1(\Omega)$ can be written as the gradient of a scalar function $p \in H_0^1(\Omega)$, i.e. $\langle \boldsymbol{f}, \boldsymbol{v}\rangle_{\boldsymbol{H}^{-1},\boldsymbol{H}_0^1(\Omega)} = 0$ for all $\boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega)$ implies that $\boldsymbol{f} = \nabla p$, $p \in L^2(\Omega)$.

◇

Remark 12. The definition of the adjoint and Theorem 2 are also, similar to Remark 4, the central "ingredients" for e.g. the very useful Fredholm alternative (see e.g. theorem 2.27 in [48]). ◇

Nonlinear Equality Constraints As in the finite-dimensional case, the theory developed for the case of linear equality constraints carries over to nonlinear equality constraints provided some regularity condition holds. This is summarized in the following result:

Theorem 3. *(Tangents and normals to a level set (**Graves-Lyusternik** theorem), thm. 5.35 [20])*
Let X and Y be Banach spaces and let $\mathcal{A}$ be given by

$$\mathcal{A} = \{u \in X : h(x) = 0\},$$

where the map $h : X \to Y$ is continuously differentiable near $x \in \mathcal{A}$. If $h'(x)$ is surjective, then $T_{\mathcal{A}}(x)$ and $N_{\mathcal{A}}(x)$ are the linear subspaces given by

$$T_{\mathcal{A}}(x) = \{\delta x \in X : \langle h'(x), \delta x\rangle = 0\}, \quad N_{\mathcal{A}}(x) = \left(h'(x)\right)^* Y^*$$

and $T_{\mathcal{A}}(x) = N_{\mathcal{A}}(x)^a$, $N_{\mathcal{A}}(x) = T_{\mathcal{A}}(x)^a$.

This is essentially the same conclusion as in the finite-dimensional case, i.e. if the derivative of the nonlinear mapping is surjective (recall the counterexample in Remark 6), the only admissible variations are the ones on which the linearization of the constraint vanishes, and the derivative of the function(al) vanishing on all these directions then implies that there is some multiplier such that

$$\mathcal{F}'(x) + \left(h'(x)\right)^* \lambda = 0.$$

[12] More precisely, even though the expression $\operatorname{grad} q$ above does not a priori make "classical" sense for $q \in L^2(\Omega)$, the image of q under the gradient operator is, by the construction of the adjoint, simply **defined** as the unique element in $\boldsymbol{g} \in \boldsymbol{H}^{-1}(\Omega)$ satisfying $\langle \boldsymbol{g}, \boldsymbol{u}\rangle_{\boldsymbol{H}^{-1}(\Omega),\boldsymbol{H}_0^1(\Omega)} = \int_\Omega q\big(-\operatorname{div}(\boldsymbol{u})\big)\,\mathrm{d}\boldsymbol{x} \forall \boldsymbol{u} \in \boldsymbol{H}_0^1(\Omega)$.

Remark 13. Note that the surjectivity assumption in Theorem 3 implicitly subsumes what is required for Theorem 2. In fact, by the assumption that h is C^1, h' is defined everywhere on X (and not just densely). Furthermore, since $h'(x)$ is assumed surjective, the graph of $h'(x)$ is simply the product space $X \times Y$, and therefore obviously a closed subset of $X \times Y$ (its complement being the empty set, which is always open). Similarly, $\mathrm{Range}(h'(x))$ equals the whole space Y and is therefore again closed, from which the remaining conclusions follow immediately. $\diamond$

4.3 Linear and Nonlinear Inequality Constraints

In many applications, one not only has to deal with equality constraints but with possible additional inequality constraints. The primary example in the phasefield context is of course the box-constraints $0 \leq \phi \leq 1$ that may be imposed on the phasefield, or, in combination with an equality constraint in the multi-phase case, the restriction to the Gibbs-simplex, $0 \leq \phi^\alpha \leq 1$, $\sum_{\alpha=1}^N \phi^\alpha = 1$. Another physically important example is given by elasto-plastic applications, in which the admissible stresses are assumed to be restricted by a yield criterion of the form $f(\boldsymbol{\sigma}) \leq \sigma^y$, with σ^y possibly depending on additional internal parameters.

In contrast to the equality constrained case, which leads to first-order optimality conditions in terms of orthogonality to certain linear subspaces as in Equations (4.3) and (4.11), the Euler-Lagrange equations for inequality constrained problems lead to first-order conditions in the form of inequalities. This in turn leads to a characterization involving cones instead of subspaces. For this reason, it is necessary to find a suitable replacement for the fundamental theorem of linear algebra (resp. the closed graph theorem), here primarily in the form of Farka's lemma.

As for the equality constraints, the first case considered will be the finite-dimensional one in combination with linear inequality constraints, a setting which already contains the essential ideas. It will then shorty be sketched how these can be extended to some more complex situations.

Linear Inequality Constraints in the Finite-Dimensional Case Consider again the minimization of a smooth objective function f defined on $\boldsymbol{R}^n$, but now subjected to $m \leq n$ linear inequality constraints:

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{b}_i \cdot \boldsymbol{x} \leq c_i, \ i = 1, ..., m. \end{cases} \tag{4.14}$$

Just as for the equality constrained case, the FONC for a local minimizer $\bar{\boldsymbol{x}}$ satisfying the constraints is given by the Euler-Lagrange equation

$$\nabla f(\bar{\boldsymbol{x}}) \cdot \delta \boldsymbol{x} \geq 0 \tag{4.15}$$

with $\delta \boldsymbol{x}$ ranging over the "admissible" directions. The fundamental difference with respect to the previous situation lies precisely in the form of this set, which is now not simply given by the kernel of the matrix $\boldsymbol{B}$ composed of the $\boldsymbol{b}_i$. Instead, it is clear that there is actually more freedom in choosing the $\delta \boldsymbol{x}$ as $\mathrm{Ker}(\boldsymbol{B})$ is now only a subset of the admissible variations[13]. In fact, for each inequality constraint one has to distinguish two possible situations:

1. $\bar{\boldsymbol{x}}$ satisfies $\boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} < c_i$, i.e. the i-th constraint is **inactive**. By continuity, the same will hold for any $\boldsymbol{x}$ sufficiently close to $\bar{\boldsymbol{x}}$ or equivalently for any $\bar{\boldsymbol{x}} + \delta \boldsymbol{x}$ with $|\delta \boldsymbol{x}|$ sufficiently small. In this case, at least locally, the constraint effectively does not impose any actual restriction on the problem and can be disregarded.

2. $\bar{\boldsymbol{x}}$ satisfies $\boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} = c_i$, i.e. the i-th constraint is **active** or **saturated**. In this case, the only admissible variations are such that $\boldsymbol{b}_i \cdot \delta \boldsymbol{x} \leq 0$, which, instead of a hyperplane, now specifies an entire half-space.

[13] If $\bar{\boldsymbol{x}}$ satisfies $\boldsymbol{B}\bar{\boldsymbol{x}} \leq \boldsymbol{c}$, $\bar{\boldsymbol{x}} + \delta \boldsymbol{x}$ obviously also satisfies $\boldsymbol{B}(\bar{\boldsymbol{x}} + \delta \boldsymbol{x}) = \boldsymbol{B}\bar{\boldsymbol{x}} + \boldsymbol{B}\delta \boldsymbol{x} = \boldsymbol{B}\bar{\boldsymbol{x}} \leq \boldsymbol{c}$ for any $\delta \boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$.

Partitioning the constraints into the inactive ones,

$$\mathcal{I}_I(\bar{\boldsymbol{x}}) = \{i \in 1, ..., n : \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} < c_i\} \tag{4.16}$$

and the active ones,

$$\mathcal{I}_A(\bar{\boldsymbol{x}}) = \{i \in 1, ..., n : \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} = c_i\}, \tag{4.17}$$

the set of admissible variations is therefore given by the set $T_K(\bar{\boldsymbol{x}})$ of $\delta\boldsymbol{x}$ such that

$$T_K(\bar{\boldsymbol{x}}) = \{\delta\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{b}_i \cdot \delta\boldsymbol{x} \leq 0 \quad \forall i \in \mathcal{I}_A(\bar{\boldsymbol{x}})\},$$

i.e. an intersection of (closed) half-spaces.

Definition 3. (Cones and polar cones)

- A **cone** (also sometimes called a **pointed cone**) is any subset C such that, if $\boldsymbol{x} \in C$, then $\alpha\boldsymbol{x} \in C$ for all $\alpha \geq 0$.
- Given any subset M of $\mathbb{R}^n$, the **(negative) polar cone** M^- is the set of vectors $\boldsymbol{y}$ such that

  $$\boldsymbol{y} \cdot \boldsymbol{x} \leq 0 \quad \forall \boldsymbol{x} \in M.$$

  Similarly, its negative, the **dual cone** (or **positive polar cone**) is the set $M^+ = -M^-$ defined by

  $$\boldsymbol{y} \cdot \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in M.$$

With this definition, it is clear[14] that the admissible variations $T_K(\bar{\boldsymbol{x}})$ form a cone.

The optimality condition (4.15) corresponds, by the very definitions, to $f'(\bar{\boldsymbol{x}}) \in T_K^+$ or, using the more common notation, to

$$-\nabla f(\bar{\boldsymbol{x}}) \in T_K^-(\bar{\boldsymbol{x}}), \tag{4.18}$$

which is thus the conclusion replacing $\nabla f(\bar{\boldsymbol{x}}) \in \mathrm{Ker}(\boldsymbol{B})^\perp$ from the equality-constrained case. In order to obtain a more explicit formulation, it remains to replace the characterization $\mathrm{Ker}(\boldsymbol{B})^\perp = \mathrm{Range}(\boldsymbol{B}^T)$ by an analogous statement describing $T_K^-(\bar{\boldsymbol{x}})$. This is the purpose of the following

Lemma 1. ***(Farka's lemma**, cor. 2.29 [58])*
Let $\boldsymbol{B}$ be an $m \times n$-matrix and let

$$K = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{0}\}.$$

Then (see also Figure 4.2a)

$$K^- = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y} = \boldsymbol{B}^T\boldsymbol{\mu},\ \boldsymbol{\mu} \in \mathbb{R}^m,\ \boldsymbol{\mu} \geq \boldsymbol{0}\}. \tag{4.19}$$

Proof. As for the fundamental theorem of linear algebra, one implications is a simple consequence of the definition of the transposed matrix[15]. In fact, denoting the set on the right-hand side of (4.19) by $\tilde{K}$, one has

$$(\boldsymbol{x}, \boldsymbol{B}^T\boldsymbol{\mu})_{\mathbb{R}^m} = (\boldsymbol{B}\boldsymbol{x}, \boldsymbol{\mu})_{\mathbb{R}^n} \leq 0 \quad \forall \boldsymbol{x} \in K, \boldsymbol{y} \in \tilde{K},$$

as $\boldsymbol{B}\boldsymbol{x}$ has only non-positve and $\boldsymbol{\mu}$ only non-negative entries by the definitions of K and $\tilde{K}$, i.e. $\tilde{K} \subset K^-$. For the other implication, observe that $\tilde{K}$ is a closed convex cone. If there is a $\boldsymbol{z} \in K^-$

[14] Either by "geometrical insight" or by noting that the defining condition $\boldsymbol{b}_i \cdot \delta\boldsymbol{x} \leq 0$ is stable under multiplication by non-negative scalars.

[15] There it is the conclusion that $\mathrm{Range}(\boldsymbol{B}^T)\boldsymbol{\mu} \subset \mathrm{Ker}(\boldsymbol{B})^\perp$, as for any $\boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{B})$ $\boldsymbol{x} \cdot \boldsymbol{B}^T\boldsymbol{\mu} = \boldsymbol{B}\boldsymbol{x} \cdot \boldsymbol{\mu} = \boldsymbol{0}$. In contrast, the analoguous statement to the other conclusion $\mathrm{Ker}(\boldsymbol{B})^\perp \subset \mathrm{Range}(\boldsymbol{B}^T)$ is trickier as it requires a separation property similar to the one used here.

such that $\boldsymbol{z} \notin \tilde{K}$, $\boldsymbol{z}$ can therefore be strictly separated from $\tilde{K}$, i.e. there is some vector $\boldsymbol{l} \in \mathbb{R}^n$ such that

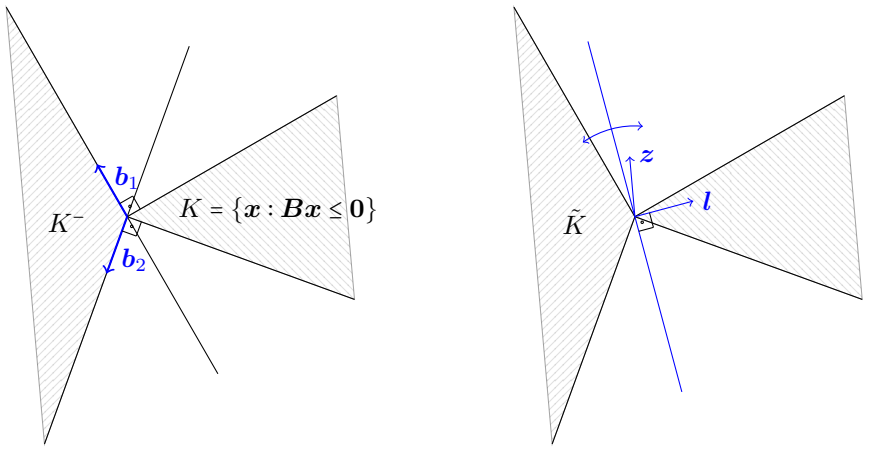
$$\boldsymbol{l} \cdot \boldsymbol{z} > \alpha \geq \boldsymbol{l} \cdot \boldsymbol{y} \quad \forall \boldsymbol{y} \in \tilde{K}.$$

As each $\boldsymbol{y} \in \tilde{K}$ is of the form $\boldsymbol{B}^T \boldsymbol{\mu}$ with $\boldsymbol{\mu} \geq \boldsymbol{0}$, one thus has

$$\boldsymbol{B}\boldsymbol{l} \cdot \boldsymbol{\mu} \leq \alpha \quad \forall \boldsymbol{\mu} \geq \boldsymbol{0},$$

which is only possible if $\boldsymbol{B}\boldsymbol{l} \leq \boldsymbol{0}$ (otherwise just let the entry in $\boldsymbol{\mu}$ corresponding to a positive entry in $\boldsymbol{B}\boldsymbol{l}$ tend to $+\infty$). In addition, as $\boldsymbol{\mu} = \boldsymbol{0}$ is admissible, this implies $\alpha \geq 0$. Since $\boldsymbol{B}\boldsymbol{l} \leq \boldsymbol{0}$ is just the same as saying $\boldsymbol{l} \in K$, one would therefore have $\boldsymbol{z} \cdot \boldsymbol{l} > \alpha \geq 0$, but this contradicts $\boldsymbol{z} \in K^-$. □

Remark 14. While the proof of Lemma 1 may seem somewhat technical through the use of a separation theorem, the geometric idea is actually rather simple. The separation theorem and the first part of the theorem imply that there is a half-space containing $\tilde{K} \subset K^-$ and, in addition, the existence of a vector in the **other** half-space, which, just as all elements of $\tilde{K}$, forms an angle of at least 90° with any vector in K. As $\tilde{K}$ always contains the vectors $\boldsymbol{b}_i$, i.e. the ones which are precisely orthogonal to one of the subspaces defining K, there is no way of arranging a half-plane such that it contains both $\tilde{K}$ and that a vector on the other side does not make an acute angle with K (see Figure 4.2b). ◇



(a) (b)

Figure 4.2: The (negative) polar cone in Farkas lemma (a) and the geometrical idea underlying its proof (b).

Applying the lemma to $-\nabla f(\bar{\boldsymbol{x}})$ with the matrix $\boldsymbol{B}_{\mathcal{I}_A(\bar{\boldsymbol{x}})}$ constructed (row-wise) from the vectors $\boldsymbol{b}_i^T$ with $i \in \mathcal{I}_A(\bar{\boldsymbol{x}})$ replacing $\boldsymbol{B}$ leads to

$$-\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{B}_{\mathcal{I}_A(\bar{\boldsymbol{x}})}^T \boldsymbol{\mu}_{\mathcal{I}_A(\bar{\boldsymbol{x}})} = \sum_{i \in \mathcal{I}_A(\bar{\boldsymbol{x}})} \mu_i \boldsymbol{b}_i,$$

where each of the μ_i appearing above has to be non-negative, $\mu_i \geq 0, i \in \mathcal{I}_A(\bar{\boldsymbol{x}})$.

As the notation involving the index set $\mathcal{I}_A(\bar{\boldsymbol{x}})$ is somewhat cumbersome and only meaningful in the discrete case, it is common to use an equivalent formulation which is obtained by noting that $\boldsymbol{B}_{\mathcal{I}_A(\bar{\boldsymbol{x}})}^T \boldsymbol{\mu}_{\mathcal{I}_A(\bar{\boldsymbol{x}})} = \boldsymbol{B}^T \boldsymbol{\mu}$, with $\boldsymbol{B}$ the analogue of $\boldsymbol{B}_{\mathcal{I}_A(\bar{\boldsymbol{x}})}$ but containing all the vectors $\boldsymbol{b}_i$, provided all $\mu_i, i \in \mathcal{I}_I(\bar{\boldsymbol{x}})$ are set to zero. These are precisely the indices though for which $c_i - \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} \geq 0$, whereas the indices in $\mathcal{I}_A(\bar{\boldsymbol{x}})$, i.e. the set where the μ_i may be non-zero, are given by

$c_i - \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}} = 0$. Both conditions can compactly be combined by introducing the **complementarity conditions** $\mu_i(c_i - \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}}) = 0$, $i = 1, ..., m$. Finally, as both μ_i and $c_i - \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}}$ are required to be non-negative and thus the sum $\sum_{i=1}^{m} \mu_i(c_i - \boldsymbol{b}_i \cdot \bar{\boldsymbol{x}})$ can only vanish if all terms are zero, the characterization of $\nabla f(\bar{\boldsymbol{x}})$ can equivalently be expressed by

$$\begin{cases} -\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{B}^T \boldsymbol{\mu}, \\ \boldsymbol{c} - \boldsymbol{B}\bar{\boldsymbol{x}} \geq \boldsymbol{0}, \\ \boldsymbol{\mu} \geq \boldsymbol{0}, \\ \boldsymbol{\mu} \cdot (\boldsymbol{c} - \boldsymbol{B}\bar{\boldsymbol{x}}) = 0. \end{cases} \tag{4.20}$$

Nonlinear Inequality Constraints in the Finite-Dimensional Case As for the equality-constrained case, the conclusions for the linear inequalites can be carried over to the nonlinear case through a simple linearization of the (active) constraints provided some constraint qualification conditions hold. More precisely, for a constraint-set $\mathcal{A} = \{\boldsymbol{x} : g_i(\boldsymbol{x}) \leq 0, i = 1, ..., m\}$ prescribed through $m \leq n$ inequality constraint, the first-order necessary condition is, as in Equation (4.8), that $\nabla f(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} \geq 0$ for all $\delta\boldsymbol{x} \in T_{\mathcal{A}}$ i.e. $-\nabla f(\bar{\boldsymbol{x}}) \in T_{\mathcal{A}}(\bar{\boldsymbol{x}})^-$ where the Boulignand cone $T_{\mathcal{A}}(\bar{\boldsymbol{x}})$ is defined as in Equation (4.7). What one would like to conclude from this is that any minimizer $\bar{\boldsymbol{x}}$ subject to $g(\boldsymbol{x}) \leq \boldsymbol{0}$ is of the form

$$-\nabla f(\bar{\boldsymbol{x}}) = \sum_{i \in \mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})} \mu_i \nabla g_i(\bar{\boldsymbol{x}}) \tag{4.21}$$

with $\mu_i \geq 0$, where $\mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})$ denotes the set of active constraints, i.e. the indices for with $g_i(\bar{\boldsymbol{x}}) = 0$. All this requires is to show that the Bouligand cone appearing in the characterization of $\nabla f(\bar{\boldsymbol{x}})$ for any local minimizer is the same as the linearizing cone

$$L(\bar{\boldsymbol{x}}, K) := \{\delta\boldsymbol{x} : \nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} \leq 0, i \in \mathcal{I}_{\mathcal{A}}\}$$

consisting of those directions which, in first order, do not lead to an increase in the values of the g_i which are at their maximal admissible value 0 at $\bar{\boldsymbol{x}}$. If such is the case, the conclusion follows immediately from Farka's lemma, as one then has $T_{\mathcal{A}}(\bar{\boldsymbol{x}}) = \{\delta\boldsymbol{x} : \boldsymbol{B}\delta\boldsymbol{x} \leq \boldsymbol{0}\}$ with $\boldsymbol{B} = \left(\nabla \boldsymbol{g}_{\mathcal{I}_{\mathcal{A}}}(\bar{\boldsymbol{x}})\right)^T$ and by Lemma 1 $T_{\mathcal{A}}(\bar{\boldsymbol{x}})^- = \{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{B}^T\boldsymbol{\mu}, \boldsymbol{\mu} \geq \boldsymbol{0}\} = \{\boldsymbol{y} : \boldsymbol{y} = \nabla \boldsymbol{g}_{\mathcal{I}_{\mathcal{A}}}(\bar{\boldsymbol{x}})\boldsymbol{\mu}, \boldsymbol{\mu} \geq \boldsymbol{0}\}$ and thus together with the first order necessary condition in Equation (4.21).

Linking the two cones is again a question of being able to ensure that, any $\delta\boldsymbol{x}$ in $L(\bar{\boldsymbol{x}}, K)$ can actually be obtained as the limit of the directions showing towards points actually lying in $\mathcal{A}$. This will hold if for any for any $\delta\boldsymbol{x}$ such that $\nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} \leq 0$ for all $i \in \mathcal{I}_{\mathcal{A}}$, one can find an actually admissible point $\boldsymbol{x}$ in $\mathcal{A}$ such that $\delta\boldsymbol{x} = \boldsymbol{x} - \bar{\boldsymbol{x}} + o(\|\delta\boldsymbol{x}\|)$ and is therefore again a question of stability.

There are a variety of different conditions ensuring this (see e.g. [58], [13] or [27] for a more detailed discussion). A relatively strong condition ensuring this is the linear independence constraint qualification condition, requiring that the gradients $\nabla g_i(\bar{\boldsymbol{x}})$ for $i \in \mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})$ be linearly independent. A weaker condition also based on the gradients themselves is the Mangasarian-Fromovitz constraint qualification condition, requiring that there be some vector $\boldsymbol{d}$ such that $\nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} < 0$ for all $i \in \mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})$. Another popular condition, which in addition applies even when the g_i are not necessarily differentiable, is Slater's constraint qualification condition, which requires for the g_i, $i \in \mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})$ to be convex functions together with the existence of some point $\boldsymbol{y}$ such that $\boldsymbol{g}_{\mathcal{I}_{\mathcal{A}}(\bar{\boldsymbol{x}})}(\boldsymbol{y}) < \boldsymbol{0}$.

Remark 15. Note that, with respect to any active constraint i, this is not an issue for a direction $\delta\boldsymbol{x}$ such that $\nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} < 0$ as the differentiability of g_i automatically implies that $g_i(\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}) < 0$ for t sufficiently small, i.e. the constraint will definitely be satisfied as $\bar{\boldsymbol{x}}$ is approached along this direction. The only potential difficulties therefore arises for those $\delta\boldsymbol{x}$ for which $\nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} = 0$. As for such directions, one always has (regardless of any qualification condition) $g_i(\bar{\boldsymbol{x}} + t\delta\boldsymbol{x}) =$

$g_i(\bar{\boldsymbol{x}}) + t\nabla g_i(\bar{\boldsymbol{x}}) \cdot \delta\boldsymbol{x} + o(t) = o(t)$, one can ensure feasibility of the a very nearby (i.e. $o(t)$ close) point by a "tiny nudge" in any direction such that $\nabla g_i(\bar{\boldsymbol{x}}) \cdot \boldsymbol{d} < 0$, provided this does not lead to a violation of the remaining active inequality constraints. The basic idea underlying the conditions above is to ensure the existence of a direction $\boldsymbol{d}$ such that this is always possible. ◇

Remark 16. The results above can be generalized quite significantly in various directions. Firstly, one can replace the condition $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$ with more abstract conditions. Secondly, many of the ideas also carry over to the infinite-dimensional setting under some additional topological assumptions, with the adjoint again replacing the transpose matrix. This will not be discussed in detail here, and the reader is referred to e.g. [37] and [14] for a discussion in relation with Lagrange multipliers or e.g. [22] for some discussion on generalization of Farka's lemma above. A very well-written introduction to some closely related background can also be found in [7], in particular concerning various extensions of classical results from functional analysis to situations involving convex cones instead of the entire spaces.

◇

Chapter 5

Lagrange Functions and Legendre-Fenchel Duality

5.1 Lagrange Functions

The results for the minimization of $f(\boldsymbol{x})$ subject to the equality-constraint $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{c}$ and $g(\boldsymbol{x}) \leq \boldsymbol{0}$ from Section 4.1 can conveniently be combined and summarized by introducing the **Lagrange-function** or **Lagrangian**

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda} \cdot \big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x}). \tag{5.1}$$

In fact, it can be observed that, provided one restrict $\boldsymbol{\mu}$ to be non-negative, simply imposing that the gradient $\nabla_{\boldsymbol{x}}$ of L with respect to $\boldsymbol{x}$ should vanish,

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla f(\boldsymbol{x}) + \nabla \boldsymbol{h}(\boldsymbol{x})\boldsymbol{\lambda} + \nabla \boldsymbol{g}(\boldsymbol{x})\boldsymbol{\mu} \stackrel{!}{=} \boldsymbol{0}, \tag{5.2}$$

one recovers the same form of relation between the gradient of f and the multipliers which was obtained in the previous chapter. In addition, again demanding that the gradient of L with respect to $\boldsymbol{\lambda}$ should vanish,

$$\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c} \stackrel{!}{=} \boldsymbol{0}, \tag{5.3}$$

one recovers the equality constraint. Finally, imposing that the $\nabla_{\boldsymbol{\mu}} L$ should be non-positive leads to

$$\nabla_{\boldsymbol{\mu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{g}(\boldsymbol{x}) \stackrel{!}{\leq} \boldsymbol{0} \tag{5.4}$$

and therefore the original inequality constraint on $\boldsymbol{g}(\boldsymbol{x})$.
Even though this has, up to this point, no deeper meaning, it is obvious that provided that this is a very convenient "mnemonic" for obtaining the correct structure of the first-order optimality conditions derived to a relatively lengthy argument based on orthogonaly relations resp. relations between cones and their polar cones in the previous section. As will be sketched below, there is in fact a close relation between Lagrangian and the original constrained optimization problem and the usefulness of the Lagrangian goes far beyond providing a simple means of "guessing" the correct optimality conditions.

The basis for this approach is contained in the following observations: Whenever the equality constraint is satisfied for some $\boldsymbol{x}$, the term in $\boldsymbol{\lambda}$ vanishes. Since L does then not depend on $\boldsymbol{\lambda}$ for such an $\boldsymbol{x}$ and taking the supremum over all $\boldsymbol{\lambda}$ thus has no effect, $\sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x})$. Secondly, a similar argument can be done in terms of $\boldsymbol{\mu}$ and $\boldsymbol{g}(\boldsymbol{x})$. If $\boldsymbol{x}$ is such that $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$, the term $\boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x})$ is necessarily less than or equal to zero. If some entry $g_i(\boldsymbol{x})$ is strictly negative, the supremum over all $\mu_i \geq 0$ is obviously achieved for $\mu_i = 0$. In

constrast, if $g_i(\boldsymbol{x}) = 0$, L does not depend on μ_i and any μ_i will therefore do for achieving the supremum $\sup_{\mu_i \geq 0} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$. Both cases together therefore show that if $\boldsymbol{x}$ satisfies the inequality constraint, taking the supremum over the $\boldsymbol{\mu} \geq \mathbf{0}$ selects $\boldsymbol{\mu}$ such that the second part $\boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x}) = 0$ of the complementarity conditions holds and, if this is the case, also eliminates the last term of the Lagrangain in Equation (5.1).
Whenever $\boldsymbol{x}$ satisfies both constraints, one therefore has

$$\sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}). \tag{5.5}$$

In contrast, when the equality constraint is not satisfied and thus $\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c} \neq \mathbf{0}$ there exists some $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that $\boldsymbol{\lambda} \cdot \big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) > 0$, and thus by "scaling" $\boldsymbol{\lambda}^*$, one has $\sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = +\infty$. Similarly, if any entry of $\boldsymbol{g}(\boldsymbol{x})$ is greater than 0, it suffices to let the corresponding component μ_i tend to $+\infty$ to show that $\sup_{\boldsymbol{\mu} \geq 0} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is then also $+\infty$.
In summary, taking the supremum in $\boldsymbol{\lambda}$ and $\boldsymbol{\mu} \geq \mathbf{0}$ to define the **primal function** $L_P(\boldsymbol{x})$, this function satisfies

$$L_P(\boldsymbol{x}) := \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} f(\boldsymbol{x}) & \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c} = \mathbf{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \\ +\infty & \text{else}, \end{cases} \tag{5.6}$$

and, since the infinite values are obviously of no interest with respect to a minimization, it follows further that the original problem of minimizing $f(\boldsymbol{x})$ subject to the constraints can equivalenty be expressed in terms of the **primal problem**

$$\inf_{\boldsymbol{x} \in \boldsymbol{R}^n} L_P(\boldsymbol{x}) = \inf_{\boldsymbol{x} \in \boldsymbol{R}^n} \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{5.7}$$

While this minimization problem is now in principle a free minimization problem in $\boldsymbol{x}$, this is an essentially formal difference, which, in this form, is hard to put to any practical use. The idea is instead to look at an - a priori different - problem obtained by exchanging the order of the inf and sup, i.e. by instead considering the **dual problem**

$$\sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{5.8}$$

where the **dual function** is defined by

$$L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} \{f(\boldsymbol{x}) + \boldsymbol{\lambda} \cdot \big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \boldsymbol{\mu} \cdot \boldsymbol{g}(\boldsymbol{x})\}. \tag{5.9}$$

Remark 17. A first point to be noted is that the **dual problem**, while still a problem potentially subject to the constraint $\boldsymbol{\mu} \geq \mathbf{0}$ has, at least in terms of $L_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$, a significantly simpler structure than the original minimization problem for $f(\boldsymbol{x})$. In fact, whereas $\boldsymbol{x}$ is potentially subject two both a nonlinear equality constraint and a nonlinear inequality constraint, there is no constraint on $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ is solely subject to the simple constraint $\boldsymbol{\mu} \geq \mathbf{0}$.
The latter observation is of course only really advantageous if the inequality constraints on the original variable $\boldsymbol{x}$ are of a more complex form. This is for example not the case for the restriction of the phasefield to the Gibbs-simplex which will be discussed in more detail in Chapter 6, since the inequality restrictions on the primal unknown are also of the form $\boldsymbol{\phi} \geq \mathbf{0}$. Nevertheless, the practical difficulty of the restriction to the Gibbs-simplex is not primarily due to the inequality constraints themselves (this could be handled using a simple truncation), but their coupling through an additional sum-constraint, an aspect which never occurs in a dual problem. ◇

A further indication why the dual problem may be useful is given by the following two simple observations, which do not require any assumptions on $f(\boldsymbol{x})$, $\boldsymbol{h}(\boldsymbol{x})$ or $\boldsymbol{g}(\boldsymbol{x})$:

Lemma 2. *(see e.g. chapter 4 [58])*
The dual function is concave and one always has

$$\sup_{\{(\boldsymbol{\lambda},\boldsymbol{\mu}):\boldsymbol{\mu}\geq\mathbf{0}\}} L_D(\boldsymbol{\lambda},\boldsymbol{\mu}) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} L_P(\boldsymbol{x}). \tag{5.10}$$

Proof. Firstly, $f(\boldsymbol{x}) + \boldsymbol{\lambda}\cdot\big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \boldsymbol{\mu}\cdot\boldsymbol{g}(\boldsymbol{x})$ is linear and thus also (even though "border-line") concave in both $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. As the infimum of an arbitrary family of concave functions is concave, the first claim follows.
By the definition of the supremum, it is also clear that

$$L_P(\boldsymbol{x}) = \sup_{\{(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}}):\tilde{\boldsymbol{\mu}}\geq\mathbf{0}\}} \{f(\boldsymbol{x}) + \tilde{\boldsymbol{\lambda}}\cdot\big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \tilde{\boldsymbol{\mu}}\cdot\boldsymbol{g}(\boldsymbol{x})\} \geq f(\boldsymbol{x}) + \boldsymbol{\lambda}\cdot\big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \boldsymbol{\mu}\cdot\boldsymbol{g}(\boldsymbol{x})$$

regardless of the choice of $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}\geq\mathbf{0}$. Taking the infimum on both sides then shows that

$$\inf_{\boldsymbol{x}\in\mathbb{R}^n} L_P(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}\in\mathbb{R}^n} \{f(\boldsymbol{x}) + \boldsymbol{\lambda}\cdot\big(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c}\big) + \boldsymbol{\mu}\cdot\boldsymbol{g}(\boldsymbol{x})\} = L_D(\boldsymbol{\lambda},\boldsymbol{\mu}),$$

for every $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}\geq\mathbf{0}$, and thus also $\inf_{\boldsymbol{x}\in\mathbb{R}^n} L_D(\boldsymbol{x}) \geq \sup_{\{(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\mu}}):\tilde{\boldsymbol{\mu}}\geq\mathbf{0}\}} L_D(\boldsymbol{\lambda},\boldsymbol{\mu})$. □

Remark 18. Provided one is able to derive an expression for L_D, one is therefore in the very favorable setting of having to maximize a concave function subject to at most a simple non-negativity constraint. In addition, even if equality does not hold in (5.10), the solution of the dual problem does always provide a lower bound for that of the primal one. ◇

Remark 19. Assuming there is an actual relation with the primal problem, the dual problem also provides an intuitive explanation for the relevance of the three conditions in Equations (5.2), (5.3) and (5.4). Since the definition of $L_D(\boldsymbol{\lambda},\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$ in Equation (5.9) is based on a free minimization of $\boldsymbol{x}$ given $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, the derivative with respect to $\boldsymbol{x}$ should vanish. Similarly, as there is no restriction on $\boldsymbol{\lambda}$ in the dual problem (5.8), one might expect for the analogous condition to hold for $\boldsymbol{\lambda}$. Finally, with $\boldsymbol{\mu}$ being restricted to be non-negative, realizing the maximum of L_D with respect to $\boldsymbol{\mu}$ does not necessarily require for $\nabla_{\boldsymbol{\mu}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = \boldsymbol{g}(\boldsymbol{x})$ to be zero, but should there be any positive entry $g_i(\boldsymbol{x})$, one could hope to further increase L by increasing μ_i.
It should be kept in mind though that, despite its intuitive appeal at first sight, this simple argument hides a critical point, namely that even though the inner minimization in $\boldsymbol{x}$ is indeed a free one, the value of the minimizer will depend on the paramters $(\boldsymbol{\lambda},\boldsymbol{\mu})$, i.e. the maximization of the dual function L_D with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ is in fact based on the function

$$L_D(\boldsymbol{\lambda},\boldsymbol{\mu}) = L\big(\boldsymbol{x}(\boldsymbol{\lambda},\boldsymbol{\mu}),\boldsymbol{\lambda},\boldsymbol{\mu}\big).$$

That it still makes sense requires a reasoning similarly to Remark 3, namely that despite the dependence of $\boldsymbol{x}$ on the parameters, the optimality satisfied by $\boldsymbol{x}$ ensures that L itself is in first order not affected by this dependence as the contributions $\frac{\partial L}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\lambda}}$ and $\frac{\partial L}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\mu}}$ arising in the derivatives with respect to the multipliers drop out by the optimality condition on $\boldsymbol{x}$.
The fact that a "blind" differentiation of the Lagrangian - i.e. ignoring any potential interplay between the various variables - is therefore still expected to deliver the correct results is clearly a major simplification, and likely one of the main reasons of its popularity in the engineering community. The same type of underlying "variational consistency" is, as in Remark 3, also fundamental to the ubiquitous changes of unknowns in thermodynamics[1]. ◇

Even thought the lower bound in Lemma 2 can by itself be quite useful, the most desirable case is of course when both values actually do coincide. This motivates the following

[1] The phasefield method owing much of its early success to its successful application in a thermodynamically based setting where such dependencies can legitimately be "ignored", this is a point which unfortunaly seems to be partially forgotten in the phasefield community.

Definition 4. (Saddle point)
A pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ is called a saddle point of the function $L(\boldsymbol{x}, \boldsymbol{\lambda})$ if one has

$$L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \leq L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}). \tag{5.11}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \geq \mathbf{0}$.

The next result is then a simple consequence of the Inequality (5.10):

Theorem 4. *(thms. 4.8 and 4.9 [58])*
If the Lagrangian $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ has a saddle point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$, then $\bar{\boldsymbol{x}}$ is a solution of the primal problem, $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ is a solution of the dual problem, and one has

$$\max_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} L_P(\boldsymbol{x}).$$

Conversely, assume that this relation holds true with finite values on both sides. Then, for every solution $\bar{\boldsymbol{x}}$ of the primal problem and every solution $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ of the dual problem, the point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ is a saddle point of the Lagrangian.

Proof. If $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ is a saddle point of L, it follows from Equation (5.11) that one has

$$L_P(\bar{\boldsymbol{x}}) = \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \leq \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = L_D(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$$

and thus also

$$\inf_{\boldsymbol{x} \in \mathbb{R}^n} L_P(\boldsymbol{x}) \leq L_P(\bar{\boldsymbol{x}}) \leq L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \leq L_D(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \leq \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}).$$

As the opposite inequality $\inf_{\boldsymbol{x} \in \mathbb{R}^n} L_P(\boldsymbol{x}) \geq \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$ always holds as seen above, the inequalities necessarily hold as equalities,

$$\inf_{\boldsymbol{x} \in \mathbb{R}^n} L_P(\boldsymbol{x}) = L_P(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = L_D(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{5.12}$$

Thus both the primal and dual problem do admit at least one solution $\bar{\boldsymbol{x}}$ and $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ and their values coincide, proving the first claim.

Conversely, if one has $\sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} L_P(\boldsymbol{x})$ and $\bar{\boldsymbol{x}}$ resp. $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ are solutions to the primal resp. dual problem, one clearly has

$$L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = L_P(\bar{\boldsymbol{x}}) = L_D(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \leq L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ with $\boldsymbol{\mu} \geq \mathbf{0}$. Inserting $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ into the left-most expression and $\bar{\boldsymbol{x}}$ into the right-most one, it follows in addition that $L_P(\bar{\boldsymbol{x}}) = L_D(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$. □

Remark 20. Note that this theorem does not assert the existence of a saddle point[2]. Often, conditions assuring this are examined sligthly more indirectly in terms of the **duality gap**, i.e. the difference between the inf and the sup obtained from the primal and dual problem[3] or in relation with the Legendre-Fenchel transform introduced in the next section (see e.g. [26] and [14] for some links between the two approaches to duality). ◇

[2] A fairly general existence theorem is e.g. given in [26].

[3] This of course amounts to the same as there being a saddle point is then clearly the same thing as this difference vanishing.

Nevertheless, it indicates one reason for the usefulness of the Lagrangian. Whereas the primal problem is by construction equivalent to the original minimization problem, this reformulation essentially has no practical use by itself. More precisely, if the constraints are satisfied for some $\boldsymbol{x}$, the values of the multipliers are, beyond the complementarity conditions for $\boldsymbol{\mu}$ essentially arbitrary. If the constraints are not satisfied, the inner supremum results in the value $+\infty$, which one does certainly not want to achieve in any practical minimization algorithm. It may thus not seem like a very reasonable idea to move the multipliers into the corresponding direction. In contrast, the statement implies that if there is a saddle point and one can find a solution $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ of the dual problem, one has, at least, made the solution of the primal problem significantly simpler since it then reduces to a free minimization of $L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ for the given values of the multipliers. In addition, provided this calculation can be justified through a sufficiently smooth dependce of the minimizer $\boldsymbol{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ given some current estimate $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ in the definition of the dual function L_D in Equation (5.9), one has $\frac{\partial L}{\partial \boldsymbol{x}} = \boldsymbol{0}$ (as $\boldsymbol{x}$ is a free minimizer for the given multipliers), and the **total** derivatives of L with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ therefore reduce to the partial ones despite this implicit dependence. It follows that

$$\nabla_{\boldsymbol{\lambda}} L\big(\boldsymbol{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu}\big) = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{c} \quad \text{and} \quad \nabla_{\boldsymbol{\mu}} L\big(\boldsymbol{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu}\big) = \boldsymbol{g}(\boldsymbol{x}).$$

This motivates for example various alternative descent/ascent methods in terms of the two sets of unknowns $\boldsymbol{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, since, provided $\boldsymbol{x}$ is at least an approximate minimizer for the given multipliers, adjusting them such as to increase L "as if" $\boldsymbol{x}$ were fixed therefore does indeed make sense.

Remark 21. Even though often not primarily motivated by such considerations, the Lagrangian point of view also provides additional insight into e.g. some popular projection-based algoritms such as e.g. fractional-step algorithms in fluid dynamics and the return mapping algorithm in elasto-plasticity. In addition, the Lagrangian is also a very helpful ingredient for Newton-type schemes for constrained problems and the analysis of second-order necessary conditions for constrained optimization problems (see e.g [58], [13] and [46] for a more detailed discussion). ◇

Remark 22. It should be noted that the discussion above is for the most part completely independent of whether the underlying spaces are finite-dimensional or not. In particular, the definitions and lemmas can essentially be applied verbatim to the case when function spaces are involved, since e.g. the proofs of the concavity and the inequalities in Equation (5.10) in Lemma 2 as well as in Theorem 4 are all based purely on relations implied by the inf- and sup-operation (for Lemma 2 combined with the concavity of linear operators), which are completely indepedendent of any particularly favorable properties of $\boldsymbol{R}^n$ and neither rely on any particular structure of the sets involved in the inf-sup-operations. For the corresponding definitions and proofs in this more general setting, a classical and very readable reference is [26]. ◇

Remark 23. The dual function $L_D(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is an example of a more general kind of function frequently arising in parametric optimization involving a function $\varphi(\boldsymbol{x}, \boldsymbol{u})$ of the variable $\boldsymbol{x}$ and the "parameter" $\boldsymbol{u}$. Based on $\varphi(\boldsymbol{x}, \boldsymbol{u})$, one can introduce the (optimal) value function $v(\boldsymbol{u}) := \inf_{\boldsymbol{x} \in \mathbb{R}^n} \varphi(\boldsymbol{x}, \boldsymbol{u})$ corresponding to a minimizer (if any) of φ given the value of the parameter $\boldsymbol{u}$. More generally, one might also restrict the domain of $\boldsymbol{x}$ in the minimization process just to subsets the underlying space, i.e. by setting $v(\boldsymbol{u}) := \inf_{\boldsymbol{x} \in X} \varphi(\boldsymbol{x}, \boldsymbol{u})$, or even to subset $X(\boldsymbol{u})$ depending themselves on the parameter.
It is clear that the study of the behavior of $v(\boldsymbol{u})$ in terms of $\boldsymbol{u}$ (e.g. regularity or differentiability properties), has an inherent interest beyond the Lagrangian setting above, but is in this particular setting also very instructive in terms of an additional interpretation of the Lagrange-multipliers as "sensitivities" of the objective functions with respect to the constraints. For further discussions in the finite-dimensional setting, the reader is e.g. referred to [46], [13], [58] and [27]. For the infinite-dimensional setting in function spaces, the reader may consult [26] or for a more detailed but also more technical discussion in a quite general setting [14]. ◇

5.2 The Legendre-Transform

An a priori somewhat different but in fact closely related approach to duality theory is provided through the **Legendre-transform**, which is in particular heavily relied upon in the thermodynamic setting. In its most basic form, the Legendre transform (or **conjugate**) of a function $f : \mathbb{R} \to (-\infty, +\infty]$ is a second function $f^* : \mathbb{R} \to (-\infty, +\infty]$ defined as[4]

$$f^*(x^*) := \sup_{x \in \mathbb{R}} \{x^* x - f(x)\}, \tag{5.13}$$

i.e. the function f^* which, for an arbitrary "slope" x^*, assign to f the maximal difference between the line $y(x) = x^* x$ and the graph of f. One can reiterate this procedure on $f^*(x^*)$ by defining the **bi-conjugate** (**bi-dual**) function

$$f^{**}(x) := \sup_{x^* \in \mathbb{R}} \{x^* x - f^*(x^*)\}, \tag{5.14}$$

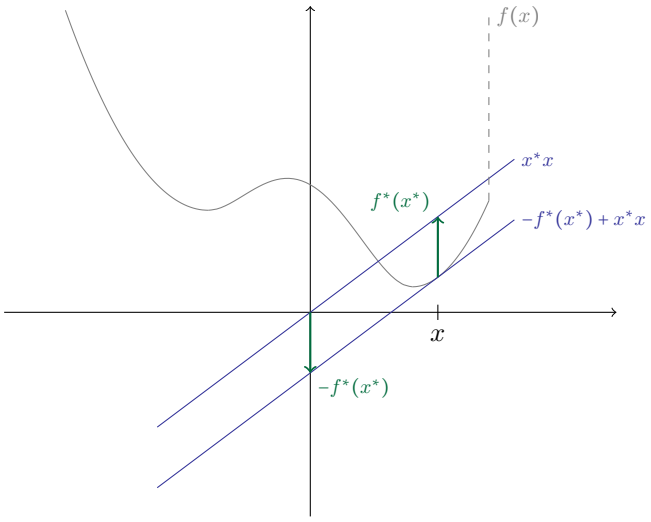


Figure 5.1: Illustration of the Legendre transform

The motivation for these definition is illustrated in Figure 5.1: $f^*(x^*)$ being the largest difference between $y(x) = x^* x$ and $y = f(x)$, the line $y(x) = -f^*(x^*) + x^* x$ necessarily passes below the graph (in the sense $y(x) \leq f(x)$ for all x). In addition, for the particular point chosen in Figure 5.1, this line is actually exact at x, meaning that $f(x) = -f^*(x^*) + x^* x$, with a slope given precisely by $f'(x)$ as it would otherwise cut the graph. The Legendre transform can therefore be interpreted as implicitly constructing tangent lines to the graph of a function for each given slope, at least provided this is possible without cutting the graph.

In addition, comparing the graphs of $-f^*(x^*) + x^* x$ for various values of x^* in Figure 5.2, this translated line actually is the one achieving the maximum value for the given value of x as in the definition of f^{**} since, by the very definition of f^*, all other lines $-f^*(x^*) + x^* x$ necessarily

[4]Note that the value $+\infty$ is explicitly admissible here for f. One could in principle also think about admitting $-\infty$, but excluding this value on the one hand simplifies certain statements and on the other hand does not really eliminate any interesting functions, since a convex function which is $-\infty$ at some point can at most be finite at a single point before jumping to $+\infty$ as convexity enforces an infinite slope (similar to the jump to infinity in the right of Figure 5.1).

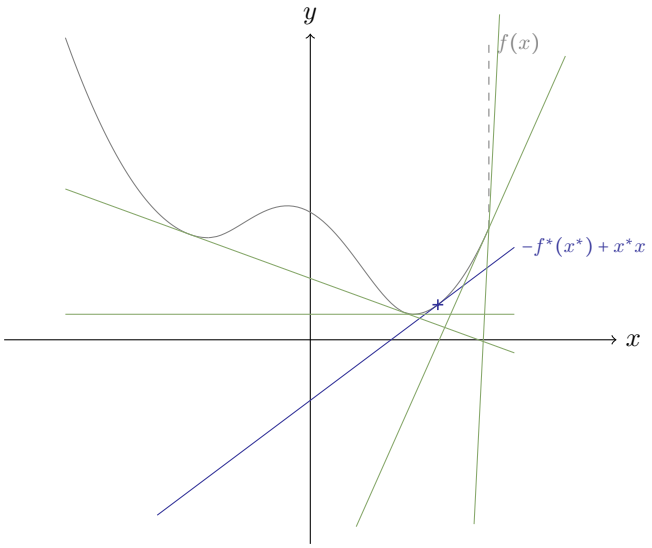


Figure 5.2: Illustration of the inverse Legendre transform

pass below the point $(x, f(x))$. This shows that that one can, at least in certain situations[5], recover the value $f(x)$ through $f^*(x^*)$.
As it turns out, the arguments underlying the theory for this type of transform have virtually nothing to do with f being defined on $\mathbb{R}$. More generally, the same procedure can be applied for a multivariate function $f : \mathbb{R}^n \to (-\infty, +\infty]$ by replacing the line (x, x^*x) with the hyperplane $(\boldsymbol{x}, \boldsymbol{x}^* \cdot \boldsymbol{x})$ in $\mathbb{R}^{n+1}$, $f^*(\boldsymbol{x}^*) = \sup_{\boldsymbol{x}\in\mathbb{R}^n}\{\boldsymbol{x}^* \cdot \boldsymbol{x} - f(\boldsymbol{x})\}$, or, if X is e.g. a Banach space[6] with dual X^*, and $f : X \to (-\infty, +\infty]$ a given functional, one can similarly define

$$f^*(x^*) := \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\}, \tag{5.15}$$

where $\langle x^*, x\rangle$ denotes the natural pairing between X and X^*. Note in particular the important relation

$$x^* \cdot x \le f(x) + f^*(x^*), \tag{5.16}$$

valid for arbitrary $x \in X$ and $x^* \in X^*$. One can reiterate the procedure - but by again using X instead of $(X^*)^* = X^{**}$ - by defining the **bi-dual function**[7]

$$f^{**}(x) := \sup_{x^*\in X^*}\{\langle x^*, x\rangle - f^*(x^*)\}, \tag{5.17}$$

whose primary motivation stems from the following remarkable property:

Theorem 5. *(Fenchel-Moreau, thm. 1.11 [16])*
Assume that $f : X \to (-\infty, +\infty]$ is convex, lower semicontinuous and not identically equal to $+\infty$. Then $f^{**} = f$.

[5]This will be made more precise below. As might be guessed from Figure 5.1 and the discussion thus far, it turns out that the crucial condition is that there is in fact a line passing below the graph of f whose value coincides with $f(x)$, i.e. a so-called subgradient (see in particular Props. 2 and 3 below).

[6]Note that the norm of X is not directly used in any of the results below, i.e. one can consider, as in e.g. [26] or [14], even more general spaces.

[7]The distinction between X^{**} and X is of course irrelevant if X is reflexive, i.e. in particular in $\mathbb{R}^n$ or the Hilbert-space setting. It is only when the inclusion of X in X^{**} is strict that this restriction to X becomes important, as one is primarily interested in the relation between f^{**} and f, the latter one a priori only being defined on X.

Even if the assumptions above do not hold, one always has the following fundamental estimate:

Lemma 3. *Let $f : X \to (-\infty, +\infty]$ be arbitrary. Then $f^{**}(x) \leq f(x)$.*

Proof. This is a direct consequence of taking the supremum over the Equality (5.16) in the form $\langle x^*, x\rangle - f^*(x^*) \leq f(x)\ \forall x \in X, x^* \in X^*$. □

Remark 24. Before stating any further properties leading up to and/or derived from the Fenchel-Moreau-theorem, it is important to note that, besides the standard mathematical definition above, there are, in particular in the physical setting, various alternative definitions of Legendre-type transformations. A very similar one to (5.15) (again primarily useful if f is convex, see below) is given by

$$\hat{f}(x^*) = \inf_{x\in X}\{f(x) - \langle x^*, x\rangle\} = -\sup_{x\in X}\{\langle x^*, x\rangle - f(x)\} = -f^*(x^*) \tag{5.18}$$

enjoying similar properties, but with $\hat{f}$ now being a concave function (as in infimum of linear and thus concave functions in x^*). Under the same assumptions on f as in Thm. 5, the inversion formula (5.17) above can be rewritten as

$$f(x) = f^{**}(x) = \sup_{x^*\in X^*}\{\langle x, x^*\rangle - f^*(x^*)\} = \sup_{x^*\in X^*}\{\langle x^*, x\rangle + \hat{f}(x^*)\}. \tag{5.19}$$

If f itself is concave instead of convex, a more useful definiton is obtained by setting e.g.

$$\tilde{f}(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle + f(x)\} = \sup_{x\in X}\{\langle x^*, x\rangle - (-f)(x)\} = (-f)^*(x^*), \tag{5.20}$$

leading again to a convex function (note that $(-f)^* \neq -f^*$, which would be concave!). Again based on the conversion formula (5.17), under the appropriate assumptions, one has

$$f(x) = -((-f)^*)^* = -\sup_{x^*\in X^*}\{\langle x^*, x\rangle - (-f)^*(x^*)\} = \inf_{x^*\in X^*}\{(-f)^*(x^*) - \langle x^*, x\rangle\}. \tag{5.21}$$

Similar modifications are clearly also possible when reverting the signs of e.g. x^* in (5.15) or (5.18) as this just applies a change in sign of the respective arguments.

◇

At first sight, the definitions above differ from the usual procedure in the physical literature (here for simplicity again in the one-dimensional setting) of introducing the variable x^* by setting $x^* := \frac{\mathrm{d}f}{\mathrm{d}x}$ and then "defining" the Legrendre transform of f e.g. as

$$f^*(x^*) := x^*x - f(x), \tag{5.22}$$

where the sole dependence of g on x^* is justified by remarking that $\frac{\partial}{\partial x}(x^*x - f(x)) = x^* - \frac{\mathrm{d}f}{\mathrm{d}x} = 0$ as $x^* = \frac{\mathrm{d}f}{\mathrm{d}x}$. An "inversion formula" is then recovered by simply rearranging Equation (5.22) to $f(x) = x^*x - f^*(x^*)$, where, by a simple differentiation of (5.22), x in addition satisfies $x = \frac{\mathrm{d}f^*}{\mathrm{d}x^*}$. This is a somewhat tricky argument though, as x^*, if **defined** as $\frac{\mathrm{d}f}{\mathrm{d}x}$, is clearly a function of x, $x^* = x^*(x)$, and f^* therefore in fact still an explicit function of x instead of x^*, $f^* = f^*(x)$, meaning that the argument above is only valid if one chooses to "forget" this (explicit) dependence.
The definition (5.15) on the one hand avoids this pitfall as f^* is evidently truly independent of x as the original variable is eliminiated through the supremum operation, and is, on the other hand, more general as it does not require any differentiability of f.

If f is strictly convex and smooth, one can easily establish a link between the two approaches though. In fact, under this assumption, the supremum in (5.15) will be taking in a single point x, which is characterized through the Euler-Lagrange equation $\frac{\partial}{\partial x}\big(x^*x - f(x)\big) = x^* - \frac{\partial f}{\partial x}(x) = \mathbf{0}$, i.e. the relation $x^* = \frac{\partial f}{\partial x}$ above. The subtle but important difference in interpretation here is that this does not define x^* - which is given - as a function of x, but rather the point x realizing the supremum as an implicit function of x^*. A more proper way of writing (5.22) would therefore be as

$$f^*(x^*) = x^*x(x^*) - f\big(x(x^*)\big),$$

which is then obviously a function of x^* alone. Assuming the dependence of x on x^* is in addition differentiable, g satisfies

$$\frac{\mathrm{d}f^*}{\mathrm{d}x^*} = x(x^*) + \Big(x^* - \frac{\mathrm{d}f}{\mathrm{d}x}\big(x(x^*)\big)\Big) \cdot \frac{\mathrm{d}x}{\mathrm{d}x^*}.$$

As the last term drops out due to $x(x^*)$ satisfying $x^* = \frac{\mathrm{d}f}{\mathrm{d}x}\big(x(x^*)\big)$, one thus recovers the previous relation $x(x^*) = \frac{\partial f^*}{\partial x^*}$, now as an explicit relation in terms of f^*.
Similarly, assuming that f^* is also strictly convex and smooth, the supremum in (5.17) will also be achieved in a single point x, whose Euler-Lagrange equation is given by (note x is again just an arbitrary **given** slope here) $\frac{\partial}{\partial x^*}\big(x^*x - f^*(x^*)\big) = x - \frac{\partial f^*}{\partial x^*}(x^*) = 0$, which then defines x^* as an implicit function of x. Using this and the fact that under the given assumptions $f^{**} = f$, one has $f(x) = f^{**}(x) = x^*(x)x - f^*\big(x^*(x)\big)$, and, if the mapping $x \mapsto x^*(x)$ is actually differentiable, $\frac{\mathrm{d}f}{\mathrm{d}x}(x) = x^*(x) + \Big(x - \frac{\mathrm{d}f^*}{\mathrm{d}x^*}\big(x^*(x)\big)\Big)\frac{\mathrm{d}x^*}{\mathrm{d}x}$. The last term again drops out by the definition of $x^*(x)$, i.e. one recovers the formula $x^*(x) = \frac{\mathrm{d}f}{\mathrm{d}x}(x)$.

This type of differential relation can be substantially generalized to situations where f involving less smoothness[8] This requires replacing the classical derivative of f with an appropriate generalized notion of differentiability, which, in the convex (resp. concave) setting is given by the following

Definition 5. (Subdifferential)
Let $f : X \to (-\infty, +\infty]$. f is said to be **subdifferentiable** (in the sense of convex analysis) at a point x in X if $f(x)$ is finite and there is some x^* in X^* such that the relation

$$f(y) \geq f(x) + \langle x^*, y - x\rangle \tag{5.23}$$

holds for all $y \in X$, i.e. if the hyperplane $\big(x, \langle x^*, x\rangle\big)$ in $X \times \mathbb{R}$ with slope x^* passing through the point $\big(x, f(x)\big)$ lies below the graph of f. The set of all (if any) such slopes at a given point x is called the **(convex) subdifferential** ∂f of f at x. If $f(x)$ is not finite, ∂f is defined to be empty, i.e.

$$\partial f(x) = \begin{cases} \varnothing & \text{if } f(x) \text{ is not finite ,} \\ \{x^* \in X^* : f(y) \geq f(x) + \langle x^*, y - x\rangle \ \forall y \in X\} & \text{else,} \end{cases}$$

and f is thus subdifferentiable at x iff $\partial f(x) \neq \varnothing$.
Similarly, the **concave subdifferential** $\partial^\cap f$ of f is obtained by reversing the sign of the inequality in Equation (5.23), and therefore corresponds to the slopes of all hyperplanes lying above the graph of f and passing through $\big(x, f(x)\big)$.

For convex functions, an important relation with the more basic notion of Gâteaux-differentiability is given by the following

[8] Even though it is well-known that the convexity itself actually does imply a certain degree of regularity, at least within the interior of the domain of f.

Proposition 1. *(Subdifferentiability and Gâteaux-differentiability, prop. I.5.3 [26])*
Let $f : X \to (-\infty, +\infty]$ be convex. If f is Gâteaux-differentiable at x, it is subdifferentiable at x. Conversely, if f is continuous and finite at x and has only one subgradient, then f is Gâteaux-differentiable at x. In both cases, the relation $\partial f(x) = \{f'(x)\}$ holds.

With this definition at hand, the generalization of the relation $x^* = \frac{\partial f}{\partial x}$ above is contained in the following

Proposition 2. *(see e.g. prop. 5.1. [26])*
Let f be a function of $X \to (-\infty, +\infty]$ and f^* its conjugate, and further assume that f is not identically equal to $+\infty$. Then $x^* \in \partial f(x)$ iff

$$f(x) + f^*(x^*) = \langle x^*, x \rangle. \tag{5.24}$$

Proof. Assume that (5.24) holds. The first - somewhat technical - observation is that this implies that both $f(x)$ and $f^*(x^*)$ are finite. In fact, as $\langle x^*, x \rangle$ is always finite[9] and f has by assumption at least one point where it is finite, f^* is never equal to $-\infty$. The sum of $f(x)$ and $f^*(x^*)$ being finite and opposite infinities not being possible, both need to be finite. If the relation (5.24) holds, the hyperplane $\big(x, -f^*(x^*) + \langle x^*, y \rangle\big)$ passes through $\big(x, f(x)\big)$ (i.e. is exact at x). As Equation (5.24) together with the inequality (5.16) imply that x in fact realizes the supremum in the definition of f^*, it holds that, for any $y \in X$, $f^*(x^*) = \langle x^*, x \rangle - f(x) \geq \langle x^*, y \rangle - f(y)$, and thus by rearrangement both $f(y) \geq -f^*(x^*) + \langle x^*, y \rangle \ \forall y \in X$ (i.e. the graph of f lies entirely above this hyperplane) and $f(y) \geq f(x) + \langle x^*, y - x \rangle \ \forall y \in X$ (i.e. the defining inequality (5.23) for a subgradient holds).
Conversely, if $x^* \in \partial f(x)$, by the same inequality (5.24), one has $\langle x^*, x \rangle - f(x) \geq \langle x^*, y \rangle - f(y)$ for all $y \in X$. Taking the supremum over y shows that

$$\langle x^*, x \rangle - f(x) = \sup_{y \in X} \{\langle x^*, y \rangle - f(y)\} = f^*(x^*).$$

□

Despite the suggestive arrangement of the duality relation (5.24), it is not quite symmetric in f and f^*, as f^* is the transform of f, but f need not be the one of f^* unless the inversion formula (5.17) holds. In fact, applying Proposition 2 starting from f^* instead of f a priori only shows that

$$x \in \partial f^*(x^*) \Leftrightarrow f^*(x^*) + f^{**}(x) = \langle x^*, x \rangle. \tag{5.25}$$

Whether or not a "symmetric" version of Proposition 2 holds thus clearly depends on the relation between f and f^{**}. This is clarified in the following

Proposition 3. *(Cor. 5.2 [26], Prop. 2.118 [14])*
Let $f : X \to (-\infty, +\infty]$ be a (possibly non-convex) function. Then the following holds:

1. *If $x^* \in \partial f(x)$, then $x \in \partial f^*(x^*)$.*
2. *If f is subdifferentiable at x, then $f^{**}(x) = f(x)$.*
3. *If $f^{**}(x) = f(x)$, then $\partial f(x) = \partial f^{**}(x)$ (which is empty if both are equal to $\pm\infty$) and the stronger statement*

$$x^* \in \partial f(x) \Leftrightarrow f(x) + f^*(x^*) = \langle x^*, x \rangle \Leftrightarrow x \in \partial f^*(x^*), \tag{5.26}$$

i.e. the statement of 5.24 togehter with its "dual" (instead of 1), holds. In addition, if the common value $f^{**}(x) = f^*(x)$ is finite, the variational characterizations

$$\partial f(x) = \operatorname{argmax}_{x^* \in X^*} \{\langle x^*, x \rangle - f^*(x^*)\}. \tag{5.27}$$

[9] By assumption, $x^* \in X^*$ and is thus a **continuous** linear functional.

and

$$\partial f^*(x^*) = \operatorname{argmax}_{x\in X}\{\langle x^*, x\rangle - f(x)\}. \tag{5.28}$$

hold for the subgradients of f and f^*.

Proof. The proof is mostly a variation of the arguments in the proof of Proposition 2 combined with Lemma 3.
As shown above, $x^* \in \partial f(x)$ iff Equation (5.24) holds, in which case $f(x^*)$ is necessarily finite, and the hyperplane $-f(x)+\langle x^*, x\rangle$ is exact for f^* at x^*. That the graph of $f^*(x^*)$ lies above this hyperplane now follows from Equation (5.16) since $\langle x^*, x\rangle - f(x) \leq f^*(x^*)$ holds for all $x^* \in X^*$. As to the second point, if $\partial f \neq \varnothing$, by Prop. 2, there exists some $x^* \in X^*$ such that

$$f(x) = \langle x^*, x\rangle - f^*(x^*) \leq \sup_{x\in X}\{\langle x^*, x\rangle - f^*(x^*)\} = f^{**}(x),$$

whereas the reverse equality always holds by Lemma 3. Finally, as $x^* \in \partial f(x)$ iff $f(x) = \langle x^*, x\rangle - f^*(x^*)$ and $x^* \in \partial f^{**}(x)$ iff $f^{**}(x) = \langle x^*, x\rangle - f^*(x^*)$, $f(x) = f^{**}(x)$ finite implies that both conditions necessarily hold at the same time. Due to the equality of $f(x)$ and $f^{**}(x)$ as well as $\partial f(x)$ and $\partial f^{**}(x)$, Equation (5.26) follows directly from Proposition 2. As already seen in the proof of this proposition, $x^* \in \partial f(x)$ iff x realizes the supremum in the definition of f^*, which, by the symmetry in the case above, happens iff $x \in \partial f^*(x)$, which in turn happens iff x^* realizes the supremum in the definition of $f^{**}(x) = f(x)$. □

Remark 25. Note that, more generally, if, for some $x \in X$, the value $f^{**}(x^*)$ is finite, then the variational characterization

$$\partial f^{**}(x) = \operatorname{argmax}_{x^*\in X^*}\{\langle x^*, x\rangle - f^*(x^*)\}$$

holds for the (potentially empty) subgradient of f^{**} at $\boldsymbol{x}$ based on considering the duality between f^{**} and $f^{***} = f^*$ (which always holds)[10]. If $f^{**}(x) \neq f(x)$, there is a priori no need for their subgradients to coincide, and the implication $x \in \partial f^*(x^*) \Rightarrow x^* \in \partial f(x)$ need **not** hold as, by Equation (5.25), the hyperplane $\big(x, -f^*(x^*) + \langle x^*, \boldsymbol{x}\rangle\big)$ is exact at $f^{**}(x)$, but not at $f(x)$. In fact, as f^{**} is the pointwise supremum of all affine functions lying below the graph of f, the subgradient $\partial f(x)$ has to be empty at such points. Otherwise any subgradient would be based by definition on such an affine function which in addition passes through $\big(x, f(x)\big)$, therefore precluding $f^{**}(x) < f(x)$. ◇

A particularly important example for a non-smooth situation where full duality holds in the optimization setting is given by the **indicator function** of a nonempty closed convex subset K,

$$I_K(x) = \begin{cases} 0 & x \in K, \\ +\infty & \text{else.} \end{cases}$$

This can be shown to be a convex, lower semicontinuous function due to the assumptions on K. As the supremum defining the Legendre-transform is always equal to $-\infty$ if $x \notin K$, it is easy to see that

$$I_K^*(x^*) = \sup_{x\in K}\{\langle x^*, x\rangle\} = \Pi_K(x^*),$$

where, for any subset S of $\mathbb{R}^n$, the function $\Pi_S(x^*) := \sup_{x\in S}\{\langle x^*, x\rangle\}$ is the so-called **support function** of the set S. By the basic properties of the Legendre-transform, Π_K is again a convex lower semicontinous function if K is closed an convex, and, from Thm. 5, $\Pi_K^*(x) = I_K(x)$. The

[10] Note that the proof of this statement in [14] is slightly misleading, as applying their eq. 2.229 with f^* instaed of f^{**} replacing f would only imply the equivalences (5.25) instead of the equivalence with $x^* \in \partial f^{**}(x)$.

subgradient of I_K is emtpy whenever $x \notin K$ (as the value of I_K is not finite there), and, directly applying the condition (5.23) as the set of all $x^* \in X^*$ such that

$$I_K(y) \geq I_K(x) + \langle x^*, y - x \rangle \overset{I_K(x)=0}{=} \langle x^*, y - x \rangle.$$

Since this inequality is trivially satisfied for $y \notin K$ as $I_K(y) = +\infty$, the only relevant condition to check is what happens if $y \in K$ (and thus $I_K(y) = 0$), and the subgradient is thus given by all x^* such that $0 \geq \langle x^*, y - x \rangle$ for all $y \in K$. This situation can be summarized by defining the **normal cone** to K at an arbitrary x as

$$N_K(x) = \begin{cases} \{x^* \in X^* : \langle x^*, x - y \rangle \geq 0 \ \forall y \in K\} & x \in K, \\ \varnothing & \text{else}, \end{cases} \tag{5.29}$$

with which one has $\partial I_K(x) = N_K(x)$. Based on Prop. 2, any of the subgradients of $\Pi_K(x^*)$ is characterized by the equality $I_K(x) + \Pi_K(x^*) = \langle x^*, x \rangle$, i.e. $\sup_{y \in K}\{\langle x^*, y \rangle\} = \langle x^*, x \rangle - I_K(x)$. If $x \notin K$, the right-hand side equals $-\infty$, and as K is nonempty this equality cannot hold. All potential subgradients thus have to lie in K. If $x \in K$, the indicator-function drops out, and the subgradient of Π_K at x^* is given by the set of all x such that $\sup_{y \in K}\{\langle x^*, y \rangle\} = \langle x^*, x \rangle$, i.e. in accordance with Equation (5.28) as $\partial \Pi_K(x^*) = \operatorname{argmax}_{x \in K}\{\langle x^*, x \rangle\}$ and therefore the vector(s) in K furthest away from the origin "along" x^*.

Remark 26. Note that the first conclusion is in principle another very efficient way of obtaining the first-order necessary (and sufficient) conditions for convex constrained optimization problems and that the supremum in the definition of the primal function $L_P(\boldsymbol{x}) = \sup_{\{(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\mu} \geq \mathbf{0}\}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ in Equation (5.6) has precisely the same effect as adding the indicator function for the admissible set to f.

In fact, it is easy to see that that the necessary and sufficient condition for a convex minimization problem is given by $0 \in \partial f(x)$, since this means that there is a horizontal "plane" passing everywhere below (in the sense of $\leq$) the graph and touching it at the point x, and x therefore has to be a minimizer (see the horizontal line in Figure 5.1). If f is smooth on K, one has[11] $\partial(f + I_K)(x) = \{f'(x)\} + \partial I_K$, and requiring that 0 lie in this set therefore shows that for any minimizer on must have $-f'(x) \in N_K(x)$.

Neverthelees, the Lagrangian formulation, while more susceptible to failure unless a constraint qualification condition is satsified, has the major advantage of directly delivering a "good guess" of what the normal cone looks like based on an algebraic characterization of the constraint set, whereas this question is left open in the characterization of the normal cone in Equation (5.29)[12]. Some further links between the two approaches will be summarized in the next section. ◇

[11] Note that the equality $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$ is not generally true but holds under some mild regularity condition of f, see e.g. [26] or [20].

[12] It should be kept in mind that, similar to Remark 6, the normal cone is the actually relevant set, which can often - but not always - be characterized using the gradient of the functions defining the constraint.

Part III

Applications

The phasefield method is primarily an approach used for the modeling and simulation of phase transformations within the material sciences. In particular, it has been and continues to be heavily employed as a tool to gain a deeper understanding of the effects of different process conditions on the resulting microstructure during solidification processes. Additional - somewhat more recent - applications include the study of the evolution of purely solid microstructures when e.g. subjected to varying external loads or, more simply, during an aging process.

While many of the earlier phasefield models were already able to provide a qualitative understanding of the evolution of microstructures, it was soon realized that there are some serious practical difficulties when trying to obtain more quantitative insights. These are primarily due to the very idea underlying the phasefield method in combination with limited computational resources.
One the one hand, the lengthscale ϵ_p associated with a physical interface region between two different phases is usually several orders of magnitude smaller than the lengthscale L associated with the microstructure itself. Even though this make the volume of the interfaces essentially negligeable, these are nevertheless regions associated with a high energy density due to the mismatch in the atomic arrangement between the materials on both sides. This energy contribution can therefore normally not be neglected but can, at the scale L of the microstructure, be in a good approximation be associated with a surface energy density $\gamma^{\alpha\beta}$ on the "almost" two-dimensional interface. The major advantage of this approximation is that the total energy (resp. entropy) of the microstructure can then be described in terms of volumetric contributions from the various pure phases in terms of their bulk-properties and the effective surface energy densities $\gamma^{\alpha\beta}$ hiding the highly complex physics within the true interfaces. In addition, being an effective macroscopic property, these surface energy densities may in particular be experimentally accessible and can therefore serve as input for numerical models for the evolution of the microstructure.
On the other hand, numerically modeling the evolution of a problem defined by moving (sharp) interfaces is quite challenging. The core of the phasefield method when applied to such a situation lies in "partially undoing" this sharp interface limit by reintroducing a small but finite transition region of width $\sim \epsilon$ between the various phases. Given that the phasefield functional can be based on the **measured** effective (macroscopic) surface energy densities $\gamma^{\alpha\beta}$, one has the great advantage that it is neither necessary for this artifical length scale ϵ to match with the actual width ϵ_p of the physical interface, nor to include a complex microscopic model for this transition region.
Nevertheless, ϵ can of course not be chosen arbitrarily large as it needs to be sufficiently small in order for the phasefield functional to provide an accurate approximation of the energy associated with the sharp interface setting. As the convergence of the "pure" phasefield functional to the surface energy basically relies on the phasefield profile within the interface converging to the classical one-dimensional steady-state profiles (see Section 6.2.2), in combination with additional driving forces arising in the coupled models[13], this essentially imposes at least three restrictions. The first one is purely geometrical and requires for the interface width to be significantly smaller than the (smallest) radius of curvature. Otherwise, the "tangential" derivatives of the phasefield (w.r.t. to an assumed sharp interface surface lying e.g. at the 0.5 isoline) may become competetive with the "normal" ones as one moves away from this line, thus leading to a distorted profile. The second and third are somewhat related and dependend upon the additional energetic contributions and their associated driving forces. On the one hand, even in the absence of any curvature, it is usually necessary for the interface width to be significantly smaller than the total length scale of the problem as otherwise the contribution of the - a priori to a degree arbitrary - interpolation of the given bulk energy densities within the interface region can make up a significant part of the total energy and thus lead to a large distortion of the energetics unless the

[13] The same issue arises can even arise in the absence of additional energetic contributions in the pure phasefield case due to additional constraints such as a volume constraint.

interpolated energy happens to coincide very closely with the "true" energy.
On the other hand, it is also necessary for the strength of the pure phasefield terms (i.e. the ones arising from a and w) to be locally significantly larger than these additional driving forces in order to avoid an excessive deformation of the interface profile.
Which of the three constraints on ϵ is more restrictive clearly depends upon the particular problem. The most obvious idea for avoiding any such issues is of course to simply choose a smaller interface width ϵ as compared to the size of the microstructure. It is clear that if one does so using a given grid-spacing, a reduction in ϵ will entail higher numerical errors in resolving the transition of the phasefield within the increasingly narrow interfaces. Even though this procedure is free of any computational cost (and will actually usually even decrease it) as it amounts to simply reducing a parameter and will improve the approximation at a continuous level, the numercial difficulties will at some point overshadow this improvement. Once this point is reached, the only viable alternative is then an increase of the resolution, which, unlike the modficiation of ϵ, will always entail an increase in the computational effort. Finding the "sweet spot" for a given problem, i.e. the point where the most accurate results are - due to a good balance between numerical and continuous (in terms of ϵ) approximation errors - obtained for a given cost is a fairly difficult problem, which also depends heavily on the interplay of various practical factors. Some of these as well as their interplay with some of the numerical difficulties will be discussed in Chapter 6, with a strong focus on an obstacle-potential based setting.
In contrast to these primarily numerical factors, the way any additional energy contributions $f(\boldsymbol{\phi}, \boldsymbol{c}, T, ...)$ affect the precision of the results for a given ϵ can often be influenced quite heavily by the particular manner they are modeled within the interface region. Since easing the restriction on the numerical interface width can lead to a significant decrease of the required resolution and therefore the necessary computational effort, this has been the subject of a fairly extensive research effort over the past decades and will be the primary focus of Chapter 7.

Chapter 6

Pure Phasefield Problems

The decisive role of the phasefield method within a modeling context is its ability to represent effects associated with surface energies (or entropies). Two very useful properties in this respect are that (at least when approaching the sharp-interface limit) this ability has relatively little to do with the presence of additional energetic contributions provided these are well-behaved as $\epsilon \to 0$, and that these additional influences can be included by simply "adding" them to the essential gradient energy density and bulk-potential contributions. While this statement needs to be put into perspective as it is in practice often necessary to use artifically enlarged interfaces due to limited computational resources[1], it explains the basic additive structure with respect to the driving forces underlying the common phasefield modeling approach. In addition, it highlights the crucial importance of these "standard" phasefield contributions in general.
Before discussing the more complex multiphysics problems alluded to in Chapter 3, it is therefore instructive to first consider the simpler setting in which the phasefield variable is the only unknown. On the one hand, this will serve to introduce some more details on the particular phasefield model used in this thesis. On the other hand, while coupling the phasefield and other fields - even those governed by more "classical" equations which are better understood at both a theoretical and numerical level - can often lead to additional challenges by itself, there is also a number of relatively generic challenges associated with phasefield-type problems, in particular from a numerical point of view. Two of these - related with the bound-constraints in the obstacle case - will be discussed in more detail in this chapter.

Remark 27. The "pure" phasefield-setting in this chapter is chosen primarily due to its simplicity and since there is no point in adding any additional complexity through couplings for the discussion below. Nevertheless, the minimization of surface energy is an interesting phenomenon by itself, and there is in fact a number of relevant applications which can be modeled using the phasefield variable alone. Two particularly interesting ones are the equilibrium shape of droplets on substrates or fibers, for which the interested reader is referred to e.g. [9] and [1]. ◇

Remark 28. A sizeable part of the research focus in the phasefield community in the last decades has been oriented towards improving the accuracy of phasefield methods despite the practical need for artificially enlarged interfaces (often in terms of thin interface limits), in particular for coupled multiphysics problems. This can often be achieved through an improved modeling of the precise form of additional energy contributions within the interfacial regions (two such examples will be considered in Sections 7.1 and 7.2). Even though this can lead to a quite complex description, this added complexity has very little direct influence on the discussions in this chapter. ◇

[1] Using a small but finite interface width will generally lead to deviations from the desired sharp-interface limits. These are often related with the problem of **excess energies** and/or high **Cahn-numbers** (the ratio of the interface width to the "radius" or, in the non-circular case, a measure thereof).

Remark 29. There are also applications where the phasefield method is not primarily used due to its ability to capture effects associated with surface energies, but instead as a purely numerical tool with the main purpose of replacing a physical problem involving sharp interfaces (whether moving or not) with a diffuse approximation thereof. This includes in particular models where the surface-minimizing property of the standard phasefield model is eliminated through an appropriately chosen "counter-term" with the purpose of obtaining a method implicitly tracking sharp interface motions (see e.g. [72], [67] and the references therein for some background and applications). Even though this can be of obvious interest from a numerical point of view, as it in principle allows for working on fixed grids (in particular simple Cartesian ones) even in the presence of complex and potentially moving interface, such problems will not be considered here. ◇

Section 6.1 will first provide a quick outline of the basic variational framework in the multiphase case, corresponding essentially to a simplified version of the one in [52] underlying most the work in this thesis to a pure phasefield setting. Before returning to this more complex case in Section 6.3, the discussion in Section 6.2 will again focus on the simplest possible setting consisting of a reduced (i.e. expressed solely in terms of a single phasefield ϕ) two-phase version of this general setting. After introducing some standard background and results, the main focus of this section will be a relatively detailed analysis of the impact of the 0-1-bounds on the phasefield values in the discrete case as one of the central "ingredients" of the basic phasefield model in the obstacle case. More precisely, a discrete equivalent of the basic analytical one-dimensional phasefield profile (corresponding to an indisturbed flat interface) and the associated energetics will be derived for the case of the obstacle potential. One the one hand, this allows for a very simple interpretation of some commonly observed effects (such as the discrete gradient energy always being larger than the continuous one). On the other hand, the resulting expressions can conveniently be expanded in terms of the discretization parameter Δx, from which a number of interesting facts can simply be "read off". In particular, even though the discrete interface width is only first-order accurate (and always more narrow than the continuous one), both the energetics and the discrete profile itself are second-order convergent.
Some additional issues related to the multiphase setting will then be discussed in Section 6.3. Subsection 6.3.1 will first recall the basic equations to be fulfilled and different choices for the dynamics. This is followed in Subsection 6.3.3 by a discussion of some numerical and algorithmical aspects arising when dealing with multiphase problems.

Remark 30. Parts of this chapter consist of elementary background for the phasefield equation. It was nevertheless chosen to introduce this background here instead of the actual background part since it is on the one hand mostly directly relevant for the discussion of the main points and on the other hand, explicitly being based on the phasefield equation itself, considerably more specific than the relatively general and consequently abstract previous considerations.
The intention is also not to provide a full discussion of general phasefield problems (which, in particular in the multiphase case, would be a very complex and difficult task), but simply to recall some elementary facts before focusing on some particular but quite relevant topics encountered when using an obstacle potential. ◇

6.1 The Basic Phasefield Functional

The key idea underlying the phasefield method for multiphase problems is essentially the same as the one already outlined for the simpler two-phase setting in section 2 and consist in the use of a vectorial **order parameter** $\boldsymbol{\phi} = (\phi^\alpha)_{1\leq\alpha\leq N}$, with each ϕ^α providing a smooth approximation of the characteristic function of the given phase, i.e. with $\phi^\alpha = 0$ indicating that one is outside

the phase and $\phi^\alpha = 1$ indicating that one is within the phase[2]. The requirement for a smooth transition between these two cases enforces the existence of a transition region or **interface region** whenever two or more phases meet, i.e. mixed regions where several phases are considered to coexist. Within the coexistence regions, it is quite natural to interpret ϕ^α as representing the individual phase or volume fractions, and thus to impose the sum condition $\sum_{\alpha=1}^N \phi^\alpha = 1$. In addition, it may be desirable or even necessary to, as in the two-phase case, impose the additional restriction $0 \leq \phi^\alpha \leq 1$ on the phase fractions, i.e. to restrict the phasefield values to lie within the N-dimensional Gibbs-simplex

$$\mathcal{GS}^N = \Sigma_1^N \cap [0,1]^N = \{\boldsymbol{\phi} \in \mathbb{R}^N : \sum_{\alpha=1}^N \phi^\alpha = 1,\ 0 \leq \phi^\alpha \leq 1\ \forall \alpha\}, \tag{6.1}$$

where Σ_1^N denotes the hyperplane of all N-dimensional vectors whose entries sum up to 1.

Remark 31. Note that the conditions in (6.1) are partially redundant as $0 \leq \phi^\alpha\ \forall \alpha$ and the sum-constraint automatically imply that $\phi^\alpha = 1 - \sum_{\beta \neq \alpha} \phi^\beta \leq 1\ \forall \alpha$. As this redundancy can sometimes lead to unnecessary complications - in particular when investigating Lagrange-multipliers associated with the constraints - it is therefore often convenient to use the equivalent but simpler description of $\mathcal{GS}^N$ through[3]

$$\mathcal{GS}^N = \Sigma_1^N \cap \mathbb{R}_+^N = \{\boldsymbol{\phi} \in \mathbb{R}^N : \sum_{\alpha=1}^N \phi^\alpha = 1,\ 0 \leq \phi^\alpha\ \forall \alpha\}. \tag{6.2}$$

◇

While there are often additional difficulties in the modeling of the physics within these coexistence regions, they are also the basis for the approximation of the surface energies (or, in other settings e.g. the entropy) through a **volume** integral

$$\int_\Omega \epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi})\, \mathrm{d}\boldsymbol{x} \tag{6.3}$$

over the entire (fixed) domain. This is in contrast to the "sharp-interface" representation where this modeling is not necessary, but the surface energy needs to be evaluated in terms of surface integrals over surfaces whose position in time is usually an unknown to be determined as a part of the problem to be solved. In Equation (6.3), a and w are generalizations of the gradient energy density and bulk potential to the multiphase case, whose role within a variational framework is the same as in the simpler two-phase case, namely a penalizes gradients (indicated by the additional $\nabla\boldsymbol{\phi}$-argument) while w penalizes values deviating from the bulk-values 0 and 1, and ϵ is a parameter for controlling their relative strenghts and thus the resulting interface width.
In order for the integral in Equation (6.3) to provide the desired smooth approximation of the surface energy (or entropy) in a multiphase problem, a and w still need to be chosen in an appropriate manner. As, even for multiphase problems, the surface energies in the sharp interface setting are associated with the interfaces between just two coexisting neighboring regions (with regions where more than two phases meet making up lower-dimensional objects such as lines or points), it is quite natural that the multiphase generalizations are usually based upon various modifications of a simple summation procedure over two-phase interactions, such that, in an idealized situation consisting of flat two-phase interphases alone, one would recover the known results from that simpler setting. This approach still leaves a large degree of flexibility

[2]Note that in the two-phase setting, where it suffices to use a single order paramter, other choices such as ± 1 or $\pm\frac{1}{2}$ for indicating one or the other phase are also popular. While one could in principle also use other values than 0 and 1 in the multiphase case, this choice turns out to be a particularly convenient one.

[3]Here it is important to maintain the lower bound 0 instead of the upper one as soon as $N \geq 3$. For example, the vector $\boldsymbol{\phi} = (-0.2, 0.6, 0.6)^T$ would satisfy both the sum-constraint and $\phi^\alpha \leq 1\ \forall \alpha$.

in constructing particular functionals, which primarily differ in their treatment of triple- and multiphase regions. Combined with the additional degree of freedom in the precise form of the postulated phasefield dynamics - even if based on a gradient type flow - this has lead to a variety of different phasefield models, each with their own advantages and disadvantages.

The one primarily used in this thesis - and the current default within the **Pace3D**-framework - is based on the formulation originally introduced in [52] and [71] within an entropy framework for solidification problems. Considering for simplicity the corresponding energy formulation, the gradient energy density in the isotropic case is defined by

$$a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) = \sum_{\beta>\alpha} \gamma^{\alpha\beta} |\boldsymbol{q}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})|^2, \tag{6.4}$$

with the **generalized antisymmetric gradient vectors**

$$\boldsymbol{q}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) = \phi^\alpha \nabla\phi^\beta - \phi^\beta \nabla\phi^\alpha. \tag{6.5}$$

This amounts to a summation over energy contributions stemming from two-phase interactions in terms of the respective $\gamma^{\alpha\beta}$ and $\boldsymbol{q}^{\alpha\beta}$ corresponding to each α-β-pairing.
Extensions to anisotropic gradient energy densities can then be obtained in much the same way as in the two-phase case by setting

$$a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) = \sum_{\beta>\alpha} A^{\alpha\beta}(\boldsymbol{q}^{\alpha\beta}) = \sum_{\beta>\alpha} \gamma^{\alpha\beta} \left(a^{\alpha\beta}\right)^2 (\boldsymbol{q}^{\alpha\beta}) |\boldsymbol{q}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})|^2 \tag{6.6}$$

and a prefactor $a^{\alpha\beta}$ homogeneous of degree 0 and thus depending only on the orientation of $\boldsymbol{q}^{\alpha\beta}$, but not its norm (i.e. is a function of the normal vector $\boldsymbol{n}^{\alpha\beta} = \frac{\boldsymbol{q}^{\alpha\beta}}{|\boldsymbol{q}^{\alpha\beta}|}$ only).

Remark 32. Note that, while sometimes convenient as it makes the desired quadratic dependence of a on the norm of the gradient vectors easily visible, the definition of the prefactors as a function of the direction only leads to some tricky issues when the norm of $\boldsymbol{q}^{\alpha\beta}$ approches zero as this leaves the normal direction - and thus the $a^{\alpha\beta}$ - undefined unless $a^{\alpha\beta}$ does not actually depend on the direction, i.e. unless the gradient energy density is isotropic.
This does not entail any issues in terms of the "total" a-function or the $A^{\alpha\beta}$, since both a and $\frac{\partial a}{\partial \boldsymbol{q}^{\alpha\beta}}$ (resp. $A^{\alpha\beta}$ and $\frac{\partial A^{\alpha\beta}}{\partial \boldsymbol{q}^{\alpha\beta}}$) will converge to zero with $|\boldsymbol{q}^{\alpha\beta}| \to 0$ (being homogeneous of degree 2 resp. 1 in $\boldsymbol{q}^{\alpha\beta}$) regardless of the direction along which one approaches this point, but poses a problem when trying to work with the $a^{\alpha\beta}$ themselves as in [75]. ◇

As outlined above, a generalized w-formulation could be obtained by introducing an additional summation into the two-phase formulation, i.e. by setting

$$\tilde{w}_{mw}(\boldsymbol{\phi}) = 9 \sum_{\beta>\alpha} \gamma^{\alpha\beta} (\phi^\alpha)^2 (\phi^\beta)^2$$

in the well-case resp.

$$\tilde{w}_{mo}(\boldsymbol{\phi}) = \frac{16}{\pi^2} \sum_{\beta>\alpha} \gamma^{\alpha\beta} \phi^\alpha \phi^\beta$$

in the obstacle-case. Unfortunately, using this simple extension, there is a tendency towards the appearence of additional phases throughout the entire interface region, and in particular also in regions that one would naturally want to consist of just the two phases corresponding to neighboring bulk regions. For this reason, an additional penalty term is used in [52] to associate an additional cost with the presence of more than two phases by setting

$$w_{mw}(\boldsymbol{\phi}) = 9 \sum_{\alpha<\beta} \gamma^{\alpha\beta} (\phi^\alpha)^2 (\phi^\beta)^2 + \sum_{\alpha<\beta<\delta} \gamma^{\alpha\beta\delta} (\phi^\alpha)^2 (\phi^\beta)^2 (\phi^\delta)^2 \tag{6.7}$$

resp.

$$w_{mo}(\boldsymbol{\phi}) = \frac{16}{\pi^2} \sum_{\alpha<\beta} \gamma^{\alpha\beta} \phi^\alpha \phi^\beta + \sum_{\alpha<\beta<\delta} \gamma^{\alpha\beta\delta} \phi^\alpha \phi^\beta \phi^\delta. \tag{6.8}$$

The penalty parameters $\gamma^{\alpha\beta\delta}$ are thus used to add an additional energy contribution over triple-phase interactions with an analogous form as the previous two-phase energy contributions.

Remark 33. Note that within a pure two-phase α-β region with $\phi^\alpha = 1 - \phi^\beta$, $\boldsymbol{q}^{\alpha\beta}$ reduces to $-\nabla\phi^\alpha = \nabla\phi^\beta$, i.e. the gradient energy distribution in such regions reduces precisely to the same form as in the basic two-phase case. Similarly, as all triple-phase terms vanish within two-phase regions and the summation over the two-phase pairings is based on the expressions obtained by replacing ϕ and $1 - \phi$ with ϕ^α resp. ϕ^β (or vice versa) in the two-phase case, w will also reduce to the desired form. ◇

Remark 34. There are of course many conceivable ways of generalizing the gradient energy density $a(\nabla\phi)$ from the two-phase setting to the multiphase case. The construction used above as a summation over two-phase interactions can be motivated by the observation that the surface energy densities $\gamma^{\alpha\beta}$ correspond to energies between pairings of phases. In addition, in the anisotropic case, the surface energy associated with an interface only depends on the orientation between the adjacent phases and not the absolute orientation of either one of them. This is closely related to another somewhat delicate issue within multiphase regions, namely that of defining appropriate "normal vectors" and angles between phases (the choice $\boldsymbol{n} = \pm\frac{\nabla\phi}{|\nabla\phi|}$ being a fairly obvious one in the two-phase case).

While the choice $\boldsymbol{n}^\alpha = \pm\frac{\nabla\phi^\alpha}{|\nabla\phi^\alpha|}$ is still a natural one for the normal vector associated with the phase α "as a whole", the a priori intuitive choice of defining the orientation between two phases based directly on two such normals may not necessarily lead to the best results. Considering the simple sharp interface example of three straight interfaces meeting in a triple point, it is clear that the angle between every pair of phases is well-defined and constant everywhere except at a single singular point, at which all angles admit an obvious limit though. Within the diffuse interface setting, this triple point is represented by an entire triple point region with small but finite extension. It is clearly desirable with respect to this example for the definition of the relative orientation between any two phases to remain as stable as possible upon the transition from a two-phase to a three-phase (or, more generally, multiphase) region as this may otherwise lead to non-negligeable deviations in the energetics unless the ratio of the interface width to the domain size is chosen sufficiently small[4]. As the vectors $-\frac{\nabla\phi^\alpha}{|\nabla\phi^\alpha|}$ of the individual phases will generally follow a smooth transition from one (approximately) constant normal vector to the other during the transition through the triple-point region, due to an increasing influence of the respective third phase, this may be difficult to achieve based on these vectors themselves.

The choice of the generalized gradient vectors $\boldsymbol{q}^{\alpha\beta}$ replacing the original $\nabla\phi$ corresponds to one possible way of defining a better suited normal $\boldsymbol{n}^{\alpha\beta} = \frac{\boldsymbol{q}^{\alpha\beta}}{\|\boldsymbol{q}^{\alpha\beta}\|}$ between the individual phase pairings, another popular choice being[5] $\boldsymbol{n}^{\alpha\beta} = \frac{\nabla\phi^\beta - \nabla\phi^\alpha}{\|\nabla\phi^\beta - \nabla\phi^\alpha\|}$ (see e.g. [61] and [74]). ◇

Assuming in line with the purpose of this chapter that any additional energetic contribution are prescribed in the form of a simple weighted interpolation of **given** phase-specific contributions[6] f^α, another question to be adressed is how the nonlinear interpolation functions[7] $h(\phi)$

[4] Based on geometric considerations, one would expect the influence of these multiphase regions to decrease very quickly with ϵ. Whereas interfaces translate to regions with a volume $\sim \epsilon$, the volumetric contribution of "edges" and "points" are expected to be roughly of the order $\sim \epsilon^2$ resp. $\sim \epsilon^3$. Nevertheless, these regions can also be associated with high energy densities such as e.g. when modeling an elastic problem with corners and edges where one expects singularities in the solution, making these estimates too optimistic.

[5] Note that the choice of sign here in the second case is such that it agrees with the one for $\boldsymbol{q}^{\alpha\beta}$ in two-phase regions.

[6] Regarding the motivation of using f^α here see remark 36.

[7] For the simplest linear case $h(\phi) = \phi$ this is obvious.

from the two-phase setting can be generalized while maintaining their principle properties.
Denoting the interpolation weights for the individual phases by $h^\alpha(\boldsymbol{\phi})$, the first such property is that $h^\alpha(\boldsymbol{e}^\beta) = \delta^{\alpha,\beta}$, i.e. the interpolation weights in a bulk-region should be one for the bulk phase and zero for all other phases. A second desirable property is the monotonic increase of $h^\alpha(\boldsymbol{\phi})$ with ϕ^α. Finally, the interpolation weights should also satisfy the consistency condition $\sum_\alpha h^\alpha(\boldsymbol{\phi}) = 1$, which is necessary to ensure that the interpolation of any property e which does not in fact depend on the phases, $f^\alpha = f\ \forall \alpha$, satisfies $\sum_\alpha f^\alpha h^\alpha(\boldsymbol{\phi}) = f$.
A simple approach ensuring these properties under very basic conditions is the following (see e.g. also [63]). It consists in taking any monotonically increasing (scalar) function[8] $\tilde{h}(\phi)$ satisfying $\tilde{h}(0) = 0$ and $\sum_\beta \tilde{h}(\phi^\beta) > 0$ for $\sum_\beta \phi^\beta = 1$ and defining the interpolation weights $h^\alpha(\boldsymbol{\phi})$ through a normalization by

$$h^\alpha(\boldsymbol{\phi}) = \frac{\tilde{h}(\phi^\alpha)}{\sum_\beta \tilde{h}(\phi^\beta)}. \tag{6.9}$$

Remark 35. The third condition $\sum_\alpha h^\alpha(\boldsymbol{\phi}) = 1$ is then satisfied by construction. Together with $\tilde{h}(0) = 0$, this normalization automatically also guarantees that the bulk weights $h^\alpha(\boldsymbol{e}^\alpha)$ satisfy $h^\alpha(\boldsymbol{e}^\alpha) = 1$. That the second condition also holds for all $\boldsymbol{\phi}$ in the Gibbs-simplex follows from a simple direct calculation showing that

$$\frac{\partial h^\alpha(\boldsymbol{\phi})}{\partial \phi^\alpha} = \frac{\left(\sum_\beta \tilde{h}(\phi^\beta)\right)\tilde{h}'(\phi^\alpha) - \tilde{h}(\phi^\alpha)\frac{\partial}{\partial \phi^\alpha}\sum_\beta \tilde{h}(\phi^\beta)}{\left(\sum_\beta \tilde{h}(\phi^\beta)\right)^2} = \frac{\sum_{\beta\neq\alpha} \tilde{h}(\phi^\beta)}{\left(\sum_\beta \tilde{h}(\phi^\beta)\right)^2}\tilde{h}'(\phi^\alpha).$$

The denominator is obviously non-negative and $\tilde{h}'(\phi^\alpha) \geq 0$ by monotonicity. Finally, the numerator is also non-negative for $0 \leq \phi^\beta \leq 1$ for all $\beta \neq \alpha$ due to the monotonicity condition together with $\tilde{h}(0) = 0$, therefore showing that $\frac{\partial h^\alpha(\boldsymbol{\phi})}{\partial \phi^\alpha} \geq 0$. ◇

Combining the different components above, one obtains a phasefield functional of the form

$$\mathcal{F}_\epsilon(\boldsymbol{\phi}) = \int_\Omega \epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi}) + f(\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{x}, \tag{6.10}$$

which is to be minimized under the relevant constraints on the phasefield $\boldsymbol{\phi}$. Besides the (relatively simple) local sum-constraint $\sum_{\alpha=1}^N \phi^\alpha = 1$ in the case of the multi-well potential or its somewhat more complex counterpart in terms of the Gibbs-simplex $\mathcal{GS}^N$ in the multi-obstacle case, a further quite common restriction (considered e.g. in [29]) is to fix the total volume $\int_\Omega \phi^\alpha\,\mathrm{d}\boldsymbol{x} \stackrel{!}{=} V^\alpha$ of one or several phases α.

Remark 36. Here, the choice of the letter f (resp. $\mathcal{F}$) should not be assigned any deeper physical or thermodynamic meaning. Since the functional in this particularly simple setting has no dependence on any of the standard thermodynamic quantities, this is just a matter of notational convenience here since this leads to a set of equations which is closer to the one in [52] (where the - then actual free energy - enters the phasefield equation through the term $-\frac{1}{T}\frac{\partial f}{\partial \boldsymbol{\phi}}$) and the discussion in section 7.1 which also relies on the actual free energy. ◇

[8] Some popular examples in the two-phase setting will be listed in section 6.2.

6.2 Two-Phase Problems

Before returning to the multi-phasefield model introduced in the previous section in some additional detail, this section will first discuss a number of pure phasefield applications in a pure two-phase setting. Interpreting the variables ϕ^1 and ϕ^2 as the phase fractions of the first and second phase, it is natural to impose the constraint $\phi^1 + \phi^2 = 1$ as a special case of the one in Section 6.1. This implies that each phase fraction is known as soon as the other one is, such that these problems can, as in the discussion in the introduction, be completely described in terms of a single order parameter ϕ alone, taken here for definiteness as $\phi := \phi^1$.
It is clear that this leads to a number of significant simplifications as compared to the more general multi-phasefield model from the outlined above. Firstly, this reduces the vectorial phasefield equations to a single scalar one for ϕ. Secondly - and this is partially related to the particular phasefield model used here - the generalized gradient vector $\boldsymbol{q}^{12}$ reduces back to the (negative of) the basic gradient vector $-\nabla\phi$, i.e. unless the gradient energy density is anisotropic, there is no nonlinearity in the "spatial" part of the functional. Thirdly, unless additional constraints are present, the representation of the admissible set becomes essentially trivial. As the sum-constraint is already fully incorporated into the formulation, it simply drops out in the double-well case, while the restriction to the Gibbs-simplex in the double-obstacle case reduces to the box-constraint $\phi(\boldsymbol{x}) \in [0, 1]$.

Remark 37. Even though it may seem that the interest of the two-phase setting (as a particularly simple special case of the more general multiphase-setting) is quite limited, this is somewhat misleading. On the one hand, there are in fact a number of interesting applications of the phasefield method (such as capillary rise problems and topology optimization problems) which, while not a priori restricted to two phases, in practice often arise naturally in precisely this setting. On the other hand, even though the simplifications above as compared to the multiphase models are certainly significant, many of the essential challenges arising in the general case (high resolution requirements due to the small length-scale ϵ, lack of convexity, bound-constraints, treatment of strong anisotropies, ...) are actually inherited - though possibly in an exacerbated fashion - from the ones in the two-phase case[9]. ◇

The problems to be considered in this section will, similar to Section 3, be based on slight variations of the functional

$$\mathcal{F}_\epsilon(\phi) = \int_\Omega \epsilon a(\nabla\phi) + \frac{1}{\epsilon} w(\phi) + f(\phi)\,\mathrm{d}\boldsymbol{x}, \tag{6.11}$$

or, with a in the simplest case given by $a(\nabla\phi) = \gamma|\nabla\phi|^2$, its isotropic version

$$\mathcal{F}_\epsilon(\phi) = \int_\Omega \gamma\epsilon|\nabla\phi|^2 + \frac{1}{\epsilon} w(\phi) + f(\phi)\,\mathrm{d}\boldsymbol{x}. \tag{6.12}$$

The bulk potential may be either given by the **double well potential**

$$w_{dw}(\phi) = 9\gamma\phi^2(1-\phi)^2 \tag{6.13}$$

or the **double obstacle potential**

$$\tilde{w}_{ob}(\phi) = w_{ob}(\phi) + 1_{[0,1]}(\phi), \tag{6.14}$$

where

$$w_{ob}(\phi) = \frac{16}{\pi^2}\gamma\phi(1-\phi), \tag{6.15}$$

[9] A notable exception here is given by the triple-phase terms in the bulk potential w, which, depending on the size of the penalty parameters, may lead to an additional stiffness in the equations absent in the presence of only two phases.

is the smooth part of the obstacle potential and $1_{[0,1]}$ is the indicator function

$$1_{[0,1]}(\phi) = \begin{cases} 0 & , 0 \leq \phi \leq 1, \\ +\infty & , \text{else.} \end{cases}$$

of the interval $[0,1]$, whose role it is to maintain the phasefield values within the interval $[0,1]$. Alternatively, one can also impose this restiction directly in terms of an additional constraint $0 \leq \phi \leq 1$, and, since $1_{[0,1]}$ vanishes for all such values, work with the smooth part $w_{ob}(\phi)$ only instead of $\tilde{w}_{ob}(\phi)$.

Remark 38. Whereas the restriction of the ϕ-values to lie within $[0,1]$ is a necessity when using w_{ob}, in the case of the double-well potential one may or may not choose to impose this additional constraint. As these bounds entail some notable complications in the solution of the resulting phasefield problems below, which would only seem to be counterbalanced by large "pure" bulk regions (i.e. regions with $\phi = 0$ or $\phi = 1$) when using double obstacle potential, this possibility will not further be considered here[10]. ◇

In the simplest cases, $f(\boldsymbol{\phi})$ is given by an interpolation

$$f(\phi) = \Delta f h(\phi) = (f^1 - f^2)h(\phi), \tag{6.16}$$

describing the energy difference $\Delta f = f^1 - f^2$ associated with transforming from the first to the second phase. $h(\phi)$ is some appropriate interpolation function, whose choice provides some additional freedom in modeling the energy contributions within two-phase regions.

Classical choices[11] (see e.g. [52]) together with their derivatives are shown in Table 6.2. Unlike h_0, for which the contribution of the f^α for each phase are directly proportional to the phase fractions, the energy contributions for the other two functions remain closer to that of the predominant phase, with a steeper transition around $\phi = \frac{1}{2}$ and therefore a sharper transition region in an energetic sense than the phasefield variable itself.

	$h(\phi)$	$h'(\phi)$	$h''(\phi)$
h_0	ϕ	1	0
h_1	$\phi^2(3-2\phi)$	$6\phi(1-\phi)$	$6-12\phi$
h_2	$h_2(\phi) = \phi^3(6\phi^2 - 15\phi + 10)$	$30\phi^2(1-\phi)^2$	$60\phi(1-\phi)(1-2\phi)$

6.2.1 The Steady-State and Dynamic Phasefield Equations

From the functional $\mathcal{F}_\epsilon$ above and using the appropriate version of the functional derivative $\frac{\delta \mathcal{F}_\epsilon}{\delta \phi} = \frac{\partial \mathcal{F}}{\partial \phi} - \nabla \cdot \left(\frac{\partial \mathcal{F}_\epsilon}{\partial \nabla \phi} \right)$, it is now seeminlgy straightforward to (re)derive the first-order necessary conditions to be satisfied by any potential minimizer ϕ of $\mathcal{F}_\epsilon$ and, based upon the postulate of a gradient flow and under some additional assumptions, their "dynamic" version.
At least in the obstacle case, there is an additional constraint in terms of $\phi \in [0,1]$ though. From this and the discussion in Section 4.3 one expects the appearence of two additional multipliers in the resulting phasefield equation. Even though these do have a very fundamental impact, they are often simply left out in the more applied literature.

[10] For coupled problems though, allowing negative ϕ-values in combination with simple interpolation schemes for the material properties (e.g. weighing the material properties by the ϕ-values of the respective phase) may lead to very serious issues such as e.g. negative diffusivities or stiffnesses. One way of of dealing with this is of course to simply add the bound-constraints on the phasefield variable. Another one - avoiding the additional complications this constraint entails - is to modify the basic interpolation schemes outside the range $[0,1]$ such as to ensure physically valid values of the interpolated quantities.

[11] In relation with the variational approach, it should be noted that $H^1(\Omega) \hookrightarrow L^q$ with $1 \leq q < \infty$ if $n = 2$ and $1 \leq q \leq 6$ if $n = 3$, i.e. integrability is ensured without knowing a priori whether ϕ is particularly well-behaved or not.

Remark 39. This may partly be due to the fact that their action can in principle be "hidden" behind an appropriate projection operator[12], but the practical usefulness of this approach is highly problem-dependent. While this can be very convenient, in particular when using an explicit time-stepping scheme involving a purely local projection operations, this need not be the case anymore when using other time-discretizations or if one adds additional constraints (see e.g. Remark 42).

◇

Since analyzing the impact of box-constraint is - at least at a purely formal level disregarding regularity issues and the precise meaning of this constraint[13] - mostly basic, it therefore seems worthwhile to at least indicate the basic reasoning from which the resulting steady-state equations and their complete form can in principle be derived. To keep the discussion as simple as possible, it will in addition be assumed that $a(\nabla\phi) = \gamma|\nabla\phi|^2$, i.e. that the phasefield energy is isotropic[14]. Taking the directional derivative of the phasefield functional $\mathcal{F}_\epsilon$ leads to the first-order necessary condition

$$\mathcal{F}_\epsilon'(\phi;\psi) = \int_\Omega 2\gamma\epsilon\nabla\phi\cdot\nabla\psi + \frac{1}{\epsilon}w'(\phi)\psi + f'(\phi)\psi\,\mathrm{d}\boldsymbol{x} \geq 0,$$

or, assuming sufficient regularity, an immediate integration by parts modifies this into

$$\mathcal{F}_\epsilon'(\phi;\psi) = \int_{\partial\Omega} 2\gamma\epsilon\frac{\partial\phi}{\partial\boldsymbol{n}}\psi\,\mathrm{d}\boldsymbol{s} + \int_\Omega \Big(-2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi)\Big)\psi\,\mathrm{d}\boldsymbol{x} \geq 0$$

for all admissible directions ψ. At this point, there is now a strong difference between the (unconstrained when expressed in the reduced form) double-well case and the double-obstacle one. In the former case, ψ is (up to some regularity requirements due to the underlying function space) essentially arbitrary. Since in particular both ψ and $-\psi$ are admissible as long as ψ is so, one has both $\mathcal{F}_\epsilon(\phi;\psi) \geq 0$ and $\mathcal{F}_\epsilon(\phi;-\psi) = -\mathcal{F}_\epsilon(\phi;\psi) \geq 0$, and thus

$$\mathcal{F}_\epsilon'(\phi;\psi) = \int_{\partial\Omega} 2\gamma\epsilon\frac{\partial\phi}{\partial\boldsymbol{n}}\psi\,\mathrm{d}\boldsymbol{s} + \int_\Omega \Big(-2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi)\Big)\psi\,\mathrm{d}\boldsymbol{x} = 0 \quad \forall\psi\in H^1(\Omega). \tag{6.17}$$

From this it is, at least formally, quite easy to obtain a more explicit characterization of the minimizer. First "testing" this equality with functions ψ vanishing on the boundary, one can conclude that one must have $-2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) = 0$ in Ω for this to hold. Provided this is the case, the volume integral vanishes in Equation (6.17), and by varying the boundary values of ψ, it then follows that $\frac{\partial\phi}{\partial\boldsymbol{n}}$ will also have to be zero. In summary, one thus has to satisfy

$$\begin{cases} -2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) = 0 & \text{in } \Omega, \\ \frac{\partial\phi}{\partial\boldsymbol{n}} = 0 & \text{on } \partial\Omega, \end{cases} \tag{6.18}$$

or, in the general anisotropic case,

$$\begin{cases} -\epsilon\nabla\cdot\Big(\frac{\partial a}{\partial\nabla\phi}\Big) + \frac{1}{\epsilon}w'(\phi) + f'(\phi) = 0 & \text{in } \Omega, \\ \frac{\partial a}{\partial\boldsymbol{n}} = 0 & \text{on } \partial\Omega. \end{cases} \tag{6.19}$$

In contrast, when subject to the box-constraints, ψ and $-\psi$ are usually not both admissble directions[15], and even a formal argument requires a little more care.

[12] This topic will be discussed in some more detail in Section 6.3, where, even in the discrete and fully explicit case, this projection can become somewhat more complex depending on the precise form of dynamics chosen.

[13] On a less formal level, dealing with an obstacle-type constraint is in fact a quite involved question. This will not be discussed in any detail here, but some of the more relevant issues will nevertheless be pointed out in Remark 40 below.

[14] This is mostly irrelevant, but avoids a tedious discussion for relating the signs of $\frac{\partial a}{\partial\boldsymbol{n}}$ and $\frac{\partial\phi}{\partial\boldsymbol{n}}$.

[15] Unlike for the double-well case without constraints or e.g. equality constraints such as the volume-constraints, the admissible directions in combination with the inequality constraints do not form a linear space but a cone (see Chapter 4.3 for a basic discussion).

If ϕ is at least locally continuous and $0 < \phi < 1$ at some point inside the domain, there would, by this continuity, exist some neighborhood on which ϕ would continue to satisfy this inequality and one could thus take (a sufficient amount of) local variations ψ for which both ψ and $-\psi$ would be admissible. From this one can deduce that

$$-2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) = 0 \text{ if } 0 < \phi < 1.$$

If one instead has $\phi = 0$ in some region, the only admissible variations ψ will locally have to satisfy (in the appropriate sense) $\psi \geq 0$, from wich it then only follows that

$$\mu^- := -2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) \geq 0.$$

Applying the same argument to those "regions" where $\phi = 1$, it similarly follows that

$$-\mu^+ := -2\gamma\epsilon\Delta\, + \frac{1}{\epsilon}w'(\phi) + f'(\phi) \leq 0.$$

As both μ^+ and μ^- are non-negative, these three conditions can be summarized to

$$-2\gamma\epsilon\Delta\,\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) + \mu^+ - \mu^- = 0 \tag{6.20}$$

combined with the complementarity conditions

$$\mu^+(1-\phi) = 0 \quad \text{and} \quad \mu^-\phi = 0 \tag{6.21}$$

enforcing that μ^+ resp. μ^- has to vanish unless the associated constraint is active, i.e. $1-\phi = 0$ or $\phi = 0$.

That this conclusion is not quite as simple as in the unconstrained case also leads to some complications when trying to "focus" on the boundary. In fact, in contrast to the previous argument, it does not follow that the volume integral simply drops out. Instead, using Equation (6.20), one can only conclude that

$$\mathcal{F}'_\epsilon(\phi;\psi) = \int_{\partial\Omega} 2\gamma\epsilon\frac{\partial\phi}{\partial\boldsymbol{n}}\psi\,\mathrm{d}\boldsymbol{s} + \int_\Omega \left(\mu^- - \mu^+\right)\psi\,\mathrm{d}\boldsymbol{x} \geq 0$$

with $\mu^\pm \geq 0$ together with the additional restrictions in Equation (6.21). This in principle also makes the argument pertaining to the correct boundary conditions a little more complicated. The simplest case is again the one if there is some neighborhood of the boundary where $0 < \phi < 1$, as then both μ^+ and μ^- vanish in this region and one can additionally (locally) choose ψ freely, leading, as before, to $\frac{\partial a}{\partial\boldsymbol{n}} = 0$. If in contrast $\phi = 0$ or $\phi = 1$ in some neighborhood of the boundary, either μ^+ or μ^- need not (and in combination with w_{ob} normally will not) vanish. Nevertheless, since this means that one locally has a bulk-region, there is no need to pass through an indirect argument using the test-functions as this implies through an obvious limit that $\nabla\phi = 0$ on the boundary and thus obviously also $\frac{\partial\phi}{\partial\boldsymbol{n}} = 0$.

What is tricky though is the case when the boundary is not locally in a bulk-region, but ϕ tends to one of the critical values 0 or 1 as one approaches the boundary. Even though μ^+ and μ^- and therefore the volume-integral would locally vanish, one would face a sign restriction on the variations of ψ on the boundary[16] and would thus a priori only be able to conclude that $\frac{\partial\phi}{\partial\boldsymbol{n}} \geq 0$

[16]Note that this does not follow from e.g. the constraint $0 \leq \phi \leq 1$ a.e. in Ω by itself, since the boundary is a set of measure zero. It is only in combination with the additional regularity - such as the one imposed through the space $H^1(\Omega)$ - that this as well as talking about the values of ψ on the boundary becomes meaningful.

if $\phi \to 0$ resp. $\frac{\partial \phi}{\partial \boldsymbol{n}} \leq 0$ if $\phi \to 1$ as one approaches the boundary[17]. A strict inequality in these conclusions is in contradiction with the assumptions on the profile though, as this to first order means that any point in the boundary with $\phi = 0$ would have to be approached from **below** and any point with $\phi = 1$ from **above**. Combining all these observations, the first-order necessary conditions can finally be summarized to

$$\begin{cases} -2\gamma\epsilon\Delta\phi + \frac{1}{\epsilon}w'(\phi) + f'(\phi) + \mu^+ - \mu^- = 0 & \text{in } \Omega, \\ 0 \leq \phi \leq 1,\ \mu^{\pm} \geq 0,\ \mu^+(1-\phi) = 0 \text{ and } \mu^-\phi = 0, & \\ \frac{\partial \phi}{\partial \boldsymbol{n}} = 0 & \text{on } \partial\Omega, \end{cases} \tag{6.22}$$

respectively in the general potentially anisotropic case

$$\begin{cases} -\epsilon\nabla \cdot \left(\frac{\partial a}{\partial \nabla \phi}\right) + \frac{1}{\epsilon}w'(\phi) + f'(\phi) + \mu^+ - \mu^- = 0 & \text{in } \Omega, \\ 0 \leq \phi \leq 1,\ \mu^{\pm} \geq 0,\ \mu^+(1-\phi) = 0 \text{ and } \mu^-\phi = 0. & \\ \frac{\partial a}{\partial \boldsymbol{n}} = 0 & \text{on } \partial\Omega, \end{cases} \tag{6.23}$$

Remark 40. It needs to be stressed that, while in principle consistent, the argument above is purely formal. Making it rigorous requires quite a bit of additional work and raises a number of - often also practically important - questions.
Besides the derivation itself, one central question is of course the sense in which the various quantities and equations should be interpreted and is thus essentially a matter of function spaces. This point is more delicate in problems involving obstacles than for other more "global" constraints or unconstrained problems. In particular, whereas the solutions of many classical equations in the applied physics based on a Laplace-type operator are known to be as regular as the data permits (and in particular C^∞ in the interior of the domain if the right-hand side is so), the same is not true for obstacle-type problems. In fact, even though both obstacles here are given by constants (and are thus extremely regular), even the "optimal" unperturbed one-dimensional phasefield profile has, as will be recalled in Remark 47, a jump in its second derivative upon crossing from the bulk to an interface region. For this reason, it is clear that one should certainly not expect for Equation (6.23) to hold in the classical sense.
An also in practice very important question is therefore whether it does at least hold in the strong (a.e.) sense in e.g. $L^2(\Omega)$, or whether it has to be interpreted in an even weaker sense. Even though $\phi \in H^1(\Omega)$ does at least (the trace-operator being well-defined) ensure a "continuity through hypersurfaces", the same does not a priori hold for the gradient of ϕ. On the one hand, jumps in the gradient of ϕ would be expected to significantly increase numerical errors due to a reduction in the convergence order of any standard numerical scheme. On the other hand, Equation (6.23) could then, through the second derivative in its first term, contain surface measures. As both w' and any reasonably chosen f' are bounded for $0 \leq \phi \leq 1$, this contribution could only be counterbalanced by equally measure-valued multipliers $\boldsymbol{\mu}^{\pm}$. While this has no precise meaning in the discrete case, any convergent numerical scheme relying (explicitly or implicitly) on these multipliers would have to mimic this behavior as the discretization is refined. This requires for the discrete counterpart of $\mu^{\pm}$ to locally tend to ∞ in some sense, and is thus highly likely to cause additional numerical difficulties.
Fortunately, both theoretical predictions (see e.g. [28] for some regularity results for obstacle problems) and computational practice[18] indicate that this issue does usually not arise. Nevertheless, the treatment of the transition region between the interface and the bulk remains a

[17] This is a technically somewhat difficult point as $\frac{\partial \phi}{\partial \boldsymbol{n}}$ need not be "intrinsically" defined on hypersurfaces as $\phi \in H^1(\Omega)$ by itself does not provide sufficient regularity for this. A simple example illustration of this in the one-dimensional case will be discussed in Remark 46 (the interested reader is referred to [48] for a much more general discussion of this topic). The central assumption made here is that $\mu^{\pm}$ are both integrable functions and not measure-valued.

[18] In the sense that calculated phasefield profiles do indeed exhibit a behavior consistent with a continuous transition of $\nabla\phi$ to $\mathbf{0}$ as one approaches the bulk.

somewhat delicate point from a numerical point of view[19].
For some related discussion in the phasefield context, the reader is referred to e.g. [60]. More general background within a more abstract setting (including e.g. also the question of the regularity of free boundaries) can, among many other sources, be found in [43], [28], [37] and [6] and the references therein. ◇

Remark 41. It should also be noted that, within the bulk-regions, if $f'(0) = f'(1) = 0$, μ^+ resp. μ^- are given by the constant value $\frac{16\gamma}{\pi^2\epsilon}$ as the only "active" term in Equation (6.23) is given by the derivative of the bulk-potential w in 0 or 1[20]. In order to eliminate this dependence on ϵ, it can therefore sometimes be convenient (as in [60]) to simply rescale the multipliers by instead using $\hat{\mu}^\pm := \epsilon\mu^\pm$, i.e. modifying Equation (6.23) to

$$-\epsilon\nabla\cdot\Big(\frac{\partial a}{\partial\nabla\phi}\Big) + \frac{1}{\epsilon}w'(\phi) + f'(\phi) + \frac{1}{\epsilon}\big(\hat{\mu}^+ - \hat{\mu}^-\big) = 0$$

subject to the same complementarity conditions as $\mu^\pm$. ◇

Remark 42. If additional constraints are added, the situation also becomes much more involved at a practical level, a very common and quite representative example being the volume-constraint on the phases. Even though this is in principle quite simple at a theoretical level - the integral constraint being very robust with respect to e.g. regularity issues and requiring only a very low-dimensional multiplier - the phasefield equation now involves both a nonlinear but **local** inequality constraint, and a linear but **global** constraint on the phasefield values.
This is in particular a difficulty for numerical schemes avoiding the use of multipliers through a projection-based approach. While the projection operations associated with each constraint separately are relatively straightforward to deal with (at least provided the outer algorithm can make use of the strict locality of the box-constraint), the actually required projection would be a both global and nonlinear one, leading to a very complex and expensive operation.
In practical terms, this can be dealt with in a quite satisfactory manner by instead satisfying a slight modification of the actual first-order necessary condition introducing an additional modified driving force as in [53] (see also [29]). Through an additional simplification (in the obstacle case), this can be turned into a quite notably more efficient algorithm based on purely sequential operations. It should be kept in mind though that these modification are primarily algorithmically motivated, and the solution thus corresponds to a (usually very accurate for the time-stepping scheme actually used in [53]) approximation of the one minimizing the underlying functional. ◇

In contrast to the conditions characterizing a minimizer of $\mathcal{F}_\epsilon$ in the "steady-state" case above, the dynamics of the problem are not directly implied by the functional and the constraints alone. The standard approach in the Allen-Cahn case is that of a (potentially weighted) L^2-gradient flow, i.e. in the unconstrained case one posulates $\frac{\partial\phi}{\partial t} \sim -\frac{\delta\mathcal{F}_\epsilon}{\delta\phi}$. Assuming for the proportionality to be of the form $\tau\epsilon$ and combined with the expression for the gradient in Equation (6.19), one obtains the dynamic counterpart

$$\tau\epsilon\frac{\partial\phi}{\partial t} = \epsilon\nabla\cdot\Big(\frac{\partial a}{\partial\nabla\phi}\Big) - \frac{1}{\epsilon}w'(\phi) - f'(\phi) \tag{6.24}$$

to Equation (6.19) in the double-well case.

Remark 43. The prefactor $\tau\epsilon$ is the "default" version within the **Pace3D**-framework, with τ potentially depending on e.g. $\nabla\phi$ in order to also enable the incorporation of anisotropic effects into the dynamics in addition to the ones in the gradient energy density. The additional use of

[19] One consequence of the expected jump in the second derivative being that one does not expect a formally fourth (or even higher order) scheme to actually achieve this increase in accuracy.

[20] Even if f' does not vanish for $\phi \in \{0,1\}$ - as when using $h(\phi) = \phi$ - the other term is generally largely dominant.

ϵ in contrast does not by itself have any direct effect in this context as one could equally well rescale time. As an alternative to the use of τ, one could also use its inverse $m := \frac{1}{\tau}$ as a prefactor on the right-hand side. In contrast to the multiphase situation in Section 6.3, this distinction is of little interest for two-phase problems. ◇

The situation is again somewhat more complicated when using the obstacle-potential (or more precisely, when enforcing the box-constraint on ϕ) as it naturally leads to counter-forces for any change in the phasefield which would lead outside the admissible set. Arguing either in a manner similar to the one for the steady-state case or simply "generalizing" the expression in Equation (6.23) by adding a time-derivative compatible with the FONC, one obtains the evolution equation

$$\begin{cases} \tau\epsilon\frac{\partial\phi}{\partial t} = \epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi}\right) - \frac{1}{\epsilon}w'(\phi) - f'(\phi) - \left(\mu^+ - \mu^-\right), \\ 0 \le \phi \le 1,\ \mu^\pm \ge 0,\ \mu^+(1-\phi) = 0 \text{ and } \mu^-\phi = 0. \end{cases} \tag{6.25}$$

Remark 44. Both Equations (6.24) and (6.25) still need to be complemented, as before, with the boundary condition $\frac{\partial a}{\partial \boldsymbol{n}} = 0$. ◇

6.2.2 The One-Dimensional Case and the "Standard" Phasefield Profiles

Before continuing the discussion above, it is instructive to take a look at the simple (parameterized families of) analytical solutions for the steady-state problems in the well and obstacle case in their simplest one-dimensional form, as these on the one hand indicate some qualitative properties expected in the more general case and on the other hand are the profiles underlying the link between the diffuse and sharp interface models as well as the thin interface asymptotics. In this setting, one is thus looking for a function ϕ satisfying

$$-2\gamma\epsilon\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} + \frac{1}{\epsilon}\frac{\mathrm{d}w}{\mathrm{d}\phi}(\phi) + \frac{\mathrm{d}f}{\mathrm{d}\phi}(\phi) = 0 \tag{6.26}$$

within the interface region[21], with w being either the double-well or double-obstacle bulk potential. An important difference between these two cases is that $\frac{\partial w_{do}}{\partial\phi} = \frac{16\gamma}{\pi^2}(1-2\phi)$ is an affine function - thus allowing for applying the standard linear theory, at least within the interface region and provided $\frac{\partial f}{\partial\phi}$ is sufficiently simple - whereas $\frac{\partial w_{dw}}{\partial\phi}$ is a cubic function of ϕ, and requires other tools for determining an analytical solution.
The standard approach for the latter case is based on the existence of a first integral for (6.26), a fact which has its own inherent interest from an energetic point of view also in the obstacle case. More precisely, a multiplication of Equation (6.26) with $\frac{\mathrm{d}\phi}{\mathrm{d}x}$ and using[22] $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}\left|\frac{\mathrm{d}\phi}{\mathrm{d}x}\right|^2$ as well as $\frac{\mathrm{d}w}{\mathrm{d}\phi}(\phi)\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}w(\phi)$ and $\frac{\mathrm{d}f}{\mathrm{d}\phi}(\phi)\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}f(\phi)$[23] shows that

$$\left(-2\gamma\epsilon\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}\frac{1}{\epsilon}\frac{\mathrm{d}w}{\mathrm{d}\phi}(\phi) + \frac{\mathrm{d}f}{\mathrm{d}\phi}(\phi)\right)\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(-\gamma\epsilon\left|\frac{\mathrm{d}\phi}{\mathrm{d}x}\right|^2 + \frac{1}{\epsilon}w(\phi) + f(\phi)\right) = 0$$

and thus that

$$-\gamma\epsilon\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)^2 + \frac{1}{\epsilon}w(\phi) + f(\phi) = \text{const} \tag{6.27}$$

[21] The restriction to this region avoids the issue of the multipliers $\mu^\pm$ in the obstacle case.

[22] More generally, one could use the relation $-\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial a}{\partial(\frac{\mathrm{d}\phi}{\mathrm{d}x})}\right)\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(a(\frac{\mathrm{d}\phi}{\mathrm{d}x}) - \frac{\partial a}{\partial(\frac{\mathrm{d}\phi}{\mathrm{d}x})}\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)$. The use of an anisotropic a-function seems of little interest in the one-dimensional case though.

[23] Valid as long as the only dependence on x in w resp. f arises solely through that of $\phi(x)$.

within any region where one does not have to consider the bound constraints, i.e. within the region $0 < \phi(x) < 1$ in the obstacle case and everywhere in the domain in the double-well case. Besides the very useful information contained in Equation (6.27) from an energetic point of view, an advantage (at least in the nonlinear double-well case) over the original second-order differential equation is that it directly implies

$$\sqrt{\gamma\epsilon}\frac{\mathrm{d}\phi}{\mathrm{d}x} = \pm\sqrt{\frac{1}{\epsilon}w(\phi) + f(\phi) - \text{const.}}$$

This gives, within each region of monotonicity of ϕ, rise to a separable first-order ODE and thus turns the problem of determining a solution to the original equation into the potentially simpler one of determining an indefinite integral, whose solution one can then try to match to the appropriate boundary conditions.

For some particular cases, this can in fact be done analytically. Assuming for example that w is given by the double-well potential, that there is no driving force and that the domain is an infinite one, it turns out that the constant in Equation (6.27) can be taken to be zero, and that the solutions to the resulting nonlinear ODE $\sqrt{\gamma\epsilon}\frac{\mathrm{d}\phi}{\mathrm{d}x} = \sqrt{\frac{1}{\epsilon}w(\phi)}$ are known to be given by[24]

$$\phi(x) = \frac{1}{2}\Big(1 \pm \tanh\big(\frac{3}{2\epsilon}x + c\big)\Big) \tag{6.28}$$

valid for $-\infty < x < +\infty$.

Remark 45. In practical computations, one of course never deals with actually infinite domains. Even though the solution (6.28) is then not valid, the deviations from this ideal profile are usually very small due to the exponential convergence to the bulk-values 0 and 1. Similarly, in higher-dimensional and multi-interface and/or multiphase simulations involving several interfaces, these "tails" will in priniple lead to an interaction between all phases and all interfaces. As long as there is a reasonable distance between the small but "crucial" parts of the transition regions, the energetic effect of these long-range interactions is typically negligeable as compared to other sources of deviations such as curvature-induced chanes, discretization errors and the changes in the energetics and profiles induced by multiphase regions.

◇

The situation in the double-obstacle case is both simpler and more complex. In fact, using the expression for $\frac{\mathrm{d}w_{do}}{\mathrm{d}\phi}$, the original second-order PDE (6.26) reduces, provided $\frac{\mathrm{d}f}{\mathrm{d}\phi}$ vanishes, to $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} + \frac{16}{\pi^2\epsilon^2}\phi = \frac{8}{\pi^2\epsilon^2}$ within the interface region. It is well-known that the general solution to this is equation is given by the superposition of a particular solution $\phi_p(x)$, which can here be chosen as $\phi_p(x) = \frac{1}{2}$, and a homogeneous solution of the form $\phi_h(x) = \tilde{c}_1\sin\big(\frac{4}{\pi\epsilon}x\big) + \tilde{c}_2\cos\big(\frac{4}{\pi\epsilon}x\big)$ corresponding to the two roots of the characteristic equation $\lambda^2 + \frac{16}{\pi^2\epsilon^2} = 0$. Appropriately choosing the constants, this can equivalently be rewritten as

$$\phi(x) = \frac{1}{2}\Big(1 \pm \tilde{c}_1\sin\big(\frac{4}{\pi\epsilon}x + \tilde{c}_2\big)\Big). \tag{6.29}$$

[24] One way of obtaining this is through a partial fraction decomposition. Inserting the definition of w and restricting the focus to the range $0 < \phi < 1$ leads to $\frac{\mathrm{d}\phi}{\mathrm{d}x} = \pm\frac{3}{\epsilon}\phi(1-\phi)$. After a separation of the variables, ϕ thus has to satisfy

$$\int \frac{\mathrm{d}\phi}{\phi(1-\phi)} = \int \frac{1}{\phi} + \frac{1}{1-\phi}\mathrm{d}\phi = \ln(\phi) + \ln(1-\phi) = \ln\Big(\frac{\phi}{1-\phi}\Big) = \pm\frac{3}{\epsilon}\int \mathrm{d}x + \tilde{c} = \pm\frac{3x}{\epsilon} + \tilde{c}.$$

Taking exponentials on both sides simplifies this to $\frac{\phi}{1-\phi} = e^{\pm\frac{3x}{\epsilon}+\tilde{c}}$, or, solving for ϕ,

$$\phi = \frac{e^{\pm\frac{3x}{\epsilon}+\tilde{c}}}{ce^{\pm\frac{3x}{\epsilon}+\tilde{c}}+1} = \frac{1}{2}\left(\frac{e^{\pm\frac{3x}{\epsilon}+\tilde{c}}+1}{ce^{\pm\frac{3x}{\epsilon}+\tilde{c}}+1} + \frac{e^{\pm\frac{3x}{\epsilon}+\tilde{c}}-1}{ce^{\pm\frac{3x}{\epsilon}+\tilde{c}}+1}\right) = \frac{1}{2}\left(1 + \frac{e^{\pm\frac{3x}{\epsilon}+\tilde{c}}-1}{ce^{\pm\frac{3x}{\epsilon}+\tilde{c}}+1}\right).$$

From this, the result follows in combination with $\tanh(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}} = \frac{e^{2y}-1}{e^{2y}+1}$, defining $c = \pm\frac{\tilde{c}}{2}$ and using the antisymmetry of the tanh-function.

In order to obtain an acceptable solution describing the transition between two bulk-regions, it remains to adjust the constants $\tilde{c}_1$ and $\tilde{c}_2$. The right guess ensuring this (renaming $\tilde{c}_2$ to c for consistency with the double-well solution in Equation (6.28)) is to set

$$\phi(x) = \frac{1}{2}\Big(1 \pm \sin\big(\frac{4}{\pi\epsilon}x + c\big)\Big), \quad x \in \big(-c - \frac{\pi^2}{8}\epsilon, -c + \frac{\pi^2}{8}\epsilon\big), \tag{6.30}$$

prolongated by the respective bulk values 0 and 1 outside this inner region.

Remark 46. Even though it is intuitively clear that the profile in Equation (6.30) is indeed the correct one as it is the one matching the "most smoothly" with the bulk-regions, this is not at all obvious based on Equation (6.26) and the condition of having to reach the values of 0 and 1. In fact, the (here implicit) condition that ϕ should at least lie in H^1 implies, in the one-dimesional setting, that ϕ at least has to be continuous. Based on Equation (6.29), this can be achieved using any constant $\tilde{c}_1 \geq 1$ and then simply truncating the profile to the bulk-values at the points where it starts going outside the bounds. Fixing for convenience the constant $\tilde{c}_2$ to zero (corresponding to centering the interface at $x = 0$) and choosing for definiteness the increasing profile, this still leaves a one-parameter family of potential solutions $\phi(x) = \frac{1}{2}\Big(1 + \tilde{c}_1 \sin\big(\frac{4}{\pi\epsilon}x\big)\Big)$ based on the ODE withing the interface and the continuity condition alone, and choosing the correct one therefore requires some additional information.

The safest way to argue here is by simply considering the total phasefield energy as function of the parameterization in terms of $\tilde{c}_1$ and to choose the relevant solution accordingly. A simple calculation shows that, given $\tilde{c}_1 \geq 1$, the interface defined as the region between the points where the profile first touches 0 resp. 1 will then extend from $x_l := -\frac{\epsilon\pi}{4}\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)$ to $x_r := \frac{\epsilon\pi}{4}\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)$. Further inserting this profile into the definitions of the gradient and bulk energy densities leads to

$$\epsilon a\big(\frac{\mathrm{d}\phi}{\mathrm{d}x}\big) = \gamma\frac{4}{\pi^2\epsilon}\tilde{c}_1^2\cos^2\big(\frac{4}{\pi\epsilon}x\big) \quad \text{and} \quad \frac{1}{\epsilon}w(\phi) = \gamma\frac{4}{\pi^2\epsilon}\Big(1 - \tilde{c}_1^2\sin^2\big(\frac{4}{\pi\epsilon}x_2\big)\Big),$$

based on which a simple calculation[25] shows that

$$\mathcal{F}_\epsilon(\tilde{c}_1) = \int_{x_l}^{x_r} \epsilon a\big(\frac{\mathrm{d}\phi}{\mathrm{d}x}\big) + \frac{1}{\epsilon}w(\phi)\,\mathrm{d}x = \frac{2\gamma}{\pi}\Big(\sqrt{\tilde{c}_1^2 - 1} + \sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)\Big).$$

Even though the first term is increasing in $\tilde{c}_1$ and the second one decreasing, differentiation this expression shows that

$$\frac{\mathrm{d}\mathcal{F}_\epsilon(\tilde{c}_1)}{\mathrm{d}\tilde{c}_1} = \frac{2\gamma}{\pi}\frac{\tilde{c}_1 - \frac{1}{\tilde{c}_1}}{\sqrt{\tilde{c}_1^2 - 1}},$$

i.e. the contribution by the first term is dominant and the energy is increasing for all $\tilde{c}_1 > 1$. The lowest energy is thus indeed achieved for the minimal admissible $\tilde{c}_1 = 1$, i.e. the previously obtained profile.

This is also a somewhat tedious argument as it in particular relies heavily on explicitly calculating the total energy. A much faster argument is to proceed analogously as in the previous section, which is in the current case much simpler to make rigorous as there are no issues with the

[25] Adding both contributions and using $\cos^2(\theta) - \sin^2(\theta) = \cos(2\theta)$ shows that the integral can be rewritten as

$$\mathcal{F}_\epsilon(\tilde{c}_1) = \frac{4\gamma}{\pi^2\epsilon}\int_{x_l}^{x_r}\Big(\tilde{c}_1^2\cos\big(\frac{8}{\pi\epsilon}x\big) + 1\Big)\,\mathrm{d}x = \frac{\gamma}{\pi}\Big(\tilde{c}_1^2\sin\Big(2\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)\Big) + 2\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)\Big).$$

This simplifies further by inserting $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ to get $\tilde{c}_1^2\sin\Big(2\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)\Big) = 2\tilde{c}_1\cos\Big(\sin^{-1}\big(\frac{1}{\tilde{c}_1}\big)\Big)$, from which the result follows in combination with $\cos\big(\sin^{-1}(\theta)\big) = \sqrt{1-\theta^2}$.

regularity of the domain (which is a nonempty interval here) and the profile is known (and very smooth) within the interface and the interior of its complement[26]. ◇

Remark 47. Despite the fact that the analytical solution ϕ is C^1 and even C^∞ within the interface and its complements, the use of this monotonous transition profile requires "violating" Equation (6.26), and thus, by the complementary condition, re-introducing the multipliers $\mu^\pm$. The only regions where ϕ need not satisfy this equation are the ones where $\phi = 0$ or $\phi = 1$, leading to $(\mu^-, \mu^+) = (\frac{16\gamma}{\pi^2\epsilon}, 0)$ (resp. the other way around). In addition, it is easy to see that it is not possible to match this inner solution in a C^2-manner with the (constant) outer regions as - by the nature of the sin-function - the only points where $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = \frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2} = 0$ are the ones where the homogeneous solution itself vanishes and therefore $\phi(x) = \frac{1}{2} \notin \{0, 1\}$. Whereas the second derivative in the outside region is obviously zero, the limits from the interior are given by $\pm\frac{8}{\pi^2\epsilon^2}$, showing in particular that the phasefield equation with the obstacle potential above does not allow for a classical solution unless it is sinusoidal everywhere or if there is no interface at all (i.e. $\phi(x) \equiv 0$ or 1). ◇

Remark 48. The jump in the second derivatives of course also has numerical consequences. Taking the analytical profile and applying e.g. a centered finite difference scheme for the second derivative at a point at the transition between the bulk and the interface, the discrete estimate for the (non-existent) second-order derivative at this point will convergence to the average of the second derivatives on the left and right of the transition point[27] and thus, one of them being zero, to half the second derivative corresponding to the limit from the interior. This is of course not a problem per se as there not being a "correct" value to converge to at this point, the average is actually a quite pleasing limit. Nevertheless, a similar effect will happen for any point whose stencil crosses the purported transition point. Taking for definiteness the transition between a zero bulk-region and the interior, a point just outside the interface will overestimate the (vanishing) second derivative at this point, whereas one just on the inside will underestimate it, with the degree of both effects depending on the relative positions. Starting a discrete simulation from the analytical profile, one would therefore in particular expect the value of this

[26] Assuming the actual domain Ω to be of the form $\Omega = (-R, R)$ with some sufficiently large R such that it fully contains the interface (the energetic contributions in the bulk vanishing), the first order necessary condition in its weak form is given by

$$\int_{-R}^{R} 2\gamma\epsilon\frac{\mathrm{d}\phi}{\mathrm{d}x}\frac{\mathrm{d}\psi}{\mathrm{d}x} + \frac{1}{\epsilon}w'(\phi)\psi\,\mathrm{d}x = \int_{(-R,x_l)\cup(x_r,R)} \frac{1}{\epsilon}w'(\phi)\psi\,\mathrm{d}x + \int_{x_l}^{x_r} 2\gamma\epsilon\frac{\mathrm{d}\phi}{\mathrm{d}x}\frac{\mathrm{d}\psi}{\mathrm{d}x} + \frac{1}{\epsilon}w'(\phi)\psi\,\mathrm{d}x \geq 0$$

for all admissible directions ψ. Due to the smoothness of ϕ in the interface, the integration by part (corresponding to the use of Gauss's theorem in the multidimensional case) leading to

$$\begin{aligned}&\int_{(-R,x_l)\cup(x_r,R)} \frac{1}{\epsilon}w'(\phi)\psi\,\mathrm{d}x + 2\gamma\epsilon\Big(\frac{\mathrm{d}\phi}{\mathrm{d}x}(x_r^-)\psi(x_r) - \frac{\mathrm{d}\phi}{\mathrm{d}x}(x_l^+)\psi(x_l)\Big) - \int_{x_l}^{x_r}\Big(-2\gamma\epsilon\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} + \frac{1}{\epsilon}w'(\phi)\Big)\psi\,\mathrm{d}x\\ = &\int_{(-R,x_l)\cup(x_r,R)} -\frac{16\gamma}{\pi^2\epsilon}\psi\,\mathrm{d}x + 2\gamma\epsilon\Big(\frac{\mathrm{d}\phi}{\mathrm{d}x}(x_r^-)\psi(x_r) - \frac{\mathrm{d}\phi}{\mathrm{d}x}(x_l^-)\psi(x_l)\Big) \geq 0\end{aligned}$$

in the interface is now easily justified provided ψ is moderately smooth and one takes the limits "from the interior" for the values of $\frac{\mathrm{d}\phi}{\mathrm{d}x}$. As there are no gradients involved in the expression for the bulk-region, one can then simply take e.g. a sufficiently narrow hat-function of the appropriate sign centered around each endpoint for ψ to conclude that $\frac{\mathrm{d}\phi}{\mathrm{d}x}(x_l)$ cannot be positive and $\frac{\mathrm{d}\phi}{\mathrm{d}x_r}$ cannot be negative, contradicting the assumed shape of ϕ unless $\tilde{c}_1 = 1$.

[27] This is easily seen by a two-sided Taylor-expansion, which, while not legitimiate **through** the discontinuity of the second derivative at the transitions point, is valid in both the left and right region of the endpoints. Fixing the position x_i as one of two transition points x_l and x_r, a simple Taylor expansion using the left resp. right limit of the second derivative gives $\phi_{i-1} = \phi(x_i - \Delta x) = \phi(x_i) - \phi'(x_i)\Delta x + \frac{1}{2}\phi''(x_i^-)(\Delta x)^2 + \mathcal{O}\big((\Delta x)^3\big)$ and $\phi_{i+1} = \phi(x_i + \Delta x) = \phi(x_i) + \phi'(x_i)\Delta x + \frac{1}{2}\phi''(x_i^+)(\Delta x)^2 + \mathcal{O}\big((\Delta x)^3\big)$, from which the evaluation of the discrete (3-point stencil based) second derivative is obtained as $\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{(\Delta x)^2} = \frac{1}{2}\big(\phi''(x_i^-) + \phi''(x_i^+)\big) + \mathcal{O}(\Delta x)$.

first inner point to start to drop[28]. Similarly, the last inner point near $\phi = 1$ is expected to rise. This can serve to provide an intuitive explanation for the observation (see Figure 6.1 and the next section) that the resulting discrete phasefield profile is generally slightly more narrow and steeper than the continuous one if the last "inner" point is close to the transition to the bulk.

A further quite interesting observation (discussed in more detail in Appendix A) is that, at least without driving forces, there are generically (i.e. except for a discrete sequence of ratios of $\frac{\epsilon}{\Delta x}$) two possible solutions to the discrete first-order necessary conditions corresponding to Equation (6.22) with only one of them actually minimizing the discrete energy. For low resolutions, these differ quite notably as is also illustrated in Figure 6.1. There, N corresponds to the total distance in grid-spacings between the last discrete ϕ-value at 0 and the first one at 1, there thus being a total of $N-1$ points within the interior of the discrete interface (i.e. with $0 < \phi_i < 1$) and a total of $N+1$ when the two endpoints at 0 resp. 1 are included.
The discrete solution corresponding to $N = 4$ experiences precisely the effect just outlined above. As the difference stencil of the last discrete point at $\frac{x}{\epsilon} = 1$ (an integer here for the choice $\epsilon = 2$) is roughly at a distance of $\frac{\Delta x}{2}$ from the outer limit of the continuous profile, it will significantly underestimate the "true" curvature at this point and is thus forced upwards (into the upper bound at 1) as this weakens the contribution by the second derivative as compared to the bulk potential one.
In contrast, the solution corresponding to $N = 5$ is in a sense "numerically optimal" as the last discrete position is almost perfectly aligned with the right endpoint of the discrete profile (and only very slightly outside the actual interface indicated by the vertical dashed line). For this reason, the evaluation of the second derivative of the analytical solution at the last inner point is essentially unaffected by the jump discontinuity and thus quite precise (up to the standard $\mathcal{O}\big((\Delta x)^2\big)$ error, see Footnote 28). Even though the (negative) second derivative "just outside" the interface at the last point is highly overestimated, this only amounts (in line with the above) to roughly half of what is required to counterbalance the forcing by the bulk potential term. This leads to an underestimation of the corresponding multiplier μ^+ by again roughly $\frac{1}{2}$, but this is not sufficient to move the right-most point away from one. The jump of the second derivative is thus essentially "hidden" in the multipliers, explaining the very close numerical agreement for the profile itself. In this respect, it can also be noted that the error in the multiplier is much lower for the more narrow profile.
While this may therefore seem very favorable from an approximation point of view, the discussion in the next section and the Appendix A show that on energetic grounds the numerics should a priori favor the **other** profile with $N = 4$ as it has a (slightly) lower total "discrete" energy (of $\approx 0.981\gamma$ as compared to $\approx 0.984\gamma$ for $N = 5$). ◇

[28] The discrete second-order derivatives are accurate to second order at all points not in the neighborhood of the transition region and thus "almost" in equilibrium with the local $w'(\phi)$-values. Even though a closer look reveals that the discrete three-point stencil applied to the analytical profile also underestimates the magnitude of the actual second derivative by a factor $2\frac{\cos(\Delta x)}{\Delta x} = 1 - \frac{2}{4!}(\Delta x)^2 + \mathcal{O}\big((\Delta x)^4\big)$, this is a priori nowhere near the reduction by a factor of up to 2 at the transition points. Nevertheless, this (perhaps) surprisingly turns out to have an effect of roughly the same order of magnitude (see Remark 66 for a more detailed discussion).

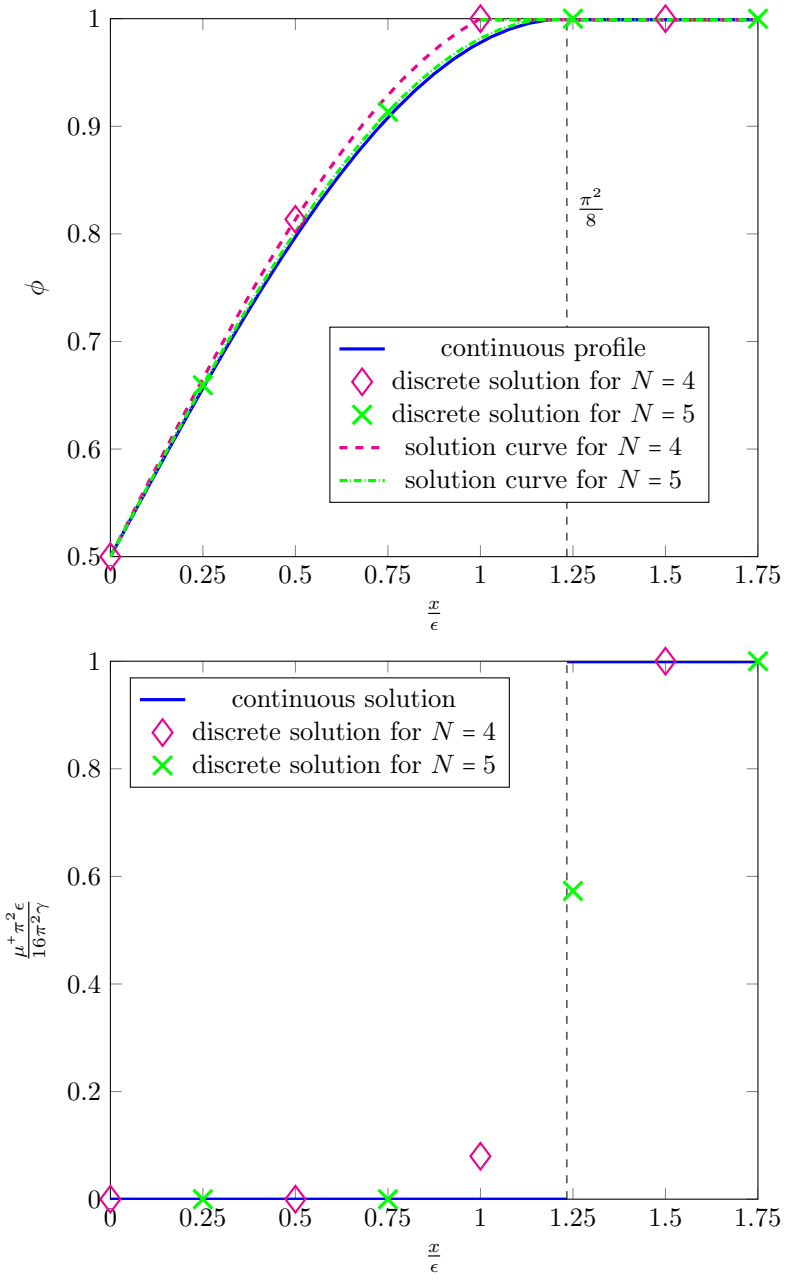


Figure 6.1: Comparison of the right half of the ϕ-profiles and their corresponding multiplier μ^+ for the continuous and the two (!) discrete solutions of the discrete analogue of the first-order necessary conditions in Equation (6.22) for $\epsilon = 2\Delta x$ (see Remark 48 and the next Subsection 6.2.3 for a precise definition of N).
For the discrete profiles, the dashed resp. dash-dotted lines in addition show the underlying sinusoidals on which the solution values lie (see Equation (6.48)). Note that while the curve for $N = 5$ is closer to the analytical profile, it is actually the energetically unfavorable one.

6.2.3 Some Analytics for the Discrete Case

The Discrete One-Dimensional Minimizers

A similar though somehwat heavier calculation for the obstacle potential can in fact, again relying on the linearity, be reproduced in the discrete setting and sheds some light on the differences between the continuous and discrete behaviour within the simple one-dimensional setting. Using a standard centered difference discretization for the Laplacian and discretizing w using the values at the cell centers, the discrete one-dimensional phasefield equation in the absence of driving forces is given by

$$-2\gamma\epsilon\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{(\Delta x)^2}+\frac{16}{\pi^2\epsilon}\gamma(1-2\phi_i)=0. \tag{6.31}$$

By linearity, one can again decompose the solution as the sum of the same particular solution $\phi_i^p=\frac{1}{2}$ as in the continuous case and the homogeneous solution ϕ_i^h satisfying

$$\phi_{i+1}^h-2\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)\phi_i^h+\phi_{i-1}^h=0.$$

From the standard theory of difference equations (see e.g. [59]), the solution can be derived in terms of the solutions λ of the characteristic equation[29]

$$\lambda^2-2\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)\lambda+1=\Big(\lambda-\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)\Big)^2+\left(1-\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)^2\right)=0,$$

which, for $\left(\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\right)^2\leq 2$, are given by

$$\alpha+\boldsymbol{i}\beta=\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)\pm\boldsymbol{i}\sqrt{1-\Big(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)^2}$$

where $\boldsymbol{i}=\sqrt{-1}$ and $\alpha^2+\beta^2=1$. The general homogeneous solution is therefore given by

$$\phi_i^h=c_1\cos(\kappa i)+c_2\sin(\kappa i)$$

with

$$\cos(\kappa)=1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2}=1-\frac{1}{2}\Big(\frac{4\Delta x}{\pi\epsilon}\Big)^2 \tag{6.32}$$

$$\sin(\kappa)=\sqrt{1-(1-\frac{8(\Delta x)^2}{\pi^2\epsilon^2})^2}=\sqrt{2\frac{8(\Delta x)^2}{\pi^2\epsilon^2}-\Big(\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)^2}=\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2}-1}. \tag{6.33}$$

Due to the translation invariance of the problem, one may for simplicity assume that $\phi_0=0$, and that the difference equation holds from ϕ_1 up to ϕ_{N-1}. As this imposes $\phi_0^h=-\frac{1}{2}$, this fixes the first constant as $c_1=-\frac{1}{2}$. Similarly, using the condition

$$\phi_N^h=-\frac{1}{2}\cos(\kappa N)+c_2\sin(\kappa N)\stackrel{!}{=}\frac{1}{2} \tag{6.34}$$

[29] The approach here is similar to the one in the continuous case. Instead of the ansatz $\phi_h(x)=e^{\lambda x}$ based upon which one can obtain the characteristic equation underlying the homogeneous solution in the continuous double-well case, one instead postulates $\phi_i=\lambda^i$ (with generally complex λ), from which one then obtains the characteristic equation by extracting the highest common power of λ. In the case with two complex-conjugate λ's below, it is then convenient to rewrite λ^i as $e^{i\ln\lambda}=\cos\big(\Re(\ln\lambda)\big)+\boldsymbol{i}\sin\big(\Im(\ln\lambda)\big)$. Further rewriting λ in polar form as $\lambda=|\lambda|\big(\cos(\kappa)+\boldsymbol{i}\sin(\kappa)\big)$ and arguing precisely as in the continuous case then leads to the representation used here.

at the right endpoint leads to

$$c_2 = \frac{1+\cos(\kappa N)}{2\sin(\kappa N)} = \frac{1}{2}\cot\left(\frac{\kappa N}{2}\right) \tag{6.35}$$

unless κN is a multiple of π and fixes the second constant in terms of the (as of yet unknown!) number of points N, leading to the discrete profile

$$\phi_i = \frac{1}{2} - \frac{1}{2}\cos(\kappa i) + \frac{1}{2}\cot\left(\frac{\kappa N}{2}\right)\sin(\kappa i). \tag{6.36}$$

Remark 49. If κN is multiple of π, equation (6.34) does not impose any restriction on c_2 as $\cos(\kappa N) = 1$ and $\sin(\kappa N) = 0$, implying that this equation is satisfied for any c_2. This particular case (for $\kappa N = \pi$) turns out to be a very interesting one as will be seen towards the end of this section. In order to avoid mixing two quite different lines of argument, the following paragraph will first consider the case where $\frac{\pi}{\kappa}$ is noninteger. ◇

The case $\frac{\pi}{\kappa}$ noninteger All which remains to be done is therefore to determine the width of the numerical interface in terms of N. A first point to be noted is that the assumption of ϕ_1 being the first inner interface point already imposes an upper on N as this requires

$$\phi_1 = \frac{1}{2} - \frac{1}{2}\cos(\kappa) + \frac{1}{2}\cot\left(\frac{\kappa N}{2}\right)\sin(\kappa) > 0.$$

Since $\sin(\kappa) > 0$, this translates to

$$\cot\left(\frac{\kappa N}{2}\right) > \frac{\cos(\kappa)-1}{\sin(\kappa)} = -\tan\left(\frac{\kappa}{2}\right)$$

and thus, restricting the range of $\cot^{-1}$ to the relevant one in $(0,\pi)$ to obtain a single-valued mapping and using $\cot^{-1}\left(\frac{1}{y}\right) = \frac{\pi}{2} - \cot^{-1}(y)$ for $y > 0$ and $\tan(y) = \left(\cot(y)\right)^{-1}$, to

$$N < \frac{2}{\kappa}\left(\pi - \cot^{-1}\left(\tan\left(\frac{\kappa}{2}\right)\right)\right) = \frac{2}{\kappa}\left(\pi - \left(\frac{\pi}{2} - \frac{\kappa}{2}\right)\right) = \frac{\pi}{\kappa} + 1.$$

Since N has to be integer, this together with the strict inequality leads to the maximal admissible number

$$N_{max} = \left\lceil \frac{\pi}{\kappa} \right\rceil \tag{6.37}$$

of points within the interface, where $\lceil\cdot\rceil$ denotes "round-up" operation (i.e. $\lceil n \rceil$ is the smallest integer number greater than or equal to n).
As all $N \leq N_{max}$ would in principle permit the construction of a "geometrically admissible" interface, an additional criterion is necessary for choosing the correct one. A first such criterion is easy to obtain by readding the multipliers $\mu^{\pm}$ corresponding to the constraint $0 \leq \phi_i$ and $\phi_i \leq 1$[30],

$$-2\gamma\epsilon\frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}\gamma(1-2\phi_i) = \mu_i^- - \mu_i^+, \tag{6.38}$$

with $\mu_i^{\pm} \geq 0$ and satisfying $\mu^- = 0$ if $\phi > 0$ and $\mu^+ = 0$ if $\phi < 1$, which extens the validity of the discrete difference Equation (6.31) to the entire domain. It is obvious that within the bulk-region one has either $\mu^- = \frac{16}{\pi^2\epsilon}$ or $\mu^+ = \frac{16}{\pi^2\epsilon}$, the respective other one being zero, and thus multipliers consistent with the restriction in terms of the complementarity condition (6.21). The only interesting points for judging the first-order admissibility for a given N are therefore the

[30] Recall the reasoning for Equation (6.22), which automatically becomes rigorous in the discrete case.

transition points, where there is an additional contribution by the spatial second-order difference operator.
Inserting the profile in Equation (6.36) and looking for simplicity at the left end-point of the interface (the other endpoint being completely analogous) leads to the restriction

$$\mu_0^- = -2\gamma\epsilon\frac{\phi_1 - 2\phi_0 + \phi_{-1}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}\gamma(1-2\phi_0) = \gamma\epsilon\frac{-\cot\left(\frac{\kappa N}{2}\right)\sin(\kappa) + \cos(\kappa) - 1}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}\gamma \overset{!}{\geq} 0 \quad (6.39)$$

since $\phi_{-1} = \phi_0 = 0$ and $\phi_1 = \frac{1}{2} - \frac{1}{2}\cos(\kappa) + \frac{1}{2}\cot\left(\frac{\kappa N}{2}\right)\sin(\kappa)$. Canceling the common factor γ, it follows that N has to be such that

$$\cot\left(\frac{\kappa N}{2}\right) \leq \frac{1}{\sin(\kappa)}\left(\frac{16(\Delta x)^2}{\pi^2\epsilon^2} + \cos(\kappa) - 1\right).$$

By the definition of κ in Equations (6.32) and (6.33), this can be simplified to

$$\cot\left(\frac{\kappa N}{2}\right) \leq \frac{1}{\sin(\kappa)}\left(\frac{16(\Delta\bar{x})^2}{\pi^2} - \frac{8(\Delta\bar{x})^2}{\pi^2}\right) = \frac{8(\Delta\bar{x})^2}{\pi^2\sin(\kappa)} = \frac{1}{\sqrt{\frac{\pi^2}{4(\Delta\bar{x})^2} - 1}}$$

which represents a lower bound on N as the cot-function is decreasing from $+\infty$ at 0 to $-\infty$ at π (this being the interesting range here). More precisely, N has to satisfy $\frac{\kappa N}{2} \geq \cot^{-1}\left(\frac{1}{\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}}\right)$, or, using the inverse formula $\cot^{-1}\left(\frac{1}{y}\right) = \tan^{-1}(y)$ and as N has to be integer,

$$N \geq N_{min} := \left\lceil \frac{2}{\kappa}\tan^{-1}\left(\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}\right)\right\rceil = \left\lceil \frac{\pi}{\kappa} - 1\right\rceil \quad (6.40)$$

Any (integer) N satisfying both equations (6.37) and (6.40) is therefore compatible with the first-order optimality conditions in Equation (6.38). As it is obvious from the expressions for N_{min} and N_{max} that there are always two such integers, this shows that the existence of two potentially relevant profiles in the example in Figure 6.1 is not an exception but actually the rule.

Since the discrete energy is not convex due to the concave contribution by w and this condition is in the non-convex case only necessary but **not** sufficient for characterizing a local minimizer, this naturally raises the question as to the nature of the critical points attained by both profiles. A first insight into this matter can be obtained by comparing the discrete energies

$$\mathcal{E}(N) = \epsilon\sum_{i=0}^{N}\frac{1}{2}\gamma\left(\left(\frac{\phi_i - \phi_{i-1}}{\Delta x}\right)^2 + \left(\frac{\phi_{i+1} - \phi_i}{\Delta x}\right)^2\right) + \frac{16\gamma}{\pi^2\epsilon}\sum_{i=0}^{N}\phi_i(1-\phi_i) \quad (6.41)$$

for both potential profiles (this being the one for which Equation (6.38) is the FONC).

Remark 50. This is similar to the discussion in Remark 46, except that guessing the correct choice is more difficult in the discrete case as enforcing a continuous derivative is meaningless in this setting. ◇

Remark 51. Note that in order to obtain a discrete approximation of the continuous energy, this expression still needs to be multiplied by Δx to take the physical size of each cell into account. The first contribution to this approximation is then given by $\frac{4\gamma}{\pi^2\epsilon}N\Delta x$. As Δx converges to zero, $N\Delta x$ (as the width of the discrete interface) will converge to the width of the analytical one in Equation (6.30) (i.e. to $\frac{\pi^2\epsilon}{4}$) and the first term therefore to the desired value of γ. It further follows that the second has to (and actually does) converge to zero.
A more quantitative evaluation of the energetic deviations requires considering both terms as $N\Delta x$ will only reduce to the correct value in the limit. ◇

Even though this is a rather lengthy calculation (see the Appendix A), the discrete energy contributions in Equation (6.41) can be evaluated analytically, leading to

$$\sum_{i=1}^{N-1} \frac{1}{\epsilon} w(\phi_i) = \frac{8}{\pi^2 \epsilon} \gamma \left(\frac{1}{4} \Big(1 - \cot^2 \big(\frac{\kappa N}{2} \big) \Big) N + \frac{1}{2} \cot \big(\frac{\kappa N}{2} \big) \cot(\kappa) \right) \tag{6.42}$$

and

$$\sum_{i=1}^{N-1} \epsilon a_i = \frac{8}{\pi^2 \epsilon} \gamma \left(\frac{1}{4} \Big(1 + \cot^2 \big(\frac{\kappa N}{2} \big) \Big) N + \frac{1}{2} \frac{\cot \big(\frac{\kappa N}{2} \big)}{\sin(\kappa)} \right) = \frac{8}{\pi^2 \epsilon} \gamma \left(\frac{N}{4 \sin^2 \big(\frac{\kappa N}{2} \big)} + \frac{1}{2} \frac{\cot \big(\frac{\kappa N}{2} \big)}{\sin(\kappa)} \right), \tag{6.43}$$

or, using $1 - \cot^2(\theta) = 2 - \frac{1}{\sin^2 \left(\frac{\kappa N}{2} \right)}$ and $1 + \cot^2(\theta) = \frac{1}{\sin^2(\theta)}$, alternatively to

$$\sum_{i=1}^{N-1} \frac{1}{\epsilon} w(\phi_i) = \frac{8}{\pi^2 \epsilon} \gamma \left(\Big(\frac{1}{2} - \frac{1}{4 \sin^2 \big(\frac{\kappa N}{2} \big)} \Big) N + \frac{1}{2} \cot \big(\frac{\kappa N}{2} \big) \cot(\kappa) \right) \tag{6.44}$$

and

$$\sum_{i=1}^{N-1} \epsilon a_i = \frac{8}{\pi^2 \epsilon} \gamma \left(\frac{1}{4 \sin^2 \big(\frac{\kappa N}{2} \big)} N + \frac{1}{2} \frac{\cot \big(\frac{\kappa N}{2} \big)}{\sin(\kappa)} \right). \tag{6.45}$$

Combining both expressions and making use of the defininining properties of κ (see Appendix A for details), this can in addition be shown to lead to a total energy given by

$$\mathcal{E}(N) = \frac{4\gamma}{\pi^2 \epsilon} N + \frac{\epsilon \gamma}{(\Delta x)^2} \frac{1}{2} \cot \big(\frac{\kappa N}{2} \big) \sin(\kappa) = \frac{4\gamma}{\pi^2 \epsilon} \left(N + \cot \big(\frac{\kappa N}{2} \big) \sqrt{\frac{\pi^2 \epsilon^2}{4(\Delta x)^2} - 1} \right). \tag{6.46}$$

A further analysis (see again Appendix A) then shows that, from an energetic point of view, unless $\frac{\pi}{\kappa}$ happens to be an integer, the energetically optimal interface width is given by the lower bound on N in Equation (6.40), i.e. whenever there is more than one potentially admissible discrete interface width, one has $N_{opt} = N_{min}$, leading to the optimal profile being determined by

$$N = N_{min} = \left\lceil \frac{\pi}{\kappa} - 1 \right\rceil. \tag{6.47}$$

From this, the profile itself and the associated discrete energies follow immediately from equations (6.36), (6.42) and (6.43) (resp. (6.44) and (6.45)) as well as (6.46) using the actual number $N = N_{min}$ of points.

Remark 52. The discrete profile is, as in the continuous case, only fixed up to a reflection and an arbitrary translation as long as this does not interfere with the boundaries of the domain. Even though the choice of fixing the profile as an increasing one starting at $i = 0$ is convenient for some of the calculations above and in Appendix A, one can of course also recenter the profile at the 0.5-isoline, i.e. the point $x_m = \frac{N}{2} \Delta x$. Using

$$\cos(\kappa i) = \cos\big(\kappa(x_i - x_m) + \kappa x_m \big) = \cos\big(\kappa(x_i - x_m) \big) \cos\big(\kappa x_m \big) - \sin\big(\kappa(x_i - x_m) \big) \sin\big(\kappa x_m \big)$$

and

$$\sin(\kappa i) = \sin\big(\kappa(x_i - x_m) + \kappa x_m \big) = \sin\big(\kappa(x_i - x_m) \big) \cos\big(\kappa x_m \big) + \cos\big(\kappa(x_i - x_m) \big) \sin\big(\kappa x_m \big),$$

leads to

$$\begin{aligned} \phi_i &= \frac{1}{2} - \frac{1}{2} \cos(\kappa i) + \frac{1}{2} \cot(\frac{\kappa}{2} N) \sin(\kappa i) \\ &= \frac{1}{2} - \frac{1}{2} \Big(\cos\big(\kappa x_m \big) - \cot(\frac{\kappa}{2} N) \sin\big(\kappa x_m \big) \Big) \cos\big(\kappa(x_i - x_m) \big) \\ &\quad + \frac{1}{2} \Big(\sin\big(\kappa x_m \big) + \cot(\frac{\kappa}{2} N) \cos\big(\kappa x_m \big) \Big) \sin\big(\kappa(x_i - x_m) \big). \end{aligned}$$

As, by the definition of x_m, $\kappa x_m = \frac{\kappa N}{2}$ and $\cos\left(\frac{\kappa}{2}N\right) - \cot\left(\frac{\kappa}{2}N\right)\sin\left(\frac{\kappa}{2}N\right) = 0$, the prefactor for the cosine cancels [31], whereas the one for the sine-term simplifies to

$$\sin\left(\frac{\kappa}{2}N\right) + \cot\left(\frac{\kappa}{2}N\right)\cos\left(\frac{\kappa}{2}N\right) = \sin\left(\frac{\kappa}{2}N\right) + \frac{\cos^2\left(\frac{\kappa}{2}N\right)}{\sin\left(\frac{\kappa}{2}N\right)} = \frac{1}{\sin\left(\frac{\kappa}{2}N\right)},$$

leading to the more pleasant representation

$$\phi_i = \frac{1}{2} + \frac{1}{2}\frac{1}{\sin\left(\frac{\kappa}{2}N\right)}\sin\left(\frac{\kappa}{\Delta x}(x_i - x_m)\right). \tag{6.48}$$

The other choice of orientation (i.e. a decreasing profile) is then obviously again obtained by simply changing the sign of the second term. ◇

In contrast to the choice for $N = N_{min}$, it turns out that the second possible choice of $N = N_{max} = N_{min} + 1$ does **not** satisfy the second-order necessary conditions for the characterization of a local minimizer in terms of the phasefield, i.e. this second choice is not only not optimal in terms of N, but does also not correspond to a local minimizer (of which there might a priori be more than one) but simply to a critical point of the functional $\mathcal{E}$.
The key steps for obtaining this result are (see Appendix A for additional details) that, since the multipliers $\mu^{\pm}$ are, for this larger choice of N, strictly positive, the second-order necessary condition in fact reduces to the study of the local stability for a **given** N, and thus, by the linearity of the equations **within** the interface, to showing the positive definiteness of the system matrix corresponding to the difference stencil in Equation (6.31). As this is simply a difference operator with constant coefficients (meaning also that the stability in particular has nothing to do with the ϕ-values themselves, but only with the choice of N), this reduces to an examination of the lowest eigenvalue of the discrete second-order difference operator under vanishing Dirichlet boundary data.

Remark 53. As the analysis below will show, κ is, up to a third order error in $\frac{\Delta x}{\epsilon}$, given by $\frac{4\Delta x}{\pi\epsilon}$, whereas N is, up to a term of order $\mathcal{O}(1)$ (linked to the discepency between N_{min} and N_{max} in equations (6.40) and (6.37)) given by $\frac{\pi^2\epsilon}{4\Delta x}$, i.e. the "continuous" number of points necessary to subdivide the analytical interface width into cells of width Δx. The critical dependence of the stability on the question of whether κN is less than or greater than π is therefore quite intuitive as this essentially measures whether the discrete interface is broader or more narrow than the analytical one. Roughly speaking, the choices of N for which the last point is slightly within the the continuous interface (i.e. the ones with the more narrow profiles) are stable, whereas those for which the last point is at the end or outside the continuous interface (i.e. the slightly broader profiles) are unstable (see the previous Figure 6.1 and Figure 6.4, which shows one of the "critical" points). ◇

Remark 54. Even though the choice with $N = N_{min} + 1$ is not a stable solution for a descent-based algorithm, it can sometimes nevertheless be obtained (and in some cases actually easier so than the actual solution) in numerical simulations. This somewhat surprising observation is likely due to the fact that all eigenvectors (for this fixed N) which are **antisymmetric** with respect to the center of the interface (i.e. maintain the basic symmetry of the profile) are indeed stable ones. Even without actual stability, this "stability on symmetric profiles" can persist for quite some time when evaluated based on a numerically symmetric calculation[32]. ◇

[31] This also being clear by construction as the value at $x_i = x_m$ has to be 0.5, which can only happen if the cosine is eliminated.

[32] One example for this is in fact the example in Figure 6.1 for $\epsilon = 2\Delta x$. Starting from a sharp transition, a basic gradient-descent scheme tends to develop into the profile with $N = 5$ instead of the prediced value $N = 4$ and remains stable there (at least for a long time). This profile can be "broken down" though e.g. by perturbing the ϕ-values at the first inner point by a term of order 10^{-11}, which, with some delay, then leads to the predicted profile.

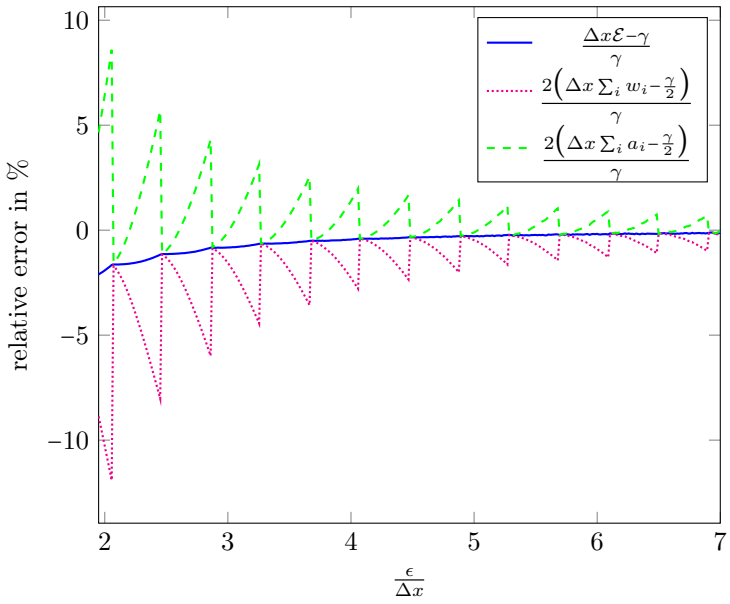


Figure 6.2: Development of the relative errors in the total energy and the bulk potential and gradient energy as a function of the ration $\frac{\epsilon}{r}$ (normalized w.r.t the exact values γ for the total energy resp. $\frac{\gamma}{2}$ for the individual contributions).

One interesting consequence which can be derived from the explicit expressions for the total energy and its two contributions is that it allows for a straightforward analysis of the energy landscape and how it affects the discrete interface width[33]. For a more intuitive approach, this energy landscape (in terms of the relative developments of the total energy $\Delta x\mathcal{E}$ and that of the contributions by the a- and w-term) is plotted in Figure 6.2.
A first observation to be made is that the total energy is in fact quite accurate and develops relatively smoothly, even for very low resolutions[34]. The contributions by the the bulk potential and gradient energy in contrast exhibit a strongly oscillating and discontinuous behavior, with significantly larger deviations than the total energy. The accuracy of the latter is therefore primarily the result of cancellations, as is also obvious from Figure 6.2.

This - as well as the tendency of the signs of the deviations - is intuitively relatively easy to understand based on the competing nature of a and w combined with Figure 6.3, which shows the relative errors in the interface width (counted as the distante between the last value at 0 and the first one at 1). There, it is obvious that the discrete interface is always - and for low resolutions quite significantly so - more narrow than the continuous one. As the role of the gradient energy term is to try to enlarge the interface width whereas that of the bulk energy potential is to make it thinner, it is clear by the decreased width that the discrete case consistenly favors the bulk potential term. This leads to a deviation from the equipartition of energy in the continuous case, with the bulk potential energy being below its continuous value $\frac{\gamma}{2}$, but partially balanced by an increase in the gradient contribution due to the necessity of a steeper interface (also recall Figure 6.1)[35].

[33] This could of course alternatively be evaluated based on a significantly more tedious simulation study.

[34] Note that the left-most point has a relative deviation of less than two percent, even though the number of points (including the respecitvely first "bulk-point" on each side) is only 4 as already seen in Figure 6.1 and also shown in Figure 6.3.

[35] The points where the gradient energy contribution drops below its reference value of $\frac{\gamma}{2}$ are those where the

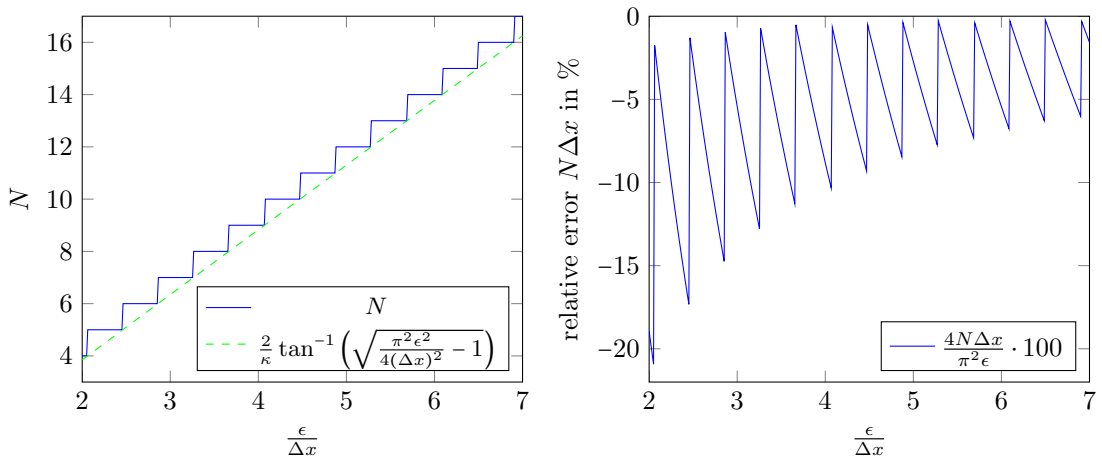


Figure 6.3: Development of the number of interface points and the relative error in the "discrete interface width" with respect to $\frac{\pi^2}{4}\epsilon$ (also see Remark 59).

Remark 55. The discontinuous nature of the problem is also evident from both figures. In particular, as the development of N in Figure 6.3 shows, the evolution of the number of interface points would almost be linear (as one would expect by simply subdividing the analytical interface width based on a spacing of Δx), but, by the necessity of the $\lceil\cdot\rceil$-operation in Equation (6.47), experiences jump discontinuities.
Each of these jumps is accompanied by a sudden drop in the gradient energy contribution and a sudden increase in the bulk potential energy as seen in Figure 6.2. In contrast, the total energy remains relatively smooth as the jumps in both contributions seem to cancel each other. ◇

Remark 56. It can also be noted that the profile shown in Figure 6.1 at the end of the previous Section 6.2.2 is, with $\frac{\epsilon}{\Delta x} = 2$, relatively close to one of the "critical" ratios associated with adding an additional discrete point in the interface. It is interesting to take a closer look at what happens just to the right of one of the discontinuity points visible in Figures 6.2 and 6.3.
The effect on the discrete ϕ-profile is examplarily shown in Figure 6.4 for a ratio of $\frac{\epsilon}{\Delta x}$ just to the right of the discontinuity point near 2, with, for comparision, the profile for a nearly double ration of $\frac{\epsilon}{\Delta x} = 4$. As can be seen there, even though the profile is quite accurate for the higher number of interface points (with $N = 9$ i.e. a total number of 8 points strictly within the interface), the match for the significantly lower ratio is also extremely good (and in fact seemingly even slightly better) despite only having $N = 5$ and thus 4 "interior" interface points. While the much better match as compared to the ratio of 2 in Figure 6.1 is not surprising given the discussion above (and the fact that there is now one additional point), it should be kept in mind that there are two factors influencing the profile. The first one, given by the discrete angular frequency κ, is completely independent of the number of actual points within the interface and is (with roughly 2% compared to 0.5% deviation from the "perfect" value $\frac{\pi\Delta x}{4\epsilon}$, see the next Subsection 6.2.3) notably more accurate for the higher ratio $\frac{\epsilon}{\Delta x} = 4$. In contrast, for $\epsilon \approx 2.061\Delta x$, the value of κ is very close to the one of the "bad" profile for $\epsilon = 2\Delta x$. The

interface width is quite accurate, i.e. only slightly more narrow than the continuous one. The solution profile at these points turns out to be quite accurate (see Remark 56). As the evaluation of the discrete gradient corresponds to evaluating an "average" gradient over the width Δx, it is to be expected that the discrete gradient energy will be slightly lower than the continuous one since this corresponds to squaring the average gradient instead of "averaging" (through the continuous integration) the square, and one always has $\left(\frac{a+b}{2}\right)^2 \leq \frac{1}{2}(a^2+b^2)$ with strict inequality unless $a = b$.

second one, depending both on κ and N, is given by the amplitude $\sin\left(\frac{\kappa N}{2}\right)^{-1}$ of the sinusoidal in Equation (6.48). It is only through a partial cancellation of both influences due to the need of fitting a slightly too "fast" sinusoidal into an almost correct interface width that one obtains the observed excellent match.

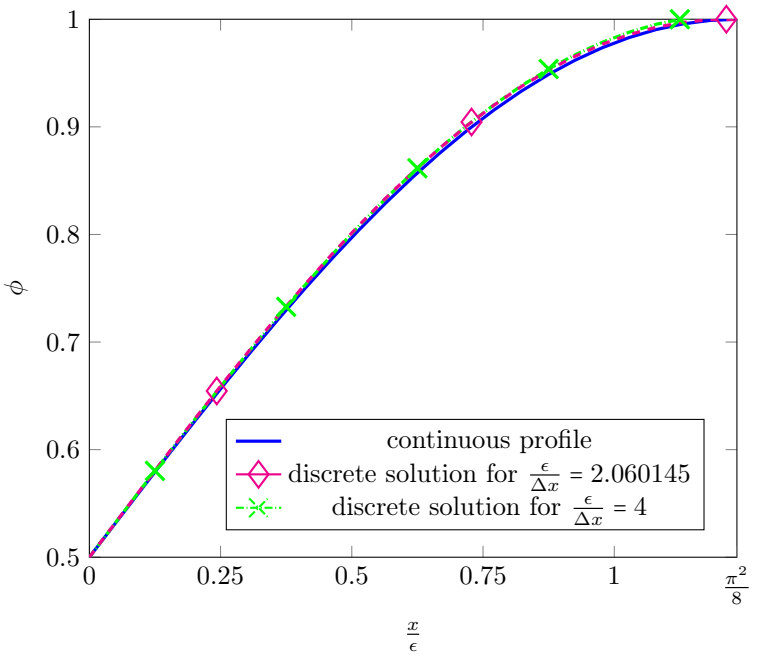


Figure 6.4: Discrete solutions together with the underlying sinusoidals for $\frac{\epsilon}{\Delta x}$ just to the right of the critical ratio 2.06014486 (with $N = 5$) and for $\frac{\epsilon}{\Delta x} = 4$ (with $N = 9$) in comparison with the continuous profile. Note that even though the resolution in the second case is almost twice as high, the deviations in the profile itself are actually sligthly larger than for the one just to the right of the critical point near $\frac{\epsilon}{\Delta x} = 2$.

A further interesting observation is that, as already discussed in Remark 48, the error in the mulitpliers $\mu^{\pm}$ is particularly high just to the right of the critical points as this will lead to an underestimation of the second derivative at the last inner point by a factor of almost 2.
Even though one might - wrongly as will be explained below - conjecture that this could lead to problems related with the undesirable pinning of the discrete interface (i.e. its inability to move despite the action of a driving force) due to the large forcing required to move this point out of the 0 or 1 bound, this seems not to be the case. Instead, as the simple numerical test-case in Figure 6.5 illustrates, these regions at the transition between two different values of N in fact also seem to be "sweet-spots" for the mobility of the interface. The figure shows the final position of an interface intially placed near the left end of the domain under the influence of a (relatively weak) driving force pulling it towards the center of the domain. As is easily visible, the "arbitrary" low-resolution profile with $\epsilon = 2\Delta x$ leads to a relatively high deviation, whereas the one with $\epsilon = 4\Delta x$ (and thus the higher numerical resolution) as expected performs significantly better.
What is a priori somewhat unexpected is that, despite its markedly lower resolution of the interface, the profile with $\Delta x = 2.061\Delta x$ (a point just slightly to the right of the jump in Figure 6.2) not only results in a very accurate profile in the absence of driving forces as seen in Figure 6.4, but also leads to a final interface position which is much more accurate than the higher-resolution

one for $\epsilon = 4\Delta x$.
It can also be observed that a similar improvement in accuracy occurs for $\epsilon = 3.667\Delta x$, corresponding to a ratio just to the right of the jump to the left of $\frac{\epsilon}{\Delta x} = 4$ in Figure 6.2, which has the same number of "interface points" as the simulation for a ratio of four. Finally, the results for $\epsilon = 3.57\Delta x$ and $\epsilon = 3.77\Delta x$ - very close to being at the same distance below and above the critical value at $\frac{\epsilon}{\Delta x} \approx 3.66615$- illustrate that this effect is quite sensitive to the choice of the ratio, but seems to favor slightly larger over slightly lower ratios.
This example illustrates that, even though the anaylsis above itself only covers the pure unperturbed one-dimensional case, the critical ratios at the transition points seem to maintain some interesting properties even in the presence of slight perturbations. As these are the values where $\kappa = \frac{\pi}{n}$, Equation (6.32) shows that these ratios are given by $\cos(\kappa) = \cos\left(\frac{\pi}{n}\right) = 1 - \frac{8(\Delta x)^2}{\pi^2\epsilon^2}$ resp. after solving for $\frac{\epsilon}{\Delta x}$ using $1 - \cos(\theta) = 2\sin\left(\frac{\theta}{2}\right)$ by

$$\frac{\epsilon}{\Delta x} = \frac{2}{\pi \sin\left(\frac{\pi}{2n}\right)}. \tag{6.49}$$

Some of these values are listed in Table 6.1.
As a moderate adjustment of a given ratio to the critical one just below it is something which can essentially be done for "free" (i.e. one expects to maintain the same number of "active" points for which actual calculations have to be performed), the vicinity of these ratios may therefore also have some practical interest for finite-difference based simulations in more complex settings, provided the interfaces are primarily aligned with the axis of the computational domain such as in e.g. some directional solidification or capillary rise problems.
In contrast, it is unlikely that they have any particular effect for settings with roughly spherically-shaped bulk regions, since the effective "grid-spacing" for the profiles along any oblique line can differ very significantly from the one along the axis (i.e. for example a factor of $\sqrt{2}$ in the two-dimensional case when an interface is oriented at an angle of 45°).

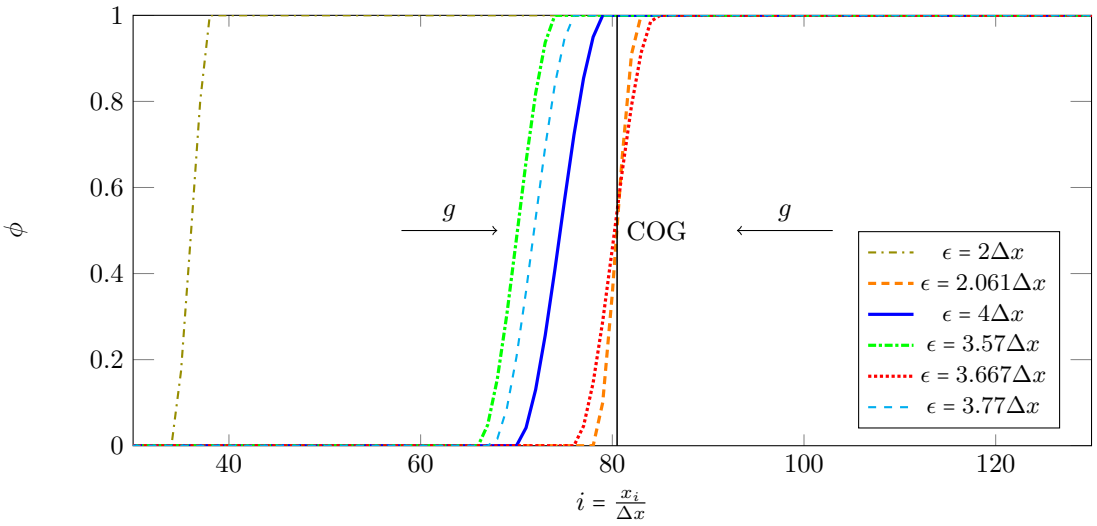


Figure 6.5: Final position of an interface initially placed on the left side under the influence of a "gravity" term pulling it towards the center of the gravity (COG) when varying ϵ for a fixed grid spacing Δx.

◇

N	5	6	7	8	9	10	11	12	13
$\left(\frac{\epsilon}{\Delta x}\right)_{min}$	2.460	2.861	3.264	3.667	4.070	4.474	4.878	5.282	5.686

Table 6.1: Some examples of the minimal values for the ratio of ϵ and Δx given by Equation (6.49) for a given number of points (rounded up at the fourth digit).

The case $\frac{\pi}{\kappa}$ integer As already alluded to in Remark 49 and seen on the example in Figure 6.5, the critical ratios for which $\frac{\pi}{\kappa}$ is integer are very particular and quite interesting. Unfortunately, much of the analysis from the previous paragraph does not carry over to the case where $\frac{\pi}{\kappa}$ is integer and requires different and slighly more complex arguments. For this reason, the discussion of this case is primarily performed in Appendix A, but the results will be summarized here.
Firstly, Equation (6.34) leaves c_2 open, which can be a priori be chosen arbitrarily as long as this is compatible with the bound constraint on the ϕ_i, $1 \leq i \leq N$. As the definition of c_2 in Equation (6.35) does not apply then, one cannot derive the upper bound $N_{max} = \lceil \frac{\pi}{\kappa} \rceil = \frac{\pi}{\kappa}$ based on the value of ϕ_1 (or ϕ_{N-1}) as above. Given that N does not explicitly enter the solution through c_2, it is neither possible to derive the lower bound for N_{min} in Equation (6.40) based on the positivity of the multiplier μ_0^- (resp. the analogous estimate for μ_N^+.)
Nevertheless, it turns out that both the bounds in terms of N_{min} and N_{max} continue to hold due to energetic considerations. Furthermore, choosing c_2 as in Equation (6.35), not only do the discrete energies for $N = N_{min}$ and $N = N_{max}$ in Equation (6.36) have the same energy, but there is actually a whole family of local minimizers with the same energy connecting these two "extreme" profiles. In fact, for $N = N_{max} = \frac{\pi}{\kappa}$, the sinusoidals with angular frequency κ are eigenfunctions of the homogeneous difference operator (i.e. dropping the constant $\frac{16}{\pi^2\epsilon}$ arising from the w-term) within the interior of the interface in Equation (6.31) with an eigenvalue of 0 subject to vanishing boundary conditions at $i = 0$ and $i = N_{max}$. As a consequence, any multiple of this function can be added to the solution corresponding to N_{max} for $c_2 = \cot\left(\frac{\kappa N_{max}}{2}\right)$, which corresponds to the symmetric solution for this number of points, without affecting the validity of Equation (6.31). In addition, while a change of c_2 to $c_2 + \delta c_2$ will affect the values of μ_0^- and by symmetry μ_N^- in Equation (6.39) by $2\gamma\epsilon\delta c_2 \frac{\sin(\kappa)}{(\Delta x)^2}$ for μ_0^- resp. with the opposite sign for μ_N^+, the common of the initial multiplier is given by $\frac{8}{\pi^2\epsilon}\gamma$ (i.e. precisely half the bulk-value in the bulk) and is therefore strictly positive. As long as $|\delta c_2| \leq \frac{\frac{4(\Delta x)^2}{\pi^2\epsilon^2}}{\sin(\kappa)} = \frac{1}{2}\tan\left(\frac{\kappa}{2}\right)$, both multipliers at $i = 0$ and $i = N$ maintain the correct sign. Finally, the values $\delta c_2 = \pm\frac{1}{2}\tan\left(\frac{\kappa}{2}\right)$ are also the ones where either ϕ_{N-1} just touches 1 resp. where ϕ_1 just touches 0, i.e. the values where a new constraint becomes active corresponding to the solution (or a translate by one cell) for the value $N = N_{max} - 1$. More precisely, as

$$\frac{1}{2}\tan\left(\frac{\kappa}{2}\right) = \frac{1}{2}\cot\left(\frac{\pi}{2} - \frac{\kappa}{2}\right) = \frac{1}{2}\cot\left(\frac{\kappa\left(\frac{\pi}{\kappa} - 1\right)}{2}\right) = \frac{1}{2}\cot\left(\frac{\kappa N_{min}}{2}\right),$$

and $\cot\left(\frac{\kappa N_{max}}{2}\right) = 0$, the family of functions

$$\phi_i = \frac{1}{2} - \frac{1}{2}\cos(\kappa i) + c_2\sin(\kappa i) \quad , 0 \leq c_2 \leq \frac{1}{2}\cot\left(\frac{\kappa N_{min}}{2}\right)$$

satisfies the first-order necessary conditions for a local minimizer, has a constant total energy and continuously transforms the "standard" profile with $c_2(N_{max})$ to the one for $c_2(N_{min})$.
This shows that there is an energetically neutral way of moving the right end of the broader interface one point to the left for obtaining the thinner interface and similarly, reverting the order of the argument for moving the left endpoint for the thinner interface one point to the left for again obtaining the broader interface, but now translated by one cell. A symmetric

discussion clearly also applies for moving the interface to the right, and then, by repetition, shows that there is an entire family of interfaces through which can move the interface, albeit with an "oscillating" deformation along the x-axis.

Remark 57. While this argument only strictly holds for the unperturbed equation at the precise critical ratios, it provides an intuitive explanation for the situation observed in Figure 6.5 since the "good" ratios are very close to a value where the interface is actually free to move without having to pass through any energetic barriers. ◇

Remark 58. Even though the total energy is constant during the transformation above, this is not true for the gradient- and bulk-contributions separately. ◇

An Asymptotic Analysis of the Profiles

Even though the expressions for the phasefield profile and discrete energy in equations (6.36) and (6.46) are exact and can easily be evaluated numercially, they are not very "readable" in this form. It is relatively straightforward to derive more tractable approximations though as Δx tends to zero.

One a priori problematic point here is that the definition $N = N_{true}$ in Equation (6.47) relies on a "round-up" operation. The number of points can therefore only be estimated up to a error of order 1, i.e.

$$N = \frac{\pi}{\kappa} - 1 + \delta, \quad \delta \in [0,1] \tag{6.50}$$

where $\Delta\bar{x} := \frac{\Delta x}{\epsilon}$ abbreviates the ratio of the grid-spacing to the length-scale ϵ of the interface. It turns out that, disregarding this round-up, the width $N\Delta x$ of the discrete interface is, up to a second order term, smaller than the continuous value $\frac{\pi^2}{4}\epsilon$ by a single grid-spacing Δx. As this is a term of the same magnitude, the round-up, even though hard to estimate, is therefore not really limiting the precision of the general first-order convergence esimate for the interface width.

Considering first the value of κ as defined in Equation (6.32), an estimate for κ can be derived based upon the expansion

$$\cos(\kappa) = 1 - \frac{1}{2}\kappa^2 + \frac{1}{4!}\kappa^4 + \mathcal{O}(\kappa^6) \overset{!}{=} 1 - \frac{1}{2}\Big(\frac{4}{\pi}\Delta\bar{x}\Big)^2$$

of the cosine around zero (this being the relevant range for κ). An estimate based upon comparing the quadratic term only then shows that $\kappa = \mathcal{O}(\Delta\bar{x})$ and then more precisely with $\frac{\kappa}{\Delta\bar{x}} = \mathcal{O}(1)$ and $\sqrt{1-y^2} = 1 + \frac{1}{2}y^2 - \mathcal{O}(y^4)$ that[36]

$$\kappa = \sqrt{\Big(\frac{4}{\pi}\Delta\bar{x}\Big)^2 + \mathcal{O}(\kappa^4)} = \frac{4}{\pi}\Delta\bar{x}\sqrt{1 + \mathcal{O}\Big(\frac{\kappa^4}{(\Delta\bar{x})^2}\Big)} = \frac{4}{\pi}\Delta\bar{x}\Big(1 + \mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big). \tag{6.51}$$

Combined with $(1+y)^{-1} = 1 - y + \mathcal{O}(y^2)$), it follows that

$$\frac{1}{\kappa} = \frac{1}{\frac{4}{\pi}\Delta\bar{x}\Big(1 + \mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big)} = \frac{\pi}{4\Delta\bar{x}}\frac{1}{1 + \mathcal{O}\Big((\Delta\bar{x})^2\Big)} = \frac{\pi}{4\Delta\bar{x}}\Big(1 - \mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big),$$

from which the formula (6.50) for N can be rewritten as

$$N = \frac{\pi^2}{4\Delta\bar{x}}\Big(1 - \mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big) - 1 + \delta = \frac{\pi^2}{4\Delta\bar{x}} - 1 + \delta - \mathcal{O}(\Delta\bar{x}), \quad \delta \in [0,1]. \tag{6.52}$$

[36] A more precise estimate will be derived in Equation (6.57).

Remark 59. Note that by the construction above, $\phi_0 \sim \phi(x_0) = \phi(0)$ is the last point where the ϕ-value is still 0, and $\phi_N \sim \phi(x_N) = \phi(N\Delta x)$ is the first point where the ϕ-value is at 1. Even though there is not really a precise definition of the interface width in the discrete case[37], the value of $N\Delta x = \frac{\pi^2}{4}\epsilon - (1-\delta)\Delta x$ can be considered to be an upper bound for any reasonable interpretation of the discrete interface width, and the discrete interface is therefore slightly more narrow (with $-1+\delta < 0$) than the continuous one. ◇

Remark 60. Note that the approximation for κ shows that[38] the angular frequency $\frac{4}{\pi\epsilon}$ of the continuous solution in Equation (6.30) is achieved up to a second-order error.
Even though the estimate for $\frac{1}{\kappa}$ is only an estimate up to the order $\mathcal{O}(\Delta\bar{x})$, it is so around a value tending to infinity. Multiplying N by Δx to obtain the physical length of the interface, this error (with the bounded $\tan^{-1}$-term) only leads to a second-order deviation, whereas the value $N\Delta x$ is generally only first-order accurate.
What is potentially surprising is that, based on Equation (6.52), the a priori fully numerically induced necessity of the round-up operation turns out to actually cancel part of this first-order error and that, if δ is very close to 1, the interface will match, then essentially up to second order, the continuous one. This is also consistent with Figure 6.4, where the remarkable agreement of the discrete profile with the continuous one, despite the low resolution, is primarily due to the fact that the "continuous" approximation for N is just slightly above 4, and therefore, together with the $\lceil\cdot\rceil$-operation leads to $\delta \approx 1$.
That the first-order error in the approximate width of the domain does not entail an error of equal order in the solution can intuitively be explained by the fact that the continuous solution enters the bulk with a slope of 0. While this is an unpleasant property in the "forward" direction (i.e. the ratio of the change of the interface width as compared to a change in the solution can be made arbitrarily high), it equally well applies in the "backward" direction, i.e. one can make the change in the profile despite a perturbation in the interface width aribtrarily small as this perturbation tends to zero. A similar argument applies in the discrete case. Even though the derivative is never actually zero, it suffices for this derivative to approach zero at a linear rate in order to be able to absorb a first-order error in the discrete interface width into a second-order error in the actual profile. ◇

Remark 61. Regardless of the asymptotics themselves, an important (and quite natural) point to be retained from the previous remark is that the interface width by itself is an unreliable criterion for judging the actual quality of a solution. This is of course also (but therefore not only) an issue when dealing with well-potentials and might partially explain why it is quite common to "define" the interface width e.g. as being the width between the points where $\phi = 0.1$ and $\phi = 0.9$ instead of two points where ϕ is actually closer to 0 and 1. ◇

Combining the estimates for κ and N, it further follows that

$$\kappa N = \frac{4}{\pi}\Delta\bar{x}\Big(1+\mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big)\Big(\frac{\pi^2}{4\Delta\bar{x}} - (1-\delta) + \mathcal{O}(\Delta\bar{x})\Big) = \pi - \frac{4\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\Big((\Delta\bar{x})^2\Big),$$

and thus combined with $\cot\left(\frac{\pi}{2}-\theta\right) = \tan(\theta) = \theta + \mathcal{O}(\theta^3)$ that

$$\cot\Big(\frac{\kappa N}{2}\Big) = \cot\left(\frac{\pi}{2} - \Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\Big((\Delta\bar{x})^2\Big)\Big)\right) = \frac{2\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\Big((\Delta\bar{x})^2\Big). \tag{6.53}$$

[37] E.g. if interpreted in a cell-centerd fashion, there would be $N-1$ "cells" with values deviating from the bulk-ones, whereas, if interpreted in a FEM-like fashion, there would be N such "elements".

[38] Multiplying κ by the "point" i leads to value

$$\kappa i = \frac{4}{\pi\epsilon} i\Delta x\big(1+\mathcal{O}\big((\Delta x)^2\big) = \frac{4}{\pi\epsilon}x_i\big(1+\mathcal{O}\big((\Delta x)^2\big)$$

for the argument of the sine and cosine in Equation (6.36).

Remark 62. ◇

A multiplication of the second expression for $\mathcal{E}(N)$ in Equation (6.46) by Δx then shows that

$$\begin{aligned}\Delta x\mathcal{E} =& \frac{4\gamma}{\pi^2}N\Delta\bar{x} + \frac{4\gamma\Delta\bar{x}}{\pi^2}\cot\big(\frac{\kappa N}{2}\big)\sqrt{\frac{\pi^2}{4(\Delta\bar{x})^2}-1} = \frac{4\gamma}{\pi^2}N\Delta\bar{x} + \frac{2\gamma}{\pi}\cot\big(\frac{\kappa N}{2}\big)\sqrt{1-\frac{4(\Delta\bar{x})^2}{\pi^2}}\\ =& \frac{4\gamma}{\pi^2}\Big(\frac{\pi^2}{4} - (1-\delta)\Delta\bar{x} + \mathcal{O}\big((\Delta\bar{x})^2\big)\Big) + \frac{2\gamma}{\pi}\Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\big((\Delta\bar{x})^2\big)\Big)\Big(1-\mathcal{O}\big((\Delta\bar{x})^2\big)\Big).\end{aligned}$$

From this, it is seen that the first-order errors through the $(1-\delta)$-terms actually cancel (up to second order) in the sum, leaving the final estimate

$$\Delta x\mathcal{E} = \gamma + \mathcal{O}\big((\Delta\bar{x})^2\big), \tag{6.54}$$

which in particular shows that the discrete energy is **second-order** convergent to the continuous value.

Remark 63. This very pleasent conclusion is in fact not obvious based on the use of a formally second-order numerical scheme. It is not a priori clear that this accuracy can actually be achieved due to the only moderate smoothness in relation with the jumps in the second derivatives at the transition points.
In continuation of Remark 51, one can also observe that, even though it is the first term in Equation (6.46) that actually converges to the desired value γ, it does so only with first-order accuracy due to the $\mathcal{O}\big(\frac{\Delta x}{\epsilon}\big)$-error in the discrete interface width $N\Delta x$. That the convergence of the energy is nevertheless quadratic is only achieved through the correction by the second term in Equation (6.46) which (up to second-order order) cancels this contribution.
It can also be noted that this does not depend on the precise form of the error in $N\Delta x$ (provided the error δN is of order 1, as e.g. the one caused by the truncation in the calculation of N). This can of course be seen by a close look at the coefficients arising in the expansions above, but is intuitively much easier to understand based upon the choice of N itself. In fact, the energetic argument underlying the choice of the "correct" number of points in the discrete interface is simply a discrete analogue of the continuous condition $\frac{\partial\mathcal{E}}{\partial N}(N) \overset{!}{=} 0$. Even though this equality need not be achieved exactly in the discrete setting, it will be so in first order. From this, the "direct" contribution $\Delta x\frac{\partial\mathcal{E}}{\partial N}\delta N$ of a perturbation in N to the discrete energy is indeed expected to be only of second-order in Δx. ◇

Remark 64. A further interesting observation is, by the expressions for the sums of the discrete bulk potential and gradient energy in equations (6.42) and (6.43), that there is - in contrast to the continuous case - not a precise equipartition of energy. In fact, one has

$$\sum_{i=1}^{N-1}\big(\epsilon a_i - \frac{1}{\epsilon}w_i\big) = \frac{4\gamma}{\pi^2\epsilon}\Big(\cot^2\big(\frac{\kappa N}{2}\big)N + \cot\big(\frac{\kappa N}{2}\big)\frac{1-\cos(\kappa)}{\sin(\kappa)}\Big),$$

showing that the gradient energy is larger than the bulk potential energy in the discrete case.
Even though one could then conjecture that the second-order accuracy of $\Delta x\mathcal{E}$ is due to a beneficial cancellation, a short calculation using the expansions above reveals that both the gradient energy and bulk energy contributions themselves (and not just their sum $\Delta x\mathcal{E}$) are also second-order accurate[39]. ◇

[39]Both equations (6.42) and (6.43) contain a term of the form $\Big(1 \pm \cot^2\big(\frac{\kappa N}{2}\big)\Big)N$, which would be the obvious candidate for a cancellation. As $\cot\big(\frac{\kappa N}{2}\big) = \mathcal{O}(\Delta x)$ by the above, this term does not actually matter in the first order. More precisely, using this estimate for the cot and $\frac{1}{\sin(\kappa)} \approx \frac{1}{\kappa}$ (or directly inserting the definition of $\sin(\kappa)$

As a final consequence of the expansions above, one can also construct a (quite accurate) continuous approximation to the discrete solution given in Equation (6.48) in terms of the paramters Δx and ϵ alone. In constrast to the exact discrete profile, this has the advantage of directly highlighting some simple - both qualitative and quantitative - points.
This requires additionally approximating the term $\frac{1}{\sin\left(\frac{\kappa N}{2}\right)}$. Based on equations (6.51) and (6.52), it follows (as already used above) that $\frac{\kappa N}{2} = \frac{\pi}{2} - \frac{2\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\big((\Delta\bar{x})^2\big)$. The elementary relation $\sin\left(\frac{\pi}{2} - \theta\right) = \cos(\theta) = 1 - \theta^2 + \mathcal{O}(\theta^4)$ near zero first leads to

$$\frac{1}{\sin\left(\frac{\kappa}{2}N\right)} = \frac{1}{1 - \frac{1}{2}\left(\frac{2\Delta\bar{x}}{\pi}(1-\delta)\right)^2 + \mathcal{O}\big(\Delta\bar{x}\big)^3\big)} = 1 + \frac{1}{2}\Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta)\Big)^2 + \mathcal{O}\big(\Delta\bar{x}\big)^3$$

and thus with $\kappa = \frac{4}{\pi\epsilon}\big(1 + \mathcal{O}\big((\Delta\bar{x})^2\big)\Delta x$ from Equation (6.51) shows that the discrete profile can also be rewritten as

$$\phi_i = \frac{1}{2} + \frac{1}{2}\Big(1 + \frac{1}{2}\Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta)\Big)^2\Big)\sin\Big(\frac{4}{\pi\epsilon}\big(1 + \mathcal{O}\big((\Delta\bar{x})^2\big)\big)(x_i - x_m)\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^3\big) \tag{6.55}$$

where $x_i \coloneqq i\Delta x$.
Expanding the "error" in the angular frequency leads to

$$\begin{aligned}\sin\Big(\frac{4}{\pi\epsilon}\big(1 + \mathcal{O}\big((\Delta\bar{x})^2\big)\big)(x_i - x_m)\Big) = &\sin\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big)\cos\Big(\mathcal{O}\big((\Delta\bar{x})^2\big)(x_i - x_m)\Big) \\ &+ \cos\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big)\sin\Big(\mathcal{O}\big((\Delta\bar{x})^2\big)(x_i - x_m)\Big).\end{aligned} \tag{6.56}$$

Since $\cos\Big(\mathcal{O}\big((\Delta\bar{x})^2\big)\Big) = 1 - \mathcal{O}\big((\Delta\bar{x})^4\big)$, the prefactor in the sine-term can actually be replaced (up to a fourth-order error) by 1, whereas the linear convergence of $\sin(\theta)$ to zero with θ in principle allows dropping the second-order term up to an error of $\mathcal{O}\big((\Delta x)^2\big)$. Together with the boundedness of the sin- and cos-function, this is already sufficient to show that the discrete profile satisfies

$$\phi_i = \frac{1}{2} + \frac{1}{2}\sin\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^2\big)$$

and is thus indeed a second-order accurate approximation of the analytical one.

Remark 65. It should again be stressed that this is not a priori obvious just because one is using a formally second-order scheme for evaluating $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$. Even though imposing e.g. the requirement

---

in Equation (6.33)), one has

$$\begin{aligned}\Delta x\sum_{i=1}^{N-1}\epsilon a_i =& \frac{8\gamma}{\pi^2}\Delta\bar{x}\Big(\frac{1}{4}\Big(1 + \cot^2\big(\frac{\kappa N}{2}\big)\Big)N + \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\frac{1}{\sin(\kappa)}\Big) = \frac{8\gamma}{\pi^2}\Delta\bar{x}\Big(\frac{1}{4}\Big(1 + \mathcal{O}\big((\Delta\bar{x})^2\big)\Big)N + \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\frac{1}{\sin(\kappa)}\Big) \\ =& \frac{8\gamma}{\pi^2}\Delta\bar{x}\Big(\frac{1}{4}\Big(1 + \mathcal{O}\big((\Delta\bar{x})^2\big)\Big)\Big(\frac{\pi^2}{4\Delta\bar{x}} - (1-\delta) + \mathcal{O}(\Delta\bar{x})\Big) + \frac{1}{2}\Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta) + \mathcal{O}\big((\Delta\bar{x})^2\big)\Big)\frac{\pi}{4\Delta\bar{x}}\Big(1 + \mathcal{O}\Big((\Delta\bar{x})^2\big)\Big)\Big) \\ =& \frac{8\gamma}{\pi^2}\Delta\bar{x}\Big(\frac{\pi^2}{16\Delta\bar{x}} - \frac{1}{4}(1-\delta) + \frac{1}{4}(1-\delta) + \mathcal{O}(\Delta\bar{x})\Big) = \frac{1}{2}\gamma + \mathcal{O}\big((\Delta\bar{x})^2\big).\end{aligned}$$

The contribution by w differs, up to the sign of the second-order term $\cot^2\left(\frac{\kappa N}{2}\right)$ only by the last factor $\cot(\kappa)$ instead of $\frac{1}{\sin(\kappa)}$. As $\cot(\kappa) = \frac{\cos(\kappa)}{\sin(\kappa)} = \frac{1}{\kappa} + \mathcal{O}\big((\Delta\bar{x})^2\big)$, the calculation for the contributions by a can essentially directly be reused to show that one also has

$$\Delta x\sum_{i=1}^{N-1}\frac{1}{\epsilon}w(\phi_i) = \frac{1}{2}\gamma + \mathcal{O}\big((\Delta\bar{x})^2\big).$$

An even quicker alternative is of course to observe that as the gradient energy contribution and the total energy are second order accurate, the bulk potential necessarily also has to be as the difference between both.

$\phi \in C^4$ for obtaining the formal estimate $\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{(\Delta x)^2} = \phi''(x_i) + \mathcal{O}\big((\Delta x)^2\big) + \mathcal{O}\big((\Delta x)^3\big)$ is in fact more restrictive than actually necessary for obtaining global second-order convergence, it should be kept in mind that, as pointed out in Remark 48, that this difference operator is only $\mathcal{O}(1)$-accurate when it straddles the transition points, corresponding to a loss of two order of accuracy. Obtaining estimates for the order of convergence under such conditions requires a much more detailed and complex analysis (see e.g. [38] for a detailed treatment of this topic). ◇

While this is interesting in itself, it is also useful to have a more precise estimate of the remaining second-order deviations from the analytical profile. As the analysis above shows, the accuracy of the approximation above is actually of the order $\mathcal{O}\big((\Delta\bar{x})^3\big)$ except for the "imprecise" second-order estimate for $\sin\Big(\mathcal{O}\big((\Delta\bar{x})^2\big)(x_i - x_m)\Big)$ due to the only second-order approximation of κ.

Improving this estimate is fortunately only a matter of solving a quadratic equation. In fact, κ being defined by $\cos(\kappa) = 1 - \frac{1}{2}\big(\frac{4\Delta x}{\pi\epsilon}\big)^2$, one can simply include the fourth-order term in the expansion of the cosine to obtain $1 - \frac{1}{2}\kappa^2 + \frac{1}{4!}\kappa^4 + \mathcal{O}(\kappa^6) \stackrel{!}{=} 1 - \frac{1}{2}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2$ resp.

$$\kappa^4 - 12\kappa^2 + 12\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}(\kappa^6) \stackrel{!}{=} 0.$$

Since $\kappa = \mathcal{O}(\Delta\bar{x})$, this can be interpreted as a quadratic equation for κ^2 perturbed by a term of order $(\Delta\bar{x})^6$, from which it follows (using $\sqrt{1+y} = 1 + \frac{1}{2}y - \frac{1}{8}y^2 + \mathcal{O}(y^3)$ near $y = 0$ and choosing the relevant sign preceding the square-root) that

$$\begin{aligned}
\kappa^2 =& 6 - 6\sqrt{1 - \frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)} \\
=& 6 - 6\left(1 - \frac{1}{2}\left(\frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)\right) - \frac{1}{8}\left(\frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)\right)^2\right) \\
=& 6\left(\frac{1}{2}\left(\frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)\right) + \frac{1}{8}\left(\frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)\right)^2\right) \\
=& \big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big) + \frac{3}{4}\left(\frac{1}{3}\Big(\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^6\big)\Big)\right)^2 = \big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \frac{1}{12}\big(\frac{4\Delta\bar{x}}{\pi}\big)^4 + \mathcal{O}\big((\Delta\bar{x})^6\big).
\end{aligned}$$

Again expanding the square-root, a more precise estimate for κ is therefore given by

$$\begin{aligned}
\kappa =& \frac{4\Delta\bar{x}}{\pi}\sqrt{1 + \frac{1}{12}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^4\big)} = \frac{4\Delta\bar{x}}{\pi}\Big(1 + \frac{1}{24}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^4\big)\Big) \\
=& \frac{4\Delta\bar{x}}{\pi} + \frac{\Delta\bar{x}}{6\pi}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^5\big),
\end{aligned} \tag{6.57}$$

allowing to obtain the improved approximation

$$\begin{aligned}
&\cos\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big)\sin\left(\Big(\frac{1}{6\pi\epsilon}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^4\big)\Big)(x_i - x_m)\right) \\
=& \frac{1}{6\pi\epsilon}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2(x_i - x_m)\cos\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big) + \mathcal{O}\big((\Delta\bar{x})^4\big)
\end{aligned}$$

for the cos-based term in Equation (6.56). Combining this with the previous analysis for Equation

(6.55) leads to the higher-order approximation[40]

$$\begin{aligned}\phi_i =& \frac{1}{2} + \frac{1}{2}\Big(1 + \frac{1}{2}\Big(\frac{2\Delta\bar{x}}{\pi}(1-\delta)\Big)^2\Big)\sin\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big) \\ &+ \frac{1}{12\pi\epsilon}\Big(\frac{4\Delta\bar{x}}{\pi}\Big)^2(x_i - x_m)\cos\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^3.\end{aligned} \tag{6.58}$$

In contrast to the previous approximation in Equation (6.55), this expansion contains both leading-order (i.e. second-order) perturbations of the analytical profile.
The first one in terms of the slightly increased amplitude of the dominant sine-term was already obtained before. It on the one hand leads to a slightly steeper profile in the region around the isoline $\phi = \frac{1}{2}$, and on the other hand to both a slightly increased (resp. decreased) ϕ-value and a slightly decreased (resp. increased) curvature near the outer limits of the interface near $\phi = 1$ (resp. $\phi = 0$) since this is where both the sin itself and thus also its second derivative attain the maximal magnitude. The second contribution in contrast will also lead to a steeper profile near the middle of the interface (the cosine being second-order close to one there), but has very little effect on the ϕ-values near the outer limits as this is where the cosine approaches zero, its primary effect on the ϕ-values itself therefore lying in the intermediate region. Nevertheless, with $(x\cos(x))'' = -2\sin(x) - x\cos(x)$, it will also have a similiar "curvature-enhancing" effect near the outer regions.

Remark 66. With respect to the interpretation of both perturbations and their relation with the **first-order** perturbation of the interface width, it is worth noting that the first contribtution could also be obtained by simply enforcing a steeper version (for simplicity with $x_m = 0$) of the analytical profile $\frac{1}{2} + \frac{1}{2}c'\sin\big(\frac{\pi}{4\epsilon}x\big)$ to match the bulk-value 1 at the perturbed point $x' = \frac{\pi^2}{8}\epsilon - \frac{1-\delta}{2}\Delta\bar{x}$, corresponding to distributing the discrepency in the width symmetrically to both sides. In fact, with

$$\sin\big(\frac{4}{\pi\epsilon}x'\big) = \sin\Big(\frac{\pi}{2} - \frac{4}{\pi\epsilon}\frac{1-\delta}{2}\Delta\bar{x}\Big) = \cos\big(\frac{2}{\pi\epsilon}(1-\delta)\Delta\bar{x}\big),$$

the equation $\frac{1}{2} + c'\sin\big(\frac{4}{\pi\epsilon}x'\big) \overset{!}{=} 1$ asymptotically reduces to

$$c'\cos\big(\frac{2}{\pi\epsilon}(1+\delta)\Delta\bar{x}\big) = c'\Big(1 - \frac{1}{2}\big(\frac{2}{\pi\epsilon}(1-\delta)\Delta\bar{x})\big)^2\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^4\big) \overset{!}{=} \frac{1}{2}$$

and thus

$$c' = \frac{1}{2\Big(1 - \frac{1}{2}\big(\frac{2}{\pi\epsilon}(1-\delta)\Delta\bar{x})\big)^2\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^4\Big)} = \frac{1}{2}\Big(1 + \frac{1}{2}\big(\frac{2}{\pi\epsilon}(1-\delta)\Delta\bar{x})\big)^2\Big) + \mathcal{O}\big(\Delta\bar{x}\big)^4\Big).$$

This is, up to a fourth-order error, the prefactor of the sin above. Note that, as already indicated before, it is crucial here that the first-order perturbation in the interface width arises in a region where the profile is to first order flat as this is the reason c' differs from 1 only in second and not in first order.
In contrast, the second contribution can be interpreted as being related to a more "mundane" discretization error in line with Footnote 28. More precisely, applying a discrete (three-point

[40] Even though the pertubation by the term $(x_i - x_m)\cos\Big(\frac{4}{\pi\epsilon}(x_i - x_m)\Big)$ a priori also appears with the slightly increased amplitude of the sin-term, this results only in a fourth-order perturbation.

stencil) second-order derivative to a sinusoidal profile, one obtains

$$\begin{aligned}
&\frac{\sin(x_{i+1}) - 2\sin(x_i) + \sin(x_{i-1})}{(\Delta x)^2} = \frac{\sin(x_i + \Delta x) - 2\sin(x_i) + \sin(x_i - \Delta x)}{(\Delta x)^2} \\
=&\frac{\sin(x_i)\cos(\Delta x) + \cos(x_i)\sin(\Delta x) - 2\sin(x_i) + \sin(x_i)\cos(\Delta x) - \cos(x_i)\sin(\Delta x)}{(\Delta x)^2} \\
=&-\sin(x_i)\frac{2\big(1-\cos(\Delta x)\big)}{(\Delta x)^2} = -\sin(x_i)\frac{2\Big(1-\big(1-\frac{1}{2}(\Delta x)^2 + \frac{1}{4!}(\Delta x)^4 - \mathcal{O}\big((\Delta x)^6\big)\Big)}{(\Delta x)^2} \\
=&-\big(1-\frac{1}{12}(\Delta x)^2\big)\sin(x_i) + \mathcal{O}\big((\Delta x)^4\big),
\end{aligned}$$

resp. from the same calculation but with a modified angular frequency for the sine,

$$\frac{\sin\big(\frac{4}{\pi\epsilon}x_{i+1}\big) - 2\sin\big(\frac{4}{\pi\epsilon}x_i\big) + \sin\big(\frac{4}{\pi\epsilon}x_{i-1}\big)}{(\Delta x)^2} = -\big(\frac{4}{\pi\epsilon}\big)^2\sin\big(\frac{4}{\pi\epsilon}x_i\big)\Big(1-\frac{1}{12}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 + \mathcal{O}\big((\Delta\bar{x})^4\big)\Big).$$

One can successfully counterbalance this deviation in the second derivative of the dominant term precisely by the cos-based term in Equation (6.58). In fact, inserting the ansatz (again taking $x_m = 0$ for simplicity) $\phi(x) = \frac{1}{2} + \frac{1}{2}\sin\big(\frac{4}{\pi\epsilon}x\big) + \delta\phi$ with $\delta\phi = \mathcal{O}\big((\Delta\bar{x})^2\big)$ into the discrete difference Equation (6.31), it follows that $\delta\phi$ should satisfy

$$\begin{aligned}
&-2\gamma\epsilon\Big(-\big(\frac{4}{\pi\epsilon}\big)^2\frac{1}{2}\big(1-\frac{1}{12}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2\big)\sin\big(\frac{4}{\pi\epsilon}x\big) - 2\gamma\epsilon\frac{\mathrm{d}^2\delta\phi}{\mathrm{d}x^2} + \frac{16}{\pi^2\epsilon}\gamma(-\sin\big(\frac{4}{\pi\epsilon}x\big) - 2\delta\phi) \\
=&-\gamma\epsilon\Big(\big(\frac{4}{\pi\epsilon}\big)^2\frac{1}{12}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2\sin\big(\frac{4}{\pi\epsilon}x\big) - 2\gamma\epsilon\frac{\mathrm{d}^2\delta\phi}{\mathrm{d}x^2} - \frac{32}{\pi^2\epsilon}\gamma\delta\phi \overset{!}{\approx} 0
\end{aligned}$$

since the first-order terms in the sin cancel by construction of the analytical profile and applying the discrete second-order derivative to $\delta\phi$ will approximate its actual second derivative in order $\mathcal{O}\big((\Delta\bar{x})^2\big)$. $\delta\phi$ should thus satisfy

$$-2\epsilon\frac{\mathrm{d}^2\delta\phi}{\mathrm{d}x^2} - \frac{32}{\pi^2\epsilon}\delta\phi \approx \frac{4}{3\pi^2\epsilon}\big(\frac{4\Delta\bar{x}}{\pi\epsilon}\big)^2\sin\big(\frac{4}{\pi\epsilon}x\big)$$

and it is easily seen that defining $\delta\phi := \frac{1}{12\pi\epsilon}\big(\frac{4\Delta\bar{x}}{\pi}\big)^2 x\cos\Big(\frac{4}{\pi\epsilon}x\Big)$ corresponding to the second correction in Equation (6.58) will actually solve this equation[41].
It can further be noted that, for the choice of κ given by Equation (6.33), one has

$$\begin{aligned}
&\frac{\sin\big(\kappa(i+1)\big) - 2\sin\big(\kappa i\big) + \sin\big(\kappa(i-1)\big)}{(\Delta x)^2} = -2\sin\Big(\kappa i\Big)\frac{1-\cos\big(\kappa\big)}{(\Delta x)^2} \\
=&-2\sin\Big(\kappa i\Big)\frac{1-\Big(1-\frac{1}{2}\big(\frac{4\Delta x}{\pi\epsilon}\big)^2\Big)}{(\Delta x)^2} = -\big(\frac{4}{\pi\epsilon}\big)^2\sin\Big(\kappa i\Big),
\end{aligned}$$

i.e. the prefactor of the sine resulting from the discrete second-order differentiation operator coincides **exactly** with the one arising from the second differentiation of the $\sin\big(\frac{4}{\pi\epsilon}x\big)$ in the continuous solution. ◇

[41] Note that while this $\delta\phi$ solves the equation exactly in the interior of the domain, it is not completely compatible with the (here implicit) boundary condition that $\delta\phi$ should vanish at the endpoints of the actual interface. This violation is only first-order in Δx (with respect to the already second-order size of $\delta\phi$ itself) though by the linear convergence of the cos-term to zero near the outer region. This can then either be absorbed through the construction of an appropriate boundary layer, or, since the dominant sin-term is maximal there, by a slight modification of the dominant term through a third-order correction of its amplitude. Regardless of the details, this does not affect the order of the error estimate in Equation (6.58).

6.3 Multi-Phasefield Problems

As discussed in the beginning of Section 6.2, one has to deal with a number of additional challenges when considering problems with more than two phases.
A first - and quite important one within the usual variational framework - is the construction of an appropriate generalization of the underlying functional to the multiphase case. While there are a variety of possible choices, the discussion here will focus on the one based on [52] outlined in Section 6.1. Together with the additional constraints through either the sum-constraint $\sum_{\alpha=1}^{N} \phi^\alpha = 1$ or the restriction of the $\boldsymbol{\phi}$-values to the Gibbs-simplex $\mathcal{GS}^N$, the functional already fixes an important part of the problem, namely the equations to be satisfied by any minimizer and thus the steady-state equations, which will quickly be rederived - in a formal manner similar to Section 6.2 - in Subsection 6.3.1.
A second important aspect, which is essentially independent of the functional (even if one postulates a gradient flow)[42] is the choice of the dynamics, i.e. how one chooses the "proportionality" with respect to $\frac{\delta \mathcal{F}_\epsilon}{\delta \boldsymbol{\phi}}$. Two popular choices also implemented within the **Pace3D**-framework will quickly be outlined in Subsection 6.3.2. Despite its practical importance, the (relatively difficult) question of the respective advantages and disadvantages of the different choices of dynamics from a physical point of view will not be discussed in any detail here. Instead, after a short outline of some general algorithmic considerations for multiphase problems in Subsection 6.3.3, in particular in the presence of a large number of phases, the focus in Subsection 6.3.3 will primarily be on practical aspects of this choice in the presence of the constraint by the Gibbs-simplex.

6.3.1 The Steady-State Equations

As in the two-phase case, a directional differentiation of the phasefield-functional in Equation (6.10) leads to the first-order necessary condition

$$\mathcal{F}'_\epsilon(\boldsymbol{\phi};\boldsymbol{\psi}) = \sum_{\alpha=1}^{N} \int_\Omega \epsilon\Big(\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \nabla \psi^\alpha + \frac{\partial a}{\partial \phi^\alpha}\psi^\alpha\Big) + \frac{1}{\epsilon}\frac{\partial w}{\partial \phi^\alpha}\psi^\alpha + \frac{\partial f}{\partial \phi^\alpha}\psi^\alpha \,\mathrm{d}\boldsymbol{x} \geq 0$$

for all admissible directions $\boldsymbol{\psi}$.
If the individual contributions to $\mathcal{E}_\epsilon$ and the functions themselves are smooth enough and with the "natural" boundary condition $\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \boldsymbol{n} = 0$ (see Remark 70 below), any minimizer in the multi-well case (without an additional restriction of the form $0 \leq \phi^\alpha \leq 1$) will therefore have to satisfy

$$\sum_{\alpha=1}^{N} \int_\Omega \underbrace{\Big(-\epsilon\nabla\Big(\cdot\frac{\partial a}{\partial \nabla \phi^\alpha}\Big) + \epsilon\frac{\partial a}{\partial \phi^\alpha} + \frac{1}{\epsilon}\frac{\partial w}{\partial \phi^\alpha} + \frac{\partial f}{\partial \phi^\alpha}\Big)}_{=:g^\alpha}\psi^\alpha \,\mathrm{d}\boldsymbol{x} \geq 0, \tag{6.59}$$

where, unlike in Equation (6.17), $\boldsymbol{\psi}$ is locally restricted to satisfy $\sum_{\alpha=1}^{N} \psi^\alpha = 0$. In the steady-state, the (L^2-)gradients $\boldsymbol{g} = (g^\alpha)_{1\leq\alpha\leq N}$ are therefore locally orthogonal to all vectors with zero average, i.e., as seen in Section 4, satisfy $\boldsymbol{g}(\boldsymbol{x}) = -\Lambda(\boldsymbol{x})\boldsymbol{e}$, where the (scalar) Lagrange-multiplier $\Lambda(\boldsymbol{x})$ does depend on $\boldsymbol{x}$ but not on the phase α and the choice of the minus-sign is only a matter of convenience. Through a simple summation, it therefore follows that Λ satisfies $\sum_{\alpha=1}^{N} -g^\alpha = N\Lambda$, and thus, inserting the expression for $\boldsymbol{g}$, that $\boldsymbol{\phi}$ has to to satisfy as in [52]

$$\begin{cases} -\epsilon\nabla\cdot\Big(\frac{\partial a}{\partial \nabla \phi^\alpha}\Big) + \epsilon\frac{\partial a}{\partial \phi^\alpha} + \frac{1}{\epsilon}\frac{\partial w}{\partial \phi^\alpha} + \frac{\partial f}{\partial \phi^\alpha} + \Lambda - \mu^\alpha = 0, \\ \sum_{\alpha=1}^{N} \phi^\alpha = 1 \end{cases} \tag{6.60}$$

[42] It is partially related to the constraint sets though.

together with the appropriate isolating boundary conditions and Λ locally given by

$$\Lambda = \frac{1}{N}\sum_{\alpha=1}^{N} \epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi^\alpha}\right) - \epsilon\frac{\partial a}{\partial\phi^\alpha} - \frac{1}{\epsilon}\frac{\partial w}{\partial\phi^\alpha} - \frac{\partial f}{\partial\phi^\alpha}. \tag{6.61}$$

If the phasefield values are additionally subject to the constraints $0 \leq \phi^\alpha \leq 1\ \forall\alpha$, the conditions are, due to the additional sum-constraint, somewhat more complex than in the two-phase setting from Section 6.2. Considering first the simpler representation of the (locally) admissible set in Equation (6.2), one formally has that $\boldsymbol{\psi}$ is an admissible local variation if $\boldsymbol{\psi}$ satisfies $\sum_{\alpha=1}^{N}\psi^\alpha = 0$ with the additional restriction that $\psi^\alpha \geq 0$ if $\phi^\alpha = 0$. As in e.g. [11], if there are at least two **free phases** with $0 < \phi^\alpha, \phi^\beta < 1$, taking any $\psi > 0$ and $\boldsymbol{\psi} = \pm\psi\boldsymbol{e}^\alpha \mp \psi\boldsymbol{e}^\beta$ - thus satisfying the sum-constraint - Equation (6.59) shows that $\pm\psi(g^\alpha - g^\beta) \geq 0$ and therefore $g^\alpha = g^\beta$. As this argument can be repeated for an arbitrary combination of free phases, it follows that

$$g^\alpha(\boldsymbol{x}) = -\Lambda(\boldsymbol{x}) \quad \forall\alpha : 0 < \phi^\alpha < 1, \tag{6.62}$$

where Λ is, just as in the simpler case when restricted to Σ_1^N only, independent of the free phases. If there is any phase α at 0, there is necessarily at least one phase β with $\phi^\beta > 0$ due to the sum-constraint. In this case, one has to restrict the sign of the variations of ϕ^α taking e.g. again $\psi > 0$ and $\boldsymbol{\psi} = \psi\boldsymbol{e}^\alpha - \psi\boldsymbol{e}^\beta$, and one can only conclude that $\psi(g^\alpha - g^\beta) \geq 0$, i.e. $g^\alpha \geq g^\beta$ if $\phi^\alpha = 0$, $\phi^\beta > 0$.
If there is any β such that $0 < \phi^\beta < 1$, by the condition above, one therefore has $g^\alpha \geq -\Lambda$. It may happen though that the (then necessarily only) phase with $\phi^\beta > 0$ actually satisfies $\phi^\beta = 1$, in which case all other phases are automatically at 0. As there are no free phases, the above definition of Λ does not apply. Instead, one only has $g^\alpha \geq g^\beta\ \forall\alpha \neq \beta$. If one in this case **chooses** $\Lambda := -g^\beta$, the necessary conditions can be written in a unified manner in terms of the KKT system

$$\boldsymbol{g} = -\Lambda\boldsymbol{e} + \boldsymbol{\mu}, \quad \boldsymbol{\mu} \geq 0, \quad \mu^\alpha\phi^\alpha = 0, \tag{6.63}$$

or, more explicitly,

$$\begin{cases} -\epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi^\alpha}\right) + \epsilon\frac{\partial a}{\partial\phi^\alpha} + \frac{1}{\epsilon}\frac{\partial w}{\partial\phi^\alpha} + \frac{\partial f}{\partial\phi^\alpha} + \Lambda - \mu^\alpha = 0, \\ \sum_{\alpha=1}^{N}\phi^\alpha = 1, \\ \mu^\alpha \geq 0, \\ \mu^\alpha\phi^\alpha = 0 \end{cases} \tag{6.64}$$

for $\alpha = 1, ..., N$.
In contrast to the simpler well-case, there is now **no** simple explicit formula for Λ as in Equation (6.61) anymore. In fact, the same summation procedure as in the well-case applied to the first line in Equation (6.63) now shows that

$$\Lambda = \underbrace{\frac{1}{N}\sum_{\alpha=1}^{N}\left(\epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi^\alpha}\right) - \epsilon\frac{\partial a}{\partial\phi^\alpha} - \frac{1}{\epsilon}\frac{\partial w}{\partial\phi^\alpha} - \frac{\partial f}{\partial\phi^\alpha}\right)}_{=:\Lambda_{mw}} + \frac{1}{N}\sum_{\alpha=1}^{N}\mu^\alpha = \Lambda_{mw} + \frac{1}{N}\sum_{\alpha=1}^{N}\mu^\alpha. \tag{6.65}$$

Since the μ^α are non-negative, the expression for Λ in Equation (6.61) is therefore only a lower bound for the actual value of Λ in the obstacle case.

Remark 67. As the a-function in [52] is generally only defined indrectly in terms of $\boldsymbol{\phi}$ and $\nabla\boldsymbol{\phi}$ based upon the generalized gradient vectors $\boldsymbol{q}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) = \phi^\alpha\nabla\phi^\beta - \phi^\beta\nabla\phi^\alpha$ from Equation (6.5), it is often more convenient to instead consider a as a function of the $\boldsymbol{q}^{\alpha\beta}$ only and then obtain the derivative with respect to ϕ^α and $\nabla\phi^\alpha$ through a simple chain rule. Since[43]

$$\frac{\partial\boldsymbol{q}^{\beta\eta}}{\partial\phi^\alpha} = \nabla\phi^\eta\delta^{\alpha\beta} - \nabla\phi^\beta\delta^{\alpha\eta} \quad \text{and} \quad \frac{\partial\boldsymbol{q}^{\beta\eta}}{\partial\nabla\phi^\alpha} = \left(\phi^\beta\delta^{\alpha\eta} - \phi^\eta\delta^{\alpha\beta}\right)\boldsymbol{I},$$

[43] Where $\delta^{\alpha\beta}$ is the usual Kronecker symbol, i.e. 1 if $\alpha = \beta$ and 0 otherwise.

the differentiation of the summation in Equation (6.6) leads to

$$\frac{\partial a}{\partial \phi^\alpha} = \frac{\partial}{\partial \phi^\alpha} \sum_\beta \sum_{\eta>\beta} A^{\beta\eta}\big(\boldsymbol{q}^{\beta\eta}\big) = \sum_\beta \sum_{\eta>\beta} \frac{\partial A^{\beta\eta}}{\partial \boldsymbol{q}^{\beta\eta}}\Big(\nabla\phi^\eta \delta^{\alpha\beta} - \nabla\phi^\beta \delta^{\alpha\eta}\Big).$$

The summation over the first term then directly reduces to $\sum_{\eta>\alpha} \frac{\partial A^{\alpha\eta}}{\partial \boldsymbol{q}^{\alpha\eta}} \cdot \nabla\phi^\eta$, whereas the second one does, after exchanging the order of summation through $\sum_\beta \sum_{\eta>\beta} (\cdot)^{\beta\eta} = \sum_\eta \sum_{\beta<\eta} (\cdot)^{\beta\eta}$, lead to $-\sum_{\beta<\alpha} \frac{\partial A^{\beta\alpha}}{\partial \boldsymbol{q}^{\beta\alpha}} \cdot \nabla\phi^\beta$. With $\boldsymbol{q}^{\alpha\beta} = -\boldsymbol{q}^{\beta\alpha}$ and $A^{\alpha\beta}\big(\boldsymbol{q}^{\alpha\beta}\big) = A^{\beta\alpha}\big(\boldsymbol{q}^{\beta\alpha}\big) = A^{\beta\alpha}\big(-\boldsymbol{q}^{\alpha\beta}\big)$ and exchanging dummy-indices, one obtains the "missing" other half of the summation and thus finally

$$\frac{\partial a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})}{\partial \phi^\alpha} = \sum_{\beta\neq\alpha} \frac{\partial A^{\alpha\beta}}{\partial \boldsymbol{q}^{\alpha\beta}} \cdot \nabla\phi^\beta. \tag{6.66}$$

The same argument applied to the differentiation with respect to $\nabla\phi^\alpha$ in turn shows that

$$\frac{\partial a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})}{\partial \nabla\phi^\alpha} = -\sum_{\beta\neq\alpha} \phi^\beta \frac{\partial A^{\alpha\beta}}{\partial \boldsymbol{q}^{\alpha\beta}}. \tag{6.67}$$

This in principle simple but still somewhat tedious reasoning occurs very frequently in the multiphase case and will for the sake of brevity not be repeated in detail again[44], but can be verified to lead to the expressions

$$\frac{\partial w_{mw}(\boldsymbol{\phi})}{\partial \phi^\alpha} = 18\phi^\alpha \sum_{\beta\neq\alpha} \gamma^{\alpha\beta}(\phi^\beta)^2 + 2\phi^\alpha \sum_{\substack{\delta>\beta \\ \beta,\delta\neq\alpha}} \gamma^{\alpha\beta\delta}(\phi^\beta)^2(\phi^\delta)^2\Big) \tag{6.68}$$

and

$$\frac{\partial w_{mo}(\boldsymbol{\phi})}{\partial \phi^\alpha} = \frac{16}{\pi^2} \sum_{\beta\neq\alpha} \gamma^{\alpha\beta}\phi^\beta + \sum_{\substack{\delta>\beta \\ \beta,\delta\neq\alpha}} \gamma^{\alpha\beta\delta}\phi^\beta\phi^\delta \tag{6.69}$$

for the derivative of the bulk-potentials in the multi-well and multi-obstacle case. ◇

Remark 68. The choice $\Lambda = -g^\beta$ for the bulk-phase β above is the most natural one when explicitly enforcing only the positivity constraint explicitly and the constraint $\phi^\alpha \leq 1$, $\alpha = 1, ..., N$ implicitly through the sum-constraint as in the description of the Gibbs-simplex in Equation (6.2). One could of course, as in Equation (6.1), also enforce this upper bound explicitly, thus introducing a multiphase equivalent of the two multipliers $\mu^\pm$ in Section 6.2, then leading to the multiphase analogon

$$\begin{cases} -\epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi^\alpha}\right) + \epsilon\frac{\partial a}{\partial\phi^\alpha} + \frac{1}{\epsilon}\frac{\partial w}{\partial\phi^\alpha} + \frac{\partial f}{\partial\phi^\alpha} + \Lambda + \mu^{+,\alpha} - \mu^{-,\alpha} = 0, \\ \mu^{\pm,\alpha} \geq 0, \\ \mu^{-,\alpha}\phi^\alpha = 0, \\ \mu^{+,\alpha}(1-\phi^\alpha) = 0 \end{cases}$$

of Equation (6.23).
This difference is of course only relevant when there is actually a bulk phase, but then has the disadvantage of introducing an indeterminacy due to the partial redundancy of the description of the constraint set. In fact, the sign restriction on the multipliers $\boldsymbol{\mu}^\pm$ in the presence of a bulk-phase could be satisfied for any Λ satisfying $g^\beta \leq -\Lambda <= \min_{\alpha\neq\beta} g^\alpha$ as for any such value, one would still have $g^\alpha = -\Lambda + \mu^{-,\alpha}$ with $\mu^{-,\alpha} \geq 0$ for all phases with $\phi^\alpha = 0$, whereas $g^\beta = -\Lambda - \mu^{+,\beta}$

[44] A slightly quicker approach is to make use of the "symmetry" of the energy with respect to the pairings and thus to replace the summation over all distinct phase-pairings with one-half the summation over all phase-pairings.

with $\mu^{+,\beta} \geq 0$. Even though this indeterminacy is not a problem in principle and the particular choice of Λ (and thus $\mu^{+,\beta}$) does not affect the actual result, this is an unnecessary complication which can easily be avoided by using the non-redundant description in Equation (6.2).
It should be noted though that this leads to a difference of the multipliers when using a multi-phase formulation with two phases as compared to the reduced two-phase formulation. The reduced formulation implictly corresponds to choosing (up to a factor of 2) the multiplier $\Lambda = -\frac{1}{2}(g^1 + g^2)$ and thus will lead to the multipliers $\mu^{\pm}$ being equal to (twice) that value instead of the choice g^β above anytime a bound-constraint is active (which then automatically implies the presence of a bulk-phase)[45]. ◇

Remark 69. Similar indeterminacies arise quite frequently in non-reduced formulations. Another simple example in relation with a volume-constraint will be seen in Section 6.3.3, whereas a more complex one in a concentration-based setting will be discussed in some detail in Section 7.1. ◇

Remark 70. Recall from Section 6.2 that imposing $\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \boldsymbol{n}$ is much less obvious in the presence of constraints than it might appear at first sight, since the boundary condition it is not a priori "up for choice" but actually has to be taken in accordance with the first-order necessary condition.
In particular, due to the multiphase setting, there is now even in the well-case an additional restriction $\sum_{\alpha=1}^N \psi^\alpha = 0$ on the variations ψ^α, which a priori only leads to the necessity of $\sum_{\alpha=1}^N \big(\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \boldsymbol{n}\big)\psi^\alpha \stackrel{!}{=} 0$ and thus $\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \boldsymbol{n} = \Lambda$ for all α.
In contrast to the interior of the domain, where "choosing" to enforce $g^\alpha = 0,\ \alpha = 1, ..., N$ (and thus $\Lambda = 0$) is potentially possible, but practically useless for making the volume integral vanish[46] (and therefore), enforcing $\Lambda = 0$ and therefore the vanishing of the conormal derivative on the boundary would seem a lot less restrictive. Nevertheless, even though this is clearly compatible with the first-order necessary condition, it is not a priori obvious whether this is a necessity or simply a selection having - with the pure bulk solution with the lowest f^α - at least one solution which is also a global minimizer. Showing that this is indeed a necessity is expected, but, similar to the discussion in Section 6.2, actually requires a non-trivial argument even in the simplest isotropic case[47] and will therefore simply be assumed here. ◇

---

[45] Note also that $g^\alpha \geq g^\beta$ in the two-phase case automatically ensures $g^\beta \leq -\Lambda = \frac{g^\alpha + g^\beta}{2} \leq g^\alpha$. A similar inequality based on the average of more than two phases however is generally **not** valid.

[46] This is basically only compatible with the solution consisting out of a single bulk-phase either in the absence of driving forces or using one of the h-functions satisfying $h'(0) = h'(1) = 0$.

[47] Based on simplifying the expression for (6.67), one has

$$\frac{\partial a}{\partial \nabla \phi^\alpha} \cdot \boldsymbol{n} = -\sum_{\beta \neq \alpha} \phi^\beta \boldsymbol{q}^{\alpha\beta} \cdot \boldsymbol{n} = -\sum_{\beta \neq \alpha} \phi^\beta \big(\phi^\alpha \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} - \phi^\beta \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}\big) \stackrel{!}{=} \Lambda, \alpha = 1, ..., N.$$

Similar to the interior of the domain, a simple summation over all phases, but now using $\sum_{\beta \neq \alpha} = 1 - \phi^\alpha$ as well as $\sum_{\beta \neq \alpha} \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} = -\frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}$ due to the sum-constraint, shows that

$$\sum_\beta \sum_{\alpha \neq \beta} \phi^\beta \big(\phi^\alpha \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} - \phi^\beta \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}\big) = \sum_\beta \phi^\beta \big((1 - \phi^\beta)\frac{\partial \phi^\beta}{\partial \boldsymbol{n}} - \phi^\beta(-\frac{\partial \phi^\beta}{\partial \boldsymbol{n}})\big) = \sum_\beta \phi^\beta \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} = -N\Lambda. \tag{6.70}$$

Making use of this relation, one can eliminate one of the summations in the condition on the normal derivative for the individual phases, since

$$-\sum_{\beta \neq \alpha} \phi^\beta \big(\phi^\alpha \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} - \phi^\beta \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}\big) = -\phi^\alpha \sum_{\beta \neq \alpha} \phi^\beta \frac{\partial \phi^\beta}{\partial \boldsymbol{n}} + \sum_{\beta \neq \alpha} \big(\phi^\beta\big)^2 \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}$$

$$= -\phi^\alpha \Big(\big(\sum_\beta \phi^\beta \frac{\partial \phi^\beta}{\partial \boldsymbol{n}}\big) - \phi^\alpha \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}}\Big) + \sum_{\beta \neq \alpha} \big(\phi^\beta\big)^2 \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}} = \phi^\alpha N\Lambda + \Big(\sum_\beta \big(\phi^\beta\big)^2\Big)\frac{\partial \phi^\alpha}{\partial \boldsymbol{n}} = \Lambda$$

and thus (with $\sum_\beta \big(\phi^\beta\big)^2 > 0$)

$$\frac{\partial \phi^\alpha}{\partial \boldsymbol{n}} = \frac{\Lambda - \phi^\alpha N\Lambda}{\sum_\beta \big(\phi^\beta\big)^2}.$$

6.3.2 The Choice of Dynamics

In practice, the phasefield model is most commonly employed as a dynamic model for the evolution of microstructures, and not in terms of a pure minimization problem. While the choice of a phasefield functional does, within a variational framework, fix the potential equilibrium points through the solutions of the steady-state equations (6.60) resp. (6.64) above, this does not imply anything specific about the evolution of the phasefield variables, besides the very broad requirement that they should converge to a (potentially local) minimizer resp. maximizer of the phasefield functional. A simple and natural way of enforcing such a behavior is to postulate a gradient-type flow. In the Allen-Cahn case, the postulate is that of a non-conservative gradient flow based on the L^2-gradient of $\mathcal{E}$, i.e. $\frac{\partial \boldsymbol{\phi}}{\partial t} \sim -\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}$, with individual models differing in the particular choice of proportionality to $\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}$.
The model in [52] corresponds to a multiphase version of the simple scalar proportionality in Equation (6.24), i.e. to postulating

$$\tau\epsilon\frac{\partial \boldsymbol{\phi}}{\partial t} = -\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} - \Lambda \boldsymbol{e}, \tag{6.71}$$

or, in combination with the Gibbs-simplex constraint,

$$\tau\epsilon\frac{\partial \boldsymbol{\phi}}{\partial t} = -\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} - \Lambda \boldsymbol{e} + \boldsymbol{\mu}, \tag{6.72}$$

where $\boldsymbol{e}$ is an N-dimensional vector of ones and $\boldsymbol{\mu}$ is subject to the same complementarity conditions as in Equation (6.64). In a more explicit form for the indiviual phases, this can also be written as

$$\tau\epsilon\frac{\partial \phi^\alpha}{\partial t} = \epsilon\nabla\cdot\left(\frac{\partial a}{\partial \nabla\phi^\alpha}\right) - \epsilon\frac{\partial a}{\partial \phi^\alpha} - \frac{1}{\epsilon}\frac{\partial w}{\partial \phi^\alpha} - \frac{\partial f}{\partial \phi^\alpha} - \Lambda, \quad \alpha = 1, ..., N. \tag{6.73}$$

resp.

$$\tau\epsilon\frac{\partial \phi^\alpha}{\partial t} = \epsilon\nabla\cdot\left(\frac{\partial a}{\partial \nabla\phi^\alpha}\right) - \epsilon\frac{\partial a}{\partial \phi^\alpha} - \frac{1}{\epsilon}\frac{\partial w}{\partial \phi^\alpha} - \frac{\partial f}{\partial \phi^\alpha} - \Lambda + \mu^\alpha, \quad \alpha = 1, ..., N. \tag{6.74}$$

together with $\sum_{\alpha=1}^N \phi^\alpha = 1$ as well as $\mu^\alpha \geq 0$ and $\mu^\alpha = 0$ if $\phi^\alpha > 0$ in the obstacle case.

One disadvantage of using a scalar-valued kinetic coefficient is that this can lead to difficulties within multiphase regions involving a mixture of highly mobile and highly immobile phases (or phase-pairings). In fact, since τ is locally the same for the evolution of all phases, an interpolation favoring the immobile interfaces may artifically slow down the evolution of the mobile ones and vice versa. This raises the question of how and to what degree one can properly interpolate between the individual mobilities such as to obtain satisfactory results.
A popular alternative, propagated in particular in [68] and used for e.g. also in [61] and [65], is again based on considering two-phase interactions, but now postulating an evolution of the form

$$\frac{\partial \phi^\alpha}{\partial t} = -\sum_{\beta\neq\alpha} m^{\alpha\beta}\Big(\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\phi^\alpha} - \frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\phi^\beta}\Big), \; \alpha = 1, ..., N. \tag{6.75}$$

This corresponds to postulating that the evolution of each phase is determinted as the sum of its individual interactions with all other phases, each such interaction being equipped with its

This can be recombined with Equation (6.70) - now in terms of a ϕ^α-weighted summation over the previous relation for the indivual phases - to obtain

$$-N\Lambda = \sum_\alpha \phi^\alpha \frac{\partial \phi^\alpha}{\partial \boldsymbol{n}} = \frac{\sum_\alpha \phi^\alpha \Lambda - \sum_\alpha (\phi^\alpha)^2 N\Lambda}{\sum_\beta \left(\phi^\beta\right)^2} = \frac{\Lambda - \sum_\alpha (\phi^\alpha)^2 N\Lambda}{\sum_\beta \left(\phi^\beta\right)^2} = \frac{\Lambda}{\sum_\beta \left(\phi^\beta\right)^2} - N\Lambda,$$

from which it follows that Λ indeed has to be zero.

own mobility $m^{\alpha\beta} = m^{\beta\alpha}$.
This may also be written more compactly as

$$\epsilon \frac{\partial \boldsymbol{\phi}}{\partial t} = -\boldsymbol{M} \frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}, \quad (\boldsymbol{M})_{\alpha\beta} = \begin{cases} \sum_{\gamma \neq \alpha} m^{\alpha\gamma} & \beta = \alpha, \\ -m^{\alpha\beta} & \text{else}, \end{cases} \tag{6.76}$$

and therefore amounts to replacing the scalar-valued kinetic coefficient on the left-hand side of Equation (6.71) with a symmetric **mobility-matrix** $\boldsymbol{M}$ premultiplying the gradient.
It should be noted that there is no multiplier Λ for the sum-constraint in Equations (6.75) resp. (6.76). This is due to the fact that the mobility matrix $\boldsymbol{M}$ above satisfies $\boldsymbol{M}\boldsymbol{e} = \boldsymbol{e}^T\boldsymbol{M} = \boldsymbol{0}$ and thus on the one hand $\boldsymbol{M} \cdot (\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} + \Lambda \boldsymbol{e}) = \boldsymbol{M} \cdot \frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}$, i.e. applying $\boldsymbol{M}$ to the gradient of the Lagrangian $\mathcal{L}(\boldsymbol{\phi}) = \mathcal{F}_\epsilon(\boldsymbol{\phi}) + \int_\Omega \Lambda(\boldsymbol{e} \cdot \boldsymbol{\phi} - 1)\,\mathrm{d}\boldsymbol{x}$ associated with the local sum constraint leads to the same result as applying $\boldsymbol{M}$ directly to the gradient of $\mathcal{F}_\epsilon$. On the other hand, as

$$\frac{\partial \sum_{\alpha=1}^N \phi^\alpha}{\partial t} = \frac{\partial \boldsymbol{e} \cdot \boldsymbol{\phi}}{\partial t} = -\frac{1}{\epsilon} \boldsymbol{e} \cdot \boldsymbol{M} \frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} = \boldsymbol{0},$$

there is no need to explicitly introduce the multiplier Λ since the phasefield will remain consistent with the sum-constraint provided the initial values were so.

Remark 71. This is analogous in nature to the choice of the mobility matrix $\boldsymbol{L}$ in [52] for the concentration- and energy-evolution, which, in addition to the multiplier Λ in the phasefield equation would a priori also require another one $\Lambda_{\boldsymbol{c}}$ for the concentration equation due to the constraint $\sum_{i=1}^K c_i = 1$. As $\boldsymbol{L}$ is chosen to be only a positive semi-definite matrix satisfying $\sum_{i=1}^K L_{ij} = \sum_{j=1}^K L_{ij} = 0$ for all rows resp. columns, the multiplier for the concentration drops out of the resulting equations while the sum-constraint nevertheless remains satisfied provided it was so initially. ◇

This is in constrast to the multipliers $\boldsymbol{\mu}$ arising due to the inequality constraints in the obstacle case, which still need to be considered explicitly such that the evolution equation has to be modified to (for the representation in (6.2) of the Gibbs-simplex)

$$\epsilon \frac{\partial \boldsymbol{\phi}}{\partial t} = -\boldsymbol{M}\Big(\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} - \boldsymbol{\mu}\Big), \quad \mu^\alpha \phi^\alpha = 0,\ \mu^\alpha \geq 0,\ \alpha = 1, ..., N. \tag{6.77}$$

Even though this formulation still has the apparent advantage of containing one multiplier less than the obstacle-version of Equation (6.71) (the sum-constraint remaining satisfied regardless of the choice of $\boldsymbol{\mu}$ as well), this simply shifts the indirect interaction of each non-zero multiplier μ^α with the other phases through the sum-constraint to a more explicit one in terms of the full matrix $\boldsymbol{M}$.

Remark 72. Note that one could in principle also rewrite Equation (6.71) resp. its obstacle-analogon in a similar manner as[48]

$$\frac{\partial \boldsymbol{\phi}}{\partial t} = -\frac{1}{\tau\epsilon}\Big(\boldsymbol{I} - \frac{1}{N}\boldsymbol{e} \otimes \boldsymbol{e}\Big)\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} \text{ resp. } \frac{\partial \boldsymbol{\phi}}{\partial t} = -\frac{1}{\tau\epsilon}\Big(\boldsymbol{I} - \frac{1}{N}\boldsymbol{e} \otimes \boldsymbol{e}\Big)\Big(\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}} - \boldsymbol{\mu}\Big),$$

where the Lagrange-multiplier Λ has been eliminated through the use of the "mobility matrix" $\boldsymbol{M} = \frac{1}{\tau\epsilon}\Big(\boldsymbol{I} - \frac{1}{N}\boldsymbol{e} \otimes \boldsymbol{e}\Big)$ corresponding to a scalar multiple of the Euclidian projection operator onto the subspace of zero-average vectors. As the diagonal entries of this projection matrix are given by $\frac{1}{\tau\epsilon}\Big(1 - \frac{1}{N}\Big) = \frac{N-1}{N\tau\epsilon} = -\sum_{\beta \neq \alpha}\big(-\frac{1}{N\tau\epsilon}\big)$, this is just a particular simple form of the one above obtained by setting $m^{\alpha\beta} = -\frac{1}{N\tau\epsilon}\ \forall \alpha \neq \beta$.

[48] This corresponds to a (partially in the obstacle case) "projected form" of the equation.

It is therefore not surprising that both approaches actually share many of the same features and difficulties, the most notable difference being that the particularly simple form in the latter case often allows for more "explicit calculations".
One point where the more complex form of the mobility matrix $\boldsymbol{M}$ leads to a notable complication is when reyling on a projection-based descent algorithm. In fact, the modified interaction of the gradients $\boldsymbol{g}$ through $\boldsymbol{M}$ also has to be taken into account during this projection operation (in line with the role of $\boldsymbol{\mu}$ in Equation (6.77)), and entails the necessity of using a more complex $\boldsymbol{M}$-weighted projection operator instead of the simpler Euclidian one. This projection, in particular some of its algorithmic implications will be discussed in Section 6.3.3.
It can also be noted that in the two-phase case, the matrix $\frac{1}{\tau\epsilon}\big(\boldsymbol{I}-\boldsymbol{e}\otimes\boldsymbol{e}\big)$ reduces to $\frac{1}{2\tau\epsilon}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, and both approaches are therefore clearly equivalent if m^{12} is chosen as $\frac{1}{2\tau}$. ◇

Remark 73. The two approaches outlined above are clearly not the only ones possible. In particular, one can in principle use any spd mobility matrix $\boldsymbol{M}$ provided one is willing to maintain the explicit multiplier Λ. Similarly, one could also choose to replace the use of a mobility matrix by that of a matrix $\boldsymbol{\tau}$ of kinetic coefficients and postulating $\boldsymbol{\tau}\frac{\partial\boldsymbol{\phi}}{\partial t} = \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\boldsymbol{\phi}} - \Lambda$ (this of course being equivalent to the previous point in the spd case by setting $\boldsymbol{M} = \boldsymbol{\tau}^{-1}$ or vice versa)[49]. The latter may be potentially useful for obtaining a somewhat "intermediate" model between the scalar kinetic coefficient and the use of a full mobility matrix if applied in the form

$$\tau^\alpha\epsilon\frac{\partial\phi^\alpha}{\partial t} = -\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\phi^\alpha} - \Lambda, \quad \alpha = 1, ..., N$$

i.e. by using a a simple diagonal matrix $\boldsymbol{\tau} = \mathrm{diag}\big((\tau^\alpha)_{1\leq\alpha\leq N}\big)$ allowing for some additional freedom in specifying the dynamics while remaining in an "almost explicit" form due to the diagonal structure of $\boldsymbol{\tau}$. ◇

6.3.3 Some Numerical and Algorithmic Considerations

A first important point to observe from a structural point of view is that the derivative conribution from $w(\boldsymbol{\phi})$ now always has a nonlinear contribution due to the triple-phase terms, even in the previously linear obstacle case. Depending on the size of the penalty parameters $\gamma^{\alpha\beta\delta}$, this contribution may be relatively large and can thus potentially also lead to additional numercial difficulties. Similarly, due to the formulation of the surface energy densities in terms of the (nonlinear) $\boldsymbol{q}^{\alpha\beta}$, the derivative contribution due to a will also always be nonlinear both in $\boldsymbol{\phi}$ and $\nabla\boldsymbol{\phi}$, even in the previously linear isotropic setting.
An additional complication, in particular in the obstacle case, arises due to the nature of the admissible set. Whereas in the two-phase case, the sum-constraint is easily absorbed into a reduced formulation and thus leads to no actual constraint in the double-well case and a simple box-constraint in the double-obstacle case, it is more common when dealing with multiphase problems to remain within a non-reduced formulation. This requires maintaining the sum-constraint as an explicit constraint. While this is generally easily dealt with in the well case, the combination with the positivity constraint is a somewhat trickier issue. In particular, there is no simple explicit formula even for the simplest (Euclidian) projector onto the Gibbs-simplex, as the combinatoric nature of the projection operation is more pronounced due to the higher complexity of the constraint set (regardless of whether one uses a reduced formulation or not).

[49] In fact, if one maintains the explicit presence of the Lagrange multiplier Λ, one could also use a spsd $\boldsymbol{\tau}$-matrix whose kernel (and, by symmetry, cokernel) coincides with $\mathrm{Span}\{\boldsymbol{e}\}$, even though this seems of somewhat limited practical interest. While $\boldsymbol{\tau}$ is then not invertible, the same choice of Λ (up to a prefactor of $\tau\epsilon$) as for Equation (6.71) ensures consistency (i.e. orthogonality to the constants) of the left- and right-hand side, while the indeterminacy in $\frac{\partial\boldsymbol{\phi}}{\partial t}$ is eliminated by the presence of the sum-constraint.

Remark 74. Even though it is in principle possible to use a reduced formulation in the multiphase setting by eliminating one of the N phasefields in terms of the remaining $N-1$, any potential benefits of doing so would primarily be restricted to the multi-well potential. The central difficulty in the multi-obstacle case is not the sum-constraint, but its coupling with the bounds on the $\boldsymbol{\phi}$-values. As this implies the risk that the phase which was eliminated in favor of the others may actually not be allowed to change according to the changes of the other ones, this does not in general avoid the issue of the sum-constraint. This is in constrast to the two-phase case, where any one phase trying to move outside one of the 0-1-bound is equivalent to the other one trying to move outside the opposite one. ◇

Before continuing the discussion on the challenges associated with "true" multi-phasefield problems with $N > 2$, it is worthwhile to first take a closer look at how the formulation above compares with the simpler description in terms of a single phasefield-variable ϕ from the previous section. While it is clear that they are in principle equivalent, the redundant representation in terms of $\phi^1 := \phi$ and $\phi^2 = 1 - \phi^1 = 1 - \phi$ leads to a number of "technicalities" which should be distinguished from the actual difficulties associated with the situation for $N > 2$.

Comparison of the "Single-Phase" to the "Two-Phase" Formulation

If one **explicitly** enforces the sum-constraint by reexpressing e.g. ϕ^2 as a function of ϕ^1, i.e. with $\phi^2 = 1 - \phi^1$, $\nabla\phi^2 = -\nabla\phi^1$ and thus also $\boldsymbol{q}^{12} = -\nabla\phi^1$, it is easily seen that the values of the a- and w-term (not the functions though!) reduce to the same expressions that would be obtained from Equation (6.16). If the reduced functions in terms of $\phi^2 = \phi^2(\phi^1)$ are denoted with a hat-symbol, i.e. for example $\hat{w}(\phi^1) = w(\phi^1, \phi^2(\phi^1))$, as a simple consequence of the chain rule and of

$$\frac{\partial \boldsymbol{q}^{12}}{\partial \phi^1} = \nabla\phi^2, \quad \frac{\partial \boldsymbol{q}^{12}}{\partial \phi^2} = -\nabla\phi^1, \quad \frac{\partial \boldsymbol{q}^{12}}{\partial \nabla\phi^1} = -\phi^2 \boldsymbol{I}, \quad \frac{\partial \boldsymbol{q}^{12}}{\partial \nabla\phi^2} = \phi^1 \boldsymbol{I}$$

the respective derivatives in the phasefield equation are given by

$$\begin{aligned}\frac{1}{\gamma^{12}}\frac{\delta \hat{a}(\phi^1)}{\delta \phi^1} =& \frac{1}{\gamma^{12}}\frac{\delta a}{\delta \phi^1} - \frac{\delta a}{\delta \phi^2} = -\nabla \cdot \Big(\frac{\partial a}{\partial \nabla\phi^1} - \frac{\partial a}{\partial \nabla\phi^2}\Big) + \Big(\frac{\partial a}{\partial \phi^1} - \frac{\partial a}{\partial \phi^2}\Big) \\ =& -\nabla \cdot \Big(\underbrace{-\phi^2}_{=1-\phi^1} \boldsymbol{q}^{12} + \phi^1 \boldsymbol{q}^{12}\Big) + \boldsymbol{q}^{12} \cdot \Big(\underbrace{\nabla\phi^2}_{=-\nabla\phi^1} - (-\nabla\phi^1)\Big) = \nabla \cdot \boldsymbol{q}^{12} = -\Delta\,\phi^1\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \hat{w}_{dw}}{\partial \phi^1} =& \frac{\partial w_{dw}}{\partial \phi^1} - \frac{\partial w_{dw}}{\partial \phi^2} = 18\gamma^{12}\big(\phi^1(\phi^2)^2 - (\phi^1)^2\phi^2\big) = 18\gamma^{12}\phi^1\phi^2(\phi^2 - \phi^1) \\ =& 18\gamma^{12}\phi^1(1-\phi^1)(1-2\phi^1)\end{aligned}$$

for the double-well potential and

$$\frac{\partial \hat{w}_{do}}{\partial \phi^1} = \frac{\partial w_{do}}{\partial \phi^1} - \frac{\partial w_{do}}{\partial \phi^2} = \frac{16}{\pi^2}\gamma^{12}\big(\phi^2 - \phi^1\big) = \frac{16}{\pi^2}\gamma^{12}(1-2\phi^1)$$

for the double-obstacle potential respectively. In addition, if $\tilde{h}$ is based on any of the "standard" two-phase interpolation functions and $\phi^1 + \phi^2 = 1$, and thus also $\tilde{h}^1(\boldsymbol{\phi}) + \tilde{h}^2(\boldsymbol{\phi}) = 1$, then the denominator drops out and the values of the derivatives reduce to $\frac{\partial h^1}{\partial \phi^1} = \tilde{h}(\phi^2)\tilde{h}'(\phi^1)$ resp. $\frac{\partial h^1}{\partial \phi^2} = -\tilde{h}(\phi^1)\tilde{h}'(\phi^2)$. As one further has $\tilde{h}'(\phi^2) = -\tilde{h}'(\phi^1)$, the derivative of the f-contribution

becomes

$$\begin{aligned}\frac{\partial \hat{f}(\phi^1)}{\partial \phi^1} =& \frac{\partial f}{\partial \phi^1}\big((\phi^1, \phi^2(\phi^1)\big) = \frac{\partial f}{\partial \phi^1} - \frac{\partial f}{\partial \phi^2} = f^1\big(\frac{\partial h^1}{\partial \phi^1} - \frac{\partial h^1}{\partial \phi^2}\big) + f^2\big(\frac{\partial h^2}{\partial \phi^1} - \frac{\partial h^2}{\partial \phi^2}\big)\\ =& f^1\big(\tilde{h}(\phi^2)\tilde{h}'(\phi^1) - \tilde{h}(\phi^1)\tilde{h}'(\phi^2)\big) + f^2\big(\tilde{h}(\phi^1)\tilde{h}'(\phi^2) - \tilde{h}(\phi^2)\tilde{h}'(\phi^1)\big)\\ =& (f^1 - f^2)\big(\underbrace{\tilde{h}(\phi^2)}_{=1-\tilde{h}(\phi^1)}\tilde{h}'(\phi^1) - \tilde{h}(\phi^1)\underbrace{\tilde{h}'(\phi^2)}_{=-\tilde{h}'(\phi^1)}\big) = (f^1 - f^2)\tilde{h}'(\phi^1),\end{aligned}$$

showing that this term as well reduces to the previous expression.

This is of course not particularly surprising as the multi-phasefield formulation is naturally intended as a generalization of the two-phase case. The simple calculations above nevertheless highlight two important points:

1. Whereas the original (isotropic) "one-phase" formulation is linear in the second-order term and only involves at most low-order polynomial expressions in the derivatives of the w- and/or f-term, its "two-phase" counterpart is a priori also nonlinear in the second order term and involves much more unpleasant nonlinearities (fractions over squares of the original h-function) in the f-term.

2. If the problem is not from the outset reduced to the previous version by using either $\phi^2 = \phi^2(\phi^1)$ or $\phi^1 = \phi^1(\phi^2)$, the cancellations occuring in the reduced formulation will instead have to be enforced through an external Lagrange mulitplier λ for the sum-constraint $\phi^1 + \phi^2 = 1$. This, by itself, poses no particular problem. Nevertheless, if one additionally chooses to (or has to) enforce the constraints $0 \leq \phi^1, \phi^2 \leq 1$, a direct extension of the procedure applied in the two-phase case using separate multipliers $\mu^{1/2,\pm}$ for the lower and upper constraints entails some additional technical difficulties. Due to the sum constraint, there will be either no active inequality constraint or **both** equality constraints will be active at the same time. In the former case, the Lagrange multiplier λ for the sum-constraint will, as expected, just equal the average of both derivatives. In the latter case though, there are a total of three multipliers (for example λ, $\mu^{1,-}$ and $\mu^{2,+}$ if one has $\phi^1 = 0$, $\phi^2 = 1$) to be determined based on only two actual unknowns ϕ^1 and ϕ^2 and one is faced with an underdetermined problem. Barring the expectional case in which both constraints are only weakly active (i.e. if both phases would remain in 0 resp. 1 even without the box-constraints), this necessarily results in a non-unique (though bounded) set of Lagrange multipliers.
   This is not a problem per se, as there is no real need to have uniqueness for the Lagrange multipliers. In particular, this indeterminancy can be dealt with through an appropriate selection mechanism and the resulting phasefield-values will be the same regardless of the particular choice made. Depending on the way λ and the $\mu^{\alpha,\pm}$ are determined, this nevertheless may require some care when e.g. working with direct solvers for the associated matrices.

Remark 75. A similar issue with respect to the multipliers arises also if one e.g. imposes an additional constraint $\int_\Omega \phi^\alpha = V^\alpha$ on the phase-volumes. These a priori two constraints require an additional set of Lagrange multipliers (χ^1, χ^2), leading to the modification

$$\tau\epsilon\frac{\partial \phi^\alpha}{\partial t} = -g^\alpha - \lambda - \chi^\alpha, \quad \alpha = 1, 2,$$

of the right-hand sides (or, when using the slightly modified approach in [53] and [29], of the form $\sum_{\beta=1}^2 \chi^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$) and the appropriate modification

$$\lambda(\boldsymbol{x}, t) = \frac{1}{2}\big(-g^1(\boldsymbol{x}, t) - g^2(\boldsymbol{x}, t) - (\chi^1 + \chi^2)\big) \tag{6.78}$$

of λ. There is again a redundancy in this case as it would be sufficient to fix just one of the volumes, e.g. that of the first phase, and then let λ "handle the rest". In particular with respect to splitting-type approaches for the above system, it is highly preferable to maintain, at least formally, both χ^1 and χ^2 as separate multipliers as this avoids enforcing the global constraint on the other phase indirectly through a multiplier which is intuitively associated with the local constraint. A natural selection criterion for eliminating the indeterminacy in this case is given by choosing as in [53] χ^1 and χ^2 such $\chi^1 + \chi^2 = 0$, since, by Equation (6.78), this leaves λ unaffected by the additional constraint. ◇

Some Choices for the Spatial Discretization

The presence of more than two phases has essentially no impact on the simplest discretization of the local terms due to the bulk potential and the driving forces within the cell-centered scheme used in the **Pace3D**-framework. In contrast to the pure two-phase setting, the derivative of the gradient energy density in the simplest isotropic case cannot be reduced to a Laplacian anymore as soon as there are more than two phases present, and one does thus not have an "obvious" choice of discretziation through the standard second-order difference quotient anymore[50].
Based on the divergence-type nature of the expression $\epsilon\nabla\cdot\left(\frac{\partial a}{\partial\nabla\phi^\alpha}\right)$, a very convenient discretization for this term is to express it as the differences of fluxes over the cell "faces", i.e. in the simplest one-dimensional setting as

$$\epsilon\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right) \approx \frac{1}{\Delta x}\left(\left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right)_{i+\frac{1}{2}} - \left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right)_{i-\frac{1}{2}}\right),$$

where the $\left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right)_{j+\frac{1}{2}}$ correspond to appropriate approximations of this derivative on these faces and the prime indicates a derivative with respect to x. As $\left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right)$ is in the isotropic case (this being the only reasonable choice in 1D) given by

$$\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)} = -2\left(\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\phi^\beta q^{\alpha\beta}\right)$$

and $q^{\alpha\beta}(\boldsymbol{\phi}, \frac{\mathrm{d}\boldsymbol{\phi}}{\mathrm{d}x}) = \phi^\alpha\frac{\mathrm{d}\phi^\beta}{\mathrm{d}x} - \phi^\beta\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}$, the most obvious choice of discretization for this expression on the faces is obtained by using a short one-cell stencil for the first-order derivatives and the average of the neighboring $\boldsymbol{\phi}$-values for the values of the phasefield itself. Defining $\boldsymbol{\phi}_{i+\frac{1}{2}} = \frac{1}{2}\left(\boldsymbol{\phi}_i + \boldsymbol{\phi}_{i+1}\right)$, this leads to the natural approximations

$$q^{\alpha\beta}_{i+\frac{1}{2}} = \phi^\alpha_{i+\frac{1}{2}}\frac{\phi^\beta_{i+1} - \phi^\beta_i}{\Delta x} - \phi^\beta_{i+\frac{1}{2}}\frac{\phi^\alpha_{i+1} - \phi^\alpha_i}{\Delta x} \tag{6.79}$$

and

$$\epsilon\left(\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial a}{\partial\left(\frac{\mathrm{d}\phi^\alpha}{\mathrm{d}x}\right)}\right)\right)_i \approx -2\epsilon\frac{1}{\Delta x}\left[\left(\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\phi^\beta_{i+\frac{1}{2}}q^{\alpha\beta}_{i+\frac{1}{2}}\right) - \left(\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\phi^\beta_{i-\frac{1}{2}}q^{\alpha\beta}_{i-\frac{1}{2}}\right)\right]. \tag{6.80}$$

In contrast, there are actually two quite natural discretizations for the second contribution from the term $\epsilon\frac{\partial a}{\partial\phi^\alpha}$, which is, in the isotropic case, given by $\epsilon\frac{\partial a}{\partial\phi^\alpha} = 2\epsilon\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}q^{\alpha\beta}\frac{\mathrm{d}\phi^\beta}{\mathrm{d}x}$. The first and most intuitive one is to evaluate it in a cell-centered fashion, i.e. by using the approximation

$$\left(\epsilon\frac{\partial a}{\partial\phi^\alpha}\right)_i \approx 2\epsilon\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\left(\phi^\alpha_i\frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} - \phi^\beta_i\frac{\phi^\alpha_{i+1} - \phi^\alpha_{i-1}}{2\Delta x}\right)\frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} \tag{6.81}$$

[50] Using higher-order stencils is of course in principle possible, but, in the obstacle case, unlikely to lead to any major improvement by the discussion in Section 6.2.3.

through central gradients only. An alternative is to try to reuse the already calculated gradients of $\boldsymbol{\phi}$ and the $q^{\alpha\beta}$ on the cell faces and thus to express $\epsilon\frac{\partial a}{\partial \phi^\alpha}$ as the average of its approximation on the left and right face as

$$\left(\epsilon\frac{\partial a}{\partial \phi^\alpha}\right)_i \approx 2\epsilon\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\frac{1}{2}\left(q^{\alpha\beta}_{i-\frac{1}{2}}\frac{\phi_i^\beta-\phi_{i-1}^\beta}{\Delta x}+q^{\alpha\beta}_{i+\frac{1}{2}}\frac{\phi_{i+1}^\beta-\phi_i^\beta}{\Delta x}\right). \tag{6.82}$$

The use of this - a priori slightly less intuitive - expression as compared to the one in Equation (6.81) is two-fold. Firstly, it can indeed partially rely upon the same face-centered quantities $q^{\alpha\beta}_{j+\frac{1}{2}}$ already required for the evaluation of the divergence-term in Equation (6.80), but at the price of introducing the same spatial buffering requirement as for the fluxes over the cell faces if one wants to avoid recalculating the same values twice. Secondly, the resulting discrete phasefield equation based on the combined discretization of the gradient energy contributions as in equations (6.80) and (6.82) together with the appropriate isolating boundary conditions corresponds itself to the derivative of a discrete energy "functional" $\mathcal{F}_{\epsilon,\Delta x}$ given by

$$\mathcal{F}_{\epsilon,\Delta x}(\boldsymbol{\phi})=\sum_i\frac{1}{2}\epsilon\left(a\big(\boldsymbol{\phi}_{i-\frac{1}{2}},\frac{\boldsymbol{\phi}_i-\boldsymbol{\phi}_{i-1}}{\Delta x}\big)+a\big(\boldsymbol{\phi}_{i+\frac{1}{2}},\frac{\boldsymbol{\phi}_{i+1}-\boldsymbol{\phi}_i}{\Delta x}\big)\right)+\frac{1}{\epsilon}w(\boldsymbol{\phi}_i)+f(\boldsymbol{\phi}_i), \tag{6.83}$$

where $a\big(\boldsymbol{\phi}_{j+\frac{1}{2}},\frac{\boldsymbol{\phi}_{j+1}-\boldsymbol{\phi}_j}{\Delta x}\big)$ abbreviates the summation $\sum_\alpha\sum_{\beta>\alpha}|q^{\alpha\beta}_{j+\frac{1}{2}}|^2$ with $q^{\alpha\beta}_{j+\frac{1}{2}}$ defined as in Equation (6.79).

This is obvious for the local terms arising due to w and f. For seeing that this is also true with respect to the spatial terms in the gradient energy contribution, it is convenient to first rewrite some of the expressions above.

Firstly, it is easy to verify that the expression for $q^{\alpha\beta}$ in Equation (6.79) can be simplified to

$$q^{\alpha\beta}_{i+\frac{1}{2}}=\frac{\phi_i^\alpha\phi_{i+1}^\beta-\phi_i^\beta\phi_{i+1}^\alpha}{\Delta x}. \tag{6.84}$$

Secondly, the combined contributions from the divergence term and the derivative of a with respect to ϕ^α can be summarized to

$$\begin{aligned}
&-2\epsilon\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\left\{\frac{1}{\Delta x}\left[\phi^\beta_{i+\frac{1}{2}}q^{\alpha\beta}_{i+\frac{1}{2}}-\phi^\beta_{i-\frac{1}{2}}q^{\alpha\beta}_{i-\frac{1}{2}}\right]+\frac{1}{2}\left(q^{\alpha\beta}_{i-\frac{1}{2}}\frac{\phi_i^\beta-\phi_{i-1}^\beta}{\Delta x}+q^{\alpha\beta}_{i+\frac{1}{2}}\frac{\phi_{i+1}^\beta-\phi_i^\beta}{\Delta x}\right)\right\}\\
=&-2\frac{\epsilon}{\Delta x}\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\left[q^{\alpha\beta}_{i+\frac{1}{2}}\left(\phi^\beta_{i+\frac{1}{2}}+\frac{1}{2}(\phi^\beta_{i+1}-\phi^\beta_i)\right)-q^{\alpha\beta}_{i-\frac{1}{2}}\left(\phi^\beta_{i-\frac{1}{2}}-\frac{1}{2}(\phi^\beta_i-\phi^\beta_{i-1})\right)\right]\\
=&-2\frac{\epsilon}{\Delta x}\sum_{\beta\neq\alpha}\gamma^{\alpha\beta}\left[\phi^\beta_{i+1}q^{\alpha\beta}_{i+\frac{1}{2}}-\phi^\beta_{i-1}q^{\alpha\beta}_{i-\frac{1}{2}}\right].
\end{aligned}$$

Differentiating the total gradient energy contribution in Equation (6.83) with respect to a particular ϕ-value ϕ_i^α, it is easy to see that it suffices to verify that for each phase-pairing individually, one has

$$-\frac{\mathrm{d}}{\mathrm{d}\phi_i^\alpha}\sum_j\frac{1}{2}\left(|q^{\alpha\beta}_{j-\frac{1}{2}}|^2+|q^{\alpha\beta}_{j+\frac{1}{2}}|^2\right)=-2\frac{\epsilon}{\Delta x}\phi^\beta_{i+1}q^{\alpha\beta}_{i+\frac{1}{2}}-\phi^\beta_{i-1}q^{\alpha\beta}_{i-\frac{1}{2}}. \tag{6.85}$$

Since ϕ_i^α is only involved in two terms in this summation, namely the ones for $j=i$, it follows that this is the same as

$$-\frac{\mathrm{d}}{\mathrm{d}\phi_i^\alpha}\frac{1}{2}\left(|q^{\alpha\beta}_{i-\frac{1}{2}}|^2+|q^{\alpha\beta}_{i+\frac{1}{2}}|^2\right)=-2\Big(q^{\alpha\beta}_{i-\frac{1}{2}}\frac{\mathrm{d}q^{\alpha\beta}_{i-\frac{1}{2}}}{\mathrm{d}\phi_i^\alpha}+q^{\alpha\beta}_{i+\frac{1}{2}}\frac{\mathrm{d}q^{\alpha\beta}_{i+\frac{1}{2}}}{\mathrm{d}\phi_i^\alpha}\Big),$$

and thus together with $\frac{\mathrm{d}q^{\alpha\beta}_{i+\frac{1}{2}}}{\mathrm{d}\phi_i^\alpha}=\frac{\phi^\beta_{i+1}}{\Delta x}$ and $\frac{\mathrm{d}q^{\alpha\beta}_{i-\frac{1}{2}}}{\mathrm{d}\phi_i^\alpha}=\frac{\mathrm{d}}{\mathrm{d}\phi_i^\alpha}\frac{\phi^\alpha_{i-1}\phi^\beta_i-\phi^\beta_{i-1}\phi^\alpha_i}{\Delta x}=-\frac{\phi^\beta_{i-1}}{\Delta x}$ based on the simplified expression in Equation (6.84) that Equation (6.85) is indeed satisfied.

In contrast, a similar property does not hold when using a discretization of the gradient energy distributions using equations (6.80) and (6.82). In fact, the existence of an underlying potential would, by Schwarz's theorem, require the symmetry of the second derivatives with respect to all unknowns. Even though it is clear that both contributions are symmetric with respect to the phase-indices $\alpha \in \{1, 2, ..., N\}$, this is not the case with respect to the $\boldsymbol{\phi}$-values at different spatial positions. Combining equations (6.80) and (6.82), the total gradient energy contribution in a cell i is given by

$$-\tilde{g}_i^\alpha = -2\epsilon \sum_{\beta \neq \alpha} \gamma^{\alpha\beta} \left\{ \frac{1}{\Delta x} \left[\phi^\beta_{i+\frac{1}{2}} q^{\alpha\beta}_{i+\frac{1}{2}} - \phi^\beta_{i-\frac{1}{2}} q^{\alpha\beta}_{i-\frac{1}{2}} \right] + \left(\phi_i^\alpha \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} - \phi_i^\beta \frac{\phi^\alpha_{i+1} - \phi^\alpha_{i-1}}{2\Delta x} \right) \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} \right\}.$$

In order to be compatible with the symmetry of the second derivatives, one would therefore have to satisfy $\frac{\partial \tilde{g}_i^\alpha}{\partial \phi_j^\alpha} = \frac{\partial \tilde{g}_j^\alpha}{\partial \phi_i^\alpha}$ for any combination of i and j, or, restricting the attention to two neighboring cells $\frac{\partial \tilde{g}_i^\alpha}{\partial \phi_{i+1}^\alpha} = \frac{\partial \tilde{g}_{i+1}^\alpha}{\partial \phi_i^\alpha}$. By the symmetry with respect to the phases, it is furthermore again clear that it suffices to check this for the contribution of a single phase-pairing, and is thus a matter of comparing

$$\frac{\partial}{\partial \phi^\alpha_{i+1}} \left\{ \frac{1}{\Delta x} \left[\phi^\beta_{i+\frac{1}{2}} q^{\alpha\beta}_{i+\frac{1}{2}} - \phi^\beta_{i-\frac{1}{2}} q^{\alpha\beta}_{i-\frac{1}{2}} \right] + \left(\phi_i^\alpha \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} - \phi_i^\beta \frac{\phi^\alpha_{i+1} - \phi^\alpha_{i-1}}{2\Delta x} \right) \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} \right\}$$

$$= \frac{1}{\Delta x} \phi^\beta_{i+\frac{1}{2}} \frac{\partial q^{\alpha\beta}_{i+\frac{1}{2}}}{\partial \phi^\alpha_{i+1}} - \phi_i^\beta \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{2\Delta x} \frac{\partial}{\partial \phi^\alpha_{i+1}} \frac{\phi^\alpha_{i+1} - \phi^\alpha_{i-1}}{2\Delta x} = \frac{1}{\Delta x} \phi^\beta_{i+\frac{1}{2}} \frac{\partial q^{\alpha\beta}_{i+\frac{1}{2}}}{\partial \phi^\alpha_{i+1}} - \phi_i^\beta \frac{\phi^\beta_{i+1} - \phi^\beta_{i-1}}{4(\Delta x)^2}$$

and

$$\frac{\partial}{\partial \phi^\alpha_{i}} \left\{ \frac{1}{\Delta x} \left[\phi^\beta_{i+\frac{3}{2}} q^{\alpha\beta}_{i+\frac{3}{2}} - \phi^\beta_{i+\frac{1}{2}} q^{\alpha\beta}_{i+\frac{1}{2}} \right] + \left(\phi_{i+1}^\alpha \frac{\phi^\beta_{i+2} - \phi^\beta_{i}}{2\Delta x} - \phi_{i+1}^\beta \frac{\phi^\alpha_{i+2} - \phi^\alpha_{i}}{2\Delta x} \right) \frac{\phi^\beta_{i+2} - \phi^\beta_{i}}{2\Delta x} \right\}$$

$$= -\frac{1}{\Delta x} \phi^\beta_{i+\frac{1}{2}} \frac{\partial q^{\alpha\beta}_{i+\frac{1}{2}}}{\partial \phi^\alpha_{i}} - \phi_{i+1}^\beta \frac{\phi^\beta_{i+2} - \phi^\beta_{i}}{2\Delta x} \frac{\partial}{\partial \phi^\alpha_{i}} \frac{\phi^\alpha_{i+2} - \phi^\alpha_{i}}{2\Delta x} = -\frac{1}{\Delta x} \phi^\beta_{i+\frac{1}{2}} \frac{\partial q^{\alpha\beta}_{i+\frac{1}{2}}}{\partial \phi^\alpha_{i}} + \phi_{i+1}^\beta \frac{\phi^\beta_{i+2} - \phi^\beta_{i}}{4(\Delta x)^2}.$$

Since $q^{\alpha\beta}_{i+\frac{1}{2}}$ depends neither on ϕ^β_{i-2} nor on ϕ^β_{i+2}, it is obvious that this equality cannot hold in general, and that there is thus no potential underlying this discretization.

Remark 76. This lack of variational interpretation in the discrete case is not necessarily an issue and does not automatically favor the other discretization. Instead, this choice (as well as that of any other discretization), also depends heavily on the interplay of various factors.
On the one hand, both discretizations are perfectly legitimate and differ only by a second-order discretization error. The results obtained by the two discretizations are therefore not expected to change beyond the level of the other errors inherent in the discretization anyway[51].
One the other hand, it does not have any real effect on the characterization of the "multipliers" in Equation (6.64) (resp. its simpler well-version) or their alternative introduction through a projection-based algorithm below. In fact, even though these are then not strictly speaking Lagrange multipliers in the sense of an underlying Lagrangian[52], they are still induced in the same form by the constraints due to the more general geometrical characterization through orthogonality and normal cone conditions for the admissible variations discussed in Sections 4 and 4.3.
It therefore often makes sense to prefer one or the other based on more practical considerations.

[51] This can also be verified numerically, with both solutions differing only very slightly.

[52] Recall that the two primary motivations for the Lagrangians are the formal simplicity with which they allow to derive necessary conditions for constrained variational problems as well as any potential benefits obtained through a saddle-point structure. Neither of these is used when arguing based on Equation (6.59).

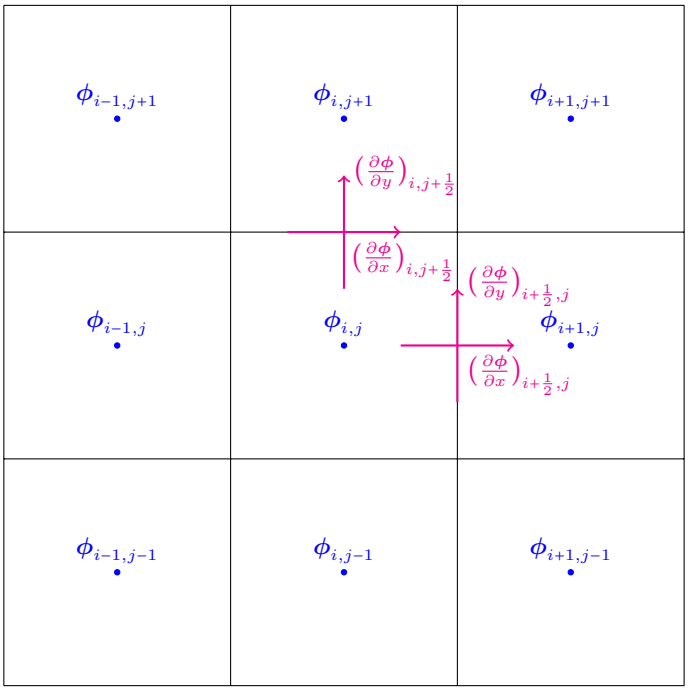


Figure 6.6: Layout of the ϕ-values and required gradients on the upper cell faces for the evaluation of the gradient energy contribution in the cell (i, j).

In particular, when using an explicit time-discretization which is typically quite "forgiving" to moderate perturbations in the equations, a much more relevant criterion is often the difference in run-time obtained for both discretizations due to the usually very high number of time-steps inherited by the stability restrictions. In contrast, other more complex schemes can often be relatively sensible to even small perturbations in the equations. For example, the common usage of line-search or trust-region algorithms in constrained optimization problems depends fundamentally on the existence of a potential, and the use of an alternative merit-function which is not actually being minimized can thus be an issue. Finally, there is also the very "mundane" issue of the effort and amount of code required for the implementation of the equations. ◇

In the isotropic case considered above, both discretizations are easily extended to the higher-dimensional setting by "splitting" the contributions of $\boldsymbol{q}^{\alpha\beta}$ onto the faces in the respective spatial direction, i.e. in two dimensions by associating the contributions from the x-component of $\boldsymbol{q}^{\alpha\beta}$ on the left and right cell faces, and those due to the y-component on the bottom and top faces. With some further modifications, they can also be extended to an anisotropic settting. In this respect, the primary difficulty faced by a cell-centered discretization in combination with fluxes discretized on the faces is that these in general then depend on the whole vector $\boldsymbol{q}^{\alpha\beta}$ resp. its orientation $\frac{\boldsymbol{q}^{\alpha\beta}}{|\boldsymbol{q}^{\alpha\beta}|}$. Whereas the layout is very convenient for calculating e.g. the x-component of $\nabla\phi$ at the face-centers in the x-direction, it is much less so for calculating the y-component at these same points as illustrated in Figure 6.6. While $\left(\frac{\partial\phi}{\partial x}\right)_{i+\frac{1}{2},j}$ and $\left(\frac{\partial\phi}{\partial y}\right)_{i,j+\frac{1}{2}}$ can as before be evaluated using a short difference as $\frac{\phi_{i+1,j}-\phi_{i,j}}{\Delta x}$ and $\frac{\phi_{i,j+1}-\phi_{i,j}}{\Delta y}$, the evaluation of the respective other component is most naturally done using a broader "averaged" stencil as $\left(\frac{\partial\phi}{\partial y}\right)_{i+\frac{1}{2},j} \approx \frac{(\phi_{i,j+1}+\phi_{i+1,j+1})-(\phi_{i,j-1}+\phi_{i+1,j-1})}{4\Delta y}$ and $\left(\frac{\partial\phi}{\partial x}\right)_{i,j+\frac{1}{2}} \approx \frac{(\phi_{i+1,j}+\phi_{i+1,j+1})-(\phi_{i-1,j}+\phi_{i-1,j+1})}{4\Delta x}$.

Based on this stencil and the same simple averaging of the ϕ-values as before for approximat-

ing $\boldsymbol{\phi}_{i+\frac{1}{2},j}$ and $\boldsymbol{\phi}_{i,j+\frac{1}{2}}$, one can then recover the full $\boldsymbol{q}^{\alpha\beta}$-vector on each of the faces. Based on Equation (6.67), the two-dimensional and potentially anisotropic extensions of the divergence-contribution in Equation (6.80) is thus given by

$$\begin{aligned}\epsilon\left(\nabla\cdot\left(\frac{\partial a}{\partial\left(\nabla\phi^{\alpha}\right)}\right)\right)_i \approx -2\epsilon\sum_{\beta\neq\alpha}\Bigg\{&\frac{1}{\Delta x}\left[\phi^{\beta}_{i+\frac{1}{2},j}\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i+\frac{1}{2},j}\right)-\phi^{\beta}_{i-\frac{1}{2},j}\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i-\frac{1}{2},j}\right)\right]\\&+\frac{1}{\Delta y}\left[\phi^{\beta}_{i,j+\frac{1}{2}}\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j+\frac{1}{2}}\right)-\phi^{\beta}_{i,j-\frac{1}{2}}\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j-\frac{1}{2}}\right)\right]\Bigg\}.\end{aligned}\tag{6.86}$$

Similarly, one can extend the additional contribution by $\frac{\partial a}{\partial\phi^{\alpha}}$ in (6.66) in a cell-centered fashion as in Equation (6.81), i.e. by calculating $\nabla\boldsymbol{\phi}_{i,j}$ through $\left(\frac{\partial\boldsymbol{\phi}}{\partial x}\right)_{i,j}\approx\frac{\boldsymbol{\phi}_{i+1,j}-\boldsymbol{\phi}_{i-1,j}}{2\Delta x}$ and $\left(\frac{\partial\boldsymbol{\phi}}{\partial y}\right)_{i,j}\approx\frac{\boldsymbol{\phi}_{i,j+1}-\boldsymbol{\phi}_{i,j-1}}{2\Delta x}$ and combining these with $\boldsymbol{\phi}_{i,j}$ to obtain $\boldsymbol{q}^{\alpha\beta}_{i,j}$, thus leading to the total contribution

$$\left(\epsilon\frac{\partial a}{\partial\phi^{\alpha}}\right)_{i,j}\approx 2\epsilon\sum_{\beta\neq\alpha}\left(\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j}\right)\right)\cdot\begin{pmatrix}\frac{\phi^{\beta}_{i+1,j}-\phi^{\beta}_{i-1,j}}{\Delta x}\\\frac{\phi^{\beta}_{i,j+1}-\phi^{\beta}_{i,j+1}}{\Delta x}\end{pmatrix}\tag{6.87}$$

corresponding to the default discretization in the **Pace3D**-framework.
Alternatively, one can also extend the more face-centered discretization in Equation (6.82) in a similar manner by instead using the already calculated approximations of $\nabla\boldsymbol{\phi}$ and $\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}$ on the cell phases and either setting

$$\begin{aligned}\left(\epsilon\frac{\partial a}{\partial\phi^{\alpha}}\right)_{i,j}\approx 2\epsilon\sum_{\beta\neq\alpha}\frac{1}{4}\Bigg[&\left(\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i-\frac{1}{2},j}\right)\right)\cdot\left(\nabla\phi^{\beta}\right)_{i-\frac{1}{2},j}+\left(\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i+\frac{1}{2},j}\right)\right)\cdot\left(\nabla\phi^{\beta}\right)_{i+\frac{1}{2},j}\\&+\left(\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j-\frac{1}{2}}\right)\right)\cdot\left(\nabla\phi^{\beta}\right)_{i,j-\frac{1}{2}}+\left(\frac{\partial A^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j+\frac{1}{2}}\right)\right)\cdot\left(\nabla\phi^{\beta}\right)_{i,j+\frac{1}{2}}\Bigg],\end{aligned}\tag{6.88}$$

or, in a somewhat cheaper manner which in addition reduces to the previous one if one is in fact isotropic, by setting

$$\begin{aligned}\left(\epsilon\frac{\partial a}{\partial\phi^{\alpha}}\right)_{i,j}\approx 2\epsilon\sum_{\beta\neq\alpha}\frac{1}{2}\Bigg[&\left(\frac{\partial A^{\alpha\beta}}{\partial q_x^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i-\frac{1}{2},j}\right)\right)\frac{\phi^{\beta}_{i,j}-\phi^{\beta}_{i-1,j}}{\Delta x}+\left(\frac{\partial A^{\alpha\beta}}{\partial q_x^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i+\frac{1}{2},j}\right)\right)\frac{\phi^{\beta}_{i+1,j}-\phi^{\beta}_{i,j}}{\Delta x}\\&+\left(\frac{\partial A^{\alpha\beta}}{\partial q_y^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j-\frac{1}{2}}\right)\right)\frac{\phi^{\beta}_{i,j}-\phi^{\beta}_{i,j-1}}{\Delta y}+\left(\frac{\partial A^{\alpha\beta}}{\partial q_y^{\alpha\beta}}\left(\boldsymbol{q}^{\alpha\beta}_{i,j+\frac{1}{2}}\right)\right)\frac{\phi^{\beta}_{i,j+1}-\phi^{\beta}_{i,j}}{\Delta y}\Bigg].\end{aligned}\tag{6.89}$$

Remark 77. Which discretization of $\frac{\partial a}{\partial\phi^{\alpha}}$ is more favorable again depends on a variety of considerations. The one in Equation (6.87) is a priori the simplest, but has the potential disadvantage of requiring an additional calculation of the $\frac{\partial\boldsymbol{A}^{\alpha\beta}}{\partial\boldsymbol{q}^{\alpha\beta}}$ in the cell-centers. In contrast, the discretizations in Equation (6.88) and (6.89) are again solely based on quantities already available on the faces. In addition, if combined with a buffering scheme, the contributions by each of the phases need only be calculated once as they are shared by the neighboring cells. This nevertheless has the disadvantage of introducing a spatial interdependence and some additional overhead through the buffering.
It should be stressed though that **none** of the formulations above is compatible with a discrete variational principle unless the contributions of the derivatives in the lateral direction in Equation (6.89) happen to drop out[53]. Even though there is no principle difficulty in first defining

[53]This is obviously the case for isotropic problems, but can also happen for some simple anisotropies such as e.g. an elliptic one described (in two dimensions) by

$$A^{\alpha\beta}(\boldsymbol{q}^{\alpha\beta})=c_x^2\left(q_x^{\alpha\beta}\right)^2+c_y^2\left(q_y^{\alpha\beta}\right)^2.$$

an "appropriate" discrete approximation of the energy $\mathcal{F}_\epsilon$ and then deriving the corresponding difference scheme through a differentiation with respect to the ϕ_i^α, this has a signifcant practical disadvantage as soon as this energy involves a gradient contribution calculated based on an averaged gradient, i.e. with a stencil spanning more than one grid spacing as this automatically entails a broader stencil.

This is most easily illustrated in an isotropic one-dimensional setting. Whereas the differentiation of an energy contribution of the form $\sum_j \left(\frac{\phi_{j+1}-\phi_j}{\Delta x}\right)^2$ with respect to ϕ_i leads to contributions by the two terms $j = i - 1$ and $j = i$ through $\frac{\phi_i-\phi_{i-1}}{\Delta x}\frac{1}{\Delta x}$ and $\frac{\phi_{i+1}-\phi_i}{\Delta x}\frac{-1}{\Delta x}$ and thus the standard second-order three-point stencil $-\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{(\Delta x)^2}$ of the negative Laplacian, an energy contribution of the form $\sum_j \frac{\phi_{j+1}-\phi_{j-1}}{2\Delta x}$ has the two contributions $\frac{\phi_i-\phi_{i-2}}{2\Delta x}\frac{1}{2\Delta x}$ and $\frac{\phi_{i+2}-\phi_i}{2\Delta x}\frac{-1}{2\Delta x}$ and thus instead leads to an undesirable five-point discretization $-\frac{\phi_{i+2}-2\phi_i+\phi_{i-2}}{(2\Delta x)^2}$ of the same operator.

For the same reason, an energetic formulation based, as in the isotropic case, on energetic contributions by the phases (but now dependent on the full $\boldsymbol{q}^{\alpha\beta}$-vectors there) would lead to the much broader stencil shown in the left of Figure 6.7, which in addition is not based on the very convenient structure of the discrete divergence operator as the difference of the fluxes through the four (resp. six in three space dimensions) cell faces. The other two more obvious choices are either the use of a purely cell-centered expression for the energy (leading to the slightly smaller stencil in the right of Figure 6.7), or the use of a "corner-centered" scheme based purely on short differences, which would allow to maintain the more narrow stencil in Figure 6.6. Both of the these latter choices are well-known to reduce to inherently unstable schemes in the isotropic case though, and are therefore likely not to be recommended.

Besides the higher computational complexity generally associated with broader stencils, they raise, in combination with the obstacle potential, an additional difficulty related with the discussion in Subsection 6.2.3. As already discussed there, by the discontinuity of the second derivative of the basic one-dimensional phasefield profile one expects a $\mathcal{O}(1)$ error in the discretization of the second derivative for any stencil crossing the transition from the interface to the bulk. While this was shown not to reduce the second-order convergence in the one-dimensional case, the degree to which this effect enters the calculation of the profile was nevertheless also seen to have a notable impact on the numerical precision. Similar effects will also arise in a higher-dimensional setting and will, with a broader stencil, affect more points depending on their relative positioning with respect to this transition[54].

Unless an energy-based discretization is necessary due to e.g. the use of a particular algorithm, a discretization of the gradient energy contributions as in Equation (6.86) and either Equation (6.87) or Equation (6.89) therefore seems to be the preferable choice.

◇

[54] Recall that the numerically most benign situation in Section 6.2.3 was the one where the stencil of the last "mobile" point with $0 < \phi_i < 1$ is essentially completely within the expected interface region, whereas the - quite high - error at the first bulk-point is only seen indirectly in the error in the multipliers for the constraint. For a five-point stencil in a given direction, the one for the last inner point will always have to stretch across the transition into the bulk by roughly one grid spacing, and is therefore likely to be associated with significant numerical errors.

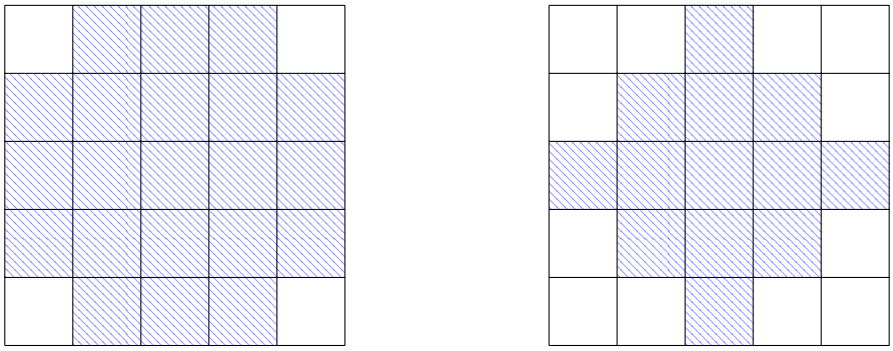


Figure 6.7: Difference stencils for a face-based (left) and cell-based (right) variational difference scheme for anisotropic gradient energy densities.

Dealing with Many Phases and the LROP Approach

In contrast to the mostly superficial difficulties from Subsection 6.3.3 (wich are relatively easy to handle with in both a reduced and non-reduced fashion), a very crucial difficulty in the computational treatment of multiphase as compared to two-phase problems lies in the a priori very high computational and memory requirements associated with the large number of phases which are in many cases necessary for the simulation of realistic microstructures (with represenative samples often consisting of thousands of grains).
Firstly, it is clear that, without further modifications, the memory requirements will increase linearly with the number N of phasefields required for the description of a given problem. Secondly, not only are there then also N equations to be solved (e.g. Equation (6.73) or Equation (6.75)), but they are in addition based on the calculation of two-phase interactions for a and even triple-phase interactions in the bulk potentials (see e.g. Equation (6.7) or (6.8)), and thus a priori lead to a cubic dependence of the calculation time on N.

A by now widely adapted manner for maintaining a roughly constant storage space and computational cost despite a large total number of phases is, at least in combination with an obstacle-type potential, the use of a **locally reduced order parameter (LROP)** approach ([40]). The underlying idea is that, based on stability considerations for multiphase-regions, one does not expect for regions with more than a few phases to be stable. If not initially present, such regions are therefore also not expected to arise naturally within the simulation, and one usually only has to deal with a very moderate maximal number l of phases which are acutally present at any given point.
This is extremely useful from a compuational point of view as it motivates replacing the storage of all phasefields with a scheme where one instead, for each computational cell, only stores the indices and the values of those phases with non-vanishing ϕ-values. While this a priori only allows reducing the memory requirements from $\mathcal{O}(N)$ to $\mathcal{O}(1)$ with respect to the number of phases, this storage scheme in addition also enables the reduction of the calculation time to a roughly constant one, regardless of the number of phases.
This is essentially due to the following two observations. On the one hand, the gradient energy contributions by construction vanish for all those phase-pairings for which at least one of the phases is locally constant at 0, as are the contributions to the equations of the other phases for all those α with $\phi^\alpha = 0$. Even though these phases themselves do experience a contribution by the other phases, the obstacle potential is designed to provide a strong counter-force to deviations of the local $\boldsymbol{\phi}$-values different from 0[55]. This in combination with the general dominance of the

[55] As well as from 1, but this is only of interest for at most one phase.

a- and w-terms in the phasefield functional is almost always (see Remark 79) sufficient to ensure that a locally "non-existent" phase which does not interact with its neighborhood through the non-local a-term will have to remain at 0. In addition, the multipliers due to the sum-constraint and the positivity constraint are such that the other phases remain unaffect by simply a priori skipping over the calculation of the new values of these phases as well as their contributions to the reminaining equations. This can be made use of in terms of a classification scheme which identifies the phases actually present within a small neighborhood (depending upon the stencil) of a given cell and completely skips all calculations for the other ones. This classification is highly efficient in combination with the storage scheme above, as this simply amounts to running through the stored $\boldsymbol{\phi}$-indices of a fixed number of neighboring cells for checking which phases need to be considered, and thus is an operation of runtime $\mathcal{O}(1)$ with respect to the phases, with a reduction of the cost of the actual computations to the same order.

Remark 78. This preclassification can - and is, at least in the sense of which cells require any calculation at all, often worth the effort even in the two-phase case - in principle also be done even if all phases are stored. While this does allow a reduction of the actual calculation time to the same level as with the LROP-based scheme above, there are still two major drawbacks as compared to the storage scheme above once the number of phases grows even moderately large. Firstly, determining which phases should actually be considered as being present is still an operation of complexity $\mathcal{O}(N)$ as one is forced to run through all possible ones. Secondly, even if this classification does not require any real calculations (essentially reducing to a comparison of $\boldsymbol{\phi}$-values with 0), it can still incur a significant cost in run-time due to both a large number of conditional statements and the comparatively slow access to main memory[56]. ◇

While the storage could in principle also be done in an adaptive fashion with the scheme adjusting according to the local requirements, having an a priori estimate of the maximal number of phases one expects to locally coexist - either based on l as above or simply by experience for a particular type of setting - allows both a simplification in the implementation and an increase in the efficiency by using a less dynamic memory layout designed to provide sufficient space for locally storing a fixed number l' of phases only. In combination with the preclassification outlined above, this leads to a very performant implementation capable of handling an essentially arbitrary total number of phases with a roughly constant computational cost (see [40] and [67] for a more detailed discussion on various aspects of this approach).

Remark 79. It is clear that the approach above is not rigorous in the sense that one is not guaranteed to actually remain below any local prefixed number $l' < N$ of phases at all times. Even in settings specifically designed to lead to an initial violation of this bound, simulations with e.g. $l' = 6$ or $l' = 8$ typically lead to practically identical simulation results after a potential (but very short) transient which quickly forces an elimination of the "excess" number of phases. One example of such a setting is starting an N-phase simulation with a simple sharp interface between two given phases. This initial jump discontinuity in the phases leads to a discrete $\delta'(x)$-type contribution and thus a term of order $\mathcal{O}\big(\frac{1}{(\Delta x)^2}\big)$ to the cells neighboring the interfaces due to the second-order derivative in the divergence-term $-\nabla\cdot\big(\frac{\partial a}{\partial\nabla\phi}\big)$. Unless ϵ is chosen unrealistically small, this term will initially dominate the contribution by the bulk-potential and will, at least for an explicit time-discretization, enforce the appearce of **all** N phases trough the Lagrange-multiplier Λ in Equation (6.73) resp. more indirectly but based on the same effect through the mobility matrix $\boldsymbol{M}$ in Equation (6.75).
This can e.g. be seen by considering a very simple isotropic one-dimensional setting for which there is a sharp transition between a bulk-phase for $\alpha = 1$ up to the cell i to a bulk-phase for $\alpha = 2$ starting at $i+1$, where, for simplicity of notation, it will additionally be assumed that all $\gamma^{\alpha\beta}$ are equal to the common value γ.

[56] This depends of course also very significantly upon the details of the particular implementation and simulation setup (i.e. for example favorable cache effects). Regardless of this, it is an issue which is strongly mitigated using the LROP approach.

Together with the initially sharp transition, the evaluation of the simplified expression for $q^{\alpha\beta}$ in Equation (6.84) leads to the only non-zero terms begin $q^{12}_{i+\frac{1}{2}} = -q^{21}_{i+\frac{1}{2}} = \frac{1}{\Delta x}$, with $q^{\alpha\beta}_{i+\frac{1}{2}} = 0$ otherwise, and $q^{\alpha\beta}_{j+\frac{1}{2}} = 0$ regardless of the phases for all $j \neq i$. For the divergence-term $\epsilon\nabla\cdot\Big(\frac{\partial a}{\partial\nabla\phi^\alpha}\Big) = -2\gamma\epsilon\sum_{\beta\neq\alpha} q^{\alpha\beta}_{i+\frac{1}{2}}$ discretized as in Equation (6.80), the only non-zero contributions arising through the divergence-part in the cell i are given by $-\frac{\gamma\epsilon}{(\Delta x)^2}$ for the phase $\alpha = 1$ and $\frac{\gamma\epsilon}{(\Delta x)^2}$ for the phase $\alpha = 2$ since both $\phi^1_{i+\frac{1}{2}}$ and $\phi^2_{i+\frac{1}{2}}$ equal $\frac{1}{2}$. From Equation (6.69), it is also obvious that one has $\frac{\partial w}{\partial\phi^\alpha}(\boldsymbol{\phi}_i) = 0$ if $\alpha = 0$ and $\frac{\partial w}{\partial\phi^\alpha}(\boldsymbol{\phi}_i) = \frac{16\gamma}{\pi^2}$ for $\alpha \neq 1$. The only remaining relevant term[57] is given by $\epsilon\frac{\partial a}{\partial\phi^\alpha} = 2\gamma\epsilon\sum_{\beta\neq\alpha} q^{\alpha\beta}\frac{\mathrm{d}\phi^\beta}{\mathrm{d}x}$. Discretizing this expression for example as in Equation (6.82), the sum on the lower face in cell i is 0 and the upper one equals $\frac{2\gamma\epsilon}{(\Delta x)^2}$ if $\alpha = 1$ or $\alpha = 2$ and 0 otherwise. Averaging over both faces and combining the result with the remaining expressions above, it follows that one has

$$\epsilon\Big(\nabla\cdot\frac{\partial a}{\partial\nabla\phi^\alpha} - \frac{\partial a}{\partial\phi^\alpha}\Big) - \frac{1}{\epsilon}\frac{\partial w}{\partial\phi^\alpha} \hat{=} \begin{cases} \Big(-\frac{\gamma\epsilon}{(\Delta x)^2} - \frac{\gamma\epsilon}{(\Delta x)^2}\Big) - 0 = -\frac{2\gamma\epsilon}{(\Delta x)^2} & ,\alpha = 1, \\ \Big(\frac{\gamma\epsilon}{(\Delta x)^2} - \frac{\gamma\epsilon}{(\Delta x)^2}\Big) - \frac{16\gamma}{\pi^2\epsilon} = -\frac{16\gamma}{\pi^2\epsilon} & ,\alpha = 2, \\ -\frac{16\gamma}{\pi^2\epsilon} & \text{else} \end{cases}$$

and thus, if none of the μ^α differs from zero (as will be seen to be the case below)

$$\Lambda = -\frac{2\gamma\epsilon}{N(\Delta x)^2} - \frac{N-1}{N}\frac{16\gamma}{\pi^2\epsilon}.$$

In total, the right-hand sides in Equation (6.73) are therefore given by

$$-g^\alpha - \Lambda = \begin{cases} \frac{N-1}{N}\Big(-\frac{2\gamma\epsilon}{(\Delta x)^2} + \frac{16\gamma}{\pi^2\epsilon}\Big) & ,\alpha = 1, \\ \frac{1}{N}\Big(\frac{2\gamma\epsilon}{(\Delta x)^2} - \frac{16\gamma}{\pi^2\epsilon}\Big) & ,\alpha \neq 1, \end{cases}$$

and are strictly positive and the same for all phases with $\phi^\alpha = 0$, leading to the appearence of all phases. ◇

Remark 80. Even though the example in the previous remark shows that there are certain risks associated with a priori assuming a result instead of performing the actual calculation, its intent is not at all to argue against the use of such an approach. Firstly, it is from a computational point of view essentially the only way one can reasonably perform simulations involving more than a few phases. As such, the use of a (somewhat more conservative) LROP-type scheme may even be justified when using a well-potential, despite the fact that one expects all phases to spread over the entire domain and thus the introduction of a persistent error due to the necessity of forcing phases with $\phi^\alpha \approx 0$ to zero. With respect to the other sources of error enforced by computational constraints - both numerical and in particular through the use of an often artificially large interface - this cut-off may in fact still lead to better results by allowing the use of higher resolutions and thus smaller interface widths.
Secondly, despite the algorithm delivering "wrong" results during one or a few time-steps, the unexpected appearance of all phases is essentially an unphysical artefact due to an interplay of the difficulties associated with modeling multiphase interactions combined with a very irregular initial setting and is therefore - even though strictly speaking incorrect - preferable from a physical point of view.
Nevertheless, one has to be very careful when skipping calculations or a priori excluding phases, as this can, depending on the setting, also either hide less well-known difficulties of the model or even lead to critical failures of the calculations[58]. ◇

[57] Unless the driving force is very strong, it plays no role in the argument here.

[58] Such failures are normally not due to a preclassification and the use of a restricted storage scheme alone, but its combintation with a second not fully justified "optimization".

Projection-Based Alogrithms

A popular approach for the time-discretization of dynamic problems subject to additional constraints is the use of projection-based schemes[59]. In these, one first obtains a prediction such as it would result from the equation obtained by partially or completely "ignoring" the constraint. Since this prediction need not be compatible with the constraints, it is then, in a second step, projected back onto the admissible set.
The basic idea of this two-step procedure is to separate the total problem into a sequence of two a priori simpler problems. The first one (i.e. obtaining the prediction) is then dependent upon the equation itself, whereas the second projection step is primarily tied to the structure of the constraint set and can therefore be performed in a relatively "generic" fashion without knowing the details of the various contributions arising in the equation itself.
While this cursory description is for the most part true in the sense that the projection operation does not have to take the precise nature of any fixed contribution to the equation (such as e.g. the actual expressions for the $g^\beta = \frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}$ if discretized in an explicit manner) into account, the combined updates must remain compatible with the underlying first-order necessary conditions characterizing the minimizer and thus, in the phasefield case, the KKT-condition in Equation (6.63). This implies that the projection must be constructed in a manner compatible with the way the prescribed forcings affect the prediction in the absence of the constraints. For this reason, while essentially independent of the "energy" underlying the gradients, it does depend crucially upon the chosen dynamics for the evolution.

This somewhat abstract difficulty as wells as some related issues are best understood by considering an example. An in practice very relevant one in the phasefield context is given by an explicit time-discretization of the differenct choices of dynamics considered in Subsection 6.3.2. For the two choices of either a scalar proportionality of the dynamics as in Equation (6.72) or the mobility-matrix based one in Equation (6.77), an "explicit" discretization leads to

$$\tau^{(n)}\epsilon\frac{\boldsymbol{\phi}^{(n+1)}-\boldsymbol{\phi}^{(n)}}{\Delta t} = \boldsymbol{r}^{(n)} - \Lambda^{(n')}\boldsymbol{e} + \boldsymbol{\mu}^{(n')} \tag{6.90}$$

resp.

$$\epsilon\frac{\boldsymbol{\phi}^{(n+1)}-\boldsymbol{\phi}^{(n)}}{\Delta t} = \boldsymbol{M}\left(\boldsymbol{r}^{(n)} + \boldsymbol{\mu}^{(n')}\right) \tag{6.91}$$

where $\tau^{(n)} := \tau\left(\boldsymbol{\phi}^{(n)}\right)$ and $\boldsymbol{r}^{(n)}$ denotes an appropriate discrete version of the negative gradient contributions from $\frac{\mathrm{d}\mathcal{F}_\epsilon}{\mathrm{d}\boldsymbol{\phi}}$ in Equation (6.74). The n' for the multipliers is used to indicate some additional freedom in the choice and interpreation of their corresponding "time-step".

One of the most popular algorithms for a constrained gradient-descent is to actually not consider (6.90) and (6.91) and the associated question of the choice of multipliers directly. Instead, the **projected gradient descent** algorithm corresponds to the the splitting approach above in its purest form, where one first generates a prediction $\hat{\boldsymbol{\phi}}^{(n+1)}$ by completely ignoring the constraints through $\hat{\boldsymbol{\phi}}^{(n+1)} = \boldsymbol{\phi}^{(n)} + \frac{\Delta t}{\tau^{(n)}\epsilon}\boldsymbol{r}^{(n)}$ resp. $\hat{\boldsymbol{\phi}}^{(n+1)} = \boldsymbol{\phi}^{(n)} + \Delta t\boldsymbol{M}\boldsymbol{r}^{(n)}$, and then projects this prediction back onto the admissible set. In order to be compatible with Equation (6.90), a direct comparison shows that the change $\delta\boldsymbol{\phi} := \boldsymbol{\phi}^{(n+1)} - \hat{\boldsymbol{\phi}}^{(n+1)}$ induced by the projection operation has to satisfy

$$\boldsymbol{\phi}^{(n+1)} = \boldsymbol{\phi}^{(n+1)} - \hat{\boldsymbol{\phi}}^{(n+1)} + \hat{\boldsymbol{\phi}}^{(n+1)} = \delta\boldsymbol{\phi} + \boldsymbol{\phi}^{(n)} + \frac{\Delta t}{\tau^{(n)}\epsilon}\boldsymbol{r}^{(n)} = \boldsymbol{\phi}^{(n)} + \frac{\Delta t}{\tau^{(n)}\epsilon}\left(\boldsymbol{r}^{(n)} - \Lambda^{(n')}\boldsymbol{e} + \boldsymbol{\mu}^{(n')}\right)$$

[59] Two particularly well-known examples for this approach are provided by the pressure-projection schemes in fluid-dynamical simulations and the popular projection-based approach in elasto-plastic simulations (where it is usually only used as a substep of a global more complex algorithm).

and thus $\delta\boldsymbol{\phi} = \frac{\Delta t}{\tau^{(n)}_\epsilon}\big(- \Lambda^{(n')}\boldsymbol{e} + \boldsymbol{\mu}^{(n')}\big)$, whereas the analogous calculation for Equation (6.91) shows that $\delta\boldsymbol{\phi}$ has to be of the form $\delta\boldsymbol{\phi} = \boldsymbol{M}\boldsymbol{\mu}^{(n')}$.
This illustrates two points. On the one hand, the difference in structure of the increments due to the projection clearly indicates that the appropriate projection operation in both cases is different, and thus the dependence on the dynamics already discussed above. The appropriate projector operator for the (scalar) kinetic coefficient is the one based on the Euclidian norm (or some multiple thereof), since[60]

$$\boldsymbol{\phi}^{(n+1)} = \min_{\boldsymbol{\phi}\in\mathcal{GS}^N} \frac{1}{2}\big|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}^{(n+1)}\big|^2 \quad \Leftrightarrow \quad \begin{cases} \boldsymbol{\phi}^{(n+1)} = \hat{\boldsymbol{\phi}}^{(n+1)} - \bar{\Lambda}\boldsymbol{e} + \bar{\boldsymbol{\mu}}, \\ \bar{\boldsymbol{\mu}} \geq 0, \bar{\mu}^\alpha \phi^\alpha = 0, \\ \boldsymbol{\phi} \in \mathcal{GS}^N, \end{cases}$$

and it thus leads to the desired structure of $\delta\boldsymbol{\phi}$. In contrast, when the dynamics are based on the mobility-matrix $\boldsymbol{M}$, the projection should be chosen as the one with respect to the semi-norm induced by the pseudo-inverse $\boldsymbol{M}^\dagger$ of $\boldsymbol{M}$, as, if $\hat{\boldsymbol{\phi}}^{(n+1)}$ has a zero average, this implies that $\boldsymbol{\phi}^{(n+1)} = \hat{\boldsymbol{\phi}}^{(n+1)} + \boldsymbol{M}\bar{\boldsymbol{\mu}}$ and thus and increment $\delta\boldsymbol{\phi}$ of the required form[61].
On the other hand, it also explains one of the main reasons for the high popularity of the projected gradient descent algorithm, namely its particularly simple formal structure. Whereas a direct discretization of the dynamics in equations (6.72) or (6.77) as for example in equations (6.90) and (6.91) requires some considerations with regards to the multipliers, these can, in the explicit case, be completely disregarded and will be generated implicitly through the projection operator.

An alternative approach - consistent with a continuous interpretation if starting from a point satisfying the constraints - would be to instead choose Λ and $\boldsymbol{\mu}$ such that the evolution is restricted to the tangent cone of the admissible set at the current point, and thus, due to the scalar proportionality on the left-hand sides of equations (6.90) resp. (6.91), a projection of the right-hand sides onto this tangent space instead of the projection of the result onto the admissible set itself. This tangent space is given by all "directions" $\boldsymbol{d}$ such that $\sum_\alpha d^\alpha = 0$ for satisfying the sum-constraint and $d^\alpha \geq 0$ if $\phi^\alpha = 0$ and arbitrary otherwise in order not to move a phase at 0 below zero.
Projecting $\boldsymbol{r}$ with respect to the Euclidian norm, i.e. minimizing $\frac{1}{2}\big|\boldsymbol{d} - \boldsymbol{r}\big|^2$ over this admissible set, the resulting direction can similar to above be written as $\boldsymbol{d} = \boldsymbol{r} - \bar{\Lambda}\boldsymbol{e} + \bar{\boldsymbol{\mu}}$. Here $\bar{\boldsymbol{\mu}}$ is subject to two "complementarity" conditions, namely $\bar{\mu}^\alpha = 0$ if $\phi^\alpha > 0$ - this is known in advance and is thus a purely artificial one for being able to use an (unnecessary) full vector of multipliers for simplifying the notation - and $\bar{\mu}^\alpha d^\alpha = 0$ for all α with $\phi^\alpha = 0$. Similarly, a projection of $\boldsymbol{M}\boldsymbol{r}$ with the respect to the semi-norm induced by $\boldsymbol{M}^\dagger$ leads to the analogous conclusion with $\boldsymbol{d} = \boldsymbol{M}(\boldsymbol{r} + \bar{\boldsymbol{\mu}})$ with $\bar{\boldsymbol{\mu}}$ subject to the same conditions.

Remark 81. This is in a sense the "most explicit" discretization possible as it amounts to a projection of the given right-hand sides $\boldsymbol{r}$ and is thus a priori completely independent of $\boldsymbol{\phi}^{(n+1)}$, and the multipliers will therefore be donoted with $n' = n$. Nevertheless, it has the disadvantage that, even though the projection of $\boldsymbol{r}$ is a feasible direction, it may only be so for a very restricted time-step. More precisely, a negative total right-hand side for a phase α with $\phi^\alpha \approx 0$ can force this phase below zero in the next time-step, unless Δt is chosen smaller than the one defined by the **break-point** where $0 = \big(\phi^\alpha\big)^{(n)} + \frac{(\Delta t)^\alpha_b}{\tau^{(n)}_\epsilon}\big(\boldsymbol{r}^{(n)} - \Lambda^{(n)} + \boldsymbol{\mu}^{(n)}\big)$ and the scheme will therefore only lead to feasible values in the next time-step if $\Delta t \leq \min_\alpha(\Delta t)^\alpha_b$. Even though the evaluation of the right-hand side can thus be made completely independent of the next time-step, this simply

[60] Due to the the convexity of the norm, the first-order condition on the right being both necessary and sufficient.
[61] In contrast to the Euclidian case, this statement is a little more technical. As will be discussed in more detail in Section 6.3.3, this projection is still well-defined despite the use of the semi-norm and the lack of actual invertibility of $\boldsymbol{M}$ does in addition not cause any major practical difficulties.

amounts to shifting the responsibility for maintaining the feasibility to the choice of Δt.
This strategy is not by itself a bad idea and is in a similar form used as part of some optimization algorithms such as e.g. the reduced gradient method [58]. The major disadvantage here though is that the choice of Δt has global effects as it will have to be the same for all cells - and thus also has to be chosen based on the global minimum of the breakpoint of all phases in all cells - in order to be admissible for an actual time-discretization. If there is no restriction to map a physical evolution (i.e. in the current setting if every cell were allowed to use their own time-step), this can be a relatively convenient way of exploring a potential solution on the "faces" of the admissible set. ◇

The two choices above are clearly not the only ones possible. In particular, one can e.g. choose to explicitly maintain an estimate Λ in the prediction step within the gradient projection scheme, i.e. to set $\hat{\boldsymbol{\phi}}^{(n+1)} = \boldsymbol{\phi}^{(n)} + \frac{\Delta t}{\tau^{(n)}\epsilon}\left(\boldsymbol{r}^{(n)} - \Lambda\right)$ and then project the result back onto the Gibbs-simplex. As the sum-constraint is always "active", this is easily shown to have no effect on the final result, regardless how good or bad the choice of Λ.
Nevertheless, using e.g. the estimate $\Lambda = \frac{1}{N}\sum_\alpha r^\alpha$ corresponding to the sum-constraint alone can be useful in practice as it automatically covers the multiwell-case and can, in a somewhat modified form, also be helpful in combination with a preclassification scheme such as in the LROP approach. In particular, this Λ is always a lower bound for the (scaled) actual multiplier resulting from the projection step[62]. As will be discussed in more detail in the next section, together with the non-negativity of $\boldsymbol{\mu}$, this ensures that any $\boldsymbol{\phi}$-value satisfying $\phi^\alpha \leq 0$ **before** the projection operation will necessarily also satisfy $\phi^\alpha = 0$ **after** the projection.
The situtation for $\boldsymbol{\mu}$ is somewhat more difficult. In fact, as long as an estimate of $\boldsymbol{\mu}$ is compatible with the complementarity condition and such that, if included in the predictor step, it does not affect the inequality constraints which are ultimately active, the precise values are again irrelevant and will be "complemented" by the projection operation in a way which leads to the same final results for $\boldsymbol{\phi}^{(n+1)}$. It is intuitively clear that this is not at all a given, since one can e.g. choose some ϕ^α at zero and force it to an arbitrarily high value in the prediction step by simply using a sufficiently large estimate for μ^α.
An admissible choice - closely related to the preclassification schemes - is the use of Λ and $\boldsymbol{\mu}$ as they would result from the projection of the gradient itself discussed above. The obvious advantage of this choice is that all phases for which this is compatible with the complementarity between d^α and μ^α (and for those phases only!), the $\boldsymbol{\phi}$-values initially at 0 will remain there and thus need not be updated.

Remark 82. As Remark 79 shows though, the correctness of this approach depends crucially upon these complementarity conditions actually being satisfied, since this example leads to a prediction where all phases at 0 have a positive right-hand side. Interpreted in terms of a "prediction step", the (counter)example in this remark can be interpreted as a partial projection of the gradient with respect to all inequality constraints for the ϕ^α at zero except for the one which is classified as active due to its appearence in the neighboring cell. Due to the positivity of the right-hand side for all these phases in the presence of the sum-constraint, this projection would require **negative** μ^α's to ensure that they remain at zero, i.e. $d^\alpha = 0$. Checking whether complementarity actually holds would require maintaining the most expensive calculations due to the necessity of disposing of the right-hand sides r^α, and thus completely eliminates any advantages in terms of the computational cost.
Similar to the discussion in Section 6.3.3, it is always, in one form or another, the ability to perform this type of speculation in "good conscience" due to the favorable properties of the obstacle-potential which leads to phasefield codes being able to handle very large problem sizes - both in terms of the domain sizes and in particular with respect to the number of unknowns per "cell" in terms of N - as compared to many a priori simpler nonlinear programming problems. ◇

[62] A similar obersvation was already made for the steady-state case in Equation (6.65).

Remark 83. Note that the main feature making the projected gradient algorithm attractive in the phasefield case is that it is based on a relatively simple projection with a purely local admissible set and can thus essentially be evaluated on-the-fly as one runs through the domain to update the phasefield values. This in particular eliminates the necessity for storing the associated multipliers Λ and $\boldsymbol{\mu}$[63].
In contrast, a straightforward extension of this two-step procedure above to situations where the dynamics are modified through e.g. either the use of some semi-implicit or implicit time-stepping scheme or a "lightweight" preconditioning of the basic gradient descent scheme would not enjoy the same favorable features. In the simplest case, this leads to a scheme where a **non-local** s.p.d. matrix $\boldsymbol{A}^{-1}$ essentially replaces the action of $\boldsymbol{M}$ on the right-hand sides. Due to this, the appropriate projection is then a non-local one with respect to the norm induced by $\boldsymbol{A}$, which is on the one hand a siginificantly more difficult projection operation and on the other hand will generally necessitate actually storing the multipliers required "internally" by the projection.
Therefore, other algorithms are generally more appropriate in this case. One of the earlier popular choices is given by a nonlinear extension of the standard SOR-smoother for linear equations through a **projected SOR (pSOR)** algorithm (see [17] and e.g. [21] for a discussion in the phasefield context). The basic idea is the use of a point-relaxation process, which consists in considering each cell at a time and determining $\boldsymbol{\phi}^{(n+1)}$ as the projection of the value $\boldsymbol{\phi}^*$ which would allow to solve the equations within this cell (or, in the nonlinear case, an approximation thereof using e.g. a single Newton-step) if the values of the phasefield in all other cells are taken as fixed. This procedure can then be further accelerated by, instead of projecting $\boldsymbol{\phi}^*$ itself, projecting the overrelaxed value $(1-\omega)\boldsymbol{\phi}^{(n)} + \omega\boldsymbol{\phi}^*$. Even though this procedure does again require the use of a weighted projection instead of the Euclidian one for consistency with the multipliers, it maintains the main advantage of the strict locality of the projections.
More recently, various alternative approaches have been considered (see e.g. [60], [12], [32], [30], [31] and [33]) which avoid the main drawback of the pSOR algorithm, namely that the strict locality of this approach, while responsible for its simplicity, can also be the primary limiting factor for its convergence rate. ◇

Some Algorithmic Aspects of the Projection onto the Gibbs-Simplex

In order to complete the description of the projected gradient algorithm above, it still remains to clarify how the projection operation of the prediction $\hat{\boldsymbol{\phi}}$ can actually be performed.
This is straightforward when the only constraint is the sum-constraint $\sum_{\alpha=1}^N \phi^\alpha \stackrel{!}{=} 1$ as it suffices to subtract the average deviation from $\hat{\boldsymbol{\phi}}$, i.e. $\mathcal{P}_{\Sigma_1}\big(\hat{\boldsymbol{\phi}}\big) = \hat{\boldsymbol{\phi}} - \frac{1}{N}\sum_{\alpha=1}^N \hat{\phi}^\alpha$.
Alternatively, at least if one has $\sum_\alpha (\phi^\alpha)_i^{(n)} = 1$, the same result can also be obtained by directly subtracting the average of the "right-hand sides" in the simplified form of the update rule (6.90) without the additional factor $\boldsymbol{\mu}$, i.e. by setting

$$\tau^{(n)}\epsilon\frac{\boldsymbol{\phi}^{(n+1)} - \boldsymbol{\phi}^{(n)}}{\Delta t} = \boldsymbol{r}^{(n)} - \Big(\frac{1}{N}\Big(\sum_\alpha \big(r^\alpha\big)^{(n)}\Big) - 1\Big)\boldsymbol{e}$$

which then actually corresponds to the use of the projection of the gradient.

In contrast, when the predicted ϕ-values need to be projected back onto

$$\mathcal{GS} = \{\boldsymbol{\phi} : \sum_\alpha \phi^\alpha = 1,\ 0 \le \phi^\alpha, \alpha = 1, ..., N\}, \tag{6.92}$$

the projection operation is generally more difficult and can in particular not be put into a convenient explicit formula.

[63] While this is unlikely to be an issue in the case of the (phase-independent) multiplier Λ, $\boldsymbol{\mu}$ is a vector consisting of phase-specific entries and therefore potentially problematic when a large number of phases are present. As one of the principal advantages of explicit schemes is (besides their simplicity) their comparatively low memory requirement, this a quite convenient property.

The Euclidian Projector As already indicated above, the necessary and sufficient condition characterizing the Euclidian projection as the minimizer $\boldsymbol{\phi} = \operatorname{argmin}_{\boldsymbol{\psi}\in\mathcal{GS}} \frac{1}{2}|\boldsymbol{\psi} - \hat{\boldsymbol{\phi}}|^2$ are given by

$$\begin{cases} \boldsymbol{\phi} = \hat{\boldsymbol{\phi}} - \Lambda \boldsymbol{e} + \boldsymbol{\mu}, \\ \boldsymbol{\mu} \geq 0, \mu^\alpha \phi^\alpha = 0, \\ \boldsymbol{\phi} \in \mathcal{GS}^N. \end{cases} \tag{6.93}$$

A summation over the entries of the first equation togehter with $\sum_\alpha \phi^\alpha = 1$ implies in particular that the mulitplier Λ for the sum-constraint is given by

$$\Lambda = \frac{1}{N} \sum_\alpha \Big(\hat{\phi}^\alpha - 1 + \sum_\alpha \mu^\alpha \Big), \tag{6.94}$$

and thus that Λ is **not** given by the simple formula corresponding to the sum-constraint alone unless the multipliers $\boldsymbol{\mu}$ happens to vanish.
Nevertheless (a point which has also already been noted before) defining $\Lambda^{(0)} := \frac{1}{N} \sum_\alpha \hat{\phi}^\alpha - 1$, the positivity of the μ^α implies that one always has $\Lambda \geq \Lambda^{(0)}$. In addition, any phase α in the set of active constraints $\mathcal{A} := \{\alpha : \phi^\alpha = 0\}$, consisting of those which are actually at 0 after the projection, satisfies $0 = \hat{\phi}^\alpha - \Lambda + \mu^\alpha$ and thus

$$\mu^\alpha = \Lambda - \hat{\phi}^\alpha \geq 0. \tag{6.95}$$

In particular, any estimate $\left(\mu^\alpha\right)^{(n)} = \Lambda^{(n)} - \hat{\phi}^\alpha$ for μ^α based on Equation (6.95) and a value $\Lambda^{(n)}$ satisfying $\Lambda^{(n)} \leq \Lambda$ is never larger than the actual value of μ^α, i.e. $\left(\mu^\alpha\right)^{(n)} \leq \mu^\alpha$.
A repeated application of these two observations makes it possible to show that the following very simply algorithm will determine the correct projection:

Algorithm 1. Input: $\hat{\boldsymbol{\phi}}$.

1. Set $\boldsymbol{\phi}^{(0)} = \hat{\boldsymbol{\phi}}$ and $\mathcal{A}^{(0)} = \varnothing$.
2. If $\boldsymbol{\phi}^{(n)} \in \mathcal{GS}^N$, stop.
3. Else:
   - Find all new active phases $\delta\mathcal{A}^{(n)} = \{\alpha \notin \mathcal{A}^{(n)} : \left(\phi^\alpha\right)^{(n)} \leq 0\}$ and update the set of active constraints: $\mathcal{A}^{(n+1)} = \mathcal{A}^{(n)} \cup \delta\mathcal{A}^{(n)}$.
   - Calculate $\delta\Lambda^{(n)} = -\frac{1}{N - |\mathcal{A}^{(n+1)}|} \sum_{\alpha\in\delta\mathcal{A}^{(n)}} \left(\phi^\alpha\right)^{(n)}$.
   - Update $\boldsymbol{\phi}$: For all $\alpha \in \mathcal{A}^{(n+1)}$ set (resp. keep) $\left(\phi^\alpha\right)^{(n+1)} = 0$. For all inactive constraints, $\alpha \notin \mathcal{A}^{(n+1)}$, set $\left(\phi^\alpha\right)^{(n+1)} = \left(\phi^\alpha\right)^{(n)} - \delta\Lambda^{(n)}$.
   - Return to step 2.

One then has the following

Lemma 4. *Given any $\hat{\boldsymbol{\phi}}$, Algorithm 1 determines the vector $\boldsymbol{\phi} = \operatorname{argmin}_{\boldsymbol{\psi}\in\mathcal{GS}} \frac{1}{2}|\boldsymbol{\psi} - \hat{\boldsymbol{\phi}}|^2$ corresponding the Euclidian projection in at most $N - 1$ steps.*
The sets $\mathcal{A}^{(n)}$ increase monotonically to the correct set $\mathcal{A}$ of active constraints, and the values of the corresponding multipliers Λ and $\boldsymbol{\mu}$ are given by

$$\Lambda = \sum_n \delta\Lambda^{(n)} = \frac{1}{N - |\mathcal{A}|}\Big(\sum_{\alpha\notin\mathcal{A}} \hat{\phi^\alpha} - 1 \Big) \quad \textit{and} \quad \mu^\alpha = \begin{cases} 0 & \alpha \notin \mathcal{A}, \\ \Lambda - \hat{\phi}^\alpha & \alpha \in \mathcal{A}. \end{cases} \tag{6.96}$$

Proof. The crucial point here are some useful monotonicity-properties. Given any prediction $\mathcal{A}^{(n)}$ of the active set which does not contain a phase falsely classified as being zero - and this is certainly true for $\mathcal{A}^{(0)} = \varnothing$ - it follows from $\phi^\alpha = 0$ if $\alpha \in \mathcal{A}^{(n)}$ that the sum-constraint also has to be satsifed by the remaining $N - |\mathcal{A}^{(n)}|$ phases. A simple summation over all of these phases in the first line of Equation (6.93) shows that $1 = \sum_{\alpha \notin \mathcal{A}^{(n)}} \phi^\alpha = \sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - (N - |\mathcal{A}^{(n)}|)\Lambda + \sum_{\alpha \notin \mathcal{A}^{(n)}} \mu^\alpha$ and thus, similar to Equation (6.94), that

$$\Lambda^{(n)} = \frac{1}{N - |\mathcal{A}^{(n)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 + \sum_{\alpha \notin \mathcal{A}^{(n)}} \mu^\alpha \Big). \tag{6.97}$$

As before, by the positivity of the μ^α, this implies that $\Lambda \geq \frac{1}{N - |\mathcal{A}^{(n)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 \Big) =: \Lambda^{(n)}$. Setting $\boldsymbol{\phi}^{(n+1)} := \hat{\boldsymbol{\phi}}^{(n)} - \Lambda^{(n)} \boldsymbol{e}$, it follows that $\boldsymbol{\phi}^{(n+1)} \geq \hat{\boldsymbol{\phi}} - \Lambda \boldsymbol{e}$, where the inequality is to be understood componentwise. In particular, for all α with $\big(\phi^\alpha\big)^{(n+1)} \leq 0$ one necessarily also has $\hat{\phi}^\alpha - \Lambda \leq \big(\phi^\alpha\big)^{(n)} - \Lambda^{(n)} \leq 0$. This implies that the constraint of any such phase is acutally active, i.e. $\alpha \in \mathcal{A}$ and $\phi^\alpha = 0$, since, assuming this is not the case, the complementary condition enforces $\mu^\alpha = 0$ and thus by the first-order necessary condition $\phi^\alpha = \hat{\phi}^\alpha - \Lambda$, contradicting the non-negativity of ϕ^α. Adding all such phases to the set of active constraints $\mathcal{A}^{(n+1)}$ at the next iteration therefore again leads to a set satisfying $\mathcal{A}^{(n+1)} \subset \mathcal{A}$.
This new estimate of the set of active constraints further leads to the new estimate

$$\Lambda^{(n+1)} = \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n+1)}} \hat{\phi}^\alpha - 1 \Big) = \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 \Big) - \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \in \delta\mathcal{A}^{(n)}} \hat{\phi}^\alpha \Big)$$

for the Lagrange multiplier for the sum-constraint, where the summation was artificially split into those phases already in $\mathcal{A}^{(n)}$ and those in $\delta\mathcal{A}^{(n)} = \mathcal{A}^{(n+1)} \setminus \mathcal{A}^{(n)}$. Using

$$\begin{aligned}\frac{1}{N - |\mathcal{A}^{(n+1)}|} =& \frac{1}{N - |\mathcal{A}^{(n)}|} \frac{N - |\mathcal{A}^{(n)}|}{N - |\mathcal{A}^{(n)}| - |\delta\mathcal{A}^{(n)}|} = \frac{1}{N - |\mathcal{A}^{(n)}|}\left(1 + \frac{|\delta\mathcal{A}^{(n)}|}{N - |\mathcal{A}^{(n)}| - |\delta\mathcal{A}^{(n)}|} \right) \\ =& \frac{1}{N - |\mathcal{A}^{(n)}|}\left(1 + \frac{|\delta\mathcal{A}^{(n)}|}{N - |\mathcal{A}^{(n+1)}|} \right),\end{aligned}$$

the first term can be rewritten as

$$\frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 \Big) = \underbrace{\frac{1}{N - |\mathcal{A}^{(n)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 \Big)}_{=\Lambda^{(n)}} + \frac{|\delta\mathcal{A}^{(n)}|}{N - |\mathcal{A}^{(n+1)}|} \underbrace{\frac{1}{N - |\mathcal{A}^{(n)}|}\Big(\sum_{\alpha \notin \mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 \Big)}_{=\Lambda^{(n)}}$$

in terms of the previous estimate $\Lambda^{(n)}$, thus showing that

$$\begin{aligned}\Lambda^{(n+1)} =& \Lambda^{(n)} + \frac{|\delta\mathcal{A}^{(n)}|}{N - |\mathcal{A}^{(n+1)}|}\Lambda^{(n)} - \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \in \delta\mathcal{A}^{(n)}} \hat{\phi}^\alpha \Big) = \Lambda^{(n)} - \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \in \delta\mathcal{A}^{(n)}} \big(\hat{\phi}^\alpha - \Lambda^{(n)} \big) \Big) \\ =& \Lambda^{(n)} - \frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \in \delta\mathcal{A}^{(n)}} \big(\phi^\alpha\big)^{(n)} \Big),\end{aligned}$$

where the last summand is precisely the definition $\delta\Lambda^{(n)} = -\frac{1}{N - |\mathcal{A}^{(n+1)}|}\Big(\sum_{\alpha \in \delta\mathcal{A}^{(n)}} \big(\phi^\alpha\big)^{(n)} \Big)$ in algorithm (1). As all $\big(\phi^\alpha\big)^{(n)}$ in $\delta\mathcal{A}^{(n)}$ are non-positive, it on the one hand is also clear that $\delta\Lambda^{(n)}$ is non-negative, and the $\Lambda^{(n)}$ indeed form an increasing sequence, and on the other hand, repeating this calculation, that $\Lambda^{(n)} = \sum_{m<n} \delta\Lambda^{(m)}$.

Finally, the algorithm is terminated once $\boldsymbol{\phi}^{(n)}$ contains no further non-positive entries except for those in $\mathcal{A}^{(n)}$. In addition, the $\boldsymbol{\phi}^{(n)}$ always satisfy the sum-constraint by the construction of $\delta\Lambda^{(n)}$ and one therefore has $\boldsymbol{\phi}^{(n)} \in \mathcal{GS}^N$. All that remains to be verified is therefore that all the $\left(\mu^\alpha\right)^{(n)}$ defined as in Equation (6.96) based on $\Lambda^{(n)}$ and $\mathcal{A}^{(n)}$ are indeed positive. This is clear though by the mononotonic increase of the $\Lambda^{(n)}$ and the fact that a phase is only added to $\mathcal{A}^{(m)}$ provided it satisfies, due to the equality $\left(\phi^\alpha\right)^{(m)} = \hat{\phi}^\alpha - \Lambda^{(m)}$ by the summation formula for the $\delta\Lambda$, $0 \geq \hat{\phi}^\alpha - \Lambda^{(m)} \geq \hat{\phi}^\alpha - \Lambda^{(n)}$ for all $n \geq m$, therefore ensuring $\left(\mu^\alpha\right)^{(n)} = \Lambda^{(n)} - \hat{\phi}^\alpha \geq 0$.
That this has to happen in at most n steps is obvious as, starting from $\mathcal{A}^{(0)} = \varnothing$, each iteration adds at least one phase with $\phi^\alpha = 0$, of which there can be at most $N-1$ for any vector in $\mathcal{GS}^N$. □

Remark 84. Note that, combining the description in Algorithm 1 and the proof of Lemma 4, there are in fact a variety of ways this algorithm can be implemented in an "equivalent" fashion. In particular, one can replace the incremental formulation in terms of the $\delta\Lambda^{(n)}$ and the $\boldsymbol{\phi}^{(n)}$ by a direct evaluation based only on $\hat{\boldsymbol{\phi}}$ and $\Lambda^{(n)} = \frac{1}{N-|\mathcal{A}^{(n)}|}\left(\sum_{\alpha\notin\mathcal{A}^{(n)}} \hat{\phi}^\alpha - 1 + \sum_{\alpha\notin\mathcal{A}^{(n)}} \mu^\alpha\right)$ as in Equation (6.97). In fact, the $\boldsymbol{\phi}^{(n)}$ are not really needed directly in the algorithm, since they only serve to judge whether $\hat{\phi}^\alpha - \Lambda^{(n)} \leq 0$ for any $\alpha \notin \mathcal{A}^{(n)}$ in order to determine the update $\delta\mathcal{A}^{(n)}$ of the set of active constraints. This can also be done directly based on this formula for $\Lambda^{(n)}$ and $\hat{\boldsymbol{\phi}}$, with $\boldsymbol{\phi}$ only being generated at the end through

$$\phi^\alpha = \begin{cases} 0 & \alpha \in \mathcal{A}^{(n)}, \\ \hat{\phi}^\alpha - \Lambda^{(n)} & \end{cases} \tag{6.98}$$

once there is no new phase added at the step n.
In addition, even though one can add all phases with $\left(\phi^\alpha\right)^{(n)} \leq 0$ to the new prediction of $\mathcal{A}$, one does not have to, and can e.g. only add the first such one that one encounters and then enter the next step of the algorithm (this being the original implementation in the **Pace3D**-framework), either in an incremental fashion using the $\delta\Lambda^{(n)}$ and the $\boldsymbol{\phi}^{(n)}$ or in a non-incremental one based directly on $\hat{\boldsymbol{\phi}}$ and $\hat{\Lambda}^{(n)}$. Which version is more favorable computationally depends on a number of practical factors. While taking all newly active constraints into account at the same time does certainly deliver the lower number of total iterations, simply "skipping" the remaining checks and restarting with an updated (and increased) estimate of Λ increases the chances of more quickly identifying another phase which will ultimately have to be added to $\mathcal{A}$ anyway and can thus also be advantageous.
Besides a pure iteration count and the number of arithmetic operations this involves, the actual run-time also depends on the time spend traversing the loops[64] in addition to the arithmetic operations themselves. ◇

Remark 85. It should also be noted that the projection above (regardless of its actual implementation) interplays very favorably with a preclassification of the phases and an LROP-based approach. More precisely, if, instead of actually using all phases in this algorithm, one a priori can already exlude a number of those since they are already "known" to be zero, one can simply restrict the projection to the reduced subvector of phases for which this is not a priori clear without any effect on the result. In practical terms, this simply corresponds to taking only those phases into account for the algorithm, and replacing the total number N of phases with the number $\tilde{N}$ of phases which are not a priori fixed at 0. ◇

[64] For example skipping over phases in $\mathcal{A}^{(n)}$ requires an additional conditional statement unless the phase is directly eliminated through an LROP-type indirect indexing. If based on this indirect indexing, eliminating a phase at 0 necessitates a rearrangement of indices, which can be an expensive operation if there are a large number of them and the one to eliminate appears very early.

The Weighted Projector Induced by the Mobility-Matrix $\boldsymbol{M}$ The projection is - both on a theoretical and practical level - somewhat more difficult when the dynamics are based on the use of the mobility-matrix $\boldsymbol{M}$ as in Equation (6.91). As discussed in the previous subsection, the compatibility of the projection-based approach (at least in the final steady state) - requires that the projection $\boldsymbol{\phi}^{(n+1)}$ and the prediction $\hat{\boldsymbol{\phi}}^{(n+1)}$ be related through $\boldsymbol{\phi}^{(n+1)} = \hat{\boldsymbol{\phi}}^{(n+1)} + \boldsymbol{M}\boldsymbol{\mu}^{(n+1)}$ with $\boldsymbol{\mu}^{(n+1)}$ subject to the same complementarity conditions with respect to the (a priori unkown) vector $\boldsymbol{\phi}^{(n+1)}$ as in the Euclidian case, at least **if** $\hat{\boldsymbol{\phi}}$ is such that it satisfies the sum-constraint[65]. In analogy with the case of invertible matrices[66] and dropping the time-indices corresponding to the "outer" algorithm, it will be seen below that this is essentially the first-order necessary condition corresponding to the minimization problem

$$\boldsymbol{\phi} = \operatorname{argmin}_{\boldsymbol{\psi}\in\mathcal{GS}^N} \frac{1}{2}(\boldsymbol{\psi}-\hat{\boldsymbol{\phi}})\cdot\boldsymbol{M}^{\dagger}(\boldsymbol{\psi}-\hat{\boldsymbol{\phi}}),$$

where $\boldsymbol{M}^{\dagger}$ is the pseudo-inverse of $\boldsymbol{M}$.

Remark 86. Recall that $\boldsymbol{M}^{\dagger}(\boldsymbol{\psi}-\hat{\boldsymbol{\phi}})$ for the Moore-Penrose pseudo-inverse is defined as the element $\boldsymbol{\zeta}$ of smallest norm satisfying $\boldsymbol{M}\boldsymbol{\zeta} = \boldsymbol{\psi}-\hat{\boldsymbol{\phi}}$. Since the kernel of $\boldsymbol{M}$ consists of the constant vectors only as long as all $m^{\alpha\beta}$ are strictly larger than 0[67], this smallest norm-condition is not even necessary as long as $\hat{\boldsymbol{\phi}}$ has the correct average as $\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}$ is then a vector of zero average, meaning that the average of $\boldsymbol{\zeta}$ plays no role for the semi-norm above. This is not true anymore if $\sum_\alpha \hat{\phi}^\alpha \neq 1$ though. ◇

A straightforward differentiation of the corresponding Lagrangian $L(\boldsymbol{\phi},\Lambda,\boldsymbol{\mu}) = \frac{1}{2}(\boldsymbol{\psi}-\hat{\boldsymbol{\phi}})\cdot\boldsymbol{M}^{\dagger}(\boldsymbol{\psi}-\hat{\boldsymbol{\phi}})+\Lambda\big(\sum_\alpha\psi^\alpha-1\big)-\boldsymbol{\mu}\cdot\boldsymbol{\psi}$ shows that the first-order necessary condition for this problem is given by $\boldsymbol{M}^{\dagger}\big(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\big) = -\Lambda\boldsymbol{e}+\boldsymbol{\mu}$ with satisfying the complementarity conditions $\boldsymbol{\mu}\geq\boldsymbol{0}$, $\mu^\alpha\phi^\alpha = 0$. Multiplying by $\boldsymbol{M}$, it follows that

$$\boldsymbol{M}\boldsymbol{M}^{\dagger}\big(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\big) = \boldsymbol{M}\big(-\Lambda\boldsymbol{e}+\boldsymbol{\mu}\big) = \boldsymbol{M}\boldsymbol{\mu}.$$

This is not quite the same as the $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}+\boldsymbol{M}\boldsymbol{\mu}$ as one has $\boldsymbol{M}\boldsymbol{M}^{\dagger} = \mathcal{P}_{\mathrm{Range}(\boldsymbol{M})}$ (see e.g. chapter 3 in [8]). By the fundamental theorem of linear algrebra (Theorem 1) with $m = n$, $\mathrm{Range}(\boldsymbol{M})\oplus\mathrm{Ker}(\boldsymbol{M}) = \boldsymbol{R}^n$ and since the kernel of $\boldsymbol{M}$ consists of the constant vectors only, it follows that $\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}$ differs from $\boldsymbol{M}\boldsymbol{\mu}$ only through a constant vector of the form $-\Lambda\boldsymbol{e}$, i.e. the the projected vector actually has to satisfy

$$\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}+\boldsymbol{M}\boldsymbol{\mu}-\Lambda\boldsymbol{e}. \tag{6.99}$$

Remark 87. It is not overly surprising that the sum constraint leads to the appearence of a multiplier Λ in a similar manner as in the Euclidian case, i.e. outside of $\boldsymbol{M}$ since $\boldsymbol{M}$ has no manner of changing the average which nevertheless has to be the correct one due to the admissible set.

[65] While this is due to the structure of $\boldsymbol{M}$ normally being ensured by the dynamics, provided the initial phasefield-vectors are compatible with this constraint, this can cause some problems in combination with a too low number of phases which can be stored in the LROP-approach. Even if the storage space is sufficient for actually storing $\boldsymbol{\phi}^{(n+1)}$, this need not be the case for $\hat{\boldsymbol{\phi}}^{(n+1)}$ which may contain some phases below zero which would then be eliminated by the projection.

[66] Recall that $\boldsymbol{M}$ is a singular matrix as it satisfies $\boldsymbol{M}\boldsymbol{e} = \boldsymbol{0}$, i.e. it vanishes on the constant vectors.

[67] This is a very natural assumption in the variational context as it is the only way of generally ensuring that $\boldsymbol{\phi}$ becoming stationary in combination with $\boldsymbol{\mu}$ as above implies that one has actually found a local minimizer. More precisely, if $\boldsymbol{M}$ only vanishes on the constant vectors, $\boldsymbol{M}(-\boldsymbol{g}+\boldsymbol{\mu}) = \boldsymbol{0}$, implies that $\boldsymbol{g} = -\Lambda\boldsymbol{e}+\boldsymbol{\mu}$ and thus the first-order necessary condition in Equation (6.63). Nevertheless, one seeming advantage of the mobility-based formulation in contrast to the one based on τ is that, in the prediction step, one can easily eliminate all interactions between certain phases by simply setting $m^{\alpha\beta} = 0$ without incurring a "division by zero". Besides its problematic implications with respect to any potential minimizing or maximizing phasefield, this simplicity is highly misleading, since it causes some very tedious issues in the projection step. The strict positivity of the $m^{\alpha\beta}$, regardless of whether or not they are very small, will therefore be a standing assumption made here.

In contrast to the Euclidian case, the correct value of Λ can be determined upon entry into the projection operation, since - regardless of the values of $\boldsymbol{\mu}$ due to $\mathrm{Range}(\boldsymbol{M}) = \mathrm{Span}\{\boldsymbol{e}\}^\perp$ - a simple summation shows that $\Lambda = \sum_\alpha \hat{\phi}^\alpha - 1$. If $\hat{\boldsymbol{\phi}}$ already satisfies the sum-contraint, it is then obvious that $\Lambda = 0$ and can thus safely be ignored. If not, it nevertheless suffices to replace $\hat{\boldsymbol{\phi}}$ by $\hat{\boldsymbol{\phi}} - \Lambda e$ within the actual projection operation in order to obtain the correct result. ◇

Remark 88. This is one of the points where vanishing values of $m^{\alpha\beta}$ can cause serious issues. Assuming e.g. there is a single phase β which is "disconnected" from the other ones, this adds an additional base vector $\boldsymbol{e}_\beta$ to $\mathrm{Ker}(\boldsymbol{M})$ and, by symmetry, to $\mathrm{Range}(\boldsymbol{M})^\perp$. The conclusion above therefore needs to be modified to $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}} + \boldsymbol{M}\boldsymbol{\mu} - \Lambda\boldsymbol{e} + \lambda^\beta \boldsymbol{e}_\beta$. As $\boldsymbol{e}_\beta \in \mathrm{Ker}(\boldsymbol{M})^\perp$, there is no influence of $\boldsymbol{\mu}$ on the β-th entry of this equation, i.e. one has $\phi^\beta = \hat{\phi}^\beta - \Lambda + \eta^\beta$. This in itself essentially leaves λ^β undefined, provided it is such that ϕ^β satisfies the non-negativity constraint. It is only in combination with the definition of the pseudo-inverse - enforcing that $\boldsymbol{M}^\dagger(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})$ be the element of smallest norm satisfying this equation - that one is indirectly able to fix both Λ and η^β.
If the "basic" prediction of Λ is such that ϕ^β remains non-negative, the smallest deviation in norm is obtained for $\eta^\beta = 0$ as this is the optimal way of distributing a difference in the sum onto a vector (the norm involved in the pseudo-inverse being the standard Euclidian one here and not a weighted one) and the previous formula for Λ remains valid. If Λ moves ϕ^β out of the constraint set though - either directly due to $\phi^\beta < 0$ or indirectly through $\hat{\phi}^\beta - \Lambda > 1$ in combination with the sum-constraint and the constraints on the other phases - η^β has to be adjusted in a manner similar to the role of $\boldsymbol{\mu}$ in the Euclidian projection, except that one now has to keep two bounds in mind.
Even though this issue can in principle be handled and the phase β is afterwards unaffected by the determination of $\boldsymbol{\mu}$, this requires a rather complex "preprocessing" operation before considering the original problem of the determination of the $\boldsymbol{\mu}$. The situation is of course much worse when it is not a priori known that there is only a single disconnected phase but when there may be one or several groups of disconnected phases. ◇

Whenever all mobilities are strictly positive, the projection can in fact be performed in a manner relatively similar to the Euclidian case. More precisely, it will be shown in Lemma 5 below that this can be achieved through

Algorithm 2. Input: $\hat{\boldsymbol{\phi}}$.

1. Set $\boldsymbol{\phi}^{(0)} = \hat{\boldsymbol{\phi}}$ and $\mathcal{A}^{(0)} = \varnothing$.
2. If $\boldsymbol{\phi}^{(n)} \in \mathcal{GS}^N$, stop.
3. Else:
   - Find all new active phases $\delta\mathcal{A}^{(n)} = \{\alpha \notin \mathcal{A}^{(n)} : \left(\phi^\alpha\right)^{(n)} \leq 0\}$ and update the set of active constraints: $\mathcal{A}^{(n+1)} = \mathcal{A}^{(n)} \cup \delta\mathcal{A}^{(n)}$.
   - Update the estimate for $\boldsymbol{\mu}$ by solving $\boldsymbol{M}_{\mathcal{A}^{(n+1)}\mathcal{A}^{(n+1)}}\boldsymbol{\mu}_{\mathcal{A}^{(n+1)}} = -\hat{\boldsymbol{\phi}}_{\mathcal{A}^{(n+1)}}$, where $(\cdot)_{\mathcal{A}^{(n+1)}}$ denotes the restriction of the respective matrices and vectors to the index-set $\mathcal{A}^{(n+1)}$.
   - Update $\boldsymbol{\phi}$: For all $\alpha \in \mathcal{A}^{(n+1)}$ set (resp. keep) $\left(\phi^\alpha\right)^{(n+1)} = 0$. For all inactive constraints, set $\left(\phi^\alpha\right)^{(n+1)} = \hat{\boldsymbol{\phi}} + \sum_{\beta\in\mathcal{A}^{(n+1)}} M^{\alpha\beta}\mu^\beta = \hat{\phi}^\alpha - \sum_{\beta\in\mathcal{A}^{(n+1)}} m^{\alpha\beta}\mu^\beta$.
   - Return to step 2.

Remark 89. Note that, in contrast to Algorithm 1, it is now necessary to explicitly keep track of the values of the multipliers $\boldsymbol{\mu}$ as they affect the remaining $\boldsymbol{\phi}$-values differently due to the non-uniform weighting by the $m^{\alpha\beta}$. This is due to the fact that, unlike for the Euclidian case, the effect of each μ^α is not the same on all remaining phases due to the weighting and can therefore not be "summarized" through the action of a single multiplier Λ. ◇

The following analogue of Lemma 4 holds:

Lemma 5. *Assume that the mobility matrix $\boldsymbol{M}$ is of the form*

$$M^{\alpha\beta} = \begin{cases} \sum_{\gamma\neq\alpha} m^{\alpha\gamma} & \alpha = \beta, \\ -m^{\alpha\beta} & else \end{cases}$$

with $m^{\alpha\beta} = m^{\beta\alpha} > 0$ for all $\alpha \neq \beta$.
Then, given any $\hat{\boldsymbol{\phi}}$ satisfying $\sum_\alpha \hat{\phi}^\alpha = 1$ or being being "replaced" by $\hat{\boldsymbol{\phi}} - \frac{1}{N}\big(\sum_{\alpha=1}^N - 1\big)\boldsymbol{e}$, Algorithm 2 determines the weighted projection $\boldsymbol{\phi} = \operatorname{argmin}_{\boldsymbol{\psi}\in\mathcal{GS}} \frac{1}{2}(\boldsymbol{\psi} - \hat{\boldsymbol{\phi}}) \cdot \boldsymbol{M}^\dagger(\boldsymbol{\psi} - \hat{\boldsymbol{\phi}})$ in at most $N-1$ steps.
The sets $\mathcal{A}^{(n)}$ increase monotonically to the correct set $\mathcal{A}$ of active constraints, each of the submatrices $\boldsymbol{M}_{\mathcal{A}^{(n)}\mathcal{A}^{(n)}}$ is invertible, and the corresponding multiplier $\boldsymbol{\mu}$ is given by $\mu^\alpha = 0$ if $\alpha \notin \mathcal{A}$ and

$$\boldsymbol{\mu}_{\mathcal{A}} = -\boldsymbol{M}^{-1}_{\mathcal{A}\mathcal{A}}\hat{\boldsymbol{\phi}}_{\mathcal{A}} \tag{6.100}$$

otherwise, all entries of this vector are non-negative, and $\boldsymbol{\phi}$ is given by $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}} + \boldsymbol{M}\boldsymbol{\mu}$.

Proof. The proof of this lemma is based on a very similar argument as the one of Lemma 4 in combination with the favorable structure of $\boldsymbol{M}$, namely verifying that the set $\mathcal{A}$ is monotonically increasing without ever falsely classifying a constraint as active and that the vector $\boldsymbol{\mu}$ in Equation (6.100) therefore has only non-negative entries and (by construction) is zero if $\phi^\alpha > 0$.
The crucial property of $\boldsymbol{M}$ here is that it, even though $\boldsymbol{M}$ itself is only a (singular) M**-matrix**, each of the submatrices $\boldsymbol{M}_{\mathcal{A}'\mathcal{A}'}$ with $\mathcal{A}'$ strictly included in $\{1,...,N\}$ is a strictly diagonally dominant matrix with only non-positive off-diagonal entries. In fact, taking any such submatrix, the diagonal entry $M^{\alpha\alpha}$ is of the from $\sum_{\beta\neq\alpha} m^{\alpha\beta}$ with all $m^{\alpha\beta} > 0$, whereas the off-diagonal entries consist of the entries $-m^{\alpha\beta}$ for all $\beta \in \mathcal{A}'$. As $\mathcal{A}'$ is assumed to be a strict subset of the set of all phase-indices, $M^{\alpha\alpha}$ is then clearly diagonally dominant as $m^{\alpha\beta} - \sum_{\mathcal{A}'\ni\beta\neq\alpha} |m^{\alpha\beta}|$ equals the sum over all $m^{\alpha\beta} \notin \mathcal{A}'$ which is strictly positive.
This implies firstly (see e.g. [55] and [76]) that $\mathcal{M}_{\mathcal{A}^{(n)}\mathcal{A}^{(n)}}$ is indeed invertible, and furthermore that all entries of $\mathcal{M}^{-1}_{\mathcal{A}^{(n)}\mathcal{A}^{(n)}}$ are non-negative. Even though this is by itself not sufficient to show that $\boldsymbol{\mu}$ is non-negative as $-\hat{\boldsymbol{\phi}}_{\mathcal{A}}$ may very well contain negative entries (i.e. "projected" phases which are initially above 0), it is sufficient to show the monotonous increase of the $\boldsymbol{\mu}^{(n)}$ in Algorithm 2 and thus, as $\boldsymbol{\mu}^{(0)}$ corresponding to $\mathcal{A}^{(0)} = \varnothing$ is all zero, that $\boldsymbol{\mu}$ indeed only consists of non-negative entries.
Assuming therefore that $\mathcal{A}^{(n)}$ does not contain any phases falsely classified as active (and thus necessarily $|\mathcal{A}^{(n)}| < N$ by the sum-constraint) and that $\boldsymbol{\mu}^{(n)} \geq 0$ - this trivially holding for $n = 0$ - it remains to show the same for the step $n+1$. By the definition of the $\boldsymbol{\phi}^{(n+1)}$ and $\boldsymbol{\mu}^{(n+1)}$ in Algorithm 2 together with the invertibility of the submatrices, it follows that $\boldsymbol{\phi}^{(n+1)}$ actually satisfies

$$\boldsymbol{\phi}^{(n+1)} = \hat{\boldsymbol{\phi}} + \boldsymbol{M}\boldsymbol{\mu}^{(n+1)}. \tag{6.101}$$

On the one hand, since, with $\mu^\alpha = 0$ for $\alpha \notin \mathcal{A}^{(n+1)}$, the summation in the product $\boldsymbol{M}\boldsymbol{\mu}^{(n+1)}$ automatically reduces to the summation over all phases in $\mathcal{A}^{(n+1)}$ and thus the formula for $\big(\phi^\alpha\big)^{(n+1)}$ if $\alpha \notin \mathcal{A}^{(n+1)}$. One the other hand, $\boldsymbol{\mu}^{(n+1)}$ is chosen precisely such that $\hat{\boldsymbol{\phi}}_{\mathcal{A}^{(n+1)}} + \boldsymbol{M}_{\mathcal{A}^{(n+1)}\mathcal{A}^{(n+1)}}\boldsymbol{\mu}^{(n+1)} = \boldsymbol{0}$ and thus implies that $\boldsymbol{\phi}^{(n+1)}_{\mathcal{A}^{(n+1)}} = \boldsymbol{0}$ as imposed in the algorithm.
Since $\boldsymbol{\mu}^{(n)}_{\{1,...,N\}\setminus\mathcal{A}^{(n)}} = \boldsymbol{0}$ by definition (and all relations are trivially true for $n = 0$), it follows by restricting Equation (6.101) to the set $\mathcal{A}^{(n+1)}$ that

$$\boldsymbol{\phi}^{(n)}_{\mathcal{A}^{(n+1)}} = \hat{\boldsymbol{\phi}}_{\mathcal{A}^{(n+1)}} + \boldsymbol{M}_{\mathcal{A}^{(n+1)}\mathcal{A}^{(n+1)}}\boldsymbol{\mu}^{(n)}_{\mathcal{A}^{(n+1)}}$$

and thus, by subtracting this from Equation (6.101), that

$$\big(\boldsymbol{\phi}^{(n+1)} - \boldsymbol{\phi}^{(n)}\big)_{\mathcal{A}^{(n+1)}} = \boldsymbol{M}_{\mathcal{A}^{(n+1)}\mathcal{A}^{(n+1)}}\big(\boldsymbol{\mu}^{(n+1)} - \boldsymbol{\mu}^{(n)}\big)_{\mathcal{A}^{(n+1)}}.$$

As $\mathcal{A}^{(n)} \subset \mathcal{A}^{(n+1)}$, $\boldsymbol{\phi}^{(n+1)}_{\mathcal{A}^{(n)}} = \boldsymbol{\phi}^{(n)}_{\mathcal{A}^{(n)}} = \mathbf{0}$ (by the definition of the $\left(\phi^{\alpha}\right)^{(n)}$) and all entries in $\delta\mathcal{A}^{(n)} = \mathcal{A}^{(n+1)} \setminus \mathcal{A}^{(n)}$ satisfy $\left(\boldsymbol{\phi}^{(n+1)} - \boldsymbol{\phi}^{(n)}\right)_{\delta\mathcal{A}^{(n)}} = \mathbf{0} - \boldsymbol{\phi}^{(n)}_{\delta\mathcal{A}^{(n)}}$ with $\boldsymbol{\phi}^{(n)}_{\delta\mathcal{A}^{(n)}} \leq \mathbf{0}$ (by the definition of $\delta\mathcal{A}^{(n)}$), it follows that the left-hand side consists of non-negative entries only. Together with the postivity of all entries of $\boldsymbol{M}^{-1}_{\mathcal{A}^{(n+1)}\mathcal{A}^{(n+1)}}$, it is then clear that the same also holds for $\left(\boldsymbol{\mu}^{(n+1)} - \boldsymbol{\mu}^{(n)}\right)_{\mathcal{A}^{(n+1)}}$ and thus, with all other entries being zero, $\boldsymbol{\mu}^{(n+1)} - \boldsymbol{\mu}^{(n)}$.
Summarizing these arguments, the sets $\mathcal{A}^{(n)}$ increase by at least one entry per step unless there are no negative entries left in the prediction of the projection. All vectors satisfy the sum-constraint (potentially after a correction of their average as above) by the nature of $\boldsymbol{M}$ and the $\boldsymbol{\mu}$ are monotonically increasing from $\mathbf{0}$ and therefore non-negative, showing that the algorithm converges to a solution of the necessary and here also sufficient condition in Equation (6.99). □

Remark 90. It has to be observed that from a dynamic point of view, the mobility approach is in general **not** compatible with a preclassification scheme. Nevertheless, it can be shown that, once a steady state is reached and if the choice of skipping the calculations was correct (this being the crucial assumption of course), this state is compatible with the first-order necessary condition in Equation (6.63).
Even though this is therefore a potential problem only for the dynamics but not from a variational point of view, this point should not be underestimated as the motivation of the mobility-based formulation is precisely the modification of the dynamics, which, if used in an LROP-type fashion, gets mixed with a very "discrete" concept in terms of the preclassification.
What can in additionally generate some difficulties with the projection-based scheme here in combination with the LROP storage scheme is that, in contrast to the Euclidian projector, is that it is not sufficient for the calculation of the correct projection that a phase will be at 0 after the projection step. Instead, even if the result itself for this phase is known, it nevertheless necessary to also know he corresponding value in the prediction $\hat{\boldsymbol{\phi}}$ as this difference has to be redistributed to the remainig phases in accordance with the weights in $\boldsymbol{M}$. Even if the storage is chosen sufficiently large to be able to store all "actually" occuring phases with non-zero values - i.e. those **after** the projection - this need not be the case before the projection as there may be a number of phases with negative values after the prediction step. Due to the discussion above, it is in this case not legitimate to simply truncate these to zero, since, even if this is their correct final value, the difference has to known for the projection operation itself. ◇

Chapter 7

Applications in the Material Sciences

The use of the phasefield method in the material sciences is based on combining its ability for capturing effects associated with surface energies with driving forces induced by an appropriate thermodynamic contribution to the phasefield functional togehter with evolution equations for the relevant additional fields such as e.g. the concentration or temperature. Within the **Pace3D**-framework, these thermodynamic contributions are generally described directly or indirectly in terms of a free energy density $f(\boldsymbol{\phi}, \boldsymbol{c}, T, ...)$, through which the underlying free energy densities for the bulk-phases are extended to the diffuse interface region.

Given that the surface and bulk free energies are basically fixed parameters based on measured quantities for the given materials, one of the key challenges in achieving accurate results based on the phasefield method therefore lies in the construction of a sufficiently accurate model of the material behavior **within** the transition region. Whereas the earlier models for the energy contributions within this region were usually based upon relatively simple interpolation procedures in terms of the available additional fields - such as the total concentration $\boldsymbol{c}$ or, in the mechanical case, the total strains $\boldsymbol{\epsilon}$ - the more recent modeling approach consists in including a larger degree of the underlying physics into the interpolation procedure.

As the chemically driven phase transformations where on the one hand historically the first setting within which this question was adressed in detail and on the other hand in many cases allows for obtaining a satisfactory model using a somewhat simpler modification than in the mechanical case, this setting will be considered first in Section 7.1. It will start with an outline of and the discussion of some practically relevant issues induced by a now well established approach for isothermal solidifation problems developed, among others, in the works of [42], [25], [19] and [56]. Through a slight adaptation, this approach can in fact be extended to include the more general non-isothermal setting underlying the model in [52].
Section 7.2, will then discuss related modeling approaches within a mechanical setting. In contrast to the chemical setting, a simple translation of the ideas in the chemical case unfortunately does not result in equally satisfactory results due to a difference in the nature of the equilibrium conditions[1]. Nevertheless, at least within a two-phase settings, a very satisfactory model based on the sharp interface mechanical jump conditions at interfaces has been developed in different forms in the works of [23], [51], [64] and [74]. Unlike in the chemical case, an extension to more than two phases involves a very fundamental difficulty closely related to the well-known singularities at singular points arising in the sharp interface setting.

[1] This is somewhat misleading though as similar difficulties can in fact arise in the chemical setting as well depending on e.g. the boundary conditions.

Even though partial extensions to the multi-phase setting were proposed in, among others, [61], [62] and [74], these models loose, within multiphase regions, many of the favorable properties valid in the two-phase setting. The discussion of the multi-phasefield models together with some of these issues as well as modificitations aiming to at least mitigate their impact will therefore be postponed until Section 7.2.4. Finally, Section 7.2.5 will outline a situation where both chemical and elastical effects arise in a coupled fashion, which will therefore be treated using a combination of both the chemical and mechanical models from the previous sections.

7.1 Quantitative Phasefield Models for Solidification

7.1.1 The "Traditional" Model

Before discussing some modifications through a more advanced modeling approach in the next section, it is instructive to recall the more traditional approach to alloy solidification problems as proposed in [52] and which parts of this model might be improved upon. As outlined in Section 3.2.1, it is based upon the entropy functional $\mathcal{S}_\epsilon(\boldsymbol{\phi}, \boldsymbol{c}, e)$ from Equation (3.3) (repeated here for convenience)

$$\mathcal{S}_\epsilon(\boldsymbol{\phi}, \boldsymbol{c}, e) = \int_\Omega s(\boldsymbol{\phi}, \boldsymbol{c}, e) - \left(\epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi})\right) \mathrm{d}\boldsymbol{x},$$

with the dependence of the thermodynamics of the material on the concentration $\boldsymbol{c}$ and the local energy density e entering through the bulk entropy density $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$. $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ can be related to the often more convenient free energy density $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ through the Legendre-type transform pair

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \inf_e \{e - Ts(\boldsymbol{\phi}, \boldsymbol{c}, e)\} \quad \leftrightarrow \quad s(\boldsymbol{\phi}, \boldsymbol{c}, e) = \inf_T \{\frac{e}{T} - \frac{f(\boldsymbol{\phi}, \boldsymbol{c}, T)}{T}\}$$

and it is thus sufficient to specify the bulk free energy density $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ in order to specify $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$. Two examples for such a free energy density considered in [52] consist of an ideal-solution type model

$$f_{id}(\boldsymbol{\phi}, \boldsymbol{c}, T) = \sum_{\alpha=1}^{N} \sum_{i=1}^{K} \left(c_i L_i^\alpha \frac{T - T_i^\alpha}{T_i^\alpha} h^\alpha(\boldsymbol{\phi}) \right) + \sum_{i=1}^{K} \left(\frac{R}{v_m} T c_i \ln(c_i) \right) - c_v T(\ln(T) - 1) \tag{7.1}$$

and a subregular solution model, which, in its simplest form, reduces to the Redlich-Kister-type regular solution free energy density

$$f_r(\boldsymbol{\phi}, \boldsymbol{c}, T) = f_{id}(\boldsymbol{\phi}, \boldsymbol{c}, T) + \sum_{\alpha=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{K} c_i c_j A_{ij}^\alpha h^\alpha(\boldsymbol{\phi}). \tag{7.2}$$

Here, L_i^α and T_i^α are the latent heats per unit volume and the melting temperatures respectively, R_g the gas constant, v_m the (constant) molar volume and c_v the specific heat per volume. In addition, the $\boldsymbol{A}^\alpha = (A_{ij}^\alpha)_{1 \le i,j \le K}$ in Equation (7.2) represent a set of binary interaction coefficients for each phase α.

In relation with the discussion above, the following observation can be made: Taking $\boldsymbol{\phi} = \boldsymbol{e}^\alpha$, the free energy density within any bulk phase α is of the form

$$f_{id}^\alpha(\boldsymbol{c}, T) = c_i L_i^\alpha \frac{T - T_i^\alpha}{T_i^\alpha} + \frac{R}{v_m} T c_i \ln(c_i) - c_v^\alpha T(\ln(T) - 1),$$

and the above definition of the free energy density in Equation (7.1) therefore corresponds to the simple weighted average $f_{id}(\boldsymbol{\phi}, \boldsymbol{c}, T) = \sum_{\alpha=1}^{N} f_{id}^\alpha(\boldsymbol{c}, T) h^\alpha(\boldsymbol{\phi})$ of the ones in the bulk-phases (a similar conclusion obviously also being valid for the regular solution case). In terms of the distinction between the phase-specific and average concentrations above, and, a priori, an analogous distinction between the phase-specific temperatures T^α and the average temperature T, this could also be interpreted as being the weighted average $f_{id}(\boldsymbol{\phi}, \boldsymbol{c}, T) = \sum_{\alpha=1}^{N} f_{id}^\alpha(\boldsymbol{c}^\alpha, T^\alpha) h^\alpha(\boldsymbol{\phi})$ with the additional assumption that, within the interface region, one has $\boldsymbol{c}^\alpha = \boldsymbol{c}$ and $T^\alpha = T$ for all phases α.
Whereas this type of model was quite successful for purely temperature dependent problems, it was realized in e.g. the works of [42] and [56] that by replacing the simple assumption on the phase-specific concentrations by an energetically more favorable one, one could significantly

reduce model artefacts within the interface region. In fact, whereas the assumption of equal temperatures is compatible with the the global equilibrium condition corresponding to the equilibration of the temperature over the entire domain (and thus in particular the equilibrium of the temperatures T^α in each of the bulk-phases), this is not the case for the assumption on the concentration fields. Instead, for the entropy functional above, the equilibrium condition for the evolution of the (average) concentration field is **not** given by the equilibration of the concentrations. Instead, it is related to the chemical driving forces $\frac{\partial \mathcal{S}}{\partial \boldsymbol{c}} = -\frac{1}{T}\frac{\partial f}{\partial \boldsymbol{c}} = -\frac{\boldsymbol{\mu}}{T}$ and thus, in combination with the equilibrium condition $T = \text{const}$ on the temperature, ultimately to the condition of the equilibration (in the appropriate sense[2]) of the chemical potentials.
This has motivated replacing the assumption of the equality of the phase-specific concentrations with a condition based upon the equality of the (reduced) chemical potentials. This condition in combination with phase-specific concentrations was first used explicitly in a two-phase setting by Kim, Kim and Suzuki in [42] and has subsequently been extended to the multiphase setting and reanalyzed in a number of papers, including [25], [56] and [19]. As first discussed in [25], this so-called called **local quasi-equilibrium** condition - which will be discussed in more detail in the next section - can in fact be interpreted as defining the free energy function in such a way that it minimizes the local free energy density $\sum_{\alpha=1}^{N} f^\alpha(\boldsymbol{c}^\alpha)h^\alpha(\boldsymbol{\phi})$ subject to a given phase-field vector $\boldsymbol{\phi}$ and average concentration $\boldsymbol{c}$. Based upon this interpretation, it is quite natural that this model will behave more favorably with respect to the issue of excess interface energy contributions, thus explaining its large popularity despite its additional complexity.

Remark 91. With regards to the literature, there are two important points which should be stressed here.
Firstly, there is a subtle but crucial difference between the earlier work [41] and the subsequent seminal paper [42] by Kim, Kim and Suzuki. Whereas the latter may be considered as the motivation and basis for the subsequent extensions by the various authors mentioned above, the analysis in the former one is, despite its large formal similarity with the discussion in this section in terms of its reliance upon implicit functions, actually much more closely related to the analysis of a variational inconsistency (see Remark 97). Secondly, despite the recurrent appearance of the grand potential energy densities Ψ^α below (a point which has been amply discussed in the literature), the underlying variational principle - at least for the natural isolating boundary conditions on the temperature and concentration field (resp. the free energy in the isothermal case) - is still given by the maximization of the entropy (resp. minimization of the free energy). While the Ψ^α appear quite naturally in this model and there is indeed a close link between the total grand potential energy density Ψ and the free energy defined in equatin (7.3), these are not the same and should not be confused. ◇

7.1.2 The "More Advanced" Free Energy Model in the Isothermal Case

Before discussing some practical implications and possible extensions in the following sections, the basic idea underlying the more advanced free energy model in the isothermal case will be outlined below. Even though the model can be considered to have originated in the work [42], the discussion will mostly follow the approach in [25], which, in the authors opinion, provides a much clearer interpretation of the resulting model.
This idea is in fact remarkably simple. As already indicated above, standard phasefield models for chemically driven solidification problems are based on using a single, "average" concentration field $\boldsymbol{c}$ only. Within interface regions with several coexistent phases, there is a priori no reason to assume for the concentrations to be the same in all phases. It is therefore - in line with the

[2] There is a slight complication in the formulation of this equilibrium condition when based upon a model using the full concentration vector $\boldsymbol{c}$ together with the sum-constraint $\sum_{i=1}^{K} c_i = 1$ instead of the more common formulation based upon a reduced concentration vector $\tilde{\boldsymbol{c}}$ consisting e.g. only out of the first $K-1$ components, the last one then being, if required, recovered as a byproduct of the sum constraint. This basically technical point will be discussed in more detail below.

interpretation of the $\boldsymbol{\phi}$-values (resp. the values $h^\alpha(\boldsymbol{\phi})$) as the volume-fraction of a given phase - quite natural to interpret the total concentration as a weighted averages of the ones in the individual phases, i.e. to postulate $\boldsymbol{c} = \sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi})$ and similarly for the total free energy density to be given by $f(\phi, \boldsymbol{c}, T) = \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi})$ as the result of the contributions by the phase-specific free energy densities f^α within the indivual phases.
These hypotheses alone do not yet fully specify the model, as, for given values of the parameters $\boldsymbol{\phi}$, $\boldsymbol{c}$ and T, there are many ways the total concentration could be distributed to the individual phases while still satisfying the constraint on their average. The simplest assumption to make - leading to models of the type in [52] - is to assume that $\boldsymbol{c}^\alpha = \boldsymbol{c}$ for all phases α. For a chemically isolated system, the underlying variational principle in terms of the minimization of the free energy functional $\mathcal{F}_\epsilon(\boldsymbol{\phi}, \boldsymbol{c}, T)$ is the redistribution of the initially present total concentration $\int_\Omega \boldsymbol{c}\, \mathrm{d}\boldsymbol{x}$ in the energetically most favorable manner. The simple but elegant idea underlying the more quantitative models is to reuse this same principle for determining the phase-specific concentrations in terms of the average one, i.e. to redistribute the (given) average concentration onto the individual phases in such a manner as to minimize the total local free energy density $\sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi})$.

Remark 92. From a physical point of view, it is clear that one would also like to, in addition to the condition of the concentration average, enforce the conditions $0 < c_i < 1$ and $0 < c_i^\alpha < 1$ on the concentration values. An important simplification as compared to the case of the phasefield equation arises from the fact that one can usually safely ignore the 0-1-bounds on the concentration values, i.e. given a total concentration satisfying $0 < c_i < 1$, one will automatically have $0 < c_i^\alpha < 1$ without having to enforce this condition explicitly. For example in the case of the ideal and regular solutions above, this will be ensured by the logarithmic contributions, whose derivatives tend to infinity as a concentration value approaches 0 or 1.
In other cases, such as e.g. when using a (significantly cheaper) quadratic approximation of the free energy contributions, this need not be the case anymore though. The same of course also holds when, due to e.g. an ill-suited time-stepping, the average concentration $\boldsymbol{c}$ falls out of these bounds. As such cases are usually due to a failure of some other part of the approximation scheme, the box-constraints will (except for some hints in the footnotes) nevertheless be systematically neglected in the following as this, in contrast to the discussion in the phasefield case with obstacle potential, adds quite some complexity with a very limited benefit. ◇

In a more explicit manner, the f-function is thus given as the solution of the parameterized minimization problem

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) := \min_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_\alpha f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\}, \tag{7.3}$$

where the admissible set $\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c})$ is, for any vector $\boldsymbol{c}$ satisfying $\sum_{i=1}^K c_i = 1$, given by[3]

$$\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}) = \left\{ (\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} : \sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c} \right\}. \tag{7.4}$$

Alternatively, one can integrate the sum-constraint $\sum_{i=1}^K c_i^\alpha = 1$ directly by eliminating one component (here e.g. the last one) as a function of the others. In terms of the **reduced**

[3] If the sum-constraint on the average concentration is not satisfied, the admissible set is obviously empty as it is then not possible to obtain $\boldsymbol{c}$ as an average of phase-specific concentrations $\boldsymbol{c}^\alpha$ satisfying this constraint.
If one in addition wants to maintain the box-constraints on the concentration values, this admissible set would have to be modified to

$$\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}) = \left\{ (\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} : \sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c},\ \sum_{i=1}^K c_i^\alpha = 1\ \forall\alpha,\ 0 \leq c_i^\alpha, i = 1, 2, ..., K, \forall\alpha \right\},$$

which, as in the phasefield case, will automatically guarantee $c_i^\alpha \leq 1$ due to the sum-constraint on the c_i^α.

concentration vectors $\tilde{\boldsymbol{c}}$ and $(\tilde{\boldsymbol{c}}^\alpha)_{1\le\alpha\le N}$ consisting of the (then independent) first $K-1$-components, i.e. $\tilde{\boldsymbol{c}} = (c_1, c_2, ..., c_{K-1})$ and similarly for the $\boldsymbol{c}^\alpha$, the minimization problem (7.3) can alternatively be expressed as

$$\tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) := \min_{(\tilde{\boldsymbol{c}}^\alpha)_{1\le\alpha\le N} \in \tilde{\mathcal{A}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}})} \left\{ \sum_\alpha \tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\}, \tag{7.5}$$

where $\tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) := f^\alpha\big((\tilde{\boldsymbol{c}}^\alpha, 1 - \sum_{i=1}^{K-1} c_i^\alpha), T\big)$ and the admissible set $\tilde{\mathcal{A}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}})$ is given by[4]

$$\tilde{\mathcal{A}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}) = \left\{ (\tilde{\boldsymbol{c}}^\alpha)_{1\le\alpha\le N} : \sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}} \right\}. \tag{7.6}$$

An important prelimiary observation for the free energy functions defined in Equation (7.3) resp. Equation (7.5) is given in the following

Lemma 6. *Provided the f^α are (strictly) convex functions of the phase-specific concentrations $\boldsymbol{c}^\alpha$, and the $\big(h^\alpha(\boldsymbol{\phi})\big)_{1\le\alpha\le N}$ represent convex weighting coefficients (i.e. satisfy $0 \le h^\alpha \le 1$, $\sum_{\alpha=1}^N h^\alpha(\boldsymbol{\phi}) = 1$), f is also a (strictly) convex function of $\boldsymbol{c}$.*
The same conclusion also holds with $\tilde{f}$ and $\tilde{\boldsymbol{c}}$ replacing f and $\boldsymbol{c}$.

Proof. This can be seen by essentially the same argument as for the convexity of inf(imal)-convolutions (see e.g. [8]). Given a convex combination $\boldsymbol{c} = \lambda\boldsymbol{c}_1 + (1-\lambda)\boldsymbol{c}_2$, $0 \le \lambda \le 1$ of the concentration vectors $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, any convex combination $\boldsymbol{c}^\alpha = \lambda\boldsymbol{c}_1^\alpha + (1-\lambda)\boldsymbol{c}_2^\alpha$ with $\big(\boldsymbol{c}_1^\alpha\big)_{1\le\alpha\le N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_1)$ and $\big(\boldsymbol{c}_2^\alpha\big)_{1\le\alpha\le N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_2)$ satisfies

$$\sum_\alpha \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \sum_\alpha \big(\lambda\boldsymbol{c}_1^\alpha + (1-\lambda)\boldsymbol{c}_2^\alpha\big) h^\alpha(\boldsymbol{\phi}) = \lambda\big(\sum_\alpha \boldsymbol{c}_1^\alpha h^\alpha(\boldsymbol{\phi})\big) + (1-\lambda)\big(\sum_\alpha \boldsymbol{c}_2^\alpha h^\alpha(\boldsymbol{\phi})\big) = \lambda\boldsymbol{c}_1 + (1-\lambda\boldsymbol{c}_2) = \boldsymbol{c},$$

$$\sum_{i=1}^K c_i^\alpha = \sum_{i=1}^K \big(\lambda(\boldsymbol{c}_1)_i^\alpha + (1-\lambda)(\boldsymbol{c}_2)_i^\alpha = \lambda\sum_{i=1}^K (\boldsymbol{c}_1)_i^\alpha + (1-\lambda)\sum_{i=1}^K (\boldsymbol{c}_2)_i^\alpha = \lambda + (1-\lambda) = 1$$

and $0 \le c_i^\alpha$, and thus $\lambda\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_1) + (1-\lambda)\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_2) \subset \mathcal{A}(\boldsymbol{\phi}, \lambda\boldsymbol{c}_1 + (1-\lambda)\boldsymbol{c}_2)$. Using this set-inclusion and the assumed convexity of f^α in terms of the $\boldsymbol{c}^\alpha$ in the definition (7.3) shows that

$$\begin{aligned}
f(\boldsymbol{\phi}, \boldsymbol{c}, T) &= \min_{\{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N} \in \mathcal{A}(\boldsymbol{\phi}, \lambda\boldsymbol{c}_1 + (1-\lambda)\boldsymbol{c}_2)\}} \left\{ \sum_\alpha f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\} \\
&\le \min_{\{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N} \in \lambda\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_1) + (1-\lambda)\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_2)\}} \left\{ \sum_\alpha f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\} \\
&= \min_{\{(\boldsymbol{c}_1^\alpha, \boldsymbol{c}_2^\alpha)_{1\le\alpha\le N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_1) \times \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_2)\}} \left\{ \sum_\alpha f^\alpha(\lambda\boldsymbol{c}_1^\alpha + (1-\lambda)\boldsymbol{c}_2^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\} \\
&\le \min_{\{(\boldsymbol{c}_1^\alpha, \boldsymbol{c}_2^\alpha)_{1\le\alpha\le N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_1) \times \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}_2)\}} \left\{ \sum_\alpha \Big(\lambda f^\alpha(\boldsymbol{c}_1^\alpha, T) + (1-\lambda) f^\alpha(\boldsymbol{c}_2^\alpha, T)\Big) h^\alpha(\boldsymbol{\phi}) \right\}.
\end{aligned}$$

[4] Now, maintaining the box-constraints would require modifying $\tilde{\mathcal{A}}$ to

$$\tilde{\mathcal{A}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}) = \left\{ (\tilde{\boldsymbol{c}}^\alpha)_{1\le\alpha\le N} : \sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}},\ 0 \le c_i^\alpha \le 1, i = 1, 2, ..., K-1, \forall\alpha \right\},$$

where, this time, it is necessary to enforce the upper bound of 1 on the c_i^α as the sum-constraints $\sum_{i=1}^K c_i^\alpha = 1$ has disappeared.

As this last minimization problem is separable[5], it can equivalently be replaced by

$$\lambda \underbrace{\min_{\{(\boldsymbol{c}_1^\alpha)_{1\le\alpha\le N}\in\mathcal{A}(\boldsymbol{\phi},\boldsymbol{c}_1)\}}\left\{\sum_\alpha f^\alpha(\boldsymbol{c}_1^\alpha,T)h^\alpha(\boldsymbol{\phi})\right\}}_{=f(\boldsymbol{\phi},\boldsymbol{c}_1,T)} + (1-\lambda)\underbrace{\{\min_{(\boldsymbol{c}_2^\alpha)_{1\le\alpha\le N}\in\mathcal{A}(\boldsymbol{\phi},\boldsymbol{c}_2)\}}\left\{\sum_\alpha f^\alpha(\boldsymbol{c}_2^\alpha,T)h^\alpha(\boldsymbol{\phi})\right\}}_{=f(\boldsymbol{\phi},\boldsymbol{c}_2,T)}$$

and thus $f(\boldsymbol{\phi},\boldsymbol{c},T) \le \lambda f(\boldsymbol{\phi},\boldsymbol{c}_1,T) + (1-\lambda)f(\boldsymbol{\phi},\boldsymbol{c}_2,T)$ and the mapping $\boldsymbol{c}\mapsto f(\boldsymbol{\phi},\boldsymbol{c},T)$ is indeed convex. □

As discussed in Subsection 5.1, the FONC for the minimizer in Equation (7.3) can e.g. be obtained by considering the Lagrangian[6]

$$L\left(\boldsymbol{\phi},\boldsymbol{c},T,(\boldsymbol{c}^\alpha)_{1\le\alpha\le N},\boldsymbol{\mu},\hat{\boldsymbol{\lambda}}\right) = \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha,T)h^\alpha(\boldsymbol{\phi}) - \boldsymbol{\mu}\cdot\Big(\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) - \boldsymbol{c}\Big) - \sum_{\alpha=1}^N \hat{\lambda}^\alpha\Big(\sum_{i=1}^K c_i^\alpha - 1\Big)$$

with $\hat{\boldsymbol{\lambda}}\in\mathbb{R}^N$. $\boldsymbol{\mu}$ could in principle be taken as an arbitrary vector in $\mathbb{R}^K$, but, since the sum-constraint $\sum_{\alpha=1}^N c_i^\alpha h^\alpha(\boldsymbol{\phi}) = c_i$ is in fact automatically fulfilled for all K components as soon as it holds for $K-1$ components[7], this would lead to an indeterminacy as $\boldsymbol{\mu}$ would be capable of "testing" a property which is already ensured through the mulitpliers $(\lambda^\alpha)_{1\le\alpha\le N}$. In order to avoid this issue, it may seem preferable to choose $\boldsymbol{\mu}$ in a suitable $(K-1)$-dimensional subspace of $\mathbb{R}^K$. One possibility here, closely related to the reduced formulation, is to choose $\boldsymbol{\mu}$ such that $\mu_K = 0$. A second, more "symmetric" choice is obtained by enforcing a zero average on $\boldsymbol{\mu}$, i.e. by requiring that $\sum_{i=1}^K \mu_i = 0$. Alternatively, and this is in many ways the most natural choice, one may also fix $\boldsymbol{\mu}$ "indirectly" by simply requiring that $\boldsymbol{\mu} = \sum_{\alpha=1}^N \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha} h^\alpha(\boldsymbol{\phi})$, which in particular enforces $\sum_{\alpha=1}^N \lambda^\alpha h^\alpha(\boldsymbol{\phi}) = 0$.

Remark 93. Each choice clearly leads to a different, though closely related, set of multipliers $(\boldsymbol{\mu},\lambda^\alpha)$ as will be discussed in Subsection 7.1.5 below. The advantage of the last choice is that it carries, in terms of the average values, the standard role of the sensitivity with respect to the sum-constraint on the total concentration. The first two choices on the other hand "shift" this responsibility to the $\boldsymbol{c}^\alpha$, therefore emphasising the role of the λ^α on the effective equation. This has no influence on the actual result, since it does not matter how the average sum-constraint on $\boldsymbol{c}$ is ultimately enforced, but favors the last description on "aesthetic" grounds. ◇

[5] It consists in the minimization of two sums of functions depending on independent arguments in independent sets. There is therefore no "competition" between the two summands, and the minimizer is obviously achieved when each of them is minimal individually.

[6] Together with the box-constraint, this needs to be modified to

$$L\left(\boldsymbol{\phi},\boldsymbol{c},T,(\boldsymbol{c}^\alpha)_{1\le\alpha\le N},\boldsymbol{\mu},\hat{\boldsymbol{\lambda}},\boldsymbol{\zeta}\right) = \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha,T)h^\alpha(\boldsymbol{\phi}) - \boldsymbol{\mu}\cdot\Big(\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) - \boldsymbol{c}\Big) - \sum_{\alpha=1}^N \hat{\lambda}^\alpha\Big(\sum_{i=1}^K c_i^\alpha - 1\Big) - \sum_{\alpha=1}^N\sum_{i=1}^K \zeta_i^\alpha c_i^\alpha$$

with $\hat{\boldsymbol{\lambda}}\in\mathbb{R}^N$ and $\boldsymbol{\zeta}\in\mathbb{R}_+^{N\times K}$.

[7] Assuming for example that it holds for the first $K-1$ components and as the concentrations are required to sum to one, one has

$$\sum_{\alpha=1}^N c_K^\alpha h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^N (1 - \sum_{i=1}^{K-1} c_i^\alpha) h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^N h^\alpha(\boldsymbol{\phi}) - \sum_{i=1}^{K-1} c_i^\alpha h^\alpha(\boldsymbol{\phi}) = 1 - \sum_{i=1}^{K-1} c_i = c_K.$$

In constrast, the modified Lagrangian for the reduced problem (7.5) is given by[8]

$$\tilde{L}\Big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T, \big(\tilde{\boldsymbol{c}}^{\alpha}\big)_{1\leq\alpha\leq N}, \tilde{\boldsymbol{\mu}}\Big) = \sum_{\alpha=1}^{N} f^{\alpha}(\tilde{\boldsymbol{c}}^{\alpha}, T) h^{\alpha}(\boldsymbol{\phi}) - \tilde{\boldsymbol{\mu}} \cdot \big(\sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) - \tilde{\boldsymbol{c}}\big) \tag{7.8}$$

where now $\tilde{\boldsymbol{\mu}} \in \mathbb{R}^{K-1}$, and the multiplier $\hat{\boldsymbol{\lambda}}$ for the phase-specific sum-constraints has vanished. As both definitions (7.3) and (7.5) only differ in the representation of the minimization problem, it is clear that the numerical values of $\tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ and $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ will actually agree for any admissible concentration, i.e. one has $f\big(\boldsymbol{\phi}, (\tilde{\boldsymbol{c}}, 1 - \sum_{i=1}^{K-1} c_i), T\big) = \tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$. Nevertheless, the different notations for both problems will be maintained in order to simplify the distinction between the two approaches.

Differentiation of L w.r.t. to the phase-specific concentration vectors $\boldsymbol{c}^{\alpha}$ leads to the system of equations

$$h^{\alpha}(\boldsymbol{\phi})\Big(\frac{\partial f^{\alpha}}{\partial \boldsymbol{c}^{\alpha}}(\boldsymbol{c}^{\alpha}, T) - \boldsymbol{\mu}\Big) - \hat{\lambda}^{\alpha}\boldsymbol{e} = h^{\alpha}(\boldsymbol{\phi})\Big(\boldsymbol{\mu}^{\alpha} - \boldsymbol{\mu}\Big) - \hat{\lambda}^{\alpha} = \boldsymbol{0},$$

together with the sum-constraint $\sum_{\alpha=1}^{N} \boldsymbol{c}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) = \boldsymbol{c}$, where $\boldsymbol{\mu}^{\alpha} := \frac{\partial f^{\alpha}}{\partial \boldsymbol{c}^{\alpha}}(\boldsymbol{c}^{\alpha}, T)$ and $\boldsymbol{e} = (1, 1, ..., 1)$ is a K-dimensional vector of ones. Similarly, with $\tilde{\boldsymbol{\mu}}^{\alpha} := \frac{\partial \tilde{f}^{\alpha}}{\partial \tilde{\boldsymbol{c}}^{\alpha}}(\tilde{\boldsymbol{c}}^{\alpha}, T)$, the FONC for a minimizer of the reduced function in Equation (7.5) is given by the set of equations

$$h^{\alpha}(\boldsymbol{\phi})\Big(\frac{\partial \tilde{f}^{\alpha}}{\partial \tilde{\boldsymbol{c}}^{\alpha}}(\tilde{\boldsymbol{c}}^{\alpha}, T) - \tilde{\boldsymbol{\mu}}\Big) = h^{\alpha}(\boldsymbol{\phi})\Big(\tilde{\boldsymbol{\mu}}^{\alpha} - \tilde{\boldsymbol{\mu}}\Big) = \boldsymbol{0}$$

together with $\sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}}$.

A first observation to be made is that in both cases the conditions on the derivatives of the phase-specific f-functions **only** pertain to those phases for which $h^{\alpha}(\boldsymbol{\phi}) \neq 0$, whereas the $\boldsymbol{c}^{\alpha}$ (resp. $\tilde{\boldsymbol{c}}^{\alpha}$) for the phases with $h^{\alpha}(\boldsymbol{\phi}) = 0$ can in principle be chosen arbitrarily within the respective admissible sets[9] without affecting either the value of f or $\sum_{\alpha=1}^{N} h^{\alpha}(\boldsymbol{\phi})\boldsymbol{c}^{\alpha}$ (resp. the reduced versions thereof). In the non-reduced case, the trivial choice $\lambda^{\alpha} = 0$ is then clearly an admissible choice for the multiplier λ^{α} for all vanishing phases, such that it suffices to focus only on those phases with $h^{\alpha}(\boldsymbol{\phi}) \neq 0$[10].

Defining the set $\mathcal{P}_p := \{\alpha \in 1, 2, ..., N : h^{\alpha}(\boldsymbol{\phi}) \neq 0\}$ of phases which are actually "present" at a given point, and for convenience scaling $\hat{\lambda}^{\alpha}, \alpha \in \mathcal{P}_R$ by $h^{\alpha}(\boldsymbol{\phi}) \neq 0$, it is therefore in practice

[8] Or, in the case with box-constraints, by

$$\begin{aligned}\tilde{L}\Big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T, \big(\tilde{\boldsymbol{c}}^{\alpha}\big)_{1\leq\alpha\leq N}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\zeta}}^{-}, \tilde{\boldsymbol{\zeta}}^{+}\Big) = &\sum_{\alpha=1}^{N} f^{\alpha}(\tilde{\boldsymbol{c}}^{\alpha}, T) h^{\alpha}(\boldsymbol{\phi}) - \tilde{\boldsymbol{\mu}} \cdot \big(\sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) - \tilde{\boldsymbol{c}}\big) \\ &- \sum_{\alpha=1}^{N} \sum_{i=1}^{K-1} (\tilde{\zeta}^{-})_i^{\alpha} c_i^{\alpha} - \sum_{\alpha=1}^{N} \sum_{i=1}^{K-1} (\tilde{\zeta}^{+})_i^{\alpha} (1 - c_i^{\alpha})\end{aligned} \tag{7.7}$$

with $\tilde{\boldsymbol{\mu}} \in \mathbb{R}^{K-1}$, $\tilde{\boldsymbol{\zeta}}^{\pm} \in \mathbb{R}_{+}^{N\times(K-1)}$. The introduction of two sets of multipliers $\tilde{\boldsymbol{\zeta}}^{\pm}$ is necessary as the inequality $c_i^{\alpha} \leq 1$ is not automatically ensured by the combination of the positivity and sum constraint anymore.

[9] This will be discussed in a little more detail in Subsection 7.1.3 when considering the derivative w.r.t. $\boldsymbol{\phi}$.

[10] Here, one could again include the box-constraints by replacing the FONC in the non-reduced case to

$$h^{\alpha}(\boldsymbol{\phi})\Big(\frac{\partial f^{\alpha}}{\partial \boldsymbol{c}^{\alpha}}(\boldsymbol{c}^{\alpha}, T) - \boldsymbol{\mu}\Big) - \hat{\lambda}^{\alpha}\boldsymbol{e} - \boldsymbol{\zeta}^{\alpha} = h^{\alpha}(\boldsymbol{\phi})\Big(\boldsymbol{\mu}^{\alpha} - \boldsymbol{\mu}\Big) - \hat{\lambda}^{\alpha} - \boldsymbol{\zeta}^{\alpha} = \boldsymbol{0},$$

which then, in addition to the sum constraint, needs to be complemented by the box-constraints as well as $\zeta_i^{\alpha} \geq 0$ and the complementarity condition $\zeta_i^{\alpha} c_i^{\alpha} = 0$. Similarly, the reduced formulation would lead to the condition

$$h^{\alpha}(\boldsymbol{\phi})\Big(\frac{\partial \tilde{f}^{\alpha}}{\partial \tilde{\boldsymbol{c}}^{\alpha}}(\tilde{\boldsymbol{c}}^{\alpha}, T) - \tilde{\boldsymbol{\mu}}\Big) - (\boldsymbol{\zeta}^{-})^{\alpha} + (\boldsymbol{\zeta}^{+})^{\alpha} = h^{\alpha}(\boldsymbol{\phi})\Big(\tilde{\boldsymbol{\mu}}^{\alpha} - \tilde{\boldsymbol{\mu}}\Big) - (\boldsymbol{\zeta}-)^{\alpha} + (\boldsymbol{\zeta}^{+})^{\alpha} = \boldsymbol{0}$$

and the additional restrictions $(\zeta^{\pm})_i^{\alpha} \geq 0$ and $(\zeta^{-})_i^{\alpha} c_i^{\alpha} = 0$, $(\zeta^{+})_i^{\alpha}(1 - c_i^{\alpha}) = 0$.
For phases with $h^{\alpha}(\boldsymbol{\phi}) = \boldsymbol{0}$, admissible choices for the multipliers would then be given by $(\lambda^{\alpha}, \boldsymbol{\zeta}^{\alpha}) = (0, \boldsymbol{0})$ (resp. $\boldsymbol{\zeta}^{-} = \boldsymbol{\zeta}^{+} = \boldsymbol{0}$).

sufficient to consider the system

$$\begin{cases} \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}(\boldsymbol{c}^\alpha, T) - \boldsymbol{\mu} - \lambda^\alpha \boldsymbol{e} = \boldsymbol{\mu}^\alpha(\boldsymbol{c}^\alpha, T) - \boldsymbol{\mu} - \lambda^\alpha \boldsymbol{e} = \boldsymbol{0}, & \alpha \in \mathcal{P}_p, \\ \sum_{i=1}^K c_i^\alpha = 1, & \alpha \in \mathcal{P}_p, \\ \sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c} \end{cases} \tag{7.9}$$

with $\lambda^\alpha = \frac{\hat{\lambda}^\alpha}{h^\alpha(\boldsymbol{\phi})}$ in the non-reduced case and

$$\begin{cases} \frac{\partial \tilde{f}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}(\tilde{\boldsymbol{c}}^\alpha, T) - \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) - \tilde{\boldsymbol{\mu}} = \boldsymbol{0}, & \alpha \in \mathcal{P}_p, \\ \sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}} \end{cases} \tag{7.10}$$

in the reduced case.
The system (7.10) with the **reduced chemical potential** or **diffusion potential** $\tilde{\boldsymbol{\mu}}$ is (except for the temporary restriction to $\alpha \in \mathcal{P}_p$) precisely the same condition also used in the models by [42], [25], [56] and [19]. As

$$\frac{\partial \tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T)}{\partial \tilde{\boldsymbol{c}}^\alpha} = \frac{\partial f^\alpha\big((\tilde{\boldsymbol{c}}^\alpha, 1 - \sum_{i=1}^{K-1} c_i^\alpha), T\big)}{\partial \tilde{\boldsymbol{c}}^\alpha} = \Big(\frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_i^\alpha}\Big)_{1 \le i \le K-1} - \frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_K^\alpha} \tilde{\boldsymbol{e}}, \tag{7.11}$$

where $\tilde{\boldsymbol{e}} = (\boldsymbol{e})_{1 \le i \le K-1}$, it follows from equations (7.10) and (7.9) that $\tilde{\boldsymbol{\mu}}$ can be obtained from $(\boldsymbol{\mu}, \lambda^\alpha)$ - independent of the specific phase α and the particular $(K-1)$-dimensional subspace chosen for $\boldsymbol{\mu}$ - through the relation

$$\begin{aligned} \tilde{\mu}_i = & \tilde{\mu}_i^\alpha = \frac{\partial \tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T)}{\partial c_i^\alpha} = \frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_i^\alpha} - \frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_K^\alpha} = \mu_i^\alpha - \mu_K^\alpha \\ = & \big(\mu_i + \lambda^\alpha\big) - \big(\mu_K + \lambda^\alpha\big) = \mu_i - \mu_K. \end{aligned} \tag{7.12}$$

Conversely, as $\frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_K^\alpha} = \mu_K + \lambda^\alpha$, the solution satisfies[11]

$$\frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_i^\alpha} = \mu_i + \lambda^\alpha = \mu_i + \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_K^\alpha} - \mu_K = \tilde{\mu}_i + \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_K^\alpha}.$$

In the simplest case $\boldsymbol{\mu}$ is, as suggested above, chosen such that $\mu_K = 0$, in which case the previous equation implies that

$$\lambda^\alpha = \frac{\partial f^\alpha}{\partial c_K^\alpha} \quad \text{and} \quad \mu_i = \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_i^\alpha} - \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_K^\alpha}. \tag{7.13}$$

Alternatively, for the choice $\sum_{i=1}^K \mu_i = 0$, a summation over all components in the first equation of (7.9) shows that, similar to the multiplier for the sum-constraint in the phasefield case,

$$\lambda^\alpha = \frac{1}{K} \sum_{i=1}^K \frac{\partial f^\alpha}{\partial c_i^\alpha} \quad \text{and} \quad \mu_i = \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_i^\alpha} - \frac{1}{K} \sum_{i=1}^K \frac{\partial f^\alpha(\boldsymbol{c}, T)}{\partial c_i^\alpha}. \tag{7.14}$$

Remark 94. It is important to stress again that the optimality systems (7.9) and (7.10) a priori only apply to those phases with $h^\alpha(\boldsymbol{\phi}) \neq 0$ whereas the values for the phases with $h^\alpha(\boldsymbol{\phi}) = 0$ are a priori arbitrary as they affect neither the weighted average of the phase-specific free energies nor that of the average concentration. Nevertheless, as some of the following calculations will require the values of the $\boldsymbol{c}^\beta$ (resp. $\tilde{\boldsymbol{c}}^\beta$) as well as the corresponding chemical potentials, the

[11] Note that, in contrast to the **differences** between the derivatives, the derivatives $\frac{\partial f^\alpha(\boldsymbol{c}^\alpha, T)}{\partial c_i^\alpha}$ themselves actually do depend upon the phase α.

standing assumption will be made that these are determined by the "natural" extension of the optimality condition above, i.e. by imposing the same optimality conditions (7.9) resp. (7.10) for all phases and not only those strictly required for actually solving the optimization problems. Despite this convention, one should still maintain the distinction between the $\mathcal{P}_p$ and the set of all phases, since it is much more convenient to a prior only solve the actual optimality systems above and then recover the $\tilde{\boldsymbol{c}}^\beta$ for all phases in $\{1,...,N\} \setminus \mathcal{P}_p$ only when actually required from the then known value of $\boldsymbol{\mu}$ (or $\tilde{\boldsymbol{\mu}}$). $\diamond$

7.1.3 The Resulting Driving Force

In constrast to the simpler models based on the "common" phase-specific concentrations $\boldsymbol{c}^\alpha = \boldsymbol{c}\ \forall \alpha$ as in e.g. [52], the evaluation of the driving force contribution $\frac{\partial f}{\partial \phi^\alpha}$ to the phasefield equation is somewhat more complex for the model above. The primary difference is that, whereas in the former case, the phase-specific concentrations are (for a given concentration $\boldsymbol{c}$) independent upon the phasefield-values and one thus has $\frac{\partial f}{\partial \phi^\alpha} = \sum_{\beta=1}^N f^\beta(\boldsymbol{c},T)\frac{\partial h^\beta}{\partial \phi^\alpha}$, the phase-specific concentration $\boldsymbol{c}^\alpha$ in the latter case depend upon the parameters of the optimization problem (7.3), i.e. $\boldsymbol{c}^\beta = \boldsymbol{c}^\beta(\boldsymbol{\phi},\boldsymbol{c},T)$. This will lead to an additional contribution to $\frac{\partial f}{\partial \phi^\alpha}$ related to $\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}$.
More precisely, one now has[12]

$$\frac{\partial f(\boldsymbol{\phi},\boldsymbol{c},T)}{\partial \phi^\alpha} = \frac{\partial}{\partial \phi^\alpha}\sum_{\beta=1}^N f^\beta(\boldsymbol{c}^\beta(\boldsymbol{\phi},\boldsymbol{c},T),T)h^\beta(\boldsymbol{\phi}) = \sum_{\beta=1}^N f^\beta \frac{\partial h^\beta}{\partial \phi^\alpha} + \frac{\partial f^\beta}{\partial \boldsymbol{c}^\beta}\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}).$$

While this would seem to require the ability to evaluate $\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}$, the variational nature of the system implicitly defining the $\boldsymbol{c}^\beta$ actually allows dispensing of this difficulty. In fact, by construction of the $\boldsymbol{c}^\beta$, one has $\frac{\partial f}{\partial \boldsymbol{c}^\beta} = \boldsymbol{\mu} + \lambda^\beta \boldsymbol{e}$ in the solution of Equation (7.9) and thus

$$\sum_{\beta=1}^N \frac{\partial f^\beta}{\partial \boldsymbol{c}^\beta}\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}) = \sum_{\beta=1}^N \left(\boldsymbol{\mu} + \lambda^\beta \boldsymbol{e}\right)\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}).$$

The first factor $\boldsymbol{\mu}$ does not depend upon the phase β though, and can thus be extracted from the sum. The second contribution $\lambda^\beta \boldsymbol{e}\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}$ in contrast vanishes automatically for each phase since $\boldsymbol{e}\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha} = \frac{\partial \boldsymbol{e}\cdot\boldsymbol{c}^\beta}{\partial \phi^\alpha} = \frac{\partial 1}{\partial \phi^\alpha} = 0$ by the sum-constraint on the phase-specific concentrations. Making use of the product rule and the constraint on the concentration average, one can then in a second step actually eliminate the $\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}$ since

$$\begin{aligned}\sum_{\beta=1}^N \frac{\partial f^\beta}{\partial \boldsymbol{c}^\beta}\cdot\frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}) =& \boldsymbol{\mu}\cdot\Big(\sum_{\beta=1}^N \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi})\Big) = \boldsymbol{\mu}\cdot\Big(\frac{\partial\big(\sum_{\beta=1}^N \boldsymbol{c}^\beta h^\beta(\boldsymbol{\phi})\big)}{\partial \phi^\alpha} - \sum_{\beta=1}^N \boldsymbol{c}^\beta\frac{\partial h^\beta(\boldsymbol{\phi})}{\partial \phi^\alpha}\Big)\\ =& \boldsymbol{\mu}\cdot\Big(\frac{\partial \boldsymbol{c}}{\partial \phi^\alpha} - \sum_{\beta=1}^N \boldsymbol{c}^\beta\frac{\partial h^\beta(\boldsymbol{\phi})}{\partial \phi^\alpha}\Big) = -\boldsymbol{\mu}\cdot\Big(\sum_{\beta=1}^N \boldsymbol{c}^\beta\frac{\partial h^\beta(\boldsymbol{\phi})}{\partial \phi^\alpha}\Big)\end{aligned}$$

and thus one actually has, readding the arguments for clarity,

$$\frac{\partial f(\boldsymbol{\phi},\boldsymbol{c},T)}{\partial \phi^\alpha} = \sum_{\beta=1}^N \Big(f^\beta\big(\boldsymbol{c}^\beta(\boldsymbol{\phi},\boldsymbol{c},T),T\big) - \boldsymbol{\mu}(\boldsymbol{\phi},\boldsymbol{c},T)\cdot\boldsymbol{c}^\beta(\boldsymbol{\phi},\boldsymbol{c},T)\Big)\frac{\partial h^\beta}{\partial \phi^\alpha}. \tag{7.15}$$

A similar calculation can also, though in a slightly simpler form, be conducted for the case of the reduced formulation based on $\tilde{\boldsymbol{c}}$. A differentiation with respect to ϕ^α taking into account the

[12] Here the convention in Remark 94 has to be kept in mind if $\frac{\partial h^\beta}{\partial \phi^\alpha}$ does not vanish for $h^\beta(\boldsymbol{\phi}) = 0$, such as for the simplest interpolation function $h^\beta(\boldsymbol{\phi}) = \phi^\beta$.

dependence of the $\tilde{\boldsymbol{c}}^\alpha$ on $\boldsymbol{\phi}$ first leads to $\frac{\partial \tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)}{\partial \phi^\alpha} = \frac{\partial}{\partial \phi^\alpha}\sum_{\beta=1}^N \tilde{f}^\beta(\tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T),T)h^\beta(\boldsymbol{\phi})$. By the characterization of the $\tilde{\boldsymbol{c}}^\beta$ in Equation (7.10), one has one the one hand $\frac{\partial \tilde{f}^\beta}{\partial \tilde{\boldsymbol{c}}^\beta}(\tilde{\boldsymbol{c}}^\beta,T) = \tilde{\boldsymbol{\mu}}$ for all β, such that $\tilde{\boldsymbol{\mu}}$ can again be moved out of the summation. On the other hand, $\sum_{\beta=1}^N \tilde{\boldsymbol{c}}^\beta h^\beta(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}}$, which can, as above, be combined with the product rule to eliminate the $\frac{\partial \tilde{\boldsymbol{c}}^\beta}{\partial \phi^\alpha}$ making use of

$$\tilde{\boldsymbol{\mu}} \cdot \sum_{\beta=1}^N \frac{\partial \tilde{\boldsymbol{c}}^\beta}{\partial \phi^\alpha} h^\beta(\boldsymbol{\phi}) = \tilde{\boldsymbol{\mu}} \cdot \Big(\frac{\partial \big(\sum_{\beta=1}^N \tilde{\boldsymbol{c}}^\beta h^\beta(\boldsymbol{\phi}) \big)}{\partial \phi^\alpha} - \tilde{\boldsymbol{c}}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha} \Big) = \tilde{\boldsymbol{\mu}} \cdot \Big(\frac{\partial \tilde{\boldsymbol{c}}}{\partial \phi^\alpha} - \tilde{\boldsymbol{c}}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha} \Big) = -\tilde{\boldsymbol{\mu}} \cdot \sum_{\beta=1}^N \tilde{\boldsymbol{c}}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$$

thus resulting in the analogue

$$\frac{\partial \tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)}{\partial \phi^\alpha} = \sum_{\beta=1}^N \Big(\tilde{f}^\beta\big(\tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T),T\big) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) \cdot \tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) \Big) \frac{\partial h^\beta}{\partial \phi^\alpha} \tag{7.16}$$

of Equation (7.15).

7.1.4 The Relation with the Grand Chemical Potential

Considering for simplicity the reduced formulation, it can be noted that since $\tilde{\boldsymbol{\mu}} = \frac{\partial \tilde{f}^\beta}{\partial \tilde{\boldsymbol{c}}^\beta}$ (see the discussion below), the contribution of each phase to the driving force in equation (7.16) corresponds precisely to the grand potential energy densities $\tilde{\Psi}^\beta$ evaluated in $\tilde{\boldsymbol{\mu}}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)$. As already indicated in Remark 91, this is a point which has been amply discussed in various forms in the literature (see e.g. [42], [56] and [19]).
One point of view - which is in particular stressed in [56] and [19] - is that it is also possible to obtain the same driving force without having to go through the somewhat more complex dependence in terms of the phase-specific concentrations. This is achieved by replacing the free energy density with the grand potential energy density $\tilde{\Psi}(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T) := \sum_{\alpha=1}^N \tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}},T)h^\alpha(\boldsymbol{\phi})$ with $\tilde{\boldsymbol{\mu}}$ considered as an indepdendent variable (instead of, as above, as $\tilde{\boldsymbol{\mu}}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)$), which suggests using the (total) grand potential energy density $\tilde{\Omega}_\epsilon(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T)$ as the relevant potential instead of the free energy density $\mathcal{F}$. In contrast, both [42] and [25] do not actually introduce the grand potential energy density directly, but simply state that the dependence on the phase-specific concentrations does, as above, lead to this type of driving force for minimizing the (original) free energy density.
Since both approaches result in the same driving force and the evolution of either $\tilde{\boldsymbol{\mu}}$ or $\tilde{\boldsymbol{c}}$ is chosen such that, up to potential differences in the choice of diffusion coefficients[13], $\tilde{\boldsymbol{c}}$ satisfies a locally conservative gradient flow driven by the gradient of the (common) chemical potential, they will in fact lead to the same result. Nevertheless, the use of two a priori quite different potentials raises a number of questions which would not seem to have been discussed in sufficient detail in the literature. The following discussion will therefore try to clarify the link - and the differences in interpretation - between these two approaches.

One can first observe - this being the standard argument underlying the use of the Lagrangian in Equation (7.8) - that the optimization problem defining $\tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)$ can equivalently be rewritten as the unconstrained minimax-problem

$$\tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) = \min_{(\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}} \sup_{\tilde{\boldsymbol{\mu}}} \Big\{ \sum_{\alpha=1}^N \tilde{f}^\alpha(\boldsymbol{\phi},\tilde{\boldsymbol{c}}^\alpha,T)h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{\mu}} \cdot \Big(\sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{c}} \Big) \Big\}.$$

As duality holds here by the strict convexity of the f^α, one may exchange the order of the two

[13] Their choice not being a matter of the underlying functional.

extremum operations, i.e. one can also write

$$\begin{aligned}\tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) &= \sup_{\tilde{\boldsymbol{\mu}}} \min_{(\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}} \Big\{ \sum_{\alpha=1}^N \tilde{f}^\alpha(\boldsymbol{\phi},\tilde{\boldsymbol{c}}^\alpha,T)h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{\mu}}\cdot\Big(\sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{c}}\Big)\Big\} \\ &= \sup_{\tilde{\boldsymbol{\mu}}} \Big\{ \min_{(\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}} \Big\{ \sum_{\alpha=1}^N \Big(\tilde{f}^\alpha(\boldsymbol{\phi},\tilde{\boldsymbol{c}}^\alpha,T) - \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}^\alpha\Big)h^\alpha(\boldsymbol{\phi})\Big\} + \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}\Big\}.\end{aligned}$$

The inner minimization problem decomposes into $\sum_{\alpha=1}^N \min_{\tilde{\boldsymbol{c}}^\alpha}\big\{f^\alpha(\tilde{\boldsymbol{c}}^\alpha,T) - \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}^\alpha\big\}h^\alpha(\boldsymbol{\phi})$, where each of the separate minimization problems in fact corresponds precisely to the definition of $\Psi^\alpha(\tilde{\boldsymbol{\mu}},T)$, and thus $\tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) = \sup_{\tilde{\boldsymbol{\mu}}}\big\{\sum_{\alpha=1}^N \tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}},T)h^\alpha(\boldsymbol{\phi}) + \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}\big\}$.
There are two things which can be noted here. Firstly, in the interior sum, **both** $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{c}}$ are actually independent parameters, and in particular the $\tilde{\Psi}^\alpha$ are indeed the phase-specific grand chemical potential densities $\tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}}^\alpha,T)$ evaluated for $\tilde{\boldsymbol{\mu}}^\alpha = \tilde{\boldsymbol{\mu}}$. Secondly, while this makes appear the average grand chemical potential density

$$\tilde{\Psi}(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T) \coloneqq \sum_{\alpha=1}^N \tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}},T)h^\alpha(\boldsymbol{\phi}), \tag{7.17}$$

it can clearly not be used to simply replace the free energy density $\tilde{f}$ since, on the one hand, there is the additional term $\tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}$, and on the other hand, there is a remaining extremum problem in terms of $\tilde{\boldsymbol{\mu}}$.
What follows though (see Remark 24 and Equation (5.19) in particular) is that the "quantitative" free energy density $\tilde{f}$ as defined above is in fact given by the (inverse) Legendre transform of $\tilde{\Psi}$ with respect to $\tilde{\boldsymbol{\mu}}$. More precisely, denoting (in analogy to Remark 24) the partial forward transform defining the $\tilde{\Psi}^\alpha$ by $\hat{\mathcal{L}}_{\tilde{\boldsymbol{c}}}(\tilde{f}^\alpha)(\tilde{\boldsymbol{\mu}}^\alpha,T)$, the free energy density can also, with the analogous notation, be characterized as

$$\tilde{f}(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) = \hat{\mathcal{L}}^{-1}_{\tilde{\boldsymbol{\mu}}}[\tilde{\Psi}](\boldsymbol{\phi},\tilde{\boldsymbol{c}},T) \coloneqq \sup_{\tilde{\boldsymbol{\mu}}}\big\{\tilde{\Psi}(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T) + \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}\big\}, \tag{7.18}$$

and is thus the inverse Legendre transform (w.r.t. $\tilde{\boldsymbol{\mu}}$) of the average grand chemical potential density for $\tilde{\boldsymbol{\mu}}^\alpha = \tilde{\boldsymbol{\mu}}$.

Remark 95. Note that this relation - while intuitively pleasing based on the standard thermodynamic relations - is neither quite standard nor entirely obvious based upon the initial definitions. Whereas the direct definition of $\tilde{\Psi}$ involves a simple averaging procedure over the phase-specific densities $\tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}},T)$ with a single parameter $\tilde{\boldsymbol{\mu}}$, $\tilde{f}$ is defined through the minimization of an average based on N phase-specific concentrations subject to an additional side condition on their average. In contrast, the $\tilde{\Psi}^\alpha$ themselves are a priori defined in terms of an **unconstrained** parameterized optimization in each of the $\tilde{\boldsymbol{c}}$, $\tilde{\Psi}^\alpha(\tilde{\boldsymbol{\mu}}) = \inf_{\tilde{\boldsymbol{c}}^\alpha}\big\{f^\alpha(\tilde{\boldsymbol{c}}^\alpha) - \tilde{\boldsymbol{\mu}}\cdot\tilde{\boldsymbol{c}}^\alpha\big\}$.
That the "effective" densities $\tilde{f}$ and $\tilde{\Psi}$ are nevertheless related in terms of a standard transform is thus not automatic but instead hinges on the internal consistency of the two definitions in terms of the common multiplier $\tilde{\boldsymbol{\mu}}$. More examples of similar relations will be given in Section 7.1.6 when considering the non-isothermal case. ◇

From this, the original optimization problem for $\tilde{\mathcal{F}}_\epsilon(\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)$ is then equivalent to

$$\min_{(\boldsymbol{\phi},\tilde{\boldsymbol{c}})} \int_\Omega \epsilon a(\boldsymbol{\phi},\nabla\boldsymbol{\phi}) + \frac{1}{\epsilon}w(\boldsymbol{\phi}) + \hat{\mathcal{L}}^{-1}_{\tilde{\boldsymbol{\mu}}}[\tilde{\Psi}](\boldsymbol{\phi},\tilde{\boldsymbol{c}},T)\,\mathrm{d}\boldsymbol{x} \tag{7.19}$$

and thus does clearly **not** correspond to the minimization of the grand potential energy

$$\Omega_\epsilon(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T) \coloneqq \int_\Omega \epsilon a(\boldsymbol{\phi},\nabla\boldsymbol{\phi}) + \frac{1}{\epsilon}w(\boldsymbol{\phi}) + \tilde{\Psi}(\boldsymbol{\phi},\tilde{\boldsymbol{\mu}},T)\,\mathrm{d}\boldsymbol{x}. \tag{7.20}$$

That the driving force of the phasefield equation based on Equation (7.18) is nevertheless of the form $\sum_{\beta=1}^{N} \tilde{\Psi}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$ above is - similarly to appearence of $\frac{\partial f}{\partial \phi}$ despite the maximization of the entropy due the expression $\frac{\partial s}{\partial \phi}(\boldsymbol{\phi}, \boldsymbol{c}, e) = -\frac{1}{T}\frac{\partial f}{\partial \phi}(\boldsymbol{\phi}, \boldsymbol{c}, T)$ discussed in Section 3.2 - a result of the variational characterization of this inverse transform. More precisely, the value $\tilde{\boldsymbol{\mu}}$ for which the supremum in Equation (7.18) is actually achieved[14] is characterized by

$$\frac{\partial}{\partial \tilde{\boldsymbol{\mu}}}\big(\tilde{\Psi}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T) + \tilde{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}}\big) = \frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{c}}) + \tilde{\boldsymbol{c}} \stackrel{!}{=} \boldsymbol{0}, \tag{7.21}$$

defining $\tilde{\boldsymbol{\mu}}$ realizing the supremum as a function of the remaining parameters, $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$. Based on this maximizer, $\tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ can then be rewritten explicitly in terms of $\tilde{\Psi}$ as

$$\begin{aligned} \tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) &= \min_{\{(\tilde{\boldsymbol{c}}^\alpha)_{1\le\alpha\le N} : \sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}}\}} \Big\{ \sum_{\alpha=1}^{N} f^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \Big\} \\ &= \Psi\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) + \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) \cdot \tilde{\boldsymbol{c}}, \end{aligned} \tag{7.22}$$

from which the driving force follows as

$$\frac{\partial \tilde{f}}{\partial \phi^\alpha}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) = \frac{\partial \Psi}{\partial \phi^\alpha}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) + \frac{\partial}{\partial \tilde{\boldsymbol{\mu}}}\big(\Psi\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) + \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) \cdot \tilde{\boldsymbol{c}}\big) \cdot \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \phi^\alpha}.$$

By the optimality condition on $\tilde{\boldsymbol{\mu}}$, the second term drops out, leaving

$$\frac{\partial \tilde{f}}{\partial \phi^\alpha}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) = \frac{\partial \Psi}{\partial \phi^\alpha}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) = \sum_{\beta=1}^{N} \tilde{\Psi}^\beta\big(\tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) \frac{\partial h^\beta}{\partial \phi^\alpha},$$

and thus an expression which, together with the relation $\tilde{\Psi}^\beta = f^\beta - \tilde{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}}$ of $\tilde{\Psi}^\beta$, is the same as in Equation (7.16)[15].

By the expression of the driving force in Equation (7.16) combined with the fact that the value (but not the function itself!) of $\tilde{f}^\beta\big(\tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) \cdot \tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ indeed coincides with $\tilde{\Psi}^\beta(\tilde{\boldsymbol{\mu}}, T)$ evaluated for $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$, it is obvious that the same result is also obtained by a straight-forward differentiation of $\tilde{\Psi}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T)$ with $\tilde{\boldsymbol{\mu}}$ considered as a fixed parameter (i.e. neglecting all potential dependencies). This simple mechanism for obtaining this driving force is also the basis for the "phenomenological" approach taken in [56] and [19]. In fact, their approach consists in postulating for $\boldsymbol{\phi}$ to follow a (minimizing) gradient-flow for the grand potential energy

[14] Note that this is a strictly concave maximization problem in $\tilde{\boldsymbol{\mu}}$.

[15] Before drawing this conclusion, there is one (somewhat hidden) point which remains to be verified, namely that the values of $\tilde{\boldsymbol{\mu}}$ in both equations as well as the values of $\tilde{\Psi}^\beta(\tilde{\boldsymbol{\mu}}, T)$ above and the $\tilde{f}^\beta\big(\tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) \cdot \tilde{\boldsymbol{c}}^\beta(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ in Equation (7.16) in fact coincide as both are a priori defined based on different conditions (one as the (free) maximizer of $\tilde{\Psi}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T) + \tilde{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}}$, the other one as the multiplier for the (constrained) optimality system (7.10)). This now follows mostly from more standard "phase-specific" thermodynamics, more precisely in the (smooth convex-concave) analogue of the third point in Proposition 3 in the case of full duality, namely that $\tilde{\boldsymbol{\mu}} = \frac{\partial \tilde{f}^\beta(\tilde{\boldsymbol{c}}^\beta, T)}{\partial \tilde{\boldsymbol{c}}^\beta}$ iff $\tilde{\Psi}^\beta(\tilde{\boldsymbol{\mu}}, T) = \tilde{f}^\beta(\tilde{\boldsymbol{c}}^\beta, T) - \tilde{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}}^\beta$ iff $\tilde{\boldsymbol{c}}^\beta = \frac{\partial \tilde{\Psi}^\beta}{\partial \tilde{\boldsymbol{\mu}}}$. Since this implies that simply **defining** $\tilde{\boldsymbol{c}}^\beta(\tilde{\boldsymbol{\mu}}, T) := \frac{\partial \tilde{\Psi}^\beta}{\partial \tilde{\boldsymbol{\mu}}}(\tilde{\boldsymbol{\mu}}, T)$ (now as a function of $\tilde{\boldsymbol{\mu}}$) with $\tilde{\boldsymbol{\mu}}$ as in Equation (7.21), one has, on the one hand, $\tilde{\Psi}^\beta(\tilde{\boldsymbol{\mu}}, T) = \tilde{f}^\beta\big(\tilde{\boldsymbol{c}}^\beta(\tilde{\boldsymbol{\mu}}, T), T\big) - \tilde{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}}^\beta(\tilde{\boldsymbol{\mu}}, T)$ as well as $\tilde{\boldsymbol{\mu}} = \frac{\partial \tilde{f}^\beta(\tilde{\boldsymbol{c}}^\beta, T)}{\partial \tilde{\boldsymbol{c}}^\beta}\big(\tilde{\boldsymbol{c}}^\beta(\tilde{\boldsymbol{\mu}}, T), T\big)$ and on the other hand by the optimality condition in Equation (7.21) that

$$\tilde{\boldsymbol{c}} = -\frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}} = -\sum_{\alpha=1}^{N} \frac{\partial \tilde{\Psi}^\alpha}{\partial \tilde{\boldsymbol{\mu}}}(\tilde{\boldsymbol{\mu}}, T) h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^\alpha(\tilde{\boldsymbol{\mu}}, T) h^\alpha(\boldsymbol{\phi}),$$

i.e. $\tilde{\boldsymbol{\mu}}$ in combination with the concentration values derived from it actually satisfies the optimality system (7.10) and thus by unicity for this system that both expressions actually coincide.

functional $\Omega_\epsilon(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T)$ in Equation (7.20). This is combined with the standard evolution equation (here using the notation similar to [19][16] together with the Einstein summation convention) for the concentration in the isothermal case,

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \Big(M_{ijk} \frac{\partial \mu_j}{\partial x_k} \Big),\ i = 1, ..., K-1 \quad \text{resp.} \quad \frac{\partial \tilde{\boldsymbol{c}}}{\partial t} = \nabla \cdot \big(\boldsymbol{M} : \nabla \tilde{\boldsymbol{\mu}} \big) \tag{7.23}$$

for the independent concentration values.
If one takes, as proposed in both works and as the use of the grand potential energy would seem to suggest, $\tilde{\boldsymbol{\mu}}$ instead of $\tilde{\boldsymbol{c}}$ as the "fundamental" dynamic variable, this makes, together with the condition on the equilibrium of the phasespecifc chemical potentials, $\tilde{\boldsymbol{c}}$ a function of $(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T)$, where $\tilde{\boldsymbol{c}}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^\alpha(\tilde{\boldsymbol{\mu}}, T) h^\alpha(\boldsymbol{\phi})$. Together with the chain-rule and the given evolution equation for $\boldsymbol{\phi}$, this permits rewriting the evolution equation for the reduced concentration as one for $\tilde{\boldsymbol{\mu}}$ through $\frac{\partial \tilde{\boldsymbol{c}}}{\partial t} = \frac{\partial \tilde{\boldsymbol{c}}}{\partial \boldsymbol{\phi}} \cdot \frac{\partial \boldsymbol{\phi}}{\partial t} + \frac{\partial \tilde{\boldsymbol{c}}}{\partial \tilde{\boldsymbol{\mu}}} \cdot \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial t}$, from which one obtains

$$\frac{\partial \tilde{\boldsymbol{\mu}}}{\partial t} = \Big(\frac{\partial \tilde{\boldsymbol{c}}}{\partial \tilde{\boldsymbol{\mu}}} \Big)^{-1} \cdot \Big(\nabla \cdot \big(\boldsymbol{M} : \nabla \tilde{\boldsymbol{\mu}} \big) - \frac{\partial \tilde{\boldsymbol{c}}}{\partial \boldsymbol{\phi}} \cdot \frac{\partial \boldsymbol{\phi}}{\partial t} \Big). \tag{7.24}$$

Remark 96. Note that, together with $-\tilde{\boldsymbol{c}} = \frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}$ and the evolution equation for $\boldsymbol{\phi}$ (e.g. from Equation (6.74)) complemented by the driving force $\frac{\partial \tilde{\Psi}}{\partial \phi^\alpha} = \sum_{\beta=1}^{N} \tilde{\Psi}^\beta(\tilde{\boldsymbol{\mu}}, T) \frac{\partial h^\beta}{\partial \phi^\alpha}(\boldsymbol{\phi})$, this is in principle a system dependent solely on $\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\mu}}$ since the parameter T is fixed by assumption. It would thus seem that one can completely forget about the concentration and in particular the phase-specific concentrations, or at least only use them as auxiliary quantities whenever convenient.
While this is essentially true[17] in the continuous case, one does well to remember that the evolution Equation (7.24) for $\tilde{\boldsymbol{\mu}}$ is designed solely for the purpose of ensuring the (conservative) evolution of $\tilde{\boldsymbol{c}}$ as in Equation (7.23) in the discrete case.
For example, when applying - as far as possible due to the term $\frac{\partial \boldsymbol{\phi}}{\partial t}$ - the simplest time-discretization in terms of an explicit Euler scheme, there are several potential pitfalls. Firstly, the evaluation of the time-derivative of $\boldsymbol{\phi}$ requires both a new and an old value $\boldsymbol{\phi}^{(n+1)}$ and $\boldsymbol{\phi}^{(n)}$ for evaluating the time-derivative $\frac{\partial \boldsymbol{\phi}}{\partial t}$, at least when implemented as a single-step scheme[18]. Based solely on Equation (7.24) and the use of an explicit scheme, it is then quite natural to firstly evaluate the matrix $\frac{\partial \tilde{\boldsymbol{c}}}{\partial \tilde{\boldsymbol{\mu}}}$ as $\frac{\partial \tilde{\boldsymbol{c}}}{\partial \tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}^{(n)}, \tilde{\boldsymbol{\mu}}^{(n)}, T\big)$, and secondly to explicitly evaluate the contribution due to the time-derivative of $\boldsymbol{\phi}$ as $\frac{\partial \tilde{\boldsymbol{c}}}{\partial \boldsymbol{\phi}}$ as

$$\sum_{\alpha=1}^{N} \sum_{\beta=1}^{N} \tilde{\boldsymbol{c}}^\beta\big(\boldsymbol{\phi}^{(n)}, \tilde{\boldsymbol{\mu}}^{(n)}, T\big) \frac{\partial h^\beta}{\partial \phi^\alpha}\big(\boldsymbol{\phi}^{(n)}\big) \frac{\big(\phi^\alpha\big)^{(n+1)} - \big(\phi^\alpha\big)^{(n)}}{\Delta t}.$$

In relation with Equation (7.23) and the known data, both of these are bad choices though. In fact, instead of discretizing the "original" evolution equation for the concentration, one would like to satisfy, still within an explicit Euler approach,

$$\frac{\tilde{\boldsymbol{c}}^{(n+1)} - \tilde{\boldsymbol{c}}^{(n)}}{\Delta t} = \nabla \cdot \big(\boldsymbol{M} : \nabla \tilde{\boldsymbol{\mu}}^{(n)} \big).$$

Treating $\tilde{\boldsymbol{\mu}}$ as the primary unknown and $\tilde{\boldsymbol{c}}$ as the secondary one, this, together with the already calculated value of $\boldsymbol{\phi}^{(n+1)}$ (and $\nabla_{\boldsymbol{D}}$ denoting a discretized version of the ∇-operator), leads to the equation

$$\tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n+1)}, \tilde{\boldsymbol{\mu}}^{(n+1)}, T\big) \stackrel{!}{=} \tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n)}, \tilde{\boldsymbol{\mu}}^{(n)}, T\big) + \Delta t \nabla_D \cdot \big(\boldsymbol{M} : \nabla_D \tilde{\boldsymbol{\mu}}^{(n)} \big). \tag{7.25}$$

[16] Note that the formulation of Plapp in [56] is slightly different in that it uses a single density field, which can be related to the single independent component via the atomic volume, which is assumed to be the same for both species of atoms. For more details, see [56] or the final note in section I.B. in [19].

[17] Except potentially for initial conditions which are often expressed more naturally in terms of the concentration.

[18] An alternative of course consists in using $\boldsymbol{\phi}^{(n)}$ and $\boldsymbol{\phi}^{(n-1)}$, but is both more complex in terms of storage (when combined with an appropriate buffering scheme) and less accurate.

As all quantities except for $\tilde{\boldsymbol{\mu}}^{(n+1)}$ are known in this equation, a simple linearization leads to the more appropriate discrete equivalent of Equation (7.24), namely

$$\begin{aligned}\tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n+1)},\tilde{\boldsymbol{\mu}}^{(n+1)},T\big) &\approx \tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n+1)},\tilde{\boldsymbol{\mu}}^{(n)},T\big) + \frac{\partial\tilde{\boldsymbol{c}}}{\partial\tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}^{(n+1)},\tilde{\boldsymbol{\mu}}^{(n+1)},T\big)\cdot\big(\tilde{\boldsymbol{\mu}}^{(n+1)}-\tilde{\boldsymbol{\mu}}^{(n)}\big)\\ &\overset{!}{=}\tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n)},\tilde{\boldsymbol{\mu}}^{(n)},T\big) + \Delta t\nabla_D\cdot\big(\boldsymbol{M}:\nabla_D\tilde{\boldsymbol{\mu}}^{(n)}\big).\end{aligned}$$

Further combining this with $\tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n+1)},\tilde{\boldsymbol{\mu}}^{(n)},T\big) = \sum_{\alpha=1}^{N}\tilde{\boldsymbol{c}}^{\alpha}\big(\tilde{\boldsymbol{\mu}}^{(n)},T\big)h^{\alpha}\big(\boldsymbol{\phi}^{(n+1)}\big)$ and the analogous expression $\tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}^{(n)},\tilde{\boldsymbol{\mu}}^{(n)},T\big) = \sum_{\alpha=1}^{N}\tilde{\boldsymbol{c}}^{\alpha}\big(\tilde{\boldsymbol{\mu}}^{(n)},T\big)h^{\alpha}\big(\boldsymbol{\phi}^{(n)}\big)$ for $\boldsymbol{c}^{(n)}$, this then leads to the discrete update-rule

$$\begin{aligned}\tilde{\boldsymbol{\mu}}^{(n+1)} :=& \Big(\frac{\partial\tilde{\boldsymbol{c}}}{\partial\tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}^{(n+1)},\tilde{\boldsymbol{\mu}}^{(n)},T\big)\Big)^{-1}\\ &\cdot\left(\sum_{\alpha=1}^{N}\tilde{\boldsymbol{c}}^{\alpha}\big(\tilde{\boldsymbol{\mu}}^{(n)},T\big)\Big(h^{\alpha}\big(\boldsymbol{\phi}^{(n)}\big)-h^{\alpha}\big(\boldsymbol{\phi}^{(n+1)}\big)\Big) + \Delta t\nabla_D\cdot\big(\boldsymbol{M}:\nabla_D\tilde{\boldsymbol{\mu}}^{(n)}\big)\right),\end{aligned} \tag{7.26}$$

which differs from the discretization above in two important points.
Firstly, the linearization through the matrix $\frac{\partial\tilde{\boldsymbol{c}}}{\partial\tilde{\boldsymbol{\mu}}}$ is used only in terms of the (as of yet unknown) values of $\tilde{\boldsymbol{\mu}}$ but using the already known value of $\boldsymbol{\phi}^{(n+1)}$. In contrast, the most straightforward discretization does the same thing, but using a less appropriate linearization based on an "outdated" value of the phasefield-vector $\boldsymbol{\phi}^{(n)}$. Whereas this linearization will be second-order accurate in the case of Equation (7.26), it will therefore only be first-order accurate when using $\boldsymbol{\phi}^{(n)}$. Secondly, using the (also already known) differences between the weight-functions $h^{\alpha}(\boldsymbol{\phi})$ instead of their linearization together with $\frac{\boldsymbol{\phi}^{(n+1)}-\boldsymbol{\phi}^{(n)}}{\Delta t}$ is both cheaper and more accurate (actually exact) than what would result from an a priori natural discretization based on Equation (7.24).
With this choice of parameters, the resulting update rule for $\tilde{\boldsymbol{\mu}}$ clearly corresponds to a single Newton-step for the discrete update rule (7.25) for the concentration. An obvious advantage of using such a "blind" single-step procedure in the update rule (7.26) as compared to actually solving Equation (7.25) is that this is certainly cheaper and requires fewer parameters as there is for example no control of the residual before accepting the new $\tilde{\boldsymbol{\mu}}$-value as sufficiently accurate. This is at the same time also potentially a severe disadvantage, since this essentially restricts the use of Equation (7.26) to situations where the changes in $\tilde{\boldsymbol{\mu}}$ between successive time-steps are small enough for this to be the case. A particular but practically quite useful setting where this never causes any issues is of course when using parabolic approximations of the free energy densities, as, with the $\tilde{\boldsymbol{c}}^{\alpha}$ being affine functions of $\tilde{\boldsymbol{\mu}}$, the linearization is actually exact. ◇

Even though taking $\tilde{\boldsymbol{\mu}}$ instead of $\tilde{\boldsymbol{c}}$ as the primary unknown (and thus a quantity which is seemingly independent of $\boldsymbol{\phi}$) would at first sight seem to justify the use of the partial differentiation of $\tilde{\Psi}$ with respect to $\boldsymbol{\phi}$ only and thus the use of the grand potential energy density as the relevant functional, this is somewhat misleading. More precisely, from an optimization point of view, the relevance of the functional is not really a question of whether one parameterizes an intermediary evolution in terms of $\tilde{\boldsymbol{c}}$ or $\tilde{\boldsymbol{\mu}}$ (there being a one-to-one correspondence between the two), but of whether the final steady-state can indeed be related to a critical point of this functional or not. This is closely related to the discussion in [41]. In fact, once evolved to a steady-state, the (reduced) chemical potential $\tilde{\boldsymbol{\mu}}$ will, regardless of the choice of parameterization in $\tilde{\boldsymbol{c}}$ or $\tilde{\boldsymbol{\mu}}$, satisfy the equation

$$-\nabla\cdot\big(\boldsymbol{M}:\nabla\tilde{\boldsymbol{\mu}}\big) = \boldsymbol{0}$$

together with the natural isolating boundary conditions $\frac{\partial\tilde{\boldsymbol{\mu}}}{\partial\boldsymbol{n}} = \boldsymbol{0}$. This equation will (under the natural coercivity condition on $\boldsymbol{M}$) enforce for $\tilde{\boldsymbol{\mu}}$ to be equal to a constant, i.e. $\tilde{\boldsymbol{\mu}}(\boldsymbol{x}) = \tilde{\boldsymbol{\eta}}$ for some $(K-1)$-dimensional vector $\tilde{\boldsymbol{\eta}}$, which is just the well-known condition of the equilibration of the

chemical potentials. As the conservative gradient flow for $\tilde{\boldsymbol{c}}$ in Equation (7.23) is, in combination with the isolating boundary conditions, specifically chosen such that the total concentration $\int_\Omega \tilde{\boldsymbol{c}}\,\mathrm{d}\boldsymbol{x}$ remains equal to its initial value $\tilde{\boldsymbol{C}}_0$, this final value $\tilde{\boldsymbol{\eta}}$ is not arbitrary, but fixed through the condition

$$\int_\Omega \tilde{\boldsymbol{c}}(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}},T)\,\mathrm{d}\boldsymbol{x} = \int_\Omega -\frac{\partial\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}}(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}},T)\,\mathrm{d}\boldsymbol{x} = \tilde{\boldsymbol{C}}_0. \tag{7.27}$$

From this, it is obvious that, even if one initially chooses $\tilde{\boldsymbol{\mu}}$ as an "independent" unknown, the final values of $\tilde{\boldsymbol{\mu}}$ are of the form $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},\tilde{\boldsymbol{C}}_0,T)$, where $\tilde{\boldsymbol{\eta}}$ depends implicitly (and in a non-local fashion) on $\boldsymbol{\phi}$ through the consistency condition (7.27) on the total concentration[19].
As the (physically obviously important) dynamics in themselves ultimately have no impact on the final values of the functional, one might just as well directly consider Ω_ϵ using this final state $\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T)$, reducing Ω_ϵ to a function of $\boldsymbol{\phi}$ and the fixed temperature alone,

$$\Omega_\epsilon\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},\tilde{\boldsymbol{C}}_0,T),T\big) = \int_\Omega \epsilon a(\boldsymbol{\phi},\nabla\boldsymbol{\phi}) + \frac{1}{\epsilon}w(\boldsymbol{\phi}) + \tilde{\Psi}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},\tilde{\boldsymbol{C}}_0,T),T\big)\,\mathrm{d}\boldsymbol{x}$$

together with Equation (7.27) as a side-condition. An actual minimizer for Ω_ϵ in terms of $\boldsymbol{\phi}$ (being the only free parameter now), would thus have to satisfy the analogue of the steady-state phasefield equation 6.64, but with the driving force contribution due to $\tilde{\Psi}$ now derived from

$$\int_\Omega \frac{\partial\tilde{\Psi}}{\partial\boldsymbol{\phi}}\cdot\delta\boldsymbol{\phi} + \frac{\partial\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}}\cdot\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi},T)\,\mathrm{d}\boldsymbol{x}$$

where $\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})$ can be determined through the differentiation

$$\begin{aligned}&\int_\Omega \frac{\partial\tilde{\boldsymbol{c}}}{\partial\boldsymbol{\phi}}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\cdot\delta\boldsymbol{\phi} + \frac{\partial\tilde{\boldsymbol{c}}}{\partial\tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\cdot\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{x}\\ =&-\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\boldsymbol{\phi}\partial\tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\cdot\delta\boldsymbol{\phi} + \frac{\partial^2\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}^2}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\cdot\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{x} = \boldsymbol{0}\end{aligned}$$

of the constraint in Equation (7.27). Since $\delta\tilde{\boldsymbol{\eta}}$ is a constant vector, this equation is easily resolved explicitly for $\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})$ as

$$\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi}) = -\Big(\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}^2}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\,\mathrm{d}\boldsymbol{x}\Big)^{-1}\cdot\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\boldsymbol{\phi}\partial\tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi},\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi},T),T\big)\cdot\delta\boldsymbol{\phi}\,\mathrm{d}\boldsymbol{x},$$

showing that an actual minimizer of Ω_ϵ would have to be based on the driving force corresponding to

$$\int_\Omega \frac{\partial\tilde{\Psi}}{\partial\boldsymbol{\phi}}\cdot\delta\boldsymbol{\phi} - \frac{\partial\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}}\cdot\Big(\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}^2}\,\mathrm{d}\boldsymbol{y}\Big)^{-1}\cdot\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\boldsymbol{\phi}\partial\tilde{\boldsymbol{\mu}}}\cdot\delta\boldsymbol{\phi}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{x}.$$

Exchanging the order of integration in the last part, this contribution simplifies to the "local" density

$$\left(\frac{\partial\tilde{\Psi}}{\partial\boldsymbol{\phi}} - \Big(\int_\Omega \frac{\partial\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}}\,\mathrm{d}\boldsymbol{y}\Big)\cdot\Big(\int_\Omega \frac{\partial^2\tilde{\Psi}}{\partial\tilde{\boldsymbol{\mu}}^2}\,\mathrm{d}\boldsymbol{y}\Big)^{-1}\cdot\frac{\partial^2\tilde{\Psi}}{\partial\boldsymbol{\phi}\partial\tilde{\boldsymbol{\mu}}}\right)\cdot\delta\boldsymbol{\phi}. \tag{7.28}$$

This is obviously quite different from the expression resulting from the first part alone, showing that steady-state solutions of the system above do **not** correspond to actual minimizers (or more precisely critical points in general) of Ω_ϵ under the side-condition on the total concentration.

[19] This is essentially the same argument as in [41].

In contrast, repeating an analogous argument for the free energy functional, with $\tilde{\boldsymbol{c}}$ parameterized in terms of $\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, T), T\big)$, leads to the condition

$$\int_\Omega \frac{\partial \tilde{f}}{\partial \boldsymbol{\phi}}\Big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big), T\Big) \cdot \delta\boldsymbol{\phi} + \frac{\partial \tilde{f}}{\partial \tilde{\boldsymbol{c}}}\Big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big), T\Big) \cdot \delta\tilde{\boldsymbol{c}}(\delta\boldsymbol{\phi})\, \mathrm{d}\boldsymbol{x}, \quad (7.29)$$

where $\delta\tilde{\boldsymbol{c}}(\delta\boldsymbol{\phi})$ (despite its a priori more complex dependence on both $\delta\boldsymbol{\phi}$ and $\delta\tilde{\boldsymbol{\mu}}(\delta\boldsymbol{\phi})$) can again be characterized from a differentiation of Equation (7.27) through the very simple relation $\int_\Omega \delta\tilde{\boldsymbol{c}}(\delta\boldsymbol{\phi})\, \mathrm{d}\boldsymbol{x} = \boldsymbol{0}$. By the equilibrium condition in terms of the concentration through the equilibration of the chemical potentials, one in addition has

$$\frac{\partial \tilde{f}}{\partial \tilde{\boldsymbol{c}}}\Big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big), T\Big) = \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T)$$

with $\tilde{\boldsymbol{\eta}}$ independent of $\boldsymbol{x}$. It follows that the second contribution in Equation (7.29) simply drops out and the driving force in Equation (7.16) is thus indeed consistent with a variational principle in terms of the free energy together with the side-condition given by the conservation of the total concentration $\tilde{\boldsymbol{C}}_0$.

This can alternatively also be seen by, instead of relying on the conservative gradient flow for $\tilde{\boldsymbol{c}}$ with respect to $\mathcal{F}_\epsilon$, including the constraint $\int_\Omega \tilde{\boldsymbol{c}}\, \mathrm{d}\boldsymbol{x} = \tilde{\boldsymbol{C}}_0$ directly into the problem by augmenting $\mathcal{F}_\epsilon$ with this constraint, i.e. by considering

$$\min_{(\boldsymbol{\phi}, \tilde{\boldsymbol{c}})} \sup_{\check{\boldsymbol{\eta}}} \Big\{ \mathcal{F}_\epsilon - \check{\boldsymbol{\eta}} \cdot \Big(\int_\Omega \tilde{\boldsymbol{c}}\, \mathrm{d}\boldsymbol{x} - \tilde{\boldsymbol{C}}_0 \Big) \Big\}.$$

Using Equation (7.18), this may also be written as

$$\min_{(\boldsymbol{\phi}, \tilde{\boldsymbol{c}})} \sup_{\check{\boldsymbol{\eta}}} \Big\{ \int_\Omega \epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi}) + \sup_{\check{\boldsymbol{\mu}}} \Big\{ \tilde{\Psi}(\boldsymbol{\phi}, \check{\boldsymbol{\mu}}, T) + \check{\boldsymbol{\mu}} \cdot \tilde{\boldsymbol{c}} \Big\} - \check{\boldsymbol{\eta}} \cdot \Big(\int_\Omega \tilde{\boldsymbol{c}}\, \mathrm{d}\boldsymbol{x} - \tilde{\boldsymbol{C}}_0 \Big) \Big\}.$$

Assuming one can interchange the minimization with respect to $\tilde{\boldsymbol{c}}$ with the suprema in $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\eta}}$, the minimization with respect to $\tilde{\boldsymbol{c}}$, with the only term depending on $\tilde{\boldsymbol{c}}$ given by $\int_\Omega \big(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\eta}}\big) \cdot \tilde{\boldsymbol{c}}\, \mathrm{d}\boldsymbol{x}$ leads to

$$\sup_{(\check{\boldsymbol{\eta}}, \check{\boldsymbol{\mu}})} \inf_{\tilde{\boldsymbol{c}}} \Big\{ \int_\Omega \tilde{\Psi}(\boldsymbol{\phi}, \check{\boldsymbol{\mu}}, T) + \big(\check{\boldsymbol{\mu}} - \check{\boldsymbol{\eta}}\big) \cdot \tilde{\boldsymbol{c}}\, \mathrm{d}\boldsymbol{x} + \check{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0 \Big\} = \sup_{(\check{\boldsymbol{\eta}}, \check{\boldsymbol{\mu}})} \left\{ \begin{cases} \int_\Omega \tilde{\Psi}(\boldsymbol{\phi}, \check{\boldsymbol{\mu}}, T)\, \mathrm{d}\boldsymbol{x} + \check{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0 & \check{\boldsymbol{\mu}} = \check{\boldsymbol{\eta}} \text{ a.e.,} \\ -\infty & \text{else} \end{cases} \right\}.$$

The second case will obviously be eliminated by the sup-operations, such that it suffices to consider the case with $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\eta}}$ which can then be reduced to a global maximization operation in the vector $\tilde{\boldsymbol{\mu}}$, $\sup_{\check{\boldsymbol{\eta}}} \Big\{ \int_\Omega \tilde{\Psi}(\boldsymbol{\phi}, \check{\boldsymbol{\eta}}, T)\, \mathrm{d}\boldsymbol{x} + \check{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0 \Big\}$. Combining this with the contributions by a and w, an alternative formulation of the global problem is thus given by

$$\min_{\boldsymbol{\phi}} \sup_{\check{\boldsymbol{\eta}}} \Big\{ \int_\Omega \epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi}) + \tilde{\Psi}(\boldsymbol{\phi}, \check{\boldsymbol{\eta}}, T)\, \mathrm{d}\boldsymbol{x} + \check{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0 \Big\} = \min_{\boldsymbol{\phi}} \sup_{\check{\boldsymbol{\eta}}} \Big\{ \Omega_\epsilon(\boldsymbol{\phi}, \check{\boldsymbol{\eta}}, T) + \check{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0 \Big\}. \quad (7.30)$$

Note that since the phasefield terms play no role with respect to the supremum in $\tilde{\boldsymbol{\mu}}$, this is basically just a "global" version of the inverse transform defining $\tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ in terms of $\tilde{\Psi}(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}, T)$ as in Equation (7.18). More precisely, the optimality condition for $\tilde{\boldsymbol{\mu}}$ in Equation (7.21) defines $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$ as a function of the given parameters $(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$. This value of $\tilde{\boldsymbol{\mu}}$ is such that, on the one hand, $\tilde{\boldsymbol{c}} = -\frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big)$ and on the other hand, from Equation (7.22) and the definition (7.5) of $\tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)$, such that the value of $\tilde{\Psi}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T), T\big) + \tilde{\boldsymbol{\mu}}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T) \cdot \tilde{\boldsymbol{c}}$ is precisely

the one of the effective free energy obtained by an energetically optimal distribution of the given average concentration onto the phase-specific ones.
In an analogous manner, the optimality condition with respect to $\tilde{\boldsymbol{\eta}}$ in Equation (7.30) is given by $\tilde{\boldsymbol{C}}_0 = -\int_\Omega \frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}, T\big)\,\mathrm{d}\boldsymbol{x}$, i.e. an integral version of $\tilde{\boldsymbol{c}} = -\frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}$, which serves to define $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T)$ with the total concentration $\tilde{\boldsymbol{C}}_0$ replacing the local one in $\tilde{\boldsymbol{\mu}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T\big)$. In addition, this value of $\tilde{\boldsymbol{\eta}}$ satisfies an integral analogue of Equation (7.22), namely one has

$$\begin{aligned}\min_{\{\tilde{\boldsymbol{c}}:\int_\Omega \tilde{\boldsymbol{c}}\,\mathrm{d}\boldsymbol{x}=\tilde{\boldsymbol{C}}_0\}} \Big\{ \int_\Omega \tilde{f}(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T)\,\mathrm{d}\boldsymbol{x}\Big\} &= \sup_{\tilde{\boldsymbol{\eta}}} \Big\{ \int_\Omega \tilde{\Psi}(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}, T)\,\mathrm{d}\boldsymbol{x} + \tilde{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{C}}_0\Big\} \\ &= \int_\Omega \tilde{\Psi}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big)\,\mathrm{d}\boldsymbol{x} + \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T) \cdot \tilde{\boldsymbol{C}}_0,\end{aligned} \tag{7.31}$$

and thus the value of $\int_\Omega \tilde{\Psi}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big)\,\mathrm{d}\boldsymbol{x} + \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T) \cdot \tilde{\boldsymbol{C}}_0$ is the one obtained from an energetically optimal redistribution of the given concentration $\tilde{\boldsymbol{C}}_0$ over the domain.
The primary interest of this observation in terms of the current discussion is that it can serve to derive an "extension" of the expression for the driving force in Equation (7.16). Whereas this expression was obtained based on the local minimization of $\tilde{f}$ in terms of the $\boldsymbol{c}^\alpha$ given the local concentration $\tilde{\boldsymbol{c}}$, Equation (7.31) defines the global optimal contribution due to the bulk free energy density through an a priori quite complex double-minimization as

$$\min_{\{\tilde{\boldsymbol{c}}:\int_\Omega \tilde{\boldsymbol{c}}\,\mathrm{d}\boldsymbol{x}=\tilde{\boldsymbol{C}}_0\}} \min_{\{(\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}:\sum_{\alpha=1}^N \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi})=\tilde{\boldsymbol{c}}\}} \Big\{ \int_\Omega \sum_{\alpha=1}^N f^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) h^\alpha(\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{x}\Big\}$$

enforcing an optimal redistribution of the given total concentration $\tilde{\boldsymbol{C}}_0$ over both the domain and the individual phases. Regardless, in the same manner as before, it follows from Equation (7.31) that the driving force contribution to this problem has the same simple structure as before. More precisely, instead of differentating the expression for $\int_\Omega \tilde{f}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}(\boldsymbol{\phi}, T), T\big)$ where $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{c}}(\boldsymbol{\phi}, T)$ is the minimizer for the left-most expression in Equation (7.31), one can also differentiate the right-most one, leading to

$$\delta\Big(\int_\Omega \tilde{f}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}(\boldsymbol{\phi}, T), T\big)\,\mathrm{d}\boldsymbol{x}\Big)(\delta\boldsymbol{\phi}) = \int_\Omega \frac{\partial \tilde{\Psi}}{\partial \boldsymbol{\phi}} \cdot \delta\boldsymbol{\phi} + \frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\eta}}} \cdot \delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})\,\mathrm{d}\boldsymbol{x} + \delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi}) \cdot \tilde{\boldsymbol{C}}_0$$

with $\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})$ denoting the increment of $\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})$ as a (non-local) function of $\delta\boldsymbol{\phi}$. As $\tilde{\boldsymbol{\eta}}$ is such that $\tilde{\boldsymbol{C}}_0 = -\int_\Omega \frac{\partial \tilde{\Psi}}{\partial \tilde{\boldsymbol{\mu}}}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}, T\big)\,\mathrm{d}\boldsymbol{x}$ though, the last two terms drop out regardless of the value of $\delta\tilde{\boldsymbol{\eta}}(\delta\boldsymbol{\phi})$, thus again reducing the derivative with respect to $\boldsymbol{\phi}$ (despite the additional global dependence of the optimal $\tilde{\boldsymbol{c}}$ on $\boldsymbol{\phi}$) to the simple local expression

$$\frac{\partial \tilde{f}}{\partial \phi^\alpha}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big) = \sum_{\beta=1}^N \tilde{\Psi}^\beta\big(\boldsymbol{\phi}, \tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T), T\big) \frac{\partial h^\beta}{\partial \phi^\alpha}(\boldsymbol{\phi}).$$

Summarizing the discussion above, it can be seen that while the $\tilde{\Psi}^\alpha$ appear very naturally as the relevant phase-specific quantities for the driving force in the phasefield equation, the link between this model and the grand potential energy functional Ω_ϵ is a somewhat tenuous one. While the use of this potential as suggested in [56] and [19] is clearly a very efficient approach for "deriving" this driving force, this should rather be considered as a purely formal device[20].

[20] In particular, as seen in Equation (7.28), a more carfeful derivation of the driving force corresponding to the minimization of Ω_ϵ under the implicit constraint of maintaining the total concentration $\int_\Omega \tilde{\boldsymbol{c}}\,\mathrm{d}\boldsymbol{x}$ enforced through the evolution equation for $\tilde{\boldsymbol{c}}$ (whether expressed in $\tilde{\boldsymbol{\mu}}$ or not) leads to a quite different expression. In contrast, as seen based on Equation (7.31), it is only in combination with the additonal contribution by $\tilde{\boldsymbol{\eta}}(\boldsymbol{\phi}, \tilde{\boldsymbol{C}}_0, T) \cdot \tilde{\boldsymbol{C}}_0$ to Ω_ϵ - and thus reverting back to the free energy - that one recovers consistency of the driving force with a global minimization property.

As such this approach is in a certain sense in competition with the more common approach of "guessing" the correct equations based upon a straightforward partial differentiation of the Lagrangian

$$L\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}, T, (\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}, \tilde{\boldsymbol{\mu}}\big) = \sum_{\alpha=1}^{N} \tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{\mu}} \cdot \big(\sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) - \tilde{\boldsymbol{c}}\big).$$

While this approach is a priori also purely formal, obtaining the relevant driving force is then just as simple as when starting from Ω_ϵ as it suffices to partially differentiate L with respect to ϕ^α, with the additional advantage of being, even at a purely formal level, easier to link to the original problem one is trying to solve[21].

Remark 97. It is important to note that the discussion above is only valid in this form when the concentration is (through the imposed conservative flow of $\tilde{\boldsymbol{c}}$) subject to isolating boundary conditions and therefore implicitly the constraint on the total concentration $\int_\Omega \tilde{\boldsymbol{c}} \,\mathrm{d}\boldsymbol{x}$. If, in contrast, one chooses to enforce a constant value of the chemical potential $\hat{\tilde{\mu}}$ on the outer boundary, it is clear that the phasefield equation together with the driving force given in Equation (7.16), will indeed minimize the grand potential energy $\Omega_\epsilon(\boldsymbol{\phi}, \hat{\tilde{\boldsymbol{\mu}}}, T)$. In fact, since the chemical potential is, up to some potential delay through the diffusion equation for $\tilde{\boldsymbol{c}}$, then fixed to the value $\hat{\tilde{\boldsymbol{\mu}}}$, there will be no additional contribution as in Equation (7.28).

At the same time, there will not anymore be a natural minimization interpretation in terms of $\mathcal{F}$, a property which is closely related to the discussion in the earlier paper [41] by Kim, Kim and Suzuki. More precisely, the analysis in [41] investigates the influence on the minimization of the free energy $\mathcal{F}(\phi, c)$ if one, instead of treating both c and ϕ (here in a reduced scalar formulation) as independent variables, considers the steady-state concentration profile for c as a function of the steady-state phasefield profile $\phi(\boldsymbol{x})$. Their argument is that any equilibrium profile of the concentration has to satisfy the equilibration of the chemical potentials μ with the prescribed one, i.e. $\frac{\partial f}{\partial c}(c, \phi) = \hat{\mu} = \text{const}$ and thus implicitly defines c as a function of ϕ (and $\hat{\mu}$). They then continue to argue that defining $F(c(\phi), \phi) := f\big(c(\phi), \phi\big) - \hat{\mu} c(\phi)$ and thus such that F satisfies $\frac{\partial F}{\partial c} = \frac{\partial f}{\partial c} - \hat{\mu} = 0$ and $\frac{\mathrm{d}F}{\mathrm{d}\phi} = \frac{\partial F}{\partial \phi} + \frac{\partial F}{\partial c}\frac{\partial c}{\partial \phi} = \frac{\partial F}{\partial \phi} = \frac{\partial f}{\partial \phi}(c(\phi), \phi)$, the driving force $\frac{\partial f}{\partial \phi}$ for the phasefield equation can equivalently be reformulated in terms of $\frac{\mathrm{d}F}{\mathrm{d}\phi}$.
While this argument is correct, it needs to be interpreted with some care in the present context[22]. In particular, the analysis in [41] should **not** be considered in the sense that the driving force for a minimization of the free energy in this case should be given by the derivative of $f(c, \phi) - \hat{\mu} c$. It is thus quite different from the analysis in the paper [42] by the same authors, where, despite the fact that it is still based on the minimization of the free energy, the use of two phase-specific concentration fields defined as functions of (ϕ, c) in terms of the equality of the phase-specific (reduced) chemical potentials does in fact lead to the driving force being given by the differences of $f^\alpha(c^\alpha) - \mu c^\alpha$[23].

Instead, it simply shows a variational inconsistency in the sense that under the constraint in terms of $\hat{\mu}$, using the driving force $\frac{\partial f}{\partial \phi}$ derived as a partial derivative from the free energy density, f will in fact not minimize the free energy but the grand chemical potential (i.e. the

[21] I.e. in particular, when the $\tilde{\boldsymbol{c}}^\alpha$ satisfy the sum-constraint, L reduces to $\sum_{\alpha=1}^N f^\alpha(\tilde{\boldsymbol{c}}^\alpha, T) h^\alpha(\boldsymbol{\phi})$ and equating the partial differentiations with respect to $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{c}}^\alpha$ to $\boldsymbol{0}$ leads to the constraint and optimality condition $\frac{\partial \tilde{f}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} = \tilde{\boldsymbol{\mu}}$ respectively.

[22] In [41], the appearence of this total differential is primarily used to enable a more standard analysis of the energetics of the transition region in the one-dimensional case.

[23] In contrast to the former paper, this is based on a purely local analysis and does not consider any additional dependence of the average concentration c on ϕ.

quantity whose total derivative $\frac{\partial f}{\partial \phi}$ represents in combination with the constraint). As such, it is essentially a restatement of the well-known fact that the minimization of the free energy is not the appropriate variational principle when dealing with an open system.
In contrast, actually minimizing the free energy under the constraint on the chemical potential would require using the total derivative of f, i.e.

$$\frac{\mathrm{d}f}{\mathrm{d}\phi}\big(\phi, c(\phi)\big) = \frac{\partial f}{\partial \phi} + \frac{\partial f}{\partial c}\frac{\partial c}{\partial \phi} = \frac{\partial f}{\partial \phi} + \hat{\mu}\frac{\partial c}{\partial \phi},$$

which could then, using the same argument as in [41], be rewritten as $\frac{\partial f}{\partial \phi} - \hat{\mu}\big(\frac{\partial^2 f}{\partial c^2}\big)^{-1}\frac{\partial^2 f}{\partial \phi \partial c}$. ◇

Remark 98. It is quite interesting to observe what happens if the analysis in [42] is combined with the one in [41] by again enforcing a **prescribed** chemical potential[24]. Firstly, adjusting the previous argument to the notation above, it was already observed that $\frac{\partial \tilde{f}}{\partial \boldsymbol{\phi}} = \sum_{\beta=1}^{N} \tilde{\Psi}^{\beta}(\hat{\boldsymbol{\mu}}, T)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}}$. Since this is, for a fixed value of $\hat{\boldsymbol{\mu}}$, clearly the same as $\frac{\mathrm{d}\tilde{\Psi}}{\mathrm{d}\boldsymbol{\phi}}$, using $\frac{\partial f}{\partial \boldsymbol{\phi}}$ as the driving force does in fact lead to the minimization of Ω_{ϵ}. Secondly, the total derivative of $\tilde{f}\big(\boldsymbol{\phi}, \tilde{\boldsymbol{c}}(\boldsymbol{\phi}, \hat{\boldsymbol{\mu}}, T), T\big)$ is then given by

$$\frac{\mathrm{d}\tilde{f}}{\mathrm{d}\boldsymbol{\phi}} = \frac{\partial \tilde{f}}{\partial \boldsymbol{\phi}} + \frac{\partial \tilde{f}}{\partial \tilde{\boldsymbol{c}}} \cdot \frac{\partial \tilde{\boldsymbol{c}}}{\partial \boldsymbol{\phi}}.$$

As in this model the average concentration $\tilde{\boldsymbol{c}}$ is by construction given by the weigthed average $\tilde{\boldsymbol{c}} = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{c}}(\hat{\tilde{\boldsymbol{\mu}}}, T) h^{\alpha}(\boldsymbol{\phi})$ of the $\tilde{\boldsymbol{c}}^{\alpha}$ corresponding to the - then given - chemical potential $\hat{\tilde{\boldsymbol{\mu}}}$, there is in fact no need to invoke an implicit function theorem for determining $\frac{\partial \tilde{\boldsymbol{c}}}{\partial \boldsymbol{\phi}}$ as these derivatives can directly be obtained from this definition. Combining this with the expresssion for $\frac{\partial \tilde{f}}{\partial \phi^{\alpha}}$ and the chain-rule for the differentiation of $\tilde{f}$, one thus obtains

$$\frac{\mathrm{d}\tilde{f}}{\mathrm{d}\phi^{\alpha}} = \sum_{\beta=1}^{N} \tilde{\Psi}^{\beta}(\hat{\boldsymbol{\mu}}, T)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} + \hat{\tilde{\boldsymbol{\mu}}} \cdot \sum_{\beta=1}^{N} \tilde{\boldsymbol{c}}^{\beta}(\hat{\tilde{\boldsymbol{\mu}}}, T)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} = \sum_{\beta=1}^{N} f^{\beta}\big(\tilde{\boldsymbol{c}}^{\beta}(\hat{\boldsymbol{\mu}}, T), T\big)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}}.$$

In contrast to the minimization of $\mathcal{F}$ subject to a fixed total concentration, the correct driving force for the minimization of $\mathcal{F}$ subject to a fixed chemical potential is therefore indeed given in terms of the $\tilde{f}^{\alpha}$. ◇

7.1.5 The Practical Evaluation of $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ and the Chemical Potential

Given that f is now only implicitly defined in terms of either the abstract minimization problem Equation (7.3) (resp. through its alternative representation $\tilde{f}$ in Equation (7.5)) or the more concrete optimality system in Equation (7.9) (resp. in Equation (7.10)), it is clear that the evaluation of the free energy function is more involved than for an explicit formulation in terms of the phase-averaged concentrations $\boldsymbol{c}$ as in e.g. [52]. Therefore, each evaluation of $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ now requires solving for the $(\boldsymbol{c}^{\alpha})_{\alpha \in \mathcal{P}_p}$ realizing the minimum in Equation (7.3). As the $f^{\alpha}(\boldsymbol{c}^{\alpha}, T)$ are in general nonlinear (though strictly convex) functions of the phase-specific concentrations $\boldsymbol{c}^{\alpha}$, this has to be done - except for particularly simple examples such as an ideal solution model - using some iterative solution approach. The natural choice here is of course a Newton-type scheme, based on either Equation (7.9) or Equation (7.10).

The Reduced Formulation

Beginning with the somewhat simpler case of the reduced formulation, each Newton step consists in, given the current estimates of $\big(\tilde{\boldsymbol{c}}^{\alpha}\big)_{\alpha \in \mathcal{P}_p}^{(n)}$ (and therefore trivially also $(c_K^{\alpha})^{(n)} = 1 -$

[24] Given that both reduced chemical potentials are enforced to be equal, the only consistent choice for them is to be equal to the prescribed one.

$\sum_{i=1}^{K-1}(c_i^\alpha)_i^{(n)})$, determining the updated concentration values $\left(\tilde{\boldsymbol{c}}^\alpha\right)_{\alpha\in\mathcal{P}_p}^{(n+1)} = \left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)} + \delta\tilde{\boldsymbol{c}}^\alpha\right)_{\alpha\in\mathcal{P}_p}$ and the associated updated multiplier $\tilde{\boldsymbol{\mu}}^{(n+1)}$ by either setting $\tilde{\boldsymbol{\mu}}^{(n)} + \delta\tilde{\boldsymbol{\mu}}$ and solving the linearized problem

$$\begin{cases} \tilde{\boldsymbol{\mu}}^\alpha\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}, T\right) + \frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}, T\right)\delta\tilde{\boldsymbol{c}}^\alpha - (\tilde{\boldsymbol{\mu}}^{(n)} + \delta\tilde{\boldsymbol{\mu}}) = \boldsymbol{0}, \\ \sum_{\alpha\in\mathcal{P}_p}\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)} + \delta\boldsymbol{c}^\alpha\right)h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}}. \end{cases} \tag{7.32}$$

Alternatively, as the original equations $\tilde{\boldsymbol{\mu}}^\alpha \overset{!}{=} \tilde{\boldsymbol{\mu}}$ and thus automatically also (7.32) are in fact linear in $\tilde{\boldsymbol{\mu}}$, one can also directly use $\tilde{\boldsymbol{\mu}}^{(n+1)}$ and solve

$$\begin{cases} \tilde{\boldsymbol{\mu}}^\alpha\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}, T\right) + \frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}, T\right)\delta\tilde{\boldsymbol{c}}^\alpha - \tilde{\boldsymbol{\mu}}^{(n+1)} = \boldsymbol{0}, \\ \sum_{\alpha\in\mathcal{P}_p}\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)} + \delta\boldsymbol{c}^\alpha\right)h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}}. \end{cases} \tag{7.33}$$

Remark 99. It is clear that[25] both the incremental (in $\tilde{\boldsymbol{\mu}}$) System (7.32) and the non-incremental System (7.33) will lead to the same value for $\tilde{\boldsymbol{\mu}}^{(n+1)}$. Nevertheless, using an incremental formulation can simplify certain modifications (such as the use of a damped Newton scheme when the convexity is lost due to additional contributions) and, as pointed out below, allows a simpler reuse of the algorithm when determining the $\boldsymbol{c}^\alpha$ for a **given** $\tilde{\boldsymbol{\mu}}$. It will therefore be preferred in the following. If desired, the non-incremental version can easily be obtained from the description below by simply replacing $\delta\boldsymbol{\mu}$ with $\tilde{\boldsymbol{\mu}}^{(n+1)}$ and the old value $\tilde{\boldsymbol{\mu}}^{(n)}$ by $\boldsymbol{0}$. ◇

Assuming for notational simplicity that $\mathcal{P}_p = \{1, 2, ..., N_p\}$, the matrix-vector form of the system (7.32) is given by[26]

$$\begin{pmatrix} \begin{pmatrix} \left(\frac{\partial^2\tilde{f}^1}{\partial(\tilde{\boldsymbol{c}}^1)^2}\right) & \boldsymbol{0} & ... & \boldsymbol{0} \\ \boldsymbol{0} & \left(\frac{\partial^2\tilde{f}^2}{\partial(\tilde{\boldsymbol{c}}^2)^2}\right) & ... & \boldsymbol{0} \\ & \ddots & \ddots & \ddots \\ \boldsymbol{0} & ... & \boldsymbol{0} & \left(\frac{\partial^2\tilde{f}^{N_p}}{\partial(\tilde{\boldsymbol{c}}^{N_p})^2}\right) \end{pmatrix} & \begin{pmatrix} -\boldsymbol{I} \\ -\boldsymbol{I} \\ \vdots \\ -\boldsymbol{I} \end{pmatrix} \\ \begin{pmatrix} h^1(\boldsymbol{\phi})\boldsymbol{I} & h^2(\boldsymbol{\phi})\boldsymbol{I} & ... & h^{N_p}(\boldsymbol{\phi})\boldsymbol{I} \end{pmatrix} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \delta\tilde{\boldsymbol{c}}^1 \\ \delta\tilde{\boldsymbol{c}}^1 \\ \vdots \\ \delta\tilde{\boldsymbol{c}}^{N_p} \end{pmatrix} \\ \delta\tilde{\boldsymbol{\mu}} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^1}^{(n)} \\ \boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^2}^{(n)} \\ \vdots \\ \boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^{N_p}}^{(n)} \end{pmatrix} \\ \boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(n)} \end{pmatrix},$$

where the residuals are given by

$$\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)} := \tilde{\boldsymbol{\mu}}^{(n)} - \left(\tilde{\boldsymbol{\mu}}^\alpha\right)^{(n)} \quad \text{and} \quad \boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(n)} := \boldsymbol{c} - \sum_{\alpha=1}^{N_p}(\tilde{\boldsymbol{c}}^\alpha)^{(n)}h^\alpha(\boldsymbol{\phi}) \tag{7.34}$$

and the dependence of the $\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}$ on $\left((\tilde{\boldsymbol{c}}^\alpha)^{(n)}, T\right)$ has been suppressed. A noteworthy feature of this system is that the upper left part is block-diagonal as each $\tilde{f}^\alpha$ only depends on the phase-specific concentrations of this particular phase, and each $(K-1)\times(K-1)$-block is formed by an s.p.d. matrix due to the convexity of the $\tilde{f}^\alpha$ w.r.t. $\tilde{\boldsymbol{c}}^\alpha$. Due to this particular structure, the system can conveniently be solved using a Schur complement approach, i.e. by first eliminating the $\delta\tilde{\boldsymbol{c}}^\alpha$ in terms of the $\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)}$ and $\delta\tilde{\boldsymbol{\mu}}$ (resp. $\tilde{\boldsymbol{\mu}}^{(n+1)}$) as

$$\delta\tilde{\boldsymbol{c}}^\alpha = \left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)} + \delta\tilde{\boldsymbol{\mu}}\right) = \left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\tilde{\boldsymbol{\mu}}^{(n+1)} - (\tilde{\boldsymbol{\mu}}^\alpha)^{(n)}\right) \quad \alpha\in\mathcal{P}_p. \tag{7.35}$$

[25] Also see Remark 101 below.

[26] Note that this system in this form does not correspond to a classical saddle point matrix. If required, it can be brought into such a form y "undoing" the division of the first set of equations through $h^\alpha(\boldsymbol{\phi})$ and changing the signs in the last row.

This leads to the reduced system

$$\sum_{\alpha\in\mathcal{P}_p}\delta\tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi})=\sum_{\alpha=1}^{N_p}\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)}+\delta\tilde{\boldsymbol{\mu}}\right)h^\alpha(\boldsymbol{\phi})=\boldsymbol{r}_{\tilde{\boldsymbol{c}}}$$

and thus, defining $\tilde{\boldsymbol{\zeta}}^\alpha:=\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}$, the $(K-1)$-dimensional Schur complement system

$$\boldsymbol{S}_{\tilde{\boldsymbol{c}}}\delta\tilde{\boldsymbol{\mu}}=\boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(n)}-\sum_{\alpha\in\mathcal{P}_p}\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)}h^\alpha(\boldsymbol{\phi})=\boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(n)}-\sum_{\alpha\in\mathcal{P}_p}\left(\tilde{\boldsymbol{\zeta}}^\alpha\right)^{(n)}h^\alpha(\boldsymbol{\phi}) \tag{7.36}$$

with the (reduced) Schur complement matrix

$$\boldsymbol{S}_{\tilde{\boldsymbol{c}}}=\sum_{\alpha\in\mathcal{P}_p}\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}h^\alpha(\boldsymbol{\phi}). \tag{7.37}$$

Once $\delta\tilde{\boldsymbol{\mu}}$ (resp. $\tilde{\boldsymbol{\mu}}^{(n+1)}$) is found, the new $\tilde{\boldsymbol{c}}^\alpha$-values are then easily recovered using Equation (7.35) in terms of the (independent) subsystems

$$\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+1)}=\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}+\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(n)}+\delta\tilde{\boldsymbol{\mu}}\right)=\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}+\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\tilde{\boldsymbol{\mu}}^{(n+1)}-(\tilde{\boldsymbol{\mu}}^\alpha)^{(n)}\right). \tag{7.38}$$

Remark 100. This procedure can in principle alternatively be interpreted as a predictor-corrector scheme. In fact, defining the auxiliary quantities

$$\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+\frac{1}{2})}:=\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}+\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\tilde{\boldsymbol{\mu}}^{(n)}-(\tilde{\boldsymbol{\mu}}^\alpha)^{(n)}\right) \tag{7.39}$$

corresponding to the predicted values of the phase-specific concentration for the **current** value $\tilde{\boldsymbol{\mu}}^{(n)}$ of $\tilde{\boldsymbol{\mu}}$, the increment $\delta\tilde{\boldsymbol{\mu}}$ and the final $\left(\boldsymbol{c}^\alpha\right)^{(n+1)}$ satisfy

$$\boldsymbol{S}_{\tilde{\boldsymbol{c}}}\delta\tilde{\boldsymbol{\mu}}=\tilde{\boldsymbol{c}}-\sum_{\alpha\in\mathcal{P}_p}\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+\frac{1}{2})}h^\alpha(\boldsymbol{\phi})=\boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(n+\frac{1}{2})}$$

and

$$\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+1)}=\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+\frac{1}{2})}+\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\delta\tilde{\boldsymbol{\mu}}.$$

It is easily verified that, after a division by $h^\alpha(\boldsymbol{\phi})$ for the phases with $h^\alpha(\boldsymbol{\phi})\neq 0$, this system is also the FONC of the minimization problem

$$\begin{cases}\text{minimize} & \frac{1}{2}\sum_{\alpha\in\mathcal{P}_p}\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+1)}-\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+\frac{1}{2})}\right)\cdot\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\left(\left(\boldsymbol{c}^\alpha\right)^{(n)}\right)\cdot\left(\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+1)}-\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+\frac{1}{2})}\right)h^\alpha(\boldsymbol{\phi})\\ \text{subject to} & \sum_{\alpha\in\mathcal{P}_p}\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n+1)}h^\alpha(\boldsymbol{\phi})=\tilde{\boldsymbol{c}}\end{cases}$$

and thus a correction step through a weigted (by the $\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}$-matrices evaluated for the old concentration values $\left(\tilde{\boldsymbol{c}}^\alpha\right)^{(n)}$) projected operation onto the subspace of the phase-specific concentration values satisfying the sum-constraint $\sum_{\alpha\in\mathcal{P}_p}\tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi})=\tilde{\boldsymbol{c}}$. ◇

The block-factorization procedure using the Schur-complement $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ instead of a direct solution of the full system at once has two advantages. On the one hand, it can take advantage of the block-diagonal structure. In particular, it is sufficient to work with the individual subblocks $\frac{\partial^2 \tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}$ alone when calculating the Schur complmenent $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ and the modified residual in Equation (7.36) as well as for the $(\tilde{\boldsymbol{c}}^\alpha)^{(n+1)}$ in equation (7.38). On the other hand, while it is very expensive in terms of memory requirements to store all the phase-specific concentrations for the entire computational domain when a large number of phases are present, one can easily store the phase-independent multiplier $\tilde{\boldsymbol{\mu}}$. At any point at which it is necessary to use the $\boldsymbol{c}^\alpha$, they can be recalculated from the **known** value $\tilde{\boldsymbol{\mu}}$, at least provided none of the other parameters (here $\boldsymbol{\phi}$ and T) have changed in between. Starting e.g. from the initial guess $(\tilde{\boldsymbol{c}}^\alpha)^{(0)} = \boldsymbol{c}$, the corresponding Newton scheme reduces precisely to the same steps as the determination of the $(\boldsymbol{c}^\alpha)^{(n+\frac{1}{2})}$ in Equation (7.39) above, as one in fact has[27]

$$(\tilde{\boldsymbol{c}}^\alpha)^{(n+1)} := (\tilde{\boldsymbol{c}}^\alpha)^{(n)} + \left(\frac{\partial^2 \tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\tilde{\boldsymbol{\mu}} - (\tilde{\boldsymbol{\mu}}^\alpha)^{(n)}\right) = (\tilde{\boldsymbol{c}}^\alpha)^{(n)} + \left(\frac{\partial \tilde{\mu}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1}\left(\tilde{\boldsymbol{\mu}} - (\tilde{\boldsymbol{\mu}}^\alpha)^{(n)}\right). \tag{7.40}$$

One can therefore simply reuse the same functionality for both problems, that of determining the phase-specific concentrations and the reduced chemical potential from $(\boldsymbol{\phi}, \boldsymbol{c}, T)$ alone or from $(\boldsymbol{\phi}, \boldsymbol{c}, T)$ and $\tilde{\boldsymbol{\mu}}$, the latter one of course being somewhat cheaper as there is in particular no need of actually inverting the $\frac{\partial^2 \tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}$-submatrices in order to determine $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$.
In matrix terms, the increments in the phase-specific concentration and $\tilde{\boldsymbol{\mu}}$ as a function of the residuals from Equation (7.34) can succinctly be summarized as[28] $\begin{pmatrix} (\delta\tilde{\boldsymbol{c}}^\alpha)_{\alpha\in\mathcal{P}_p} \\ \delta\boldsymbol{\mu} \end{pmatrix} = \boldsymbol{A}\begin{pmatrix} (\boldsymbol{r}^\alpha_{\tilde{\boldsymbol{\mu}}})_{\alpha\in\mathcal{P}_p} \\ \boldsymbol{r}_{\tilde{\boldsymbol{c}}} \end{pmatrix}$ with

$$\boldsymbol{A} = \begin{pmatrix} & & & \boldsymbol{A}_{11} & & & \begin{pmatrix} \left(\frac{\partial\tilde{\boldsymbol{\mu}}^1}{\partial\tilde{\boldsymbol{c}}^1}\right)^{-1}\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \\ \left(\frac{\partial\tilde{\boldsymbol{\mu}}^2}{\partial\tilde{\boldsymbol{c}}^2}\right)^{-1}\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \\ \vdots \\ \left(\frac{\partial\tilde{\boldsymbol{\mu}}^{N_p}}{\partial\tilde{\boldsymbol{c}}^{N_p}}\right)^{-1}\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \end{pmatrix} \\ \begin{pmatrix} -h^1\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1}\frac{\partial\tilde{\boldsymbol{\mu}}^1}{\partial\tilde{\boldsymbol{c}}^1} & -h^2\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1}\frac{\partial\tilde{\boldsymbol{\mu}}^2}{\partial\tilde{\boldsymbol{c}}^2} & \dots & -h^{N_p}\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1}\frac{\partial\tilde{\boldsymbol{\mu}}^{N_p}}{\partial\tilde{\boldsymbol{c}}^{N_p}} \end{pmatrix} & & & \boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \end{pmatrix}, \tag{7.41}$$

where the individual block-entries of the submatrix $\boldsymbol{A}_{11}$ are given by

$$\boldsymbol{A}_{11}^{\alpha\beta} = \left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)^{-1}\left(\delta^{\alpha\beta}\boldsymbol{I} - h^\beta(\boldsymbol{\phi})\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1}\left(\frac{\partial\tilde{\boldsymbol{\mu}}^\beta}{\partial\tilde{\boldsymbol{c}}^\beta}\right)^{-1}\right). \tag{7.42}$$

While this representation can be convenient for theoretical purposes[29] , its $(1,1)$-block - unlike the one corresponding to Equation (7.32) - is now a fully filled matrix due to the coupling by the

[27] Where, as usual, it is preferable to solve these systems instead of multiplying the right-hand side by the inverse of $\frac{\partial^2 \tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2} = \frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}$.

[28] This is simply a formula for the inverse of a matrix in terms of its Schur-complement in the case when the off-diagonal blocks are not transposes of each other, see e.g. [10] or [77].

[29] Yet another represenation of this inverse - which is closely linked with the predictor-corrector interpretation above - is based on the following block-factorization as a "perturbation" of the inverse of the $(1,1)$-block (see [77]),

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{R} \\ \boldsymbol{L} & \boldsymbol{C} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{A}^{-1}\boldsymbol{R} \\ \boldsymbol{I} \end{pmatrix}\boldsymbol{S}^{-1}\begin{pmatrix} -\boldsymbol{L}\boldsymbol{A}^{-1} & \boldsymbol{I} \end{pmatrix},$$

which, in terms of the given quantities here, corresponds to

$$\begin{pmatrix} \left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)^{-1}_{\alpha\in\mathcal{P}_p} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)^{-1}_{\alpha\in\mathcal{P}_p} \\ \boldsymbol{I} \end{pmatrix}\boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1}\begin{pmatrix} -h^\alpha\left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)^{-1}_{\alpha\in\mathcal{P}_p} & \boldsymbol{I} \end{pmatrix}.$$

sum constraint. For practical purposes, it is therefore preferable to work with the "sequential" representation of this inverse consisting in first determining $\delta\tilde{\boldsymbol{\mu}}$ and then simply updating the $\tilde{\boldsymbol{c}}^\alpha$ through $\delta\tilde{\boldsymbol{c}}^\alpha = \left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)^{-1}\left(\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha} + \delta\tilde{\boldsymbol{\mu}}^\alpha\right)$.

Remark 101. It should be noted that, due to the linearity in $\tilde{\boldsymbol{\mu}}$, disposing of a good "initial guess" $\tilde{\boldsymbol{\mu}}^{(0)}$ has no direct effect on the basic Newton scheme in the form of (7.32) as $\tilde{\boldsymbol{\mu}}^{(1)}$ is a function of the $(\tilde{\boldsymbol{c}}^\alpha)^{(0)}$ only. This is also easily seen based on equations (7.36) and (7.34), as the prefactor of $\tilde{\boldsymbol{\mu}}^{(n)}$ in the update formula (7.36) is (due to the independence of $\tilde{\boldsymbol{\mu}}$ on α) precisely $-\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$, i.e. one can equivalently replace Equation (7.36) by the alternative update rule

$$\boldsymbol{S}_{\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{\mu}}^{(n)} + \delta\tilde{\boldsymbol{\mu}}) = \boldsymbol{S}_{\tilde{\boldsymbol{c}}}\tilde{\boldsymbol{\mu}}^{(n+1)} = \left(\tilde{\boldsymbol{c}} - \sum_{\alpha\in\mathcal{P}_p}\tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi})\right) + \sum_{\alpha\in\mathcal{P}_p}\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\right)^{-1}\left(\tilde{\boldsymbol{\mu}}^\alpha\right)^{(n)}h^\alpha(\boldsymbol{\phi}).$$

With $\tilde{\boldsymbol{\mu}}^{(n+1)}$, the $(\delta\tilde{\boldsymbol{c}}^\alpha)_{\alpha\in\mathcal{P}_p}$ can then be recovered using the left expression in Equation (7.35). ◇

Remark 102. A natural initial guess is to simply choose the initial phase-specific concentrations to be the same as the total concentration[30], i.e. setting $(\tilde{\boldsymbol{c}}^\alpha)^{(0)} = \tilde{\boldsymbol{c}},\ \alpha = 1,...,\mathcal{P}_p$. Due to the averaging property $\sum_{\alpha\in\mathcal{P}_p}\tilde{\boldsymbol{c}}h^\alpha(\boldsymbol{\phi}) = \left(\sum_{\alpha\in\mathcal{P}_p}h^\alpha(\boldsymbol{\phi})\right)\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{c}}$, this in particular ensures that the sum-constraint on the concentrations is automatically satisfied for the initial guess, i.e. $\boldsymbol{r}_{\tilde{\boldsymbol{c}}}^{(0)} = \boldsymbol{0}$.
Due to the linearity of the second equation in either Equation (7.33) or (7.32) w.r.t. the $\delta\boldsymbol{c}^\alpha$ and a simple recursion, this property will in fact be inherited by all subsequent iterates, as, per construction, $\sum_{\alpha\in\mathcal{P}_p}\delta\tilde{\boldsymbol{c}}^\alpha h^\alpha(\boldsymbol{\phi}) = \tilde{\boldsymbol{c}} - \sum_{\alpha\in\mathcal{P}_p}(\tilde{\boldsymbol{c}}^\alpha)^{(n)}h^\alpha(\boldsymbol{\phi})$ and the weighted average of the $\boldsymbol{c}^\alpha$ does therefore not change between the n-th and $(n+1)$-th step provided it was already consistent[31].
This unfortunately does not allow for any major simplifications of the algorithm above, as, despite the ability of dropping the right-most row in Equation (7.41), the (larger) $(1,1)$-block given by $\boldsymbol{A}_{11}$ in Equation (7.42) still contains all inverse matrices appearing in (7.41). ◇

Remark 103. Complementing Remark 101, if one believes an old value of $\tilde{\boldsymbol{\mu}}$ to be a good approximation of the new one - as is e.g. the case in many time-stepping schemes, in particular if the time-step is small - one may nevertheless use this value to generate a well-educated initial guess for the $\tilde{\boldsymbol{c}}^\alpha$. To do so, it suffices to (approximately) determining the $\boldsymbol{c}^\alpha$ such that $\frac{\partial\tilde{f}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha} = \tilde{\boldsymbol{\mu}}$ which, as $\tilde{\boldsymbol{\mu}}$ is taken as fixed, only involves the small block-diagonal submatrices $\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}$ whose factorizations need to be determined anyway.
In contrast to Remark 102, this does not ensure the fulfillment of the averaging on the $(\tilde{\boldsymbol{c}}^\alpha)^{(0)}$ to $\tilde{\boldsymbol{c}}$, but, if solved exactly[32], would ensure the equality of the $\tilde{\boldsymbol{\mu}}^\alpha$ with $\tilde{\boldsymbol{\mu}}$ and thus the vanishing of $(\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}^{(0)})_{\alpha\in\mathcal{P}_p}$.

While the contribution of the first (larger) part of $\boldsymbol{A}$ to the increments does in this case vanish in the first step - i.e. it seems as if one could make full use of the sparsity of the $(1,1)$-submatrix $\left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\beta}\right)_{\alpha,\beta\in\mathcal{P}_p}$ in the first equation of the system (7.32) instead of the fully filled $\boldsymbol{A}_{11}$-matrix in Equation (7.42) - this, similar to Remark 102, typically would generally not lead to any substantial decrease in the computational effort in the first iteration step as compared to solving the full system in Equation (7.32). In fact, on the one hand, the appearence of $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ still

[30] Which, as the conserved quantity, is a priori the natural one to be stored. In relation with the discussion in Section 7.1.4, an alternative consists in - as proposed e.g. in [19] - instead storing $\tilde{\boldsymbol{\mu}}$ only and using the (discretized) evolution equation for $\tilde{\boldsymbol{\mu}}$ from Equation (7.24) for updating $\tilde{\boldsymbol{\mu}}$ instead of the concentration itself. While this approach is in certain cases quite efficient and has been used quite successfully also for very large simulation setups (see for example the works [36], [70], [39] and the references therein), it is also somewhat restrictive as will be discussed below.

[31] The same will actually happen for any initial guess after the first iteration.

[32] This is certainly not to be recommended in general! Even if the individual sub-iterations in this first step would be slightly less inexpensive (but see Remark 103 below) than an iteration on the full system, this subiteration on a **subsystem** of the actual one to be solved would be at the cost of the quadratic convergence of the Newton-scheme on the **actual** system to be solved once the $\tilde{\boldsymbol{c}}^\alpha$ and $\tilde{\boldsymbol{\mu}}$ are sufficiently close to the true ones.

requires an factorization and inversion of the submatrices $\left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)_{\alpha\in\mathcal{P}_p}$ for all phases in $\mathcal{P}_p$. On the other hand, the natural algorithmic approach for the solution of the full system consists in **first** determining $\delta\tilde{\boldsymbol{\mu}}$ in terms of the auxiliary vectors $\tilde{\boldsymbol{\zeta}}^\alpha$ as in Equation (7.36) and then using Equation (7.38) for determining the $\delta\tilde{\boldsymbol{c}}^\alpha$. This requires precisely the same steps as the first iteration step assuming, as above, that the $\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha},\ \alpha=\alpha\in\mathcal{P}_p$ vanish, except for the solution of Equation (7.38) based on the already factorized or inverted submatrices $\left(\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\right)_{\alpha\in\mathcal{P}_p}$.

◇

Remark 104. In relation with the approaches in [19] and [56], it should be kept in mind that there is a seemingly very convenient alternative solution procedure, which consists in directly choosing the (common) multiplier $\tilde{\boldsymbol{\mu}}$ as the primary unknown and expressing the remaining unknowns $\tilde{\boldsymbol{c}}^\alpha$ as direct functions of $\tilde{\boldsymbol{\mu}}$ and T.
This reduction in the number of unknowns is clearly a quite appealing feature, since it replaces the system (7.10) with one consisting of the sum-constraint as a function of $\boldsymbol{\mu}$ alone,

$$\sum_{\alpha=1}^{N}\tilde{\boldsymbol{c}}^\alpha(\tilde{\boldsymbol{\mu}},T)h^\alpha(\boldsymbol{\phi})\overset{!}{=}\tilde{\boldsymbol{c}}. \tag{7.43}$$

Furthermore, taking again a Newton scheme as the obvious choice of solving Equation (7.43), one directly obtains the update rule

$$\Big(\sum_{\alpha=1}^{N}\frac{\partial\tilde{\boldsymbol{c}}^\alpha}{\partial\tilde{\boldsymbol{\mu}}}(\tilde{\boldsymbol{\mu}}^{(n)},T)h^\alpha(\boldsymbol{\phi})\Big)\delta\tilde{\boldsymbol{\mu}}=\tilde{\boldsymbol{c}}-\sum_{\alpha=1}^{N}\tilde{\boldsymbol{c}}^\alpha(\boldsymbol{\mu}^{(n)},T)h^\alpha(\boldsymbol{\phi}). \tag{7.44}$$

This can in fact be a very efficient approach provided one disposes of a direct formulation of the $\tilde{\boldsymbol{c}}^\alpha$ as functions of $\tilde{\boldsymbol{\mu}}$ and T since it only requires the solution of a single and usually fairly small system for the $K-1$ unknowns in $\delta\tilde{\boldsymbol{\mu}}$.
It is much less practical though if this is not the case, i.e. if one is not able to derive an explicit expression for $\tilde{\boldsymbol{c}}^\alpha(\tilde{\boldsymbol{\mu}},T)$. Even though this does not preclude the use of such a scheme since one can always solve the equation

$$\frac{\partial\tilde{f}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}(\tilde{\boldsymbol{c}}^\alpha,T)=\tilde{\boldsymbol{\mu}} \tag{7.45}$$

numerically[33] and then recover the required expressions for $\frac{\partial\tilde{\boldsymbol{c}}^\alpha}{\partial\tilde{\boldsymbol{\mu}}}$ by differentiating Equation (7.45) for $\boldsymbol{c}^\alpha=\boldsymbol{c}^\alpha(\tilde{\boldsymbol{\mu}},T)$ leading to

$$\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}(\tilde{\boldsymbol{c}}^\alpha,T)\frac{\partial\boldsymbol{c}^\alpha}{\partial\tilde{\boldsymbol{\mu}}}(\tilde{\boldsymbol{\mu}},T)=\boldsymbol{I}\quad\text{resp.}\quad\frac{\partial\boldsymbol{c}^\alpha}{\partial\tilde{\boldsymbol{\mu}}}(\tilde{\boldsymbol{\mu}},T)=\left(\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\big(\tilde{\boldsymbol{c}}^\alpha(\tilde{\boldsymbol{\mu}},T),T\big)\right)^{-1}, \tag{7.46}$$

this reintroduces the necessity of the inversion of these phase-specific matrices and leads back to the formula for the reduced Schur-complement matrix $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ in Equation (7.37). Furthermore, the numerical solution of Equation (7.45) is again most naturally done using a Newton-scheme and thus through the update rule $\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n+1)}=\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)}+\delta\tilde{\boldsymbol{c}}^\alpha$ with $\delta\tilde{\boldsymbol{c}}^\alpha$ determined by

$$\frac{\partial^2\tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)\delta\tilde{\boldsymbol{c}}^\alpha=\tilde{\boldsymbol{\mu}}-\frac{\partial\tilde{f}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big), \tag{7.47}$$

reintroducing the $\boldsymbol{c}^\alpha$ and $\delta\boldsymbol{c}^\alpha$ as auxiliary unknowns. In addition, Equation (7.45) should a priori be solved exactly in order for the expression for $\frac{\partial\tilde{\boldsymbol{c}}^\alpha}{\partial\tilde{\boldsymbol{\mu}}^\alpha}$ in Equation (7.46) to be correct, and will therefore in general require several Newton steps.

[33] This is also indicated in [56].

The additional computational cost of this can of course be alleviated by deliberately allowing for inaccuracies in the solution of this system by e.g. replacing the exact solution of the equations (7.45) with a single Newton-step as in Equation (7.47), but one has to exercise some care when doing so. While simply using the relation (7.46) is in principle still possible, this will seriously impeed the convergence of the update rule (7.44). The problem with this approach is that it neglects an effect distinguishing Equation (7.47) from the derivation of Equation (7.46). Namely, whereas Equation (7.46) is based on predicting the change of $\tilde{\boldsymbol{c}}^\alpha$ to changes in $\tilde{\boldsymbol{\mu}}$ $\frac{\partial^2 f^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^{2^2}}\delta\tilde{\boldsymbol{c}}^\alpha = \delta\tilde{\boldsymbol{\mu}}$, what actually happens due the update (7.47) is that $\delta\tilde{\boldsymbol{c}}^\alpha$ changes according to

$$\frac{\partial^2 \tilde{f}^\alpha}{\partial(\tilde{\boldsymbol{c}}^\alpha)^2}\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)}, T\Big)\delta\tilde{\boldsymbol{c}}^\alpha = \delta\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\mu}} - \frac{\partial \tilde{f}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)}, T\Big),$$

and therefore is not only affected by $\delta\tilde{\boldsymbol{\mu}}$ but also by the residual $\boldsymbol{r}_{\tilde{\boldsymbol{\mu}}^\alpha}$ for Equation (7.45) at the last step. Including this effect for restoring the quadratic convergence rate then simply leads back to Equation (7.36) and thus is ultimately a somewhat cumbersome and more error-prone way of obtaining the the previous linearized system (7.32) obtained through a "mechanical" differentiation based on the direct use of the more natural (for the $\tilde{f}^\alpha$ given as functions of $\tilde{\boldsymbol{c}}^\alpha$ and T) variables. ◇

Remark 105. By the previous remark, the computational cost of calculating an increment in $\tilde{\boldsymbol{\mu}}$ due to a change in the concentrations using implicitly defined functions $\tilde{\boldsymbol{c}}^\alpha$ in terms of $\tilde{\boldsymbol{\mu}}$ and T is, at least when using the inaccurate version, roughly similar to the one of obtaining $\tilde{\boldsymbol{\mu}}$ based on a given concentration as in Section 7.1.5 based on Equation (7.10).
A very important exception to this occurs in the bulk-regions, where, given $\tilde{f}^\alpha(\tilde{\boldsymbol{c}}^\alpha, T)$ and the average concentration $\tilde{\boldsymbol{c}}$ and the temperature T, one can obtain $\tilde{\boldsymbol{\mu}}$ directly from $\tilde{\boldsymbol{\mu}} = \frac{\partial \tilde{f}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}(\tilde{\boldsymbol{c}}, T)$ and thus is an operation which is expected to be relatively cheap. This is a major advantage over a scheme (such as proposed in [19]) based on the storage of $\tilde{\boldsymbol{\mu}}$ only, unless of course one has an explicit formulation of the $\tilde{\boldsymbol{c}}^\alpha$ as functions of $\tilde{\boldsymbol{\mu}}$ allowing to make use of the same simplification through a then also relatively cheap "conversion" of $\tilde{\boldsymbol{\mu}}$ to $\tilde{\boldsymbol{c}}^\alpha$ for the bulk-phase. If such is not the case, the storage of $\tilde{\boldsymbol{\mu}}$ instead of $\tilde{\boldsymbol{c}}$ is unlikely to be an efficient approach as this enforces an iterative update scheme in some form for updating $\tilde{\boldsymbol{\mu}}$ if there are any changes in the concentration. In particular, there is no really valid way of justifying the use of a "blind" single-step scheme as for the time-discretization of Equation (7.24) in [19]. Even if the changes in $\tilde{\boldsymbol{\mu}}$ are very small, due to e.g. a small time-step, and a single Newton-step for $\delta\tilde{\boldsymbol{\mu}}$ is therefore expected to be very accurate, this requires at least an accurate estimate of $\frac{\partial\tilde{\boldsymbol{c}}^\alpha}{\partial\tilde{\boldsymbol{\mu}}}$, and therefore a sufficiently accurate estimate of $\tilde{\boldsymbol{c}}^\alpha$ in the bulk-phase for which there is none if the concentration is not stored[34].
Except for sufficiently simple free energies where one can derive the appropriate relations analytically, a formulation directly in terms of $\tilde{\boldsymbol{c}}$ or, if using a primarily $\tilde{\boldsymbol{\mu}}$-based formulation, at least the additional storage of the concentration field is therefore to be recommended. ◇

[34] One of course has to be careful about the meaning of sufficiently accurate and to take the changes of other parameters into account in order to maintain a high accuracy with such a scheme.

The Non-Reduced Formulation

Solving for the unknowns in the non-reduced formulation from Equation (7.9) is a priori a little more difficult. In this case, the linearized system is given by

$$\frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}\big((\boldsymbol{c}^\alpha)^{(n)},T\big)+\frac{\partial^2 f^\alpha}{\partial(\boldsymbol{c}^\alpha)^2}\big((\boldsymbol{c}^\alpha)^{(n)},T\big)\delta\boldsymbol{c}^\alpha-(\boldsymbol{\mu}^{(n)}+\delta\boldsymbol{\mu}+(\lambda^\alpha)^{(n+1)}\boldsymbol{e})=\boldsymbol{0} \tag{7.48}$$

together with the sum-constraint $\sum_{\alpha\in\mathcal{P}_p}\boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi})=\boldsymbol{c}$ and the sum-constraints $(\boldsymbol{e}\cdot\boldsymbol{c}^\alpha)^{(n+1)}=1$ on the phase-specific concentrations. On the one hand, this system a priori has the disadvantage of being somewhat larger due to the presence of the additional phase-specific multipliers $\boldsymbol{\lambda}$ and the additional sum-constraints on the $\boldsymbol{c}^\alpha$. On the other hand, as pointed out above, due to the partial redundancy in the sum-constraint for $\boldsymbol{c}$, one needs to impose an additional condition on $\boldsymbol{\mu}$ in order to obtain unicity of the multipliers.
Nevertheless, this formulation is also somewhat more natural in the sense that the phase-specific free energies are often more easily described using the full concentration vectors $\boldsymbol{c}^\alpha$ instead of an "artificial" reduction to $K-1$ components only.

Fortunately, this is not a very serious concern as one already disposes of a simple solution procedure for the reduced formulation and the system in Equation (7.48) together with the constraints provides all the differential information on the system required to recover the remaining quantities if desired. Both the reduced and the non-reduced case can therefore be handled in a relatively transparent fashion using the same central algorithm.
More precisely, subtracting the K'th row in the System (7.48) from the first $K-1$ ones - corresponding to multiplying the system from the left by the matrix $\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}$ - makes it possible to eliminate the common factor λ^α and reduces the remaining system to

$$\begin{aligned}&\Big(\frac{\partial f^\alpha(\boldsymbol{c}^\alpha,T)}{\partial c_i^\alpha}\Big)_{1\le i\le K-1}-\frac{\partial f^\alpha(\boldsymbol{c}^\alpha,T)}{\partial c_K^\alpha}\tilde{\boldsymbol{e}}+\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}\frac{\partial^2 f^\alpha}{\partial(\boldsymbol{c}^\alpha)^2}\big((\boldsymbol{c}^\alpha)^{(n)},T\big)\delta\boldsymbol{c}^\alpha\\&-\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}(\boldsymbol{\mu}^{(n)}+\delta\boldsymbol{\mu})=\boldsymbol{0}.\end{aligned}$$

Firstly, $\Big(\frac{\partial f^\alpha(\boldsymbol{c}^\alpha,T)}{\partial c_i^\alpha}\Big)_{1\le i\le K-1}-\frac{\partial f^\alpha(\boldsymbol{c}^\alpha,T)}{\partial c_K^\alpha}\tilde{\boldsymbol{e}}=\frac{\partial f^\alpha\Big((\boldsymbol{c}^\alpha)^{(n)},1-\sum_{i=1}^{K-1}\big(c_i^\alpha\big)^{(n)}\Big),T\Big)}{\partial\tilde{\boldsymbol{c}}^\alpha}=\tilde{\boldsymbol{\mu}}^\alpha\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)$ by Equation (7.11) and similarly $\mu_i-\mu_K=\tilde{\mu}_i$ and $\delta\mu_i-\delta\mu_K=\delta\tilde{\boldsymbol{\mu}}$. Secondly, making use of the constraint to express to δc_K^α as $-\sum_{i=1}^{K-1}\delta c_i^\alpha$ - corresponding to $\delta\boldsymbol{c}^\alpha=\begin{pmatrix}\boldsymbol{I}_{K-1}\\-\tilde{\boldsymbol{e}}\end{pmatrix}\delta\tilde{\boldsymbol{c}}^\alpha$ - the system is reduced to the $(K-1)\times(K-1)$ one

$$\tilde{\boldsymbol{\mu}}^\alpha\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)+\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}\frac{\partial^2 f^\alpha}{\partial(\boldsymbol{c}^\alpha)^2}\big((\boldsymbol{c}^\alpha)^{(n)},T\big)\begin{pmatrix}\boldsymbol{I}_{K-1}\\-\tilde{\boldsymbol{e}}\end{pmatrix}\delta\tilde{\boldsymbol{c}}^\alpha-(\tilde{\boldsymbol{\mu}}^{(n)}+\delta\tilde{\boldsymbol{\mu}})=\boldsymbol{0}. \tag{7.49}$$

This, as is to be expected, is precisely the one in Equation (7.32) that would be obtained based on directly using a reduced formulation, except that $\tilde{\boldsymbol{\mu}}^\alpha\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)$ and the matrix $\frac{\partial\tilde{\boldsymbol{\mu}}}{\partial\tilde{\boldsymbol{c}}^\alpha}$ is recovered a posteriori as

$$\tilde{\boldsymbol{\mu}}^\alpha\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)=\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}\frac{\partial f^\alpha}{\partial\boldsymbol{c}^\alpha}\big((\boldsymbol{c}^\alpha)^{(n)},T\big) \tag{7.50}$$

and

$$\frac{\partial\tilde{\boldsymbol{\mu}}^\alpha}{\partial\tilde{\boldsymbol{c}}^\alpha}\Big(\big(\tilde{\boldsymbol{c}}^\alpha\big)^{(n)},T\Big)=\begin{pmatrix}\boldsymbol{I}_{K-1} & -\tilde{\boldsymbol{e}}\end{pmatrix}\frac{\partial^2 f^\alpha}{\partial(\boldsymbol{c}^\alpha)^2}\big((\boldsymbol{c}^\alpha)^{(n)},T\big)\begin{pmatrix}\boldsymbol{I}_{K-1}\\-\tilde{\boldsymbol{e}}\end{pmatrix} \tag{7.51}$$

through two simple algebraic operations instead of a priori being included in the formulation of the f^α.

Remark 106. Even though it may seem relatively cumbersome to algebraically reproduce a caculcation which could from the outset be done by hand when constructing the free energy function, this is also closely related to a matter of "convention". In particular, it is clear that starting from a reduced formulation is in many ways the simplest approach, in particular since one typically does not face any of the issues associated with such a formulation in the obstacle case for the phasefield (i.e. where phases may not be allowed to move according to the changes of the other ones). Nevertheless, this also introduces a certain asymmetry into the formulation by having to choose which component is the one being eliminated (the K-th one clearly being a good candidate though simply in terms of memory management and simplicity of any loops). This may on the one hand be perceived as unpleasant on aesthetic grounds and on the other hand entails a strict consistency requirement.
One advantage of the purely algebraic procedure above is its high robustness with respect to these issues, since, even though it entails an in principle completely unnecessary calculation if used with a reduced formulation (the operations in Equation (7.50) and (7.51) are then easily seen to have no effect), it allows combining various preexisting model components even if based on different conventions[35] ◇

In relation with the non-reduced formulation and the reduced nature of the Newton-steps in Equation (7.49), there are now essentially two manners to proceed. If one is solely interested in an evaluation of the final concentrations, either for their own values or for evaluating the f-function itself, it is obviously sufficient to simply recover the value of the last component c_K^α in the solution of the reduced system through the sum-constraint as $1 - \sum_{i=1}^{K-1} c_i^\alpha$.
If one is instead also interested in the values of $\boldsymbol{\mu}$ and the λ^α, one can also update these after the last step of the solution procedure[36]. More precisely, as the solution satisfies (up to the chosen tolerance for the residual of course) $\boldsymbol{\mu}^\alpha = \boldsymbol{\mu} + \lambda^\alpha \boldsymbol{e}$, one can easily recover the required information from the final values of $\boldsymbol{\mu}^\alpha$ at which the solution was considered sufficiently accurate for the three choices of restriction on $\boldsymbol{\mu}$ outlined in Subsection 7.1.2 since this relation implies $\boldsymbol{e} \cdot \boldsymbol{\mu}^\alpha = \boldsymbol{e} \cdot \boldsymbol{\mu} + K\lambda^\alpha$ with a known left-hand side. If $\boldsymbol{\mu}$ is fixed by enforcing its average to equal 0, the second term drops out and one has $\lambda^\alpha = \frac{1}{K} \boldsymbol{e} \cdot \boldsymbol{\mu}^\alpha$, i.e. λ^α is simply equal to the average of the μ^α. In contrast, if $\boldsymbol{\mu}$ is fixed by the condition that $\mu_K = 0$, one obviously has $\tilde{\boldsymbol{\mu}} = \big(\mu_i\big)_{1 \le i \le K-1} - \mu_K \tilde{\boldsymbol{e}} = \big(\mu_i\big)_{1 \le i \le K-1}$, and thus $\tilde{\boldsymbol{e}} \cdot \tilde{\boldsymbol{\mu}} = \boldsymbol{e} \cdot \boldsymbol{\mu}$ as the last entry vanishes, therefore leading to $\lambda^\alpha = \frac{1}{K}\big(\boldsymbol{e} \cdot \boldsymbol{\mu}^\alpha - \tilde{\boldsymbol{e}} \cdot \tilde{\boldsymbol{\mu}}\big)$. Finally, if $\boldsymbol{\mu}$ is chosen indirectly as the h^α-weighted average of the $\boldsymbol{\mu}^\alpha = \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}$, there is no need to first determine the λ^α as $\boldsymbol{\mu}$ can be obtained directly from the $\boldsymbol{\mu}^\alpha$. The λ^α then immediately follow from the equality $\lambda^\alpha = \frac{1}{K} \boldsymbol{e} \cdot \big(\mu^\alpha - \boldsymbol{\mu}\big)$.

Remark 107. Note that despite these different choices, this has no effect on the final concentration values, and that it does also not affect the evolution of the concentration by the construction of the L_{ij}-matrix in [52]. ◇

[35] For example, it is possible to combine a chemical contribution formulated in a reduced fashion through the first $K-1$ components with other contributions where the only direct contribution arises through the K-th component. This of course has no theoretical relevance, but avoids some very tedious issues at a practical level.

[36] As the basic Newton-scheme is shifted to the reduced formulation, this is typically also where the residual would typcially be controlled in its reduced form, i.e. one never directly uses $\boldsymbol{\mu}$ and the λ^α in the algorithm itself.

7.1.6 An Extension to the Non-Isothermal Case

The "grand-chemical" model considered in the previous section seems to have previously only been used in either an isothermal setting, or, under a relatively strong simplifying assumption, in combination with a prescribed non-uniform temperature field. In the form above, it is relatively clear though how the more quantitative approach can be formulated in a non-isothermal setting. With the maximization of the entropy functional from Equation (3.3) replacing the minimization of the free energy functional in Equation (6.10) as the appropriate variational principle, it is natural to define the entropy density within an interface region by

$$s(\boldsymbol{\phi}, \boldsymbol{c}, e) := \max_{(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}, e)} \left\{ \sum_{\alpha=1}^{N} s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) h^\alpha(\boldsymbol{\phi}) \right\}, \tag{7.52}$$

using the phase-specific quantities $(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N}$ instead of the average quantities $(\boldsymbol{c}, e)$ and with the admissible set $\mathcal{A}(\boldsymbol{\phi}, \boldsymbol{c}, e)$ consisting of all $(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N}$ such that[37]

$$\begin{cases} \sum_{\alpha=1}^{N} \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c}, \\ \sum_{i=1}^{K} c_i^\alpha = 1 \quad \forall \alpha, \\ \sum_{\alpha=1}^{N} e^\alpha h^\alpha(\boldsymbol{\phi}) = e \end{cases} \tag{7.53}$$

replacing the one in Equation (7.4).
The corresponding optimality conditions can again be derived based on the modified Lagrange-function

$$\begin{aligned} L\left(\boldsymbol{\phi}, \boldsymbol{c}, e, \left(\boldsymbol{c}^\alpha\right)_{1\leq\alpha\leq N}, \left(e^\alpha\right)_{1\leq\alpha\leq N}, \boldsymbol{\eta}, \hat{\boldsymbol{\lambda}}, \beta\right) = &\sum_{\alpha=1}^{N} s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) h^\alpha(\boldsymbol{\phi}) - \boldsymbol{\beta} \cdot \Big(\sum_{\alpha=1}^{N} h^\alpha \boldsymbol{c}^\alpha - \boldsymbol{c} \Big) \\ &- \sum_{\alpha=1}^{N} \hat{\lambda}^\alpha \Big(\sum_{i=1}^{K} c_i^\alpha - 1 \Big) - \beta \Big(\sum_{\alpha=1}^{N} h^\alpha e^\alpha - e \Big) \end{aligned}$$

with $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^N$ and (a priori) $\beta \in \boldsymbol{R}$. The first-order necessary conditions can then be obtained from $\frac{\partial L}{\partial \boldsymbol{c}^\alpha} = \boldsymbol{0}$ and $\frac{\partial L}{\partial e^\alpha} = 0$, leading, after by a division through h^α for all $\alpha \in \mathcal{P}_p$, to the local quasi-equilibrium conditions

$$\begin{cases} \frac{\partial s^\alpha}{\partial \boldsymbol{c}^\alpha} = \boldsymbol{\eta} + \hat{\lambda}^\alpha \boldsymbol{e}, \\ \frac{\partial s^\alpha}{\partial e^\alpha} = \beta. \end{cases} \tag{7.54}$$

As $\frac{\partial s^\alpha}{\partial e^\alpha} = \frac{1}{T^\alpha}$ and in analogy to the equilibrium condition $T = \text{const}$ for the temperature field, it is of course not surprising that the second condition implies the equality of the phase-specific temperatures $T^\alpha = \frac{1}{\beta} =: T$ for all phases (or, more precisely, for those with $h^\alpha(\boldsymbol{\phi}) \neq 0$). Based upon the monotonicity condition $\frac{\partial s^\alpha}{\partial e^\alpha} > 0$ for the bulk entropy densities, this common value in addition is necessarily positive, and one can substitute $\boldsymbol{\eta}$ and $\hat{\lambda}^\alpha$ by $-\frac{\boldsymbol{\mu}}{T^\alpha} = -\frac{\boldsymbol{\mu}}{T}$ and $-\frac{\lambda^\alpha}{T^\alpha} = -\frac{\lambda^\alpha}{T}$ and cancel the T in the first equation above to obtain the local quasi-equilibrium conditions in the more "pleasant" form

$$T^\alpha = T \quad \text{and} \quad \boldsymbol{\mu}^\alpha = \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}(\boldsymbol{c}^\alpha, T) = \boldsymbol{\mu} + \lambda^\alpha \boldsymbol{e} \quad \forall \alpha \in \mathcal{P}_p \tag{7.55}$$

and thus in particular including the previous condition on the chemical potentials as in the isothermal case.

Remark 108. The equilibrium conditions above are clearly what one would expect in relation with the previous dicussion and based on the close analogy between the optimization problems

[37] Note that it is again assumed here that the positivity constraints on the $\boldsymbol{c}^\alpha$ can be safely neglected.

in Equations (7.52) and (7.3). On the one hand, as discussed in e.g. [56], it is the compatibility of the (local) assumption $T^\alpha = T$ with the (global) equilibrium condition $T = \text{const}$ which can be used to explain the much larger success of the early phasefield models for purely temperature-driven solidification problems as compared to concentration-driven ones. On the other hand, the equilibration of the chemical potentials is well-known to be the relevant equilibrium condition for the concentration field, whether or not the process under consideration is from the outset assumed to be isothermal.
The point of the (re-)derivation of these conditions based upon the definition (7.52) is therefore not to show that these conditions are reasonable ones to impose within the interface region, but rather to verify that they arise naturally through a definition of the relevant local density which is compatible with the underlying global variational principle. ◇

Remark 109. It should be noted that T^α and $\boldsymbol{\mu}^\alpha$ in Equation (7.55) are a priori - based upon their derivation from (7.54) - to be considered as functions of $\boldsymbol{c}^\alpha$ and e^α (and in particular $\boldsymbol{\mu}^\alpha$ as $\boldsymbol{\mu}^\alpha(\boldsymbol{c}^\alpha, e^\alpha) = -T^\alpha(\boldsymbol{c}^\alpha, e^\alpha)\frac{\partial s^\alpha}{\partial \boldsymbol{c}^\alpha}(\boldsymbol{c}^\alpha, e^\alpha)$). As it is often prefered in practice to work with more concrete temperature values T^α instead of the e^α as the "independent" variable, this is not necessarily the most convenient form and one may choose to work, using the ubiquitous changes of variables in the thermodynamic setting, with a different set of "primary" unknowns. This does of course not change the content of the equilibrium conditions, but, as it changes their representation, can affect the solution process. Some advantages and disadvantages of such changes of variables will be discussed after taking a closer look at the link of the Definition (7.52) of the entropy $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ with the Definition (7.3) from the previous section. ◇

Some Basic Thermodynamic Relations

Before briefly discussing the solution procedure for the local system (7.55) in terms of the (assumed to be given) phasefield $\boldsymbol{\phi}$ and the conserved phase-averaged quantities $\boldsymbol{c}$ and e, it is helpful to recover a number of basic thermodynamic properties resembling those from the more standard case which are in fact implied by the Definition (7.52).
The primary difficulty here, as in the isothermal case, lies in the fact that the phase-inherent quantities $(\boldsymbol{c}^\alpha, e^\alpha)$ upon which $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ is based are now only defined implicitly as the ones fulfilling the local quasi-equilibrium conditions (7.55) above, i.e. $(\boldsymbol{c}^\alpha, e^\alpha) = \left(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, e), e^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, e)\right)$. In particular due to the equilibrium condition on the T^α, a first useful step is to obtain the corresponding free energy density $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$, defined by

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \inf_e \{e - Ts(\boldsymbol{\phi}, \boldsymbol{c}, e)\}. \tag{7.56}$$

It turns out that the following intuitively pleasing but not entirely obvious characterization holds for $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$:

Lemma 7. *Let $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ be defined by Equation (7.52) and assume that the local quasi-equilibrium conditions (7.55) allow for a unique (up to the addition of an arbitrary constant vector $\xi\boldsymbol{e}$ to $\boldsymbol{\mu}$) solution which depends smoothly on $\boldsymbol{c}$ and e and that the bulk entropy densities depend smoothly on $\boldsymbol{c}^\alpha$ and e^α. Then:*

- *For any minimizer e in Equation (7.56), the common value of the phase-specific temperatures T^α for the solution of the local quasi-equilibrium conditions (7.55) coincide with the "external" paramter T.*
- *The free energy density defined in Equation (7.56) coincides with the one defined in Equation (7.3)*

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \min_{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \sum_{\alpha=1}^{N} f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}). \tag{7.57}$$

Proof.

- Firstly, under the smoothness assumption of the lemma,

$$s(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{e}) = \max_{(\boldsymbol{c}^\alpha,e^\alpha)_{1\le\alpha\le N}\in\mathcal{A}(\boldsymbol{\phi},\boldsymbol{c},e)}\left\{\sum_\alpha s^\alpha(\boldsymbol{c}^\alpha,e^\alpha)h^\alpha(\boldsymbol{\phi})\right\} = \sum_{\alpha=1}^N s^\alpha\Big(\boldsymbol{c}^\alpha(\boldsymbol{\phi},\boldsymbol{c},e),T^\alpha(\boldsymbol{\phi},\boldsymbol{c},e)\Big)h^\alpha(\boldsymbol{\phi})$$

depends smoothly on e and thus any minimizer e in definition (7.56) has to satisfy

$$\frac{1}{T} = \frac{\partial s}{\partial e} = \sum_\alpha\Big(\frac{\partial s^\alpha}{\partial \boldsymbol{c}^\alpha}\cdot\frac{\partial \boldsymbol{c}^\alpha}{\partial e} + \frac{\partial s^\alpha}{\partial e^\alpha}\frac{\partial e^\alpha}{\partial e}\Big)h^\alpha(\boldsymbol{\phi}) = \sum_\alpha\Big(-\frac{\boldsymbol{\mu}^\alpha}{T^\alpha}\cdot\frac{\partial \boldsymbol{c}^\alpha}{\partial e} + \frac{1}{T^\alpha}\frac{\partial e^\alpha}{\partial e}\Big)h^\alpha(\boldsymbol{\phi}).$$

By the equilibrium conditions, T^α and $\frac{\boldsymbol{\mu}^\alpha}{T^\alpha}$ can be replaced by a common value θ (which at this point is not yet known to be the same as T) and $\frac{\boldsymbol{\mu}+\lambda^\alpha\boldsymbol{e}}{\theta}$, leading to

$$\frac{1}{T} = \sum_\alpha\Big(-\frac{\boldsymbol{\mu}+\lambda^\alpha\boldsymbol{e}}{\theta}\cdot\frac{\partial \boldsymbol{c}^\alpha}{\partial e} + \frac{1}{\theta}\frac{\partial e^\alpha}{\partial e}\Big)h^\alpha(\boldsymbol{\phi}).$$

As θ does not depend on α and $\boldsymbol{\phi}$ does not depend on e, the summation over the last term can be simplified to

$$\frac{1}{\theta}\frac{\partial\big(\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi})\big)}{\partial e} = \frac{1}{\theta}\frac{\partial e}{\partial e} = \frac{1}{\theta}$$

and one in particular does not require to know $\frac{\partial e^\alpha}{\partial e}$ anymore. In a similar manner, the summation over the terms $\frac{\boldsymbol{\mu}}{\theta}\cdot\frac{\partial \boldsymbol{c}^\alpha}{\partial e}$ leads to $\frac{\boldsymbol{\mu}}{\theta}\cdot\frac{\partial\big(\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi})\big)}{\partial e} = \frac{\boldsymbol{\mu}}{\theta}\cdot\frac{\partial \boldsymbol{c}}{\partial e} = 0$ and therefore simply drops out. That the same actually also happens for the (phase-dependent) term in λ^α is directly based upon the sum-constraint $\sum_{i=1}^K \boldsymbol{c}^\alpha = \boldsymbol{e}\cdot\boldsymbol{c}^\alpha = 1$ and thus $\boldsymbol{e}\cdot\frac{\partial \boldsymbol{c}^\alpha}{\partial e} = \frac{\partial(\boldsymbol{e}\cdot\boldsymbol{c}^\alpha)}{\partial e} = \boldsymbol{0}$, the identical argument being true for an arbitrary constant vector $\xi\boldsymbol{e}$ added to $\boldsymbol{\mu}$. Together, this shows that $\frac{1}{T} = \frac{1}{\theta}$, i.e. that the multiplier θ defining the local quasi-equilibrium has to be the same as the given parameter T.

- Inserting the definition of s into the definition (7.56), one has

$$f(\boldsymbol{\phi},\boldsymbol{c},T) = \inf_e\left\{e - T\max_{(\boldsymbol{c}^\alpha,e^\alpha)_{1\le\alpha\le N}\in\mathcal{A}(\boldsymbol{\phi},\boldsymbol{c},e)}\left\{\sum_\alpha s^\alpha(\boldsymbol{c}^\alpha,e^\alpha)h^\alpha(\boldsymbol{\phi})\right\}\right\}. \tag{7.58}$$

Since T is, as shown above, the same as the multiplier for the constraint $\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) = e$ in the local quasi-equilibrium conditions, the entropy can, given this value, equivalently be characterized through the free (with respect to the e^α) maximization problem

$$\begin{aligned}&\max_{(\boldsymbol{c}^\alpha,e^\alpha)_{1\le\alpha\le N}:(\boldsymbol{c}^\alpha)_{1\le\alpha\le N}\in\mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c})}\left\{\sum_\alpha s^\alpha(\boldsymbol{c}^\alpha,e^\alpha)h^\alpha(\boldsymbol{\phi}) - \frac{1}{T}\Big(\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) - e\Big)\right\}\\ =&\frac{e}{T} + \max_{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N}\in\mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c})}\max_{(e^\alpha)_{1\le\alpha\le N}}\left\{\sum_\alpha s^\alpha(\boldsymbol{c}^\alpha,e^\alpha)h^\alpha(\boldsymbol{\phi}) - \frac{1}{T}\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi})\right\}\end{aligned}$$

where $\mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c})$ is the admissible set from Equation (7.4) in the last section for defining f. Extracting the (fixed) factor $-\frac{1}{T}$ and using that the e^α are now independent, the last term can be rewritten as

$$\begin{aligned}&-\frac{1}{T}\min_{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N}\in\mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c})}\min_{(e^\alpha)_{1\le\alpha\le N}}\left\{\sum_{\alpha=1}^N\Big(e^\alpha - Ts^\alpha(\boldsymbol{c}^\alpha,e^\alpha)\Big)h^\alpha(\boldsymbol{\phi})\right\}\\ =&-\frac{1}{T}\min_{(\boldsymbol{c}^\alpha)_{1\le\alpha\le N}\in\mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c})}\left\{\sum_{\alpha=1}^N\Big(\min_{e^\alpha}\{e^\alpha - Ts^\alpha(\boldsymbol{c}^\alpha,e^\alpha)\}\Big)h^\alpha(\boldsymbol{\phi})\right\}.\end{aligned}$$

The inner minimization problem is precisely the definition of the phase-specific free energy $f^\alpha(\boldsymbol{c}^\alpha, T)$ in terms of the phase-specific entropy s^α,

$$f^\alpha(\boldsymbol{c}, T) = \min_{e^\alpha} \left\{ e^\alpha - T s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) \right\}, \tag{7.59}$$

which holds due to the "standard" thermodynamic relations valid for the bulk potentials, from which it follows that

$$\begin{aligned} &\max_{(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N} : (\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_\alpha s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) h^\alpha(\boldsymbol{\phi}) - \frac{1}{T} \Big(\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) - e \Big) \right\} \\ =& \frac{e}{T} - \frac{1}{T} \min_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\}. \end{aligned}$$

Inserting this expression into Equation (7.58), e in fact cancels out and one is left with

$$\begin{aligned} f(\boldsymbol{\phi}, \boldsymbol{c}, T) &= \inf_e \left\{ e - T \left(\frac{e}{T} - \frac{1}{T} \min_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\} \right) \right\} \\ &= \min_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_{\alpha=1}^N f^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) \right\}. \end{aligned}$$

The reverse conclusion is essentially just applying the same argument in the reverse order. Using Equation (7.59), one can rewrite f as

$$\begin{aligned} f(\boldsymbol{\phi}, \boldsymbol{c}, T) &= \min_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_{\alpha=1}^N \min_{e^\alpha} \left\{ e^\alpha - T s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) \right\} h^\alpha(\boldsymbol{\phi}) \right\} \\ &= \min_{e^\alpha} \left\{ \sum_{\alpha=1}^N \left\{ e^\alpha - T \max_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) \right\} h^\alpha(\boldsymbol{\phi}) \right\}. \end{aligned}$$

The outer minimization in the e^α can be moved to the inner maximization problem by aritifially "expanding" it through a double minimization

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \min_e \min_{e^\alpha : \sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) = e} \left\{ \sum_{\alpha=1}^N \left\{ e^\alpha - T \max_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) \right\} h^\alpha(\boldsymbol{\phi}) \right\}$$

which is legitimate since, even though this adds an restriction on the sum of the previously completely free e^α, the value of sum can be chosen in any arbitrary manner to make the total expression as small as possible. Since the sum over the e^α appearing in the expression to be minimized is then obviously equal to e, one can move the $\min_{e^\alpha}$-operation into the interior, leaving

$$f(\boldsymbol{\phi}, \boldsymbol{c}, T) = \min_e \left\{ e - T \min_{e^\alpha : \sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) = e} \max_{(\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N} \in \mathcal{A}_{\boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})} \left\{ \sum_{\alpha=1}^N s^\alpha(\boldsymbol{c}^\alpha, e^\alpha) h^\alpha(\boldsymbol{\phi}) \right\} \right\},$$

and the inner problem is just the definition of $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ in Equation (7.52).

□

It turns out that the standard inversion formula for s in terms of f as

$$s(\boldsymbol{\phi}, \boldsymbol{c}, e) = \inf_T \left\{ \frac{e}{T} - \frac{f(\boldsymbol{\phi}, \boldsymbol{c}, T)}{T} \right\} \tag{7.60}$$

also continues to hold in this setting provided f is defined as in the previous section:

Lemma 8. *Let $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ be defined as in Equation (7.57) (resp. Equation (7.3)) and assume that the local quasi-equilibrium conditions (7.9) allow for a unique (up to the addition of an arbitrary constant vector $\xi \boldsymbol{e}$ to $\boldsymbol{\mu}$) solution which depends smoothly on $\boldsymbol{c}$ and T and that the bulk free energy densities depends smoothly on $\boldsymbol{c}^\alpha$ and T. Then:*

- *Any minimizer T for the problem (7.60) is characterized by the averaging relation $e = \sum_{\alpha=1}^N e^\alpha\big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\big)$ with respect to the external paramter e, where the $\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T)$ are the ones defined in the optimality system for the definition of the f-function in Equation (7.3) and*
$$e^\alpha(\boldsymbol{c}^\alpha, T) := -T^2 \frac{\partial}{\partial T}\Big(\frac{f^\alpha(\boldsymbol{c}^\alpha, T)}{T}\Big). \tag{7.61}$$
- *The two definitions of $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ through equations (7.52) and (7.60) coincide.*

Proof. • Under the same simplifying smoothness assumptions on the solution of the local quasi-equilibrium condition from Equation (7.9), any minimizer T in the definition (7.60) has to satisfy

$$\begin{aligned} 0 &= -\frac{e}{T^2} + \sum_{\alpha=1}^N \frac{\partial}{\partial T}\Big(\frac{f^\alpha\Big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\Big)}{T}\Big) h^\alpha(\boldsymbol{\phi}) \\ &= -\frac{e}{T^2} + \frac{1}{T^2}\sum_{\alpha=1}^N \Big(\frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}\big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\big)\frac{\partial \boldsymbol{c}^\alpha}{\partial T}\big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\big)\Big) h^\alpha(\boldsymbol{\phi}) \\ &\quad - \frac{\partial}{\partial T}\Big(\frac{f^\alpha}{T}\Big)\big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\big) h^\alpha(\boldsymbol{\phi}), \end{aligned}$$

where the last derivative corresponds to the one of f^α with respect to its second argument only (i.e. while holding $\boldsymbol{c}^\alpha$ fixed).
Similarly to above, using $\frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha} = \boldsymbol{\mu} + \lambda^\alpha \boldsymbol{e}$ together with $\boldsymbol{e} \cdot \frac{\partial \boldsymbol{c}^\alpha}{\partial T} = 0$ and the independence of $\boldsymbol{\mu}$ on α shows that $\sum_{\alpha=1}^N \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha} \cdot \frac{\partial \boldsymbol{c}^\alpha}{\partial T} h^\alpha(\boldsymbol{\phi}) = \boldsymbol{\mu} \cdot \frac{\partial\Big(\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi})\Big)}{\partial T} = 0$ since the $\boldsymbol{c}^\alpha$ in the definition of $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ satisfy the constraint $\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c}$. Using in addition the basic thermodynamic relation $-\frac{e^\alpha(\boldsymbol{c}, T)}{T} = \frac{\partial}{\partial T}\Big(\frac{f^\alpha(\boldsymbol{c}^\alpha, T)}{T}\Big)$ for the bulk free energy density, T thus has to be such that

$$-\frac{e}{T^2} + \sum_{\alpha=1}^N \frac{e^\alpha\big(\boldsymbol{c}^\alpha(\boldsymbol{\phi}, \boldsymbol{c}, T), T\big)}{T^2} h^\alpha(\boldsymbol{\phi}) = 0,$$

from which the claim follows.

- As was already seen in the proof of the previous lemma, the definition of $s(\boldsymbol{\phi}, \boldsymbol{c}, e)$ in Equation (7.52) is such that the phase-specific temperatures T^α defined through $\frac{1}{T^\alpha} = \frac{\partial s^\alpha}{\partial e^\alpha}$ are equal to a common value θ and that the phase-specific concentrations are such that they coincide with the ones obtained from in the definition of $f(\boldsymbol{\phi}, \boldsymbol{c}, \theta)$ in Equation (7.57) and thus correpond to the concentrations in the claim with T being replaced by θ. Furthermore, given any admissible phase-specific concentration $\boldsymbol{c}^\alpha$ and value of the multiplier T^α, it follows from "standard" thermodynamics that the values of the e^α can also be expressed as in equation (7.61)[38] in terms of $\boldsymbol{c}^\alpha$ and T^α. Evaluated for the particular choice $T^\alpha = \theta$ for all α, it follows from the sum-constraint on the e^α in the definition of (7.52) that these $e^\alpha(\boldsymbol{c}^\alpha, \theta)$ in fact satisfy $\sum_{\alpha=1}^N e^\alpha(\boldsymbol{c}^\alpha, \theta) = e$, and thus the condition in the first part of the lemma, from which it follows that $\theta = T$ under the assumption that there is a unique minimizer.

[38] This is basically the phase-specific version of the conclusion obtained from Equation (7.56).

Conversely, any minimizer in Equation (7.60) does, by the use of $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ through Equation (7.57) satisfy the equality of the chemical potentials. Furthermore, it follows from standard thermodynamics that defining $e^\alpha(\boldsymbol{c}^\alpha, T)$ as in Equation (7.61), that the $s^\alpha(\boldsymbol{c}^\alpha, e^\alpha(\boldsymbol{c}^\alpha, T))$ satisfy $\frac{1}{T} = \frac{\partial s}{\partial e^\alpha}\big(\boldsymbol{c}^\alpha, e^\alpha(\boldsymbol{c}^\alpha, T)\big)$. Combined with the necessary condition for the minimizer in Equation (7.60), it follows that this choice of $(\boldsymbol{c}^\alpha, T)$ satisfies all the conditions required for the minimizer in Equation (7.52).

□

Remark 110. Note that the reasoning above relies mostly on a combination of two arguments, namely that standard thermodynamic relations holding at a phase-specific level and the optimality conditions for the phase-averaged quantities, ensuring that the contributions due to additional implicit dependencies drop out of the differential relations. It is likely that one can similarly extend most of the well-known "basic" thermodynamic relations to the phase-average setting in a very similar manner[39]. ◇

The Resulting Driving Force

The driving force for the phasefield equation can be derived in a manner very similar to the isothermal case in Section 7.1.3. The crucial point is again to keep in mind the dependence of the phase-specific concentrations and energies for the maximizers in Equation (7.52) on the phasefield $\boldsymbol{\phi}$. Assuming a smooth dependence and extending, if necessary, the optimality conditions in Equation (7.55) to those phases with $h^\alpha(\boldsymbol{\phi}) = 0$, the differentiation of s with respect to ϕ^α leads to

$$\begin{aligned}\frac{\partial s(\boldsymbol{\phi}, \boldsymbol{c}, T)}{\partial \phi^\alpha} &= \frac{\partial}{\partial \phi^\alpha}\Big(\sum_\beta s^\beta\big(\boldsymbol{c}^\beta(\boldsymbol{\phi}, \boldsymbol{c}, e), e^\beta(\boldsymbol{\phi}, \boldsymbol{c}, e)\big)h^\beta(\boldsymbol{\phi})\Big) \\ &= \sum_\beta s^\beta\big(\boldsymbol{c}^\beta(\boldsymbol{\phi}, \boldsymbol{c}, e), e^\beta(\boldsymbol{\phi}, \boldsymbol{c}, e)\big)\frac{\partial h^\beta}{\partial \phi^\alpha} + \sum_\beta\Big(\frac{\partial s^\beta}{\partial \boldsymbol{c}^\beta} \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha} + \frac{\partial s^\beta}{\partial e^\beta}\frac{\partial e^\beta}{\partial \phi^\alpha}\Big)h^\beta(\boldsymbol{\phi}).\end{aligned}$$

By construction, it holds that $\frac{\partial s^\beta}{\partial \boldsymbol{c}^\beta} = -\frac{\boldsymbol{\mu} + \lambda^\beta \boldsymbol{e}}{T}$ and $\frac{\partial s^\beta}{\partial e^\beta} = \frac{1}{T}$. Extracting the phase-independent quantities $\boldsymbol{\mu}$ and T out of the sum and making use of $\boldsymbol{e} \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha} = 0$ due to the sum-constraint $\boldsymbol{e} \cdot \boldsymbol{c}^\beta = 1$, it follows that

$$\sum_\beta\Big(\frac{\partial s^\beta}{\partial \boldsymbol{c}^\beta} \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha} + \frac{\partial s^\beta}{\partial e^\beta}\frac{\partial e^\beta}{\partial \phi^\alpha}\Big)h^\beta(\boldsymbol{\phi}) = -\frac{\boldsymbol{\mu}}{T} \cdot \sum_\beta \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}) + \frac{1}{T}\sum_\beta \frac{\partial e^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}).$$

The derivatives of the phase-specific quantities can be eliminated by an application of the product-rule since

$$\sum_\beta \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}) = \frac{\partial \sum_\beta \boldsymbol{c}^\beta h^\beta(\boldsymbol{\phi})}{\partial \phi^\alpha} - \sum_\beta \boldsymbol{c}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha} = \frac{\partial \boldsymbol{c}}{\partial \phi^\alpha} - \sum_\beta \boldsymbol{c}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha} = -\sum_\beta \boldsymbol{c}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$$

and similarly $\sum_\beta \frac{\partial e^\beta}{\partial \phi^\alpha}h^\beta(\boldsymbol{\phi}) = -\sum_\beta e^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$, leaving $\frac{\partial s(\boldsymbol{\phi}, \boldsymbol{c}, T)}{\partial \phi^\alpha} = \sum_\beta\Big(s^\beta + \frac{\boldsymbol{\mu}}{T}\boldsymbol{c}^\beta - \frac{1}{T}e^\beta\Big)\frac{\partial h^\beta}{\partial \phi^\alpha}$. This can further be simplified as $s^\beta - \frac{1}{T}e^\beta = -\frac{1}{T}(e^\beta - Ts^\beta)$, whose value, as in Equation (7.59), coincides with $-\frac{1}{T}f^\beta(\boldsymbol{c}^\beta, T)$, now as a function of $\boldsymbol{c}^\beta$ and T. This finally leads to

$$\frac{\partial s(\boldsymbol{\phi}, \boldsymbol{c}, T)}{\partial \phi^\alpha} = \sum_\beta\Big(-\frac{1}{T}f^\beta(\boldsymbol{c}^\beta, T) + \frac{\boldsymbol{\mu}}{T}\boldsymbol{c}^\beta\Big)\frac{\partial h^\beta}{\partial \phi^\alpha} = -\frac{1}{T}\sum_\beta\big(f^\beta - \boldsymbol{\mu} \cdot \boldsymbol{c}^\beta\big)\frac{\partial h^\beta}{\partial \phi^\alpha}, \tag{7.62}$$

and thus again an expression based on the phase-specific grand chemical potential densities Ψ^β.

[39] Recall also that it was already observed in Section 7.1.4 that $f(\boldsymbol{\phi}, \boldsymbol{c}, T)$ and $\Omega(\boldsymbol{\phi}, \boldsymbol{\mu}, T)$ are related through a standard Legendre transform.

The Solution Process

As already indicated in Remark 109 and as the discussion in the previous section shows, there in fact several different ways one can choose the "primary unknowns" and thus to solve the local quasi-equilibrium conditions in the non-isothermal case.

1. The first - and most obvious one - would be to apply a standard Newton-scheme directly to the original System (7.54), i.e. given the functions $s^\alpha(\boldsymbol{c}^\alpha, e^\alpha)$ and a set of values $\left(\boldsymbol{c}^\alpha, e^\alpha\right)^{(n)}_{1\leq\alpha\leq N}$ for the phase-specific concentrations and energies as well as some value $\left(\boldsymbol{\eta}, \beta, (\hat{\lambda}^\alpha)_{1\leq\alpha\leq N}\right)^{(n)}$ for the associated multipliers, to solve for a set of corrections based on the linearization of Equation (7.54). By linearity in the multipliers $\left(\boldsymbol{\eta}, \beta, (\hat{\lambda}^\alpha)_{1\leq\alpha\leq N}\right)$, this can, as in Section 7.1.5, in effect be reduced to an iteration based on the $\left(\boldsymbol{c}^\alpha, e^\alpha\right)^{(n)}_{1\leq\alpha\leq N}$ alone, with, again by linearity, the residuals in the sum-constraints $\sum_{\alpha=1}^N \boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{c}$ and $\sum_{\alpha=1}^N e^\alpha h^\alpha(\boldsymbol{\phi}) = e$ vanishing after the first iteration. This has the advantage of on the one hand requiring only direct evaluations of the derivatives of the (assumed to be given) $s^\alpha(\boldsymbol{c}^\alpha, e^\alpha)$ and on the other hand ensuring "conservation" of the phase-averaged conserved $(\boldsymbol{c}, e)$ variables after a single Newton-step, regardless of the accuracy of the estimates for the multipliers.
   The major disadvantage is that this on the one hand requires working with the - less commonly used - functions $s^\alpha(\boldsymbol{c}^\alpha, e^\alpha)$ and on the other hand leads to a system with total of $(K+1)N$ "primary" unknowns $\left(\boldsymbol{c}^\alpha, e^\alpha\right)_{1\leq\alpha\leq N}$ in addition to the $K+N+1$ multipliers $\left(\boldsymbol{\eta}, \beta, (\hat{\lambda}^\alpha\right)$ (resp. KN primary unknowns $\left(\tilde{\boldsymbol{c}}^\alpha, e^\alpha\right)_{1\leq\alpha\leq N}$ and K unknown multipliers $(\tilde{\boldsymbol{\mu}}^\alpha, \beta)$ after a reduction to $K-1$ independent concentration values as in Section 7.1.5).

2. A quite natural alternative is suggested by the discussion in Section 7.1.5 in the isothermal case. One can directly make use of the equilibrium condition $T^\alpha = T$ and then to use the equivalent System (7.55), but instead based on the $(\boldsymbol{c}^\alpha, T^\alpha)_{1\leq\alpha\leq N}$ as primary unknowns.
   One advantage of this approach is that, since the local quasi-equilibrum conditions from Equation (7.9) in the isothermal case are implicitly also based upon the assumption that each T^α corresponds the the (given) value T, the system in Equation (7.55) can basically be considered to be the same as in the previous case. The only difference is that now the temperature T forms part of the unknowns and has to be fixed such that the averaging condition $\sum_{\alpha=1}^N e^\alpha(\boldsymbol{c}^\alpha, T) h^\alpha(\boldsymbol{\phi}) = e$ - now with the e^α as explicit functions of $\boldsymbol{c}^\alpha$ and T - holds.
   A second one is that, as the trivial constaint $T^\alpha = T$ can directly be integrated into the system. As the $(e^\alpha)_{1\leq\alpha\leq N}$ are here considered to be known for given values of $(\boldsymbol{c}^\alpha, T)$, one can reduce the number of primary unknowns to the $KN+1$ values $\left((\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N}, T\right)$ and $K+N$ additional secondary unkowns through the multipliers $\left(\boldsymbol{\mu}, (\lambda^\alpha)_{1\leq\alpha\leq N}\right)$. Using a reduced formulation, furhter eliminates unknowns, leaving only $(K-1)N+1$ primary unknowns $\left((\tilde{\boldsymbol{c}}^\alpha)_{1\leq\alpha\leq N}, T\right)$ and the $(K-1)$ values of the muliplier $\tilde{\boldsymbol{\mu}}$.
   One disadvantage - if based on storing $(\boldsymbol{c}, T)$ as the primary unknowns instead of $(\boldsymbol{c}, e)$ - is that one now needs to ensure that the (local) quasi-equilibrium systems are solved accurately in order to ensure conservation of energy. This is due to the fact that the function $e^\alpha(\boldsymbol{c}^\alpha, T)$ are now in general nonlinear in both $\boldsymbol{c}^\alpha$ and T. The same then holds for the the sum-constraint $\sum_{\alpha=1}^N e^\alpha(\boldsymbol{c}^\alpha, T) = e$, which will therefore not be solved exactly with a single correction step[40].

[40] Note that this can arise even for quite simple dependencies of the energies on the concentration and temperature including e.g. a product of a linearly (in $\boldsymbol{c}^\alpha$) interpolated specific heat capacity $(c_v)^\alpha(\boldsymbol{c}^\alpha)$ with the temperature. Even both factor are linear, the product obviously is not.

3. A third approach - which is closely related to the approach suggested in [19] - is to use a fully "dual" description, i.e. by also replacing the $\boldsymbol{c}^\alpha$ as functions of $\left(\boldsymbol{\mu}, (\lambda^\alpha)_{1\leq\alpha\leq N}, T\right)$ and (thus also, through the $\boldsymbol{c}^\alpha$, the e^α as functions of the same paramters) and then applying a Newton-scheme on this set of equations. A clear advantage is that this leads to an even further reduction of the primary unknowns to the $K+N+1$ values (resp. the K values $(\tilde{\boldsymbol{\mu}}, T)$ when using a reduced version) as the $(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N}$ are now taken as direct functions of the $\boldsymbol{\mu}^\alpha = \boldsymbol{\mu} + \lambda^\alpha \boldsymbol{e}$ and T (resp. $\tilde{\boldsymbol{\mu}}$ and T).

Which approach is more appropriate again depends quite heavily on which thermodynamic potential(s) can be considered as "given". Even though the increasing reduction in the number of unknowns above a priori seems highly favorable, it is subject to the same pitfalls as the choice of working with $\tilde{\boldsymbol{\mu}}$ as the primary unknown already discussed in Remark 104. In particular, if based on $s^\alpha(\boldsymbol{c}^\alpha, e^\alpha)$ as the basic thermodynamics potential, using $\left((\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N}, T\right)$ instead of $(\boldsymbol{c}^\alpha, e^\alpha)_{1\leq\alpha\leq N}$ as the primary unknowns is in general highly inconvenient as the solution of the subsystems $\frac{\partial s^\alpha}{\partial e^\alpha}(\boldsymbol{c}^\alpha, e^\alpha) = \frac{1}{T}$ for determining the (then dependent) e^α in terms of the (assumed given) $\boldsymbol{c}^\alpha$ and T can itself require a number of Newton-steps. Even though this is scalar equation in a single unknown, this is still likely an effort ill spent. One can of course then proceed similarly to Remark 104 and allow for controlled inaccuracies, this leading, similarly to the isothermal case based on $\tilde{\boldsymbol{\mu}}$ as the primary unkown, essentialy back to first approach. The situation is even worse when taking $(\tilde{\boldsymbol{\mu}}^\alpha, T)$ as the primary unknowns, as this then requires, for each evaluation of $\frac{\partial^2 s^\alpha}{\partial(\boldsymbol{c}^\alpha)^2}$, $\frac{\partial^2 s^\alpha}{\partial \boldsymbol{c}^\alpha \partial e^\alpha}$ and $\frac{\partial^2 s^\alpha}{\partial(e^\alpha)^2}$ solving the systems

$$\frac{\partial s^\alpha}{\partial e^\alpha}(\tilde{\boldsymbol{c}}^\alpha, e^\alpha) \stackrel{!}{=} \frac{1}{T} \quad \text{and} \quad \frac{\partial s^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}(\tilde{\boldsymbol{c}}^\alpha, e^\alpha) \stackrel{!}{=} \frac{\tilde{\boldsymbol{\mu}}^\alpha}{T}$$

to a sufficiently high accuracy.
If based on $f^\alpha(\boldsymbol{c}^\alpha, T)$ as the basic thermodynamic potential, as is currently the standard in the **Pace3D**-framework, the use of $\left((\boldsymbol{c}^\alpha)_{1\leq\alpha\leq N}, T\right)$ is in general the most convenient one, and will therefore be the one considered here.

Given that the fundamental conservative evolution equations are based on the concentration $\boldsymbol{c}$ and the energy e, the requirement is therefore to find phase-specific concentrations $\boldsymbol{c}^\alpha$ and energies e^α compatibly with the optimality conditions for Equation (7.52) given the values of the average concentration $\boldsymbol{c}$ and the average energy e. Since it is the f^α as functions of $\boldsymbol{c}^\alpha$ and T^α which are, for practical reasons, chosen as the fundamental potentials, the natural parameterization for doing so is also in terms of the concentration values $\boldsymbol{c}^\alpha$ and the - known to have to be a common value for all α - temperature T.

Given some initial guess $\boldsymbol{c}^{(0)}$ and $T^{(0)}$, it is again natural to use a Newton-scheme for the solution of the optimality system in Equation (7.55).

Remark 111. As already indicated, a parameterization in terms of T instead of e in principle has the same drawbacks already discussed in Remarks 104 and 105. Being based on f^α formulated directly in terms of T, the constructions are usually such that the conversion of a given concentration and temperature to the corresponding energy is relatively simple, and thus less problematic. It will therefore be assumed in the following that obtaining the phase-specific energies e^α as a function of $(\boldsymbol{c}^\alpha, T)$ is computationally relatively cheap.
If such is not the case, one may have to apply similar considerations as in Remark 104. ◇

Given some initial guess such e.g. $\left(\boldsymbol{c}^\alpha\right)^{(0)} = \boldsymbol{c}$, $\left(e^\alpha\right)^{(0)} = e^\alpha\left(\left(\boldsymbol{c}^\alpha\right)^{(0)}, T^{(0)}\right)$, the basic Newton-

step consists in solving

$$\begin{cases} \frac{\partial \boldsymbol{\mu}^\alpha}{\partial \boldsymbol{c}^\alpha}\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big)\delta\boldsymbol{c}^\alpha + \frac{\partial \boldsymbol{\mu}^\alpha}{\partial T}\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big)\delta T - \boldsymbol{\mu}^{(n+1)} - \big(\lambda^\alpha\big)^{(n+1)}\boldsymbol{e} \overset{!}{=} -\boldsymbol{\mu}^\alpha\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big) \\ \sum_{\alpha=1}^N \delta\boldsymbol{c}^\alpha h^\alpha(\boldsymbol{\phi}) \overset{!}{=} \boldsymbol{c} - \sum_{\alpha=1}^N \big(\boldsymbol{c}^\alpha\big)^{(n)} h^\alpha(\boldsymbol{\phi}) =: \boldsymbol{r}_{\boldsymbol{c}} \\ \sum_{\alpha=1}^n \left(\frac{\partial e^\alpha}{\partial \boldsymbol{c}^\alpha}\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big)\delta\boldsymbol{c}^\alpha + \frac{\partial e^\alpha}{\partial T}\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big)\delta T \right) h^\alpha(\boldsymbol{\phi}) \overset{!}{=} e - \sum_{\alpha=1}^N e^\alpha\Big(\big(\boldsymbol{c}^\alpha\big)^{(n)}, T^{(n)}\Big) =: r_e \end{cases}$$

Due to the same considerations as in Section 7.1.5, this system is again most easily solved by reverting to a reduced from, either by a priori using a reduced formulation or using an algebraic "a posteriori" reduction. Dropping the arguments for notational simplicity, this results in having to solve the simpler system[41]

$$\begin{cases} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\delta\tilde{\boldsymbol{c}}^\alpha + \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T}\delta T - \tilde{\boldsymbol{\mu}}^{(n+1)} \overset{!}{=} -\big(\tilde{\boldsymbol{\mu}}^\alpha\big)^{(n)} \\ \sum_{\alpha=1}^N \delta\tilde{\boldsymbol{c}}^\alpha h^\alpha \overset{!}{=} \boldsymbol{c} - \sum_{\alpha=1}^N \big(\boldsymbol{c}^\alpha\big)^{(n)} h^\alpha = \boldsymbol{r}_{\tilde{\boldsymbol{c}}} \\ \sum_{\alpha=1}^n \left(\frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\delta\tilde{\boldsymbol{c}}^\alpha + \frac{\partial e^\alpha}{\partial T}\delta T \right) h^\alpha \overset{!}{=} e - \sum_{\alpha=1}^N \big(e^\alpha\big)^{(n)} = r_e. \end{cases} \tag{7.63}$$

Since this system is still in a block-diagonal form with respect to the $\delta\tilde{\boldsymbol{c}}^\alpha$, it is again convenient to perform a block-elimination in this equation, leading, completely analogous to the discussion in Section 7.1.5, to the Schur-complement system

$$\begin{cases} \sum_{\alpha=1}^N \left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} \left(\tilde{\boldsymbol{\mu}}^{(n+1)} - \big(\tilde{\boldsymbol{\mu}}^\alpha\big)^{(n)} - \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T}\delta T \right) h^\alpha \overset{!}{=} r_{\tilde{\boldsymbol{c}}} \\ \sum_{\alpha=1}^n \left(\frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} \left(\tilde{\boldsymbol{\mu}}^{(n+1)} - \big(\tilde{\boldsymbol{\mu}}^\alpha\big)^{(n)} - \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T}\delta T \right) + \frac{\partial e^\alpha}{\partial T}\delta T \right) h^\alpha \overset{!}{=} r_e, \end{cases} \tag{7.64}$$

consisting of a total of K unkowns, $K-1$ ones in terms of $\tilde{\boldsymbol{\mu}}$ and one in terms of δT.
K is, due to computational requirements, usually quite small (typically below 4 or 5) due to the fact that the concentration evolution - unlike an obstacle-potential based phasefield equation - has to be calculated everywhere. A very natural choice for determining the values of $\delta\boldsymbol{\mu}$ and δT in Equation (7.64) is therefore simply to use any direct solver, since this is a non-singular system of typically very small size. Given the values of $\tilde{\boldsymbol{\mu}}$ and δT, it is then again an easy matter of updating the remaining phase-specific quantities $\delta\tilde{\boldsymbol{c}}^\alpha$ based on Equation (7.63), in which all phases are then uncoupled.
Once the iterations converge, one can then, if required recover the values of $\boldsymbol{\mu}$ and the λ^α as in Subsection 7.1.5, or, if not so, simply recover the concentration values c_K^α based on the sum-constraint $\sum_{i=1}^K c_i^\alpha = 1$.

Remark 112. In terms of code-reusability, it may nevertheless have some interest to, in the same spirit as in Subsection 7.1.5, perform another block-reduction on Equation (7.64). First moving all known quantities to the right-hand side, this system can be rewritten as

$$\begin{cases} \boldsymbol{S}_{\tilde{\boldsymbol{c}}}\tilde{\boldsymbol{\mu}}^{(n+1)} - \left(\sum_{\alpha=1}^N \left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T} h^\alpha \right)\delta T = \boldsymbol{r}_{\tilde{\boldsymbol{c}}} + \sum_{\alpha=1}^N \left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} \big(\tilde{\boldsymbol{\mu}}^\alpha\big)^{(n)} h^\alpha =: \hat{\boldsymbol{r}}_{\tilde{\boldsymbol{c}}}, \\ \left(\sum_{\alpha=1}^n \left(\frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} h^\alpha \right) \tilde{\boldsymbol{\mu}}^{(n+1)} + S_T\delta T = r_e + \sum_{\alpha=1}^n \frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\left(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\right)^{-1} \big(\tilde{\boldsymbol{\mu}}^\alpha\big)^{(n)} h^\alpha =: \hat{r}_e, \end{cases} \tag{7.65}$$

[41] Introducing the reduction into the term $\frac{\partial e^\alpha}{\partial \boldsymbol{c}^\alpha}$ is equally simple as in Subsection 7.1.5 using $c_K^\alpha = 1 - \sum_{i=1}^{K-1} c_i^\alpha$.

where $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ is the same Schur-complement matrix as in Equation (7.37) in Section 7.1.5 and S_T is the analogous "matrix" (this is simply a scalar value) to $\boldsymbol{S}_{\tilde{\boldsymbol{c}}}$ for the temperature increment given by

$$S_T = \sum_{\alpha=1}^{n} \left(\frac{\partial e^\alpha}{\partial T} - \frac{\partial e^\alpha}{\partial \boldsymbol{c}^\alpha} \Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T} \right) h^\alpha. \tag{7.66}$$

The two basic choices for this are to either make use of the first equation in Equation (7.64) for eliminating $\tilde{\boldsymbol{\mu}}^{(n+1)}$ as a function of δT,

$$\tilde{\boldsymbol{\mu}}^{(n+1)}(\delta T) = \boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \hat{\boldsymbol{r}}_{\tilde{\boldsymbol{c}}} + \boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \sum_{\alpha=1}^{N} \left(\Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T} h^\alpha \right) \delta T$$

or the elimination of δT as a function of $\tilde{\boldsymbol{\mu}}^{(n+1)}$ using the second equation, leading to

$$\delta T\big(\tilde{\boldsymbol{\mu}}^{(n+1)}\big) = \frac{1}{S_T} \hat{r}_e - \frac{1}{S_T} \left(\sum_{\alpha=1}^{n} \frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} h^\alpha \right) \tilde{\boldsymbol{\mu}}^{(n+1)}.$$

Inserting these two expressions into the respective other equation in the System (7.65), the first choice leads a scalar equation in δT with a modified Schur-complement

$$S_T + \left(\sum_{\alpha=1}^{n} \frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} h^\alpha \right) \boldsymbol{S}_{\tilde{\boldsymbol{c}}}^{-1} \sum_{\alpha=1}^{N} \left(\Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T} h^\alpha \right)$$

for the temperature increment, whereas the second choice leads to a modified equation for $\tilde{\boldsymbol{\mu}}^{(n+1)}$ on the modified Schur-complement

$$\boldsymbol{S}_{\tilde{\boldsymbol{c}}} + \frac{1}{S_T} \left(\sum_{\alpha=1}^{N} \Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial T} h^\alpha \right) \left(\sum_{\alpha=1}^{n} \left(\frac{\partial e^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} \Big)^{-1} h^\alpha \right) \right)$$

for the reduced chemical potential.
If the code-functionality is primarily designed with the isothermal case in mind, the former choice is likely the more convenient one, since this makes it possible to perform the solution of the System (7.63) essentially without modification to the "purely chemical" part, as one can reuse this functionality in an exterior manner for determining δT. Once δT is known, the remaining system in $\tilde{\boldsymbol{\mu}}^{(n+1)}$ and the $\delta \tilde{\boldsymbol{c}}^\alpha$ then reduces to the one in Section 7.1.5 with slightly modified right-hand sides.
A similar construction with a larger number of additional unknowns will be discussed in more detail in Section 7.2.5. ◇

Remark 113. Note that it is also possible to obtain slightly different expressions making use of some thermodynamic relations. For example, based on $e^\alpha(\boldsymbol{c}^\alpha, T) = -T^2 \frac{\partial}{\partial T}\big(\frac{f^\alpha(\boldsymbol{c}^\alpha, T)}{T} \big)$, one has

$$\begin{aligned} \frac{\partial e^\alpha}{\partial \boldsymbol{c}^\alpha} &= -T^2 \frac{\partial^2}{\partial \boldsymbol{c}^\alpha \partial T} \Big(\frac{f^\alpha(\boldsymbol{c}^\alpha, T)}{T} \Big) = -T^2 \frac{\partial}{\partial T} \Big(\frac{1}{T} \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha} \Big) = -T^2 \frac{\partial}{\partial T} \Big(\frac{\boldsymbol{\mu}^\alpha(\boldsymbol{c}^\alpha, T)}{T} \Big) \\ &= -T^2 \Big(-\frac{\boldsymbol{\mu}^\alpha(\boldsymbol{c}^\alpha, T)}{T^2} + \frac{1}{T} \frac{\partial \boldsymbol{\mu}^\alpha}{\partial T} \Big) = \boldsymbol{\mu}^\alpha(\boldsymbol{c}^\alpha, T) - T \frac{\partial \boldsymbol{\mu}^\alpha}{\partial T}. \end{aligned}$$

In terms of "readibility", an obvious definition to make use of is $c_v^\alpha = \frac{\partial e^\alpha}{\partial T}$. ◇

7.2 Quantitative Models for Solid-Solid Transformations

Similarly to the solid-liquid transformations considered in the previous section, the phasefield method has also gained considerable interest for simulating the evolution of microstructures consisting of different solid phases. Such solid-solid state transformations are, in addition to the surface energy of the various phases, also heavily influenced by stored mechanical energy contributions.
In some cases, for example during recrystallization processes, this additional stored energy distribution can be reasonably approximated as being a function of the phase only, thus allowing for a simpler model as in Section 6 with the phasefields being the only unknowns (see e.g. [78]). In other cases, the phase-transformation process itself heavily influences the mechanical state of the system, requiring a more complex treatment in terms of a coupled phasefield-mechanical model, where now there is (at least) one additional set of unknowns in terms of the displacement field $\boldsymbol{u}$.

Phasefield models for solid-solid transformations are usually again based upon an appropriately chosen energy functional in combination with a gradient-flow postulate on the dynamics of the phasefield variables. In the small deformation setting, the phasefield functional is therefore usually chosen (see e.g. [69] and [3] as some of the earlier works coupling phasefield methods with elasticity) as consisting of the standard phasefield contributions through a and w supplemented by the total elastic energy

$$\mathcal{F}_\epsilon(\boldsymbol{\phi}, \boldsymbol{u}) = \int_\Omega \epsilon a(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}) + \frac{1}{\epsilon} w(\boldsymbol{\phi}) \,\mathrm{d}\boldsymbol{x} + \mathcal{F}_{el}(\boldsymbol{\phi}, \boldsymbol{u}) \tag{7.67}$$

where, assuming for example a mixture of displacement and traction boundary conditions on the Dirichlet-part Γ_D and the Neumann-part Γ_N, $\mathcal{F}_{el}(\boldsymbol{\phi}, \boldsymbol{u})$ is given by[42]

$$\mathcal{F}_{el}(\boldsymbol{\phi}, \boldsymbol{u}) = \int_\Omega f_{el}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}(\boldsymbol{u})\big) \,\mathrm{d}\boldsymbol{x} - \rho\boldsymbol{f} \cdot \boldsymbol{u} \,\mathrm{d}\boldsymbol{x} - \int_{\Gamma_N} \boldsymbol{g} \cdot \boldsymbol{u} \,\mathrm{d}\boldsymbol{s} \tag{7.68}$$

with the (here volumetric) strain energy density $f_{el}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}(\boldsymbol{u})\big)$ depending on $\boldsymbol{u}$ only through the symmetric displacement gradient $\boldsymbol{\epsilon}(\boldsymbol{u}) = \nabla_S \boldsymbol{u}$.

In contrast to the solidification models considered in Section 7.1, the evolution of the mechanical state is not driven by a (typically comparatively slow) diffusion process, but by a wave-propagation process which tends to quickly equilibrate itself until a mechanical equilibrium is reached. As this equilibriation often takes place at a much shorter timescale than that of the evolution of the phase boundaries[43], it is common to employ a quasi-static approximation for the mechanical fields.

Remark 114. While it is well-known that the propagation of a perfectly elastic wave conserves energy, there are additional dissipative effects such as the radiation of wave energy into the environment and damping of the waves within the material. These will in reality ultimately lead to the establishment of a mechanically equilibrated state. The validity of the quasi-static approximation therefore depends upon how fast this equilibration takes place as compared to the other processes involved. ◇

[42] Note that the contribution due to $\boldsymbol{f}$ and the boundary contribution is often left out, even when considering loads through body forces and/or external stresses. While these do not change the resulting phasefield evolution equation if one does a purely formal differentiation (i.e. "forgetting" the coupling of $\boldsymbol{u}$ to $\boldsymbol{\phi}$), it is crucial for the actual validity of this calculation.

[43] This need not always be the case tough as for example martensitic phase transformations occur at a very fast rate.

Under this quasi-static assumption, $\boldsymbol{u}$ is subject to the mechanical equilibrium condition

$$\begin{cases} -\operatorname{div}\Big(\boldsymbol{\sigma}\big(\boldsymbol{\phi},\boldsymbol{\epsilon}(\boldsymbol{u})\big)\Big) = \boldsymbol{f} & \text{in } \Omega, \\ \boldsymbol{u} = \boldsymbol{U} & \text{on } \Gamma_D, \\ \boldsymbol{\sigma}\cdot\boldsymbol{n} = \boldsymbol{g} \text{ on } \Gamma_N \end{cases} \tag{7.69}$$

where $\boldsymbol{\sigma}$ denotes the Cauchy stress tensor. Note that in order to be consistent with a variational principle based on the functional $\mathcal{F}_\epsilon(\boldsymbol{\phi},\boldsymbol{u})$ from Equation (7.67) and with Equation (7.68), $\boldsymbol{\sigma}\big(\boldsymbol{\phi},\boldsymbol{\epsilon}(\boldsymbol{u})\big)$ in addition has to satisfy (see the discussion below)

$$\boldsymbol{\sigma}\big(\boldsymbol{\phi},\boldsymbol{\epsilon}(\boldsymbol{u})\big) = \frac{\partial f_{el}\big(\boldsymbol{\phi},\boldsymbol{\epsilon}\big)}{\partial\boldsymbol{\epsilon}}\big(\boldsymbol{\phi},\boldsymbol{\epsilon}(\boldsymbol{u})\big). \tag{7.70}$$

As the external loads through $\boldsymbol{f}$ and $\boldsymbol{g}$ are to be considered as given data, it is clear that the only point in which the various phasefield models differ lies in the definition of this strain energy density f_{el}. Alternatively, and which amounts up to an integration constant to the same provided[44] $\boldsymbol{\mathcal{C}}(\boldsymbol{\phi},\boldsymbol{\epsilon}) \coloneqq \frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\epsilon}}$ obeys the usual symmetry condtions compatible with $\boldsymbol{\mathcal{C}}(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \frac{\partial^2 f_{el}}{\partial\boldsymbol{\epsilon}^2}$, one may also define a stress-strain relationship $\boldsymbol{\sigma}(\boldsymbol{\phi},\boldsymbol{\epsilon})$.
Following the former approach and allowing for **eigenstrains (prestrains)** $\tilde{\boldsymbol{\epsilon}}^\alpha$, this strain energy density within an α-bulk region is given by (see e.g. [69])

$$f_{el}^\alpha(\boldsymbol{\epsilon}^\alpha) = \frac{1}{2}\boldsymbol{\epsilon}_{el}^\alpha : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}_{el}^\alpha = \frac{1}{2}\big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big) : \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big),$$

where the elastic strain $\boldsymbol{\epsilon}_{el}^\alpha$ is defined as the difference of the total and prestrains, $\boldsymbol{\epsilon}_{el}^\alpha = \boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha$.
As for the solidification problems considered in the previous section, the primary difficulty for designing an accurate phasefield model thus lies in properly extending this definition to within the interface regions. Based on the bulk-expression, the natural approach within the phasefield context is again the use of a weighted interpolation

$$f_{el}\Big(\boldsymbol{\phi},(\boldsymbol{\epsilon}^\alpha)_{1\le\alpha\le N}\Big) = \sum_{\alpha=1}^N f_{el}^\alpha(\boldsymbol{\epsilon}^\alpha)h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^N \Big(\frac{1}{2}\big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big) : \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)\Big)h^\alpha(\boldsymbol{\phi}) \tag{7.71}$$

of the elastic free energy contributions of the individual phases. This is still an incomplete description though as it remains to specify the dependence of the phase-specific strains $\boldsymbol{\epsilon}^\alpha$ on the total strain $\boldsymbol{\epsilon}$.

One natural restriction is that one would like for the relation $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$ to hold. A second one - which is typically imposed in accordance with the standard approach of representing the total quantities as an interpolation of the various phases - is that the total strain and the total stresses be related to the phase-specific ones through

$$\boldsymbol{\epsilon} = \sum_{\alpha=1}^N \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi}) \quad \text{and} \quad \boldsymbol{\sigma} = \sum_{\alpha=1}^N \boldsymbol{\sigma}^\alpha h^\alpha(\boldsymbol{\phi}). \tag{7.72}$$

This still leaves a large degree of flexibility in the construction of an effective material behavior though.

[44] Even though, at least in the linearly elastic case, it is natural here to assume that $\boldsymbol{\mathcal{C}}$ is independent of $\boldsymbol{\epsilon}$ (i.e. that the diffuse interface stress-strain relationship is linear in $\boldsymbol{\epsilon}$), this is not strictly necessary. While a nonlinear stress-strain relationship a priori would then violate the linearity of the original problem, this could still be of interest if this "violation" is carefully designed. One example of a similar situation is the use of nonlinear advection schemes for an a priori linear advection operator.

In terms of its contribution to the phasefield Equation (6.73), the minimization of $\mathcal{F}_\epsilon$ in Equation (7.67) subject to the quasi-static equilibrium conditions (7.69) a priori leads to a constrained optimization problem in terms of $\boldsymbol{\phi}$ and $\boldsymbol{u}$. Equation (7.69) is, under mild hypothesis, uniquely (or uniquely up to some irrelevant kernel) solvable for $\boldsymbol{u}$ as a function of $\boldsymbol{\phi}$. One can thus again consider a reduced formulation consisting in the minimization of $\mathcal{F}_\epsilon\big(\boldsymbol{\phi}, \boldsymbol{u}(\boldsymbol{\phi})\big)$ in terms of which the FONC for any minimzer becomes

$$\Big\langle \frac{\partial \mathcal{F}_\epsilon}{\partial \boldsymbol{\phi}}, \delta\boldsymbol{\phi} \Big\rangle + \Big\langle \frac{\partial \mathcal{F}_\epsilon}{\partial \boldsymbol{u}}, \delta\boldsymbol{u}(\delta\boldsymbol{\phi}) \Big\rangle = \langle \boldsymbol{\mu} + \Lambda, \delta\boldsymbol{\phi} \rangle$$

with $\boldsymbol{\mu}$ and Λ as in Section 6.1.
As $\boldsymbol{u}$ only enters in $\mathcal{F}_{el}(\boldsymbol{\phi}, \boldsymbol{u})$ and the system (7.69) is, under the variational consistency condition (7.70), precisely the condition $\langle \frac{\partial \mathcal{F}_{el}(\boldsymbol{\phi}, \boldsymbol{u})}{\partial \boldsymbol{u}}, \delta\boldsymbol{u} \rangle = 0$ for all admissible variations $\delta\boldsymbol{u}$. The second contribution therefore actually vanishes and one only has to consider the additional "driving force" through the direct contribution $\frac{\partial \mathcal{F}_{el}}{\partial \boldsymbol{\phi}}$ to the phasefield equation. Assuming for simplicity that neither the body force $\rho\boldsymbol{f}$ nor the imposed boundary conditions depend on $\boldsymbol{\phi}$, the only explicit dependence on $\boldsymbol{\phi}$ is therefore through the strain energy density $f_{el}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}(\boldsymbol{u})\big)$ in Equation (7.71). In the simpler models (this will be called the "traditional" phasefield models below), the dependence of f_{el} on $\boldsymbol{\phi}$ is solely in terms of the (local) values of $\boldsymbol{\phi}$ itself and not its gradients. From Equation (7.71) and the chain-rule, this leads to an additional contribution of the form

$$\frac{\partial f_{el}}{\partial \phi^\alpha} = \sum_{\beta=1}^{N} f^\beta(\boldsymbol{\epsilon}^\beta) \frac{\partial h^\beta}{\partial \phi^\alpha}(\boldsymbol{\phi}) + \sum_{\beta=1}^{N} h^\beta(\boldsymbol{\phi}) \frac{\partial f^\beta}{\partial \boldsymbol{\epsilon}^\beta} : \frac{\partial \boldsymbol{\epsilon}^\beta}{\partial \phi^\alpha}.$$

For the more complex models discussed starting from Section 7.2.2, the calculation of f_{el} relies heavily on one or several normal vectors $\boldsymbol{n}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})$ between the various phases, defined in terms of (possibly) the local values $\boldsymbol{\phi}$ and their gradients $\nabla\boldsymbol{\phi}$. It is therefore more convenient to write f_{el} and $\boldsymbol{\epsilon}^\alpha$ as functions of[45] $\big(\boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{n}^{\alpha\beta}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi})\big)_{1 \le \alpha \ne \beta \le N}$, from which one obtains the additional contributions

$$\sum_{\beta=1}^{N} \sum_{\delta=1}^{N} \left(h^\beta(\boldsymbol{\phi}) \frac{\partial f^\beta}{\partial \boldsymbol{\epsilon}^\beta} : \frac{\partial \boldsymbol{\epsilon}^\beta}{\partial \boldsymbol{n}^{\beta\delta}} \cdot \frac{\partial \boldsymbol{n}^{\beta\delta}}{\partial \phi^\alpha} - \nabla \cdot \Big(h^\beta(\boldsymbol{\phi}) \frac{\partial f^\beta}{\partial \boldsymbol{\epsilon}^\beta} : \frac{\partial \boldsymbol{\epsilon}^\beta}{\partial \boldsymbol{n}^{\beta\delta}} \cdot \frac{\partial \boldsymbol{n}^{\beta\delta}}{\partial \nabla \phi^\alpha} \Big) \right).$$

Remark 115. Note the a priori somewhat curious situation that, even though the interpretation of the $\boldsymbol{\epsilon}^\alpha$ is as the phase-inherent strains for the given phase α and thus in particular not a "phase-averaged" quantity, these will, except for the simplest models, still depend implicitly upon (at least) the $\boldsymbol{\phi}$-values. In contrast, the total strain $\boldsymbol{\epsilon}$ - even though actually interpreted as the phase-averaged quantity - is to be considered as independent of the phasefield (and the other parameters).
This situation is completely analogous to Section 7.1, where the average concentration $\boldsymbol{c}$ was considered as phase-independent, whereas the phase-specific concentration values depended upon $\boldsymbol{\phi}$ through their specification in terms of the condition of equal (reduced) chemical potentials. This is clearly an inherent feature of the common approach underlying these phasefield models, which, for obvious efficiency reasons, consists in expressing the phasefield functional and thus the corresponding equilibrium conditions through some variational principle in terms of the phase-averaged quantities as the primary unknown. The phase-specific quantities are only introduced as auxiliary "secondary" quantities through which the bulk free energy potential in the interface region is then (locally) linked to the ones of the individual phases. This is also why there is - except potentially for the bulk-regions - no reason to expect the phase-inherent quantities to

[45] Since both the definition of suitable normal vectors in the multiphase case and the definition of normals at the transition to bulk phases with $\nabla\boldsymbol{\phi} = \boldsymbol{0}$ are by themselves somewhat tricky issue, the contributions due to the dependences of the normal vectors will later on mostly be expressed only in terms of $\frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{n}^{\beta\delta}}$, leaving the specifics of the definition of those aside.

obey the same elegant relations holding for the phase-averaged ones. For example, the $\boldsymbol{\epsilon}^\alpha$ defined by the various models will in general not derive from a global displacement field $\boldsymbol{u}^\alpha$ and neither will the corresponding phase-specific equilibrium condition $-\mathrm{div}(\boldsymbol{\sigma}^\alpha) = \rho^\alpha f^\alpha$ hold throughout the domain (but it will, of course, be valid in the respective bulk-region). ◇

Remark 116. It is of course also possible to consider situations in which the density and/or body force as well as the boundary conditions in addition depend on $\boldsymbol{\phi}$. This is easily dealt with in the former case as, typically being defined in a simple local relation in terms of $\boldsymbol{\phi}$, the contribution to the phasefield equation follows from a straightforward differentiation of this dependence. In contrast, the latter case can, depending upon the complexity of the boundary condition, become significantly more technical to deal with.
If $\boldsymbol{g}$ for example only depends on $\boldsymbol{\phi}$ in a simple local manner such as e.g. when imposing $\boldsymbol{g}$ as an interpolation of normal stresses assigned to the individual phases, this only affects the phasefield equation in the sense that the natural "isolating" boundary condition needs to be replaced by one including $\frac{\partial g}{\partial \phi^\alpha}$[46]. In other cases - quite intuitive from a purely mechanical point of view - such as e.g. for a "constant-force" boundary condition, where a given force is distributed over the boundary based upon some interpolation procedure in terms of $\boldsymbol{\phi}$, one has to deal with a non-local dependence of $\boldsymbol{g}$ on $\boldsymbol{\phi}$ on the Neumann-part of the boundary. This leads to a correspondingly more complex (non-local) contribution to the phasefield boundary condition. ◇

7.2.1 "Traditional" Phasefield Models

Before considering a more recent modeling approach in a little more detail, this section will recall a few key points concerning three more classical modeling approaches, namely the Voigt-Taylor-model, the Reuss-Sachs-model and the model by Khachaturian. As the differences between these models have been amply analyzed in the literature (see e.g. [3], [23], [51] and [64]), the primary purpose of this section is to provide a brief summary of this analysis (and how it relates to the one in the previous section) as a background for the following discussion.

The Purely Elastic Case

Two popular early approaches are to either assume that the strains or the stresses of all phases are equal. In the first approach, the so-called **Voigt-Taylor model**, the assumption $\boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon}\ \forall \alpha$, the first relation in Equation (7.72) is trivially satisfied. Based on imposing $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha$ and the second relation in (7.72), this leads to the total stress

$$\boldsymbol{\sigma}_{VT}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \sum_{\alpha=1}^{N} \boldsymbol{\sigma}^\alpha(\boldsymbol{\epsilon}) h^\alpha(\boldsymbol{\phi}) = \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) : \boldsymbol{\epsilon}, \tag{7.73}$$

i.e. $\boldsymbol{\sigma}$ is determined in terms of the total strain through an effective stiffness tensor given by the (weighted) arithmetic interpolation of the phase-specific ones, $\boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha h^\alpha(\boldsymbol{\phi})$.
In addition, the total local free energy contribution is of the form

$$f_{el,VT}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \sum_{\alpha=1}^{N} \big(\frac{1}{2} \boldsymbol{\epsilon}^\alpha : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha \big) h^\alpha(\boldsymbol{\phi}) = \frac{1}{2} \boldsymbol{\epsilon} : \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) : \boldsymbol{\epsilon} = \frac{1}{2} \boldsymbol{\epsilon} : \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : \boldsymbol{\epsilon}.$$

Based on this form of f_{el} and the Equation (7.73), $\boldsymbol{\sigma}$ is easily seen to satisfy the important "compatibility" condition $\boldsymbol{\sigma}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\phi}, \boldsymbol{\epsilon})$ in Equation (7.70).
This can clearly be considered to be the mechanical analogue of the chemical model used e.g. in [52], whose underlying chemical free energy density can, as discussed above, be interpreted as a h^α-weighted interpolated of the phase-specific ones f^α under the assumption of equal phase-specific concentrations $\boldsymbol{c}^\alpha = \boldsymbol{c}\ \forall \alpha$. In addition, the total chemical potential $\boldsymbol{\mu}(\boldsymbol{\phi}, \boldsymbol{c})$ is, similar to

[46] Depending on the precise form of the implementation, this can nevertheless be somewhat tedious from a practical point of view.

the stress in Equation (7.73), given as the weighted average $\boldsymbol{\mu} = \sum_\alpha \mu^\alpha(\boldsymbol{c}) h^\alpha(\boldsymbol{\phi})$ of the phase-specific chemical potentials $\boldsymbol{\mu}^\alpha(\boldsymbol{c}) = \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}(\boldsymbol{c})$.

In the second approach, the **Reuss-Sachs model** (also called **Steinbach-Apel model** based on their paper [69]), one assumes the equality of the phase-sprecific stresses with the total one, $\boldsymbol{\sigma}^\alpha = \boldsymbol{\sigma}\ \forall \alpha$, instead of that of the strains. This leads to the second equality in (7.72) being trivially satisfied, whereas, using $\boldsymbol{\epsilon}^\alpha = \left(\boldsymbol{\mathcal{C}}^\alpha\right)^{-1} : \boldsymbol{\sigma}^\alpha = \left(\boldsymbol{\mathcal{C}}^\alpha\right)^{-1} : \boldsymbol{\sigma} = \boldsymbol{S}^\alpha : \boldsymbol{\sigma}$, the first one imposes the restriction

$$\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi}) = \Big(\sum_{\alpha=1}^{N} \boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) : \boldsymbol{\sigma} \tag{7.74}$$

and thus corresponds an effective material behavior where one interpolates the individual compliance tensors $\boldsymbol{S}^\alpha = \left(\boldsymbol{\mathcal{C}}^\alpha\right)^{-1}$ instead of the stiffness tensors. Inverting relation (7.74) leads to the explicit stress-strain relationship

$$\boldsymbol{\sigma}_{RS}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \Big(\sum_{\alpha=1}^{N} \boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \Big)^{-1} : \boldsymbol{\epsilon} = \Big(\sum_{\alpha=1}^{N} \left(\boldsymbol{\mathcal{C}}^\alpha\right)^{-1} h^\alpha(\boldsymbol{\phi}) \Big)^{-1} : \boldsymbol{\epsilon} = \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : \boldsymbol{\epsilon}, \tag{7.75}$$

with $\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) = \Big(\sum_{\alpha=1}^{N} \left(\boldsymbol{\mathcal{C}}^\alpha\right)^{-1} h^\alpha(\boldsymbol{\phi}) \Big)^{-1}$ and thus corresponds to a weighted harmonic interpolation of the stiffness tensors. In addition, using $\boldsymbol{\epsilon}^\alpha = \boldsymbol{S}^\alpha : \boldsymbol{\sigma}^\alpha = \boldsymbol{S}^\alpha : \boldsymbol{\sigma}$, the local energy contribution is given by

$$\begin{aligned} f_{el,RS}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) &= \sum_{\alpha=1}^{N} \big(\frac{1}{2} \boldsymbol{\epsilon}^\alpha : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha \big) h^\alpha(\boldsymbol{\phi}) = \frac{1}{2} \boldsymbol{\sigma} : \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \big) \Big) : \boldsymbol{\sigma} \\ &= \frac{1}{2} \Big(\big(\sum_{\alpha=1}^{N} \boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \big)^{-1} : \boldsymbol{\epsilon} \Big) : \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \big) \Big) : \Big(\big(\sum_{\alpha=1}^{N} \boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \big)^{-1} : \boldsymbol{\epsilon} \Big) \\ &= \frac{1}{2} \boldsymbol{\epsilon} : \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \big)^{-1} : \boldsymbol{\epsilon} = \frac{1}{2} \boldsymbol{\epsilon} : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : \boldsymbol{\epsilon} \end{aligned} \tag{7.76}$$

and is in particular again consistent with Equation (7.70).
In contrast to the Voigt-Taylor-model, this corresponds to a mechanical analogue (see also the discussion in [69]) of the KKS-type models considered in Section 7.1. The only slight difference is that in the chemical case one was only able to obtain equality of the reduced chemical potentials (resp. equality of the $\boldsymbol{\mu}^\alpha = \frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha}$ only up to phase-specific constants). since there is no equivalence of the additional "internal" constraint $\sum_{i=1}^{K} c_i^\alpha = 1$ in the mechanical case, one can in fact enforce full equality of the phase-specific stresses.

Remark 117. In relation with the quantitative chemical models from Section 7.1, it is further interesting to consider how the above two models above compare from an energetic point of view. In fact, the Reuss-Sachs model corresponds to the minimizer of the local strain energy density $f_{el}(\boldsymbol{\phi}, \boldsymbol{\epsilon})$ over all decompositions $\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi})$ into phaseinherent strains $(\boldsymbol{\epsilon}^\alpha)_{1 \le \alpha \le N}$ as in Equation (7.72),

$$f_{el,RS}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \min_{\{\boldsymbol{\epsilon}^\alpha : \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{\epsilon}\}} \sum_{\alpha=1}^{N} \big(\frac{1}{2} \boldsymbol{\epsilon}^\alpha : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha \big) h^\alpha(\boldsymbol{\phi}). \tag{7.77}$$

This follows from the fact that for $0 \le h^\alpha(\boldsymbol{\phi}) \le 1$, the problem (7.77) is a convex (and strictly convex for those $\boldsymbol{\epsilon}^\alpha$ with $h^\alpha(\boldsymbol{\phi}) > 0$)[47] minimization problem and that, by a simple differentiation

[47] Note that, even though convexity in each variable separately does not generally imply "global" convexity, it does so here due to the separated form of $f_{el,RS}$ in the $\boldsymbol{\epsilon}^\alpha$. In addition, the feasible set is clearly convex as any convex combination of phase-specific strains averaging to $\boldsymbol{\epsilon}$ will also do so.

of the Lagrangian,

$$L\big(\boldsymbol{\phi}, (\boldsymbol{\epsilon}^\alpha)_{1\leq\alpha\leq N}, \boldsymbol{\sigma}\big) = \sum_{\alpha=1}^{N} \big(\frac{1}{2}\boldsymbol{\epsilon}^\alpha : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha\big) h^\alpha(\boldsymbol{\phi}) - \boldsymbol{\sigma} : \big(\sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi}) - \boldsymbol{\epsilon}\big),$$

the minimizers satisfy $\big(\boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha - \boldsymbol{\sigma}\big) h^\alpha(\boldsymbol{\phi}) = \mathbf{0}$, i.e. $\boldsymbol{\sigma} = \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}^\alpha = \boldsymbol{\sigma}^\alpha$ and thus the equality of the stresses has to hold for all α with $h^\alpha(\boldsymbol{\phi}) \neq 0$[48].
Similarly, the stress-strain relationship for the Voigt-Taylor approach can be recovered from the minimization problem

$$f^*_{el,VT}\big(\boldsymbol{\phi}, \boldsymbol{\sigma}\big) = \min_{\{\boldsymbol{\sigma}^\alpha : \sum_{\alpha=1}^{N} \boldsymbol{\sigma}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{\sigma}\}} \sum_{\alpha=1}^{N} \big(\frac{1}{2}\boldsymbol{\sigma}^\alpha : \boldsymbol{S}^\alpha : \boldsymbol{\sigma}^\alpha\big) h^\alpha(\boldsymbol{\phi}). \tag{7.78}$$

corresponding to a minimization of the weighted average of the complementary strain energy densities

$$(f_{el})^*\big(\boldsymbol{\phi}, (\boldsymbol{\sigma}^\alpha)_{1\leq\alpha\leq N}\big) = \sum_{\alpha=1}^{N} (f^\alpha_{el})^*(\boldsymbol{\sigma}^\alpha) h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \big(\frac{1}{2}\boldsymbol{\sigma}^\alpha : \boldsymbol{S}^\alpha : \boldsymbol{\sigma}^\alpha\big) h^\alpha(\boldsymbol{\phi}) \tag{7.79}$$

over all decompositions satisfying $\boldsymbol{\sigma} = \sum_{\alpha=1}^{N} \boldsymbol{\sigma}^\alpha h^\alpha(\boldsymbol{\phi})$. This is again a convex function of the $(\boldsymbol{\sigma}^\alpha)_{1\leq\alpha\leq N}$ and the critical points of the corresponding Lagrangian satisfy $\big(\boldsymbol{S}^\alpha : \boldsymbol{\sigma}^\alpha - \boldsymbol{\epsilon}\big) h^\alpha(\boldsymbol{\phi}) = \mathbf{0}$ i.e. $\boldsymbol{\epsilon}^\alpha = \boldsymbol{S}^\alpha : \boldsymbol{\sigma}^\alpha = \boldsymbol{\epsilon}$ for all α with $h^\alpha(\boldsymbol{\phi}) > 0$.

Note that in both cases, even though one a priori only imposes one of the two interpolation properties in Equation (7.72) in this variational formulation, the other one is actually also trivially satisfied (the phase-specific stresses in the Reuss-Sachs model resp. the phase-specific strains in the Voigt-Taylor model all being equal by the FONC for the minimization problems).

An obvious consequence of the characterizations above is that (see e.g. also [51], [61]) one necessarily has the inequalities

$$f_{el,RS}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) \leq f_{el,VT}(\boldsymbol{\phi}, \boldsymbol{\epsilon}) \quad \text{and} \quad f^*_{el,VT}(\boldsymbol{\phi}, \boldsymbol{\sigma}) \leq f^*_{el,RS}(\boldsymbol{\phi}, \boldsymbol{\sigma})$$

as the Reuss-Sachs model is (according to Equation (7.77)) the one with the smallest possible elastic free energy density defined by (7.71) for a given strain and the interpolation property of the strains, whereas (according to Equation (7.78)), the Voigt-Taylor model has the smallest complementary strain energy density for a given stress and the interpolation property for the stresses.

◇

Remark 118. In addition, as pointed out in [61], by a well-known homogenization result, these two interpolations correspond to variational bounds on the effective behavior which would be expected from a real two-phase material.
It should be noted though that this result does **not** imply that the Voigt-Taylor model corresponds to an upper bound on the strain energy density $f_{el}\big(\boldsymbol{\phi}, (\boldsymbol{\epsilon}^\alpha)_{1\leq\alpha\leq N}\big)$ subject to the constraint of averaging to a given effective strain $\boldsymbol{\epsilon}$. In fact, it is easy to see that the effective free energy can be made arbitrarily large for a given $\boldsymbol{\epsilon}$ while maintaining the interpolation properties above[49]. ◇

[48] The other ones being basically arbitrary, but without any contribution to $f_{el,RS}$. As previously for the chemical models, these phases still have a contribution to the derivative w.r.t. $\boldsymbol{\phi}$ with the natural "definition" being the one setting $\boldsymbol{\sigma}^\alpha = \boldsymbol{\sigma}$, even if $h^\alpha(\boldsymbol{\phi}) = 0$.

[49] This is a **convex** maximization problem in the $\boldsymbol{\epsilon}^\alpha$ with an unbounded admissible set.

The Case With Eigenstrains

The approaches above can also be extended to the more general setting including eigenstrains $(\tilde{\boldsymbol{\epsilon}}^\alpha)_{1\leq\alpha\leq N}$[50], but now with, as above, $\boldsymbol{\sigma}^\alpha = \boldsymbol{C}^\alpha : (\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha)$ and $f_{el}(\boldsymbol{\phi}, \boldsymbol{\epsilon})$ as in Equation (7.71). Assuming again, as in the Voigt-Taylor case, equality of the (total) strains then leads to the phase-specific stress $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\alpha)$ and therefore a total stress given by

$$\begin{aligned}\boldsymbol{\sigma}_{VT}(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \{\tilde{\boldsymbol{\epsilon}}^\alpha\}_{1\leq\alpha\leq N}) &= \sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\alpha) h^\alpha(\boldsymbol{\phi}) \\ &= \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) : \boldsymbol{\epsilon} - \sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha : \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi}) = \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \hat{\tilde{\boldsymbol{\epsilon}}}),\end{aligned} \tag{7.80}$$

with $\hat{\tilde{\boldsymbol{\epsilon}}}(\boldsymbol{\phi}, \{\tilde{\boldsymbol{\epsilon}}^\alpha\}_{1\leq\alpha\leq N}) = \boldsymbol{\mathcal{C}}_{VT}^{-1}(\boldsymbol{\phi}) : \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha : \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big)$ (see also e.g. [3]), i.e. an effective material behavior with the same stiffness but an eigenstrain given by a $\boldsymbol{\mathcal{C}}^\alpha$-weighted average of the intrinsic ones and in particular **not** satisfying the interpolation rule $\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi})$.
If one in contrast assumes, corresponding to the model of Khachaturyan [18], the equality of the elastic strains instead of the total strains and **defining** the total strain to be given by $\boldsymbol{\epsilon} := \sum_{\alpha=1}^{N} (\boldsymbol{\epsilon}_{el} + \tilde{\boldsymbol{\epsilon}}^\alpha) h^\alpha(\boldsymbol{\phi})$, the interpolation

$$\boldsymbol{\sigma}_K(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \{\tilde{\boldsymbol{\epsilon}}^\alpha\}_{1\leq\alpha\leq N}) := \sum_{\alpha=1}^{N} \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}_{el} h^\alpha(\boldsymbol{\phi}) = \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : \boldsymbol{\epsilon}_{el} = \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi})),$$

leads to an effective material behavior which satisfies $\boldsymbol{\sigma} = \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}})$ with the "more natural" definitions of $\boldsymbol{\epsilon}$ and $\tilde{\boldsymbol{\epsilon}} := \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi})$.

Remark 119. Even though this model, in terms of the effective stress-strain relationship, may seem like the simpler - and thus in a way more elegant one (requiring only averaged quantities and no explicit phase-specific ones) - the higher simplicity is, from an energetic point of view, also its major drawback[51] (for a similar discussion, see [3]). In fact, it is obvious that, if the consistency relation (7.70) is to hold, this stress-strain-relationship results from the elastic free energy density

$$f_{el,K}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) = \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) : \boldsymbol{\mathcal{C}}_{VT}(\boldsymbol{\phi}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big), \tag{7.81}$$

and will generally not coincide with an average of the phase-specific free energy densities in (7.71) unless all eigenstrans $\tilde{\boldsymbol{\epsilon}}^\alpha$ happen to be the same. ◇

The situation for the Reuss-Sachs model is slightly simpler. Assuming the equality of the stresses, one has $\boldsymbol{\epsilon}^\alpha = \tilde{\boldsymbol{\epsilon}}^\alpha + \boldsymbol{S}^\alpha \boldsymbol{\sigma}$, $\alpha = 1, ..., N$, and, again imposing $\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha h^\alpha(\boldsymbol{\phi})$, a total strain given by

$$\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \big(\tilde{\boldsymbol{\epsilon}}^\alpha + \boldsymbol{S}^\alpha : \boldsymbol{\sigma}\big) h^\alpha(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi}) + \Big(\sum_{\alpha=1}^{N} \boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) : \boldsymbol{\sigma}.$$

Defining $\tilde{\boldsymbol{\epsilon}} = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^\alpha h^\alpha(\boldsymbol{\phi})$ and inverting this relationship shows that

$$\boldsymbol{\sigma}_{RS} = \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big)^{-1} : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}) = \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}) \quad \text{with} \quad \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) = \Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big)^{-1}$$

and thus that the effective stiffness coincides with the expected expression and that the effective eigenstrain $\tilde{\boldsymbol{\epsilon}}$ additionally obeys the standard interpolation property.

[50] Possibly depending upon additional paramters such as the concentration or temperature.

[51] This is similar in spirit to Section 7.1, where the properties of the model can improve significantly if one does **not** assume the equality of the more convenient quantities (in this case the phase-specific concentrations $\boldsymbol{c}^\alpha$) and instead, at the cost of a more complex model, the equality of derived quantities which are much more natural to assume to be the same for all phases.

From this, it follows $\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha = \boldsymbol{S}^\alpha : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}})$. Reinserting this effective stress-strain relationship into the definition of the elastic energy, one obtains the total elastic energy density given by

$$f_{el,RS}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) = \frac{1}{2}\sum_{\alpha=1}^{N}\Big(\boldsymbol{S}^\alpha : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}})\Big) : \boldsymbol{\mathcal{C}}^\alpha : \Big(\boldsymbol{S}^\alpha : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}})\Big) h^\alpha(\boldsymbol{\phi}).$$

Canceling $\boldsymbol{\mathcal{C}}^\alpha$ with $\boldsymbol{\mathcal{S}}^\alpha$ and observing that the only remaining phase-dependent term is given by $\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}^\alpha h^\alpha(\boldsymbol{\phi}) = \big(\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi})\big)^{-1}$, this further simplifies to

$$f_{el,RS}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) = \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) \tag{7.82}$$

corresponding to simply replacing the total strain $\boldsymbol{\epsilon}$ in Equation (7.76) with the average elastic strain $\boldsymbol{\epsilon}_{el} = \boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}$.

The Driving Force

The various choices above imply different driving forces $\frac{\partial f_{el}}{\partial \boldsymbol{\phi}}$ in the phasefield Equation (6.73). In the simplest cases, i.e. in the Voigt-Taylor and Khachaturyan-model, the elastic free energy density is given as a fully explicit formula in terms of the phasefield values $\boldsymbol{\phi}$ and the total strains $\boldsymbol{\epsilon}$. For the Voigt-Taylor model, a straightforward differentiation with $\boldsymbol{\epsilon}^\beta = \boldsymbol{\epsilon}$ for all β shows that

$$\frac{\partial f_{el,VT}(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \{\tilde{\boldsymbol{\epsilon}}^\beta\}_{1\le\beta\le N})}{\partial \phi^\alpha} = \sum_{\beta=1}^{N}\frac{1}{2}(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\beta) : \boldsymbol{\mathcal{C}}^\beta : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\beta)\frac{\partial h^\beta}{\partial \phi^\alpha} = \sum_{\beta=1}^{N} f_{el}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha},$$

i.e. the evolution of the phasefield is driven purely be the difference of the elastic free energy densities f^α. While reducing to the same model in the purely elastic case, the situation is different in the presence of eigenstrains for the Khatchuryan-model, since, based on equation (7.81) and $\frac{\partial f_{el,K}}{\partial \tilde{\boldsymbol{\epsilon}}} = -\boldsymbol{\mathcal{C}}_{VT} : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) = -\boldsymbol{\sigma}_K$, one has

$$\frac{\partial f_{el,K}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big)}{\partial \phi^\alpha} = \sum_{\beta=1}^{N}\frac{1}{2}\Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) : \boldsymbol{\mathcal{C}}^\beta : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) - \boldsymbol{\sigma} : \tilde{\boldsymbol{\epsilon}}^\beta\Big)\frac{\partial h^\beta}{\partial \phi^\alpha}$$

which in general can only be artificially related to a derivative involving the $f^\beta(\boldsymbol{\epsilon})$ underlying Equation (7.71) unless all eigenstrains coincide (in which case the distinction between the Voigt-Taylor and Khatchaturyan model again becomes irrelevant).
The situation for the Reuss-Sachs-model is slightly more complicated as one has, using, analogous to the Khachaturyan-model, $\frac{\partial f_{el,RS}}{\partial \tilde{\boldsymbol{\epsilon}}} = -\boldsymbol{\mathcal{C}}_{RS} : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) = -\boldsymbol{\sigma}_{RS}$,

$$\frac{\partial f_{el,RS}\big(\boldsymbol{\phi}, \boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big)}{\partial \phi^\alpha} = \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) : \Big(\frac{\partial \boldsymbol{\mathcal{C}}_{RS}}{\partial \phi^\alpha}(\boldsymbol{\phi})\Big) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) - \boldsymbol{\sigma}_{RS} : \sum_{\beta=1}^{N}\tilde{\boldsymbol{\epsilon}}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}. \tag{7.83}$$

Here, in contrast to the two previous models, $\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi})$ as a function of $\boldsymbol{\phi}$ is only defined implicitly as the (pseudo-)inverse of $\boldsymbol{\mathcal{S}}(\boldsymbol{\phi}) = \sum_{\beta=1}^{N}\boldsymbol{\mathcal{S}}^\beta h^\beta(\boldsymbol{\phi})$. This is not a serious issue though as it is well-known that[52]

$$\frac{\partial \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi})}{\partial \phi^\alpha} = -\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : \left(\sum_{\beta=1}^{N}\boldsymbol{S}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}\right) : \boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}).$$

[52] To see this, it suffices to differentiate the identity $\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}) : \Big(\sum_{\alpha=1}^{N}\boldsymbol{S}^\alpha h^\alpha(\boldsymbol{\phi})\Big) = \boldsymbol{I}^{(s)}$, which is legitimate under the standard coercivity assumption for the $\boldsymbol{\mathcal{C}}^\alpha$ resp. $\boldsymbol{\mathcal{S}}^\alpha$ (on the subspace of symmetric tensors).

While this expression in itself is somewhat cumbersome, it in fact allows for a major simplification in combination with Equation (7.83) as, by symmetry of $\boldsymbol{\mathcal{C}}_{RS}$,

$$\begin{aligned}\frac{1}{2}\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big):\Big(\frac{\partial\boldsymbol{\mathcal{C}}_{RS}}{\partial\phi^\alpha}(\boldsymbol{\phi})\Big):\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big) &= -\frac{1}{2}\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big):\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}):\left(\sum_{\beta=1}^{N}\boldsymbol{S}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}\right):\boldsymbol{\mathcal{C}}_{RS}(\boldsymbol{\phi}):\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi})\big)\\ &= -\frac{1}{2}\boldsymbol{\sigma}_{RS}:\left(\sum_{\beta=1}^{N}\boldsymbol{S}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}\right):\boldsymbol{\sigma}_{RS}.\end{aligned}$$

Since by construction $\boldsymbol{\sigma}_{RS}=\boldsymbol{\sigma}^\beta$ for all β, this is the same as $-\frac{1}{2}\sum_{\beta=1}^{N}\boldsymbol{\sigma}^\beta:\boldsymbol{\mathcal{S}}^\beta:\boldsymbol{\sigma}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}=-\sum_{\beta=1}^{N}f^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}$, thus leading to a total driving force given by

$$\frac{\partial f_{el,RS}}{\partial\phi^\alpha}\big(\boldsymbol{\phi},\boldsymbol{\epsilon},\{\tilde{\boldsymbol{\epsilon}}^\beta\}_{1\leq\beta\leq N}\big)=-\sum_{\beta=1}^{N}\Big(f_{el}^\beta+\boldsymbol{\sigma}(\boldsymbol{\phi},\boldsymbol{\epsilon}):\tilde{\boldsymbol{\epsilon}}^\beta\Big)\frac{\partial h^\beta}{\partial\phi^\alpha}. \tag{7.84}$$

Remark 120. As pointed out above (and already discussed in e.g. [69]), the Reuss-Sachs/Steinbach-Apel scheme is similar in spirit to the quantitative chemicals models from Section 7.1 with the $\boldsymbol{\epsilon}^\alpha$ taking the role of the phase-specific concentrations $\boldsymbol{c}^\alpha$ and $\boldsymbol{\sigma}$ the role of the (reduced) chemical potential $\tilde{\boldsymbol{\mu}}$. In constrast, the Voigt-Taylor model (and, up to a point, the Khatchuryan approach) is more similar in nature to the chemical model used in e.g. [52], which can be interpreted as assuming the equality of the phase-specific concentrations $\boldsymbol{c}^\alpha$ with the average one.
Even though these analogies are helpful for interpreting the mechanical models above, it needs to be realized that they are not "complete". Compared to the seemingly quite satisfactory case previously considered in Section 7.1, it was observed early on (see e.g. [23]) that the results obtained using the Reuss-Sachs model are usually not as satisfactory as those from the chemical case and in particular leads to excess interface energy contributions except for very special mechanical settings.

There are several reasons why developing successful models in the mechanical case is more difficult than in the purely chemical one. A first one is that it is more common to have to deal with an additional "internal" forcing expressed through the eigenstrains in the mechanical than in the chemical case. The more crucial one lies in the non-local dependence of f_{el} on $\boldsymbol{u}$ through its (symmtrized) gradient $\boldsymbol{\epsilon}$ and the resulting nature of the equilibrium conditions themselves.
The classical condition of the equilibration of the chemical potentials underlying the reasoning in Section 7.1 is in fact a highly restrictive one which is physically valid only for the very particular (but commonly used) settings of a **closed** system.
From a more mathematical point of view, this setting is related to the minimization of the free energy functional

$$\mathcal{F}_\epsilon(\boldsymbol{\phi},\boldsymbol{c})=\int_\Omega \epsilon a(\boldsymbol{\phi},\nabla\boldsymbol{\phi})+\frac{1}{\epsilon}w(\boldsymbol{\phi})+f_{ch}(\boldsymbol{\phi},\boldsymbol{c})\,\mathrm{d}\boldsymbol{x} \tag{7.85}$$

under the implicit assumption that $\mathcal{F}_\epsilon$ is to be minimized with respect to $\boldsymbol{c}$ under the additional constraint that the total amount of concentration of each component is to be maintained constant, $\int_\Omega c_i\,\mathrm{d}\boldsymbol{x}=\text{const}$. Even though the form of the functional (7.85) does not lend itself naturally to the application of boundary conditions on $\boldsymbol{c}$[53], the same effect can be enforced in terms of the gradient-flow appraoch to the minimization of $\mathcal{F}_\epsilon$ by "externally" postulating a conservative gradient-flow for $\boldsymbol{c}$ as discussed in Section 6.1, where the conservation of the total mass is ensured by the locally conservative flow in combination with the "natural" isolating boundary conditions for the concentration.

[53] As it is purely local in $\boldsymbol{c}$, there is a priori no reason for any minimizer in $\boldsymbol{c}$ to be sufficiently regular (e.g. $\boldsymbol{c}\in\boldsymbol{H}^1(\Omega)$) to make sense of its boundary values such that this regularity would have to be enforced "artificially".

In contrast, when modeling other physical situations (involving for example prescribed concentration values and/or chemical potential values on the outer boundary), the simple condition $\tilde{\boldsymbol{\mu}} = \text{const}$ will in general not hold anymore. Instead, the physically relevant equilibrium condition (in terms of the model in [52]) becomes

$$\begin{cases} -\nabla \cdot \left(\boldsymbol{L}_{ij}(\boldsymbol{\phi}, \boldsymbol{c}) \nabla \boldsymbol{\mu}_j(\boldsymbol{\phi}, \boldsymbol{c}) \right) = \boldsymbol{0} & \text{in } \Omega, \\ \text{boundary conditions on } \boldsymbol{c} \text{ resp. } \boldsymbol{\mu}. \end{cases} \tag{7.86}$$

On the one hand, it is clear (there for example not being any $\boldsymbol{L}_{ij}$ in $\mathcal{F}_\epsilon(\boldsymbol{\phi}, \boldsymbol{c})$) that this is not easily connected to a minmization of the functional in Equation (7.85) unless one would treat this explicitly as a constrained optimization problem subject to Equation (7.86). On the other hand, as the condition $\tilde{\boldsymbol{\mu}} = \text{const}$ would have to be replaced with the (significantly) more complex equilibrium condition in Equation (7.86) corresponding to $-\operatorname{div}\left(\boldsymbol{\sigma}(\boldsymbol{\phi}, \boldsymbol{\epsilon})\right) = \rho \boldsymbol{f}$ in the mechanical case. In terms of the diffuse interface approximation, it is easy to see that one would therefore also have to deal with the same issues as will be discussed in the mechanical case in the following sections.
The primary difference between the chemical and mechanical case therefore lies in the fact that, for the physically interesting examples, the latter one only very rarely allows for the simple solution $\boldsymbol{\sigma} = \text{const}$[54], whereas the analogous condition $\tilde{\boldsymbol{\mu}} = \text{const}$ is much more common in the chemical setting[55]. ◇

7.2.2 The Quantitative Model in the Two-Phase Case

A more recent modeling approach, first introduced in the two-phase setting by [51] and [64] (and later partially extended to the multiphase setting in [63], [62], [66] and also to situations involving plastic [35] and viscoelastic effects is based on the observation that, at an interface between two phases in the sharp interface setting, one usually neither has the equality of the strains nor that of the stresses. Instead, the spatial nature of the equilibrium conditions is also reflected in the well-known mechanical jump conditions on internal interfaces. These require on the one hand that the tangential components of the displacement gradient coincide on the interface, i.e. the jump $[\![\nabla \boldsymbol{u}]\!]$ of $\nabla \boldsymbol{u}$ upon traversing the interface is oriented along the normal direction and the displacement gradients on both sides are therefore related by the condition $\left(\nabla \boldsymbol{u}\right)^+ = \left(\nabla \boldsymbol{u}\right)^- + \boldsymbol{a} \otimes \boldsymbol{n}$. For the strain tensor usually used within the small deformation settting, this obviously implies that $[\![\boldsymbol{\epsilon}]\!] = (\boldsymbol{a} \otimes \boldsymbol{n})_S = \frac{1}{2}(\boldsymbol{a} \otimes \boldsymbol{n} + \boldsymbol{n} \otimes \boldsymbol{a})$. On the other hand, in contrast to $\nabla \boldsymbol{u}$, it is the normal components of the stress-tensor on both sides of the interface which need to coincide, i.e. $[\![\boldsymbol{\sigma}]\!] \cdot \boldsymbol{n} = \boldsymbol{0}$.

Remark 121. The idea of treating the normal and tangential components was actually introduced slightly earlier by Durga et al. under an additional simplifying assumption in the two-dimensional setting [23] and later generalized by the same authors in [24] to the three-dimensional setting without this simplification. As will be discussed in Section 7.2.3, the models in [51] and [64] differ from - and in fact improve upon - the generalization in [24] in a small but important point. For this reason, despite the fundamental importance of the work [23] in the current setting for initiating this line of thought, the other model(s) considered below should be preferred over the

[54] A notable exception being the purely elastic material chain, in which case the Reuss-Sachs approximation is in fact free of excess energy contributions, see e.g. [64].

[55] There are also a significant number of works - for example in directional solidification - in which the models from Section 7.1 are applied in situations where this does not hold. These applications usually being set up (using e.g. moving window techniques) such that one never actually achieves equilibrium, the precise advantages and disadvantages of the models are significantly harder to judge as they are always based on the much more complex dynamic case. In particular, it is common to employ an artificial dynamic contribution in terms of an antitrapping current, further increasing the complexity of the situation.

one in [24][56]. ◇

Remark 122. It is useful to keep the reasoning behind these jump conditions in mind.
The first one on $[\![\nabla \boldsymbol{u}]\!]$ is essentially imposed a priori - at least within the standard variational approach - by demanding a global $\boldsymbol{H}^1$-regularity of the solution. This implies that the trace $\gamma(\boldsymbol{u})$ of $\boldsymbol{u}$ on a smooth hypersurface Γ can be uniquely defined as an element of $\boldsymbol{H}^{\frac{1}{2}}(\Gamma)$, i.e. denoting the limiting values of $\boldsymbol{u}$ as one approaches this hypersurface from both sides by $\gamma^{\pm}(\boldsymbol{u})$, one has $\gamma^+(\boldsymbol{u}) = \gamma^-(\boldsymbol{u})$. The traces coinciding along the surface then implies the same for their tangential derivatives[57].
In constrast, the second condition on $\boldsymbol{\sigma}$ is only valid provided the right-hand side is regular enough and in particular (consistent with the intuitive reasoning) excludes the case of a "concentrated" surface force acting on Γ. Focusing on the simpler steady-state equilibrium conditions $-\operatorname{div}(\boldsymbol{\sigma}) = \boldsymbol{f}$, $\boldsymbol{f}$ being in e.g. $\boldsymbol{L}^2(\Omega)$ implies (by the very definition) that $\boldsymbol{\sigma} \in \boldsymbol{H}_{div}(\Omega)$[58]. Similar to the trace of functions in $\boldsymbol{H}^1(\Omega)$, it is a standard fact (see e.g. [34], [43] or [48]) that one can define a unique **normal** trace in $\boldsymbol{H}^{-\frac{1}{2}}(\Gamma)$ on any (sufficiently smooth) hypersurface for such functions[59], and that therefore the normal components of $\boldsymbol{\sigma}$ have to coincide as one approaches Γ from both sides. ◇

More elaborate mechanical models taking these sharp interface jump conditions at a two-phase interface into account have first been proposed by Mosler et al. [51] within an energy-based setting (a priori relying upon the jump conditions for the strains alone) and by Durga et al. [24] and Schneider et al. [64] within a more direct approach based on the jump conditions for both the strains and the stresses. In contrast to the work by [51], these models do not rely on a jump vector $\boldsymbol{a}$ (as also used below), but instead use a basis transformation into an appropriate coordinate system in which they then enforce the equality of the appropriate stress- and strain-entries. Yet another approach is taken in [74], which avoids the use of the basis transformation by using a description in terms of projecters $\mathcal{N}$ and $\mathcal{T}$ onto the "normal" and "tangent" subspaces of the symmetric second-order strain and stress tensors.
Given that all models are essentially based on the same sharp-interface jump conditions, it is not surprising that, from a modeling point of view, they are in fact all different representations of the same physical model. Nevertheless, the choice of representation clearly can have a significant impact both in terms of the computational complexity of the models as well as the effort required for their mathematical description and practical implementation. For this reason, after a quick outline of the mechanical model and its additional contributions to the phasefield equation - here using a description in terms of a jump vector $\boldsymbol{a}$ - the links between the various models will be discussed in Section 7.2.3 in a little more detail.
Further discussions, extensions and applications (often within the multiphase setting to be discussed in the next section) can e.g. be found in [61], [63], [74], [62], [35], [66], [4] and [5].

The Mechanical Model

The underlying idea is to increase the accuracy of the phasefield model (with a small but non-zero interface width) by directly incorporating the jump conditions from the sharp-interface setting. Following [51] (see also [61] for a more detailed discussion), one can therefore impose

[56] Nevertheless, as shown in [23], their model does, among other things, in fact improve significantly upon the Voigt-Taylor or Reuss-Sachs model in terms of interfacial excess energies.

[57] Then in an even weaker sense as elements of $\boldsymbol{H}^{-\frac{1}{2}}(\Gamma)$.

[58] $\boldsymbol{H}_{div}(\Omega)$ being defined by $\boldsymbol{H}_{div}(\Omega) := \{\boldsymbol{\sigma} \in \boldsymbol{L}^2(\Omega) : \operatorname{div}(\boldsymbol{\sigma}) \in \boldsymbol{L}^2(\Omega)\}$.

[59] The essential idea is based upon "extending" Gauss's divergence theorem $\int_\omega \operatorname{div}(\boldsymbol{g})\,\mathrm{d}\boldsymbol{x} = \int_{\partial\omega} \boldsymbol{g}\cdot\boldsymbol{n}\,\mathrm{d}\boldsymbol{s}$ for smooth functions by appropriately **defining** the (a priori meaningless in the non-smooth case) values of $\boldsymbol{g}$ such that this integral equality holds.
If the divergence of $\boldsymbol{\sigma}$ is for example only in $\boldsymbol{H}^{-1}(\Omega)$ (e.g. if surface forces are present), the definition of an appropriate trace is a much trickier issue (see e.g. [48]).

the condition

$$[\![\boldsymbol{F}]\!] = \boldsymbol{F}^2 - \boldsymbol{F}^1 = \boldsymbol{a} \otimes \boldsymbol{n} \quad \text{resp.} \quad [\![\boldsymbol{\epsilon}]\!] = \boldsymbol{\epsilon}^2 - \boldsymbol{\epsilon}^1 = (\boldsymbol{a} \otimes \boldsymbol{n})_S \tag{7.87}$$

on the displacement gradients $\boldsymbol{F} = \nabla \boldsymbol{u}$ resp. its symmetrization in the small deformation setting. Combined with the interpolation requirement $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^1 h^2(\phi) + \boldsymbol{\epsilon}^2 h^2(\phi)$, where, as in Section 6.2, $h^1(\phi) = h(\phi)$, $h^2 = h(1-\phi) = 1 - h(\phi)$, this results in an effective strain given by $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^1 h^1(\phi) + \big(\boldsymbol{\epsilon}^1 + (\boldsymbol{a} \otimes \boldsymbol{n})_S\big) h^2(\phi) = \boldsymbol{\epsilon}^1 + h^2(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S$, which can be solved for the strains $(\boldsymbol{\epsilon}^1, \boldsymbol{\epsilon}^2)$ as a function of $\boldsymbol{\epsilon}$ and the (as of yet arbitrary) jump vector $\boldsymbol{a}$ to give

$$\boldsymbol{\epsilon}^1(\boldsymbol{\epsilon}, \boldsymbol{a}) = \boldsymbol{\epsilon} - h^2(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S \quad \text{and} \quad \boldsymbol{\epsilon}^2(\boldsymbol{\epsilon}, \boldsymbol{a}) = \boldsymbol{\epsilon}^1 + (\boldsymbol{a} \otimes \boldsymbol{n})_S = \boldsymbol{\epsilon} + h^1(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S. \tag{7.88}$$

Inserting these expressions into the material law $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$ and the jump condition for $\boldsymbol{\sigma}$, one obtains the additional condition

$$\begin{aligned}
[\![\boldsymbol{\sigma}]\!] \cdot \boldsymbol{n} =& \Big(\boldsymbol{\mathcal{C}}^2 : \big(\boldsymbol{\epsilon}^2 - \tilde{\boldsymbol{\epsilon}}^2\big) - \boldsymbol{\mathcal{C}}^1 : \big(\boldsymbol{\epsilon}^1 - \tilde{\boldsymbol{\epsilon}}^1\big)\Big) \cdot \boldsymbol{n} \\
=& \left(\boldsymbol{\mathcal{C}}^2 : \Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\big) + h^1(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S\Big) - \boldsymbol{\mathcal{C}}^1 : \Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\big) - h^2(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S\Big)\right) \cdot \boldsymbol{n} \overset{!}{=} \boldsymbol{0}.
\end{aligned}$$

Given the subsymmetry $\mathcal{C}^\alpha_{ijkl} = \mathcal{C}^\alpha_{ijlk}$ of the stiffness-tensor[60], this can in fact be simplified to the equation

$$\left(\boldsymbol{n} \cdot \left(h^2(\phi)\boldsymbol{\mathcal{C}}^1 + h^1(\phi)\boldsymbol{\mathcal{C}}^2\right) \cdot \boldsymbol{n}\right) \cdot \boldsymbol{a} \overset{!}{=} \Big(\boldsymbol{\mathcal{C}}^1 : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\big) - \boldsymbol{\mathcal{C}}^2 : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\big)\Big) \cdot \boldsymbol{n}.$$

This is a 3×3 linear system allowing for the determination of $\boldsymbol{a}$ in terms of $\boldsymbol{\epsilon}$ as

$$\begin{aligned}
\boldsymbol{a} =& \Big(\boldsymbol{n} \cdot \big(h^2(\phi)\boldsymbol{\mathcal{C}}^1 + h^1(\phi)\boldsymbol{\mathcal{C}}^2(\phi)\big) \cdot \boldsymbol{n}\Big)^{-1} \cdot \Big(\boldsymbol{\mathcal{C}}^1 : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\big) - \boldsymbol{\mathcal{C}}^2 : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\big)\Big) \cdot \boldsymbol{n} \\
=& \Big(\boldsymbol{n} \cdot \bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) \cdot \boldsymbol{n}\Big)^{-1} \cdot \Big(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\Big) \cdot \boldsymbol{n},
\end{aligned} \tag{7.89}$$

where, adopting the notation in [74], $\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)$ denotes the "anti-arithmetic" average

$$\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) \coloneqq h^2(\phi)\boldsymbol{\mathcal{C}}^1 + h^1(\phi)\boldsymbol{\mathcal{C}}^2(\phi) = \big(1 - h(\phi)\big)\boldsymbol{\mathcal{C}}^1 + h(\phi)\boldsymbol{\mathcal{C}}^2 \tag{7.90}$$

and the $\boldsymbol{\Sigma}^\alpha \coloneqq \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$ correspond to the phase-specific stresses predicted by the Voigt-Taylor model.

Remark 123. The solvability of the system for $\boldsymbol{a}$ (resp. the invertibility of $\bar{\boldsymbol{\mathcal{C}}}^{12}$ under the standard assumptions on the $\boldsymbol{\mathcal{C}}^\alpha$ can e.g. be seen from the fact that the bilinear form $\frac{1}{2}\boldsymbol{\epsilon} : \boldsymbol{\mathcal{C}}^\alpha : \boldsymbol{\epsilon}$ is strictly convex in $\boldsymbol{\epsilon}$, which implies rank-one convexity. Together with the subsymmetries of the $\boldsymbol{\mathcal{C}}^\alpha$ and as $\bar{\boldsymbol{\mathcal{C}}}^{12}$ is a convex combination (for $0 \le h(\phi) \le 1$) of $\boldsymbol{\mathcal{C}}^1$ and $\boldsymbol{\mathcal{C}}^2$, it follows that $\boldsymbol{a} \mapsto \boldsymbol{a} \cdot \big(\boldsymbol{n} \cdot \bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) \cdot \boldsymbol{n}\big)$ is strictly convex and $\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)$ a positive definite matrix. ◇

Remark 124. Depending on the way the stiffness-tensors are arranged in memory, it may be advantageous to use any of the (equivalent given the subsymmetries) expressions

$$n_i \mathcal{C}_{ijkl} n_l = n_j \mathcal{C}_{ijkl} n_l = n_j \mathcal{C}_{ijkl} n_k = n_i \mathcal{C}_{ijkl} n_k.$$

Note that none of these is the same as $\boldsymbol{\mathcal{C}} : (\boldsymbol{n} \otimes \boldsymbol{n}) = (\boldsymbol{n} \otimes \boldsymbol{n}) : \boldsymbol{\mathcal{C}}$ though.
With respect to the commonly used Voigt-notation, it should also be noted that, as this formula

[60] By the right subsymmetric of $\boldsymbol{\mathcal{C}}$, $\mathcal{C}_{ijkl} = \mathcal{C}_{ijlk}$, applying $\boldsymbol{\mathcal{C}}$ to a non-symmetric tensor $\boldsymbol{t}$ leads to the same result as applying it to the symmetrized tensor $\boldsymbol{t}_S = \frac{1}{2}(\boldsymbol{t} + \boldsymbol{t}^T)$ since (using the Einstein summation convention)

$$\boldsymbol{\mathcal{C}} : \boldsymbol{t}_S = \frac{1}{2}(\mathcal{C}_{ijkl} t_{kl} + \mathcal{C}_{ijkl} t_{lk}) = \frac{1}{2}(\mathcal{C}_{ijkl} t_{kl} + \mathcal{C}_{ijlk} t_{kl}) \overset{\mathcal{C}_{ijkl} = \mathcal{C}_{ijlk}}{=} \mathcal{C}_{ijkl} \frac{1}{2}(t_{kl} + t_{kl}) = \boldsymbol{\mathcal{C}} : \boldsymbol{t}.$$

involves two separate contractions (and not a double-contraction with a symmetric tensor in a "reduced" vector format) there are no additional prefactors involved in the evaluation of $\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n}$. ◇

With $\boldsymbol{a}$ determined by Equation (7.89) and with $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \left(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\right)$ together with Equation (7.88), one can calculate the phase-specific stresses (again using the right subsymmetry of the stiffness) to be given by

$$\begin{aligned}\boldsymbol{\sigma}^1 =& \boldsymbol{\mathcal{C}}^1\left(\boldsymbol{\epsilon} - h^2(\phi)\left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S - \tilde{\boldsymbol{\epsilon}}^1\right) = \boldsymbol{\Sigma}^1 - h^2(\phi)\left(\boldsymbol{\mathcal{C}}^1\cdot\boldsymbol{n}\right)\cdot\boldsymbol{a}\\ =& \boldsymbol{\Sigma}^1 - h^2(\phi)\left(\boldsymbol{\mathcal{C}}^1\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right)\cdot\boldsymbol{n}\right)\end{aligned}$$

and similarly

$$\boldsymbol{\sigma}^2 = \boldsymbol{\Sigma}^2 + h^1(\phi)\left(\boldsymbol{\mathcal{C}}^2\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right)\cdot\boldsymbol{n}\right).$$

Finally, together with the definition of $\boldsymbol{\sigma}$ as the weighted average of $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$, this leads to the effective material behavior given by

$$\begin{aligned}\boldsymbol{\sigma} =& h^1(\phi)\boldsymbol{\sigma}^1 + h^2(\phi)\boldsymbol{\sigma}^2 = h^1(\phi)\boldsymbol{\Sigma}^1 + h^2(\phi)\boldsymbol{\Sigma}^2 - h^1(\phi)h^2(\phi)\left(\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\cdot\boldsymbol{n}\right)\cdot\boldsymbol{a}\\ =& \boldsymbol{\sigma}_{VT}(\boldsymbol{\epsilon},\phi) - \left(\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right)\cdot\boldsymbol{n}\right),\end{aligned} \tag{7.91}$$

in terms of the various auxiliary quantities introduced above (and with $\boldsymbol{\sigma}_{VT}(\boldsymbol{\epsilon},\phi)$ corresponding to the stress from the Voigt-Taylor model in equations (7.73) resp. (7.80)), or, after reinserting the definitions, the fully explicit formula

$$\begin{aligned}\boldsymbol{\sigma} =& h^1(\phi)\boldsymbol{\mathcal{C}}^1 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\right) + h^2(\phi)\boldsymbol{\mathcal{C}}^2 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\right)\\ & - h^1(\phi)h^2(\phi)\left(\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{n}\cdot\left(\boldsymbol{\mathcal{C}}^1 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\right) - \boldsymbol{\mathcal{C}}^2 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\right)\right)\right)\end{aligned} \tag{7.92}$$

for $\boldsymbol{\sigma}$ in terms of $\boldsymbol{\epsilon}$, $\boldsymbol{n}$ and ϕ.

Remark 125. The formulas above simplify somewhat when there are no eigenstrains, i.e. when $\tilde{\boldsymbol{\epsilon}}^1 = \tilde{\boldsymbol{\epsilon}}^2 = \boldsymbol{0}$. It is easy to see that one then obtains

$$\boldsymbol{\sigma}^1 = \boldsymbol{\mathcal{C}}^1 : \boldsymbol{\epsilon} - h^2(\phi)\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^1\right) : \left(\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{n}\cdot\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right) : \boldsymbol{\epsilon}\right)\right)$$

and

$$\boldsymbol{\sigma}^2 = \boldsymbol{\mathcal{C}}^2 : \boldsymbol{\epsilon} + h^1(\phi)\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^2\right) : \left(\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{n}\cdot\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right) : \boldsymbol{\epsilon}\right)\right)$$

and thus, using the abbreviation $\boldsymbol{\mathcal{C}}_{VT}(\phi) := h^1(\phi)\boldsymbol{\mathcal{C}}^1 + h^2(\phi)\boldsymbol{\mathcal{C}}^2 = h(\phi)\boldsymbol{\mathcal{C}}^1 + \left(1 - h(\phi)\right)\boldsymbol{\mathcal{C}}^2$ for the "arithmetic" Voigt-Talyor-type average of the stiffnesses and the various symmetries,

$$\boldsymbol{\sigma} = \left(\boldsymbol{\mathcal{C}}_{VT}(\phi) - h^1(\phi)h^2(\phi)\left(\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{n}\cdot\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\right)\right) : \boldsymbol{\epsilon}, \tag{7.93}$$

i.e. an effective behavior corresponding to the stiffness-tensor

$$\boldsymbol{\mathcal{C}}_{eff}(\phi,\boldsymbol{n}) = \boldsymbol{\mathcal{C}}_{VT}(\phi) - h^1(\phi)h^2(\phi)\left(\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi)\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{n}\cdot\left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right)\right). \tag{7.94}$$

This effective stiffness can also be very useful for numerical purposes when using an implicit time-integration scheme or when making a quasi-steady state assumption for the elastic fields. As is well known, applying any of the standard Krylov solvers for the determination of the displacement field $\boldsymbol{u}$ requires, except for the evaluation of the initial residual, only the ability to evaluate the system response to increments in $\boldsymbol{u}$, and thus here the ability to evalue $\delta\boldsymbol{\sigma} = \boldsymbol{\sigma}\big(\boldsymbol{\epsilon}(\boldsymbol{u}+\delta\boldsymbol{u})\big) - \boldsymbol{\sigma}\big(\boldsymbol{\epsilon}(\boldsymbol{u})\big)$. As long as any eigenstrains considered do not have any additional dependence[61] on $\boldsymbol{\epsilon}$, the stress-contributions due to the $\tilde{\boldsymbol{\epsilon}}^\alpha$ will cancel in $\delta\boldsymbol{\sigma}$ and the "homogeneous" version (7.93) and thus the effective stiffness $\boldsymbol{\mathcal{C}}_{eff}(\phi,\boldsymbol{n})$ in Equation (7.94) is also the relevant one for evaluating the effect of increments in $\boldsymbol{u}$, $\delta\boldsymbol{\sigma} = \boldsymbol{\mathcal{C}}_{eff}(\phi,\boldsymbol{n}) : \boldsymbol{\epsilon}(\delta\boldsymbol{u})$.
As $\boldsymbol{\mathcal{C}}_{eff}(\phi)$ can in principle be precalculated and stored once at the beginning of each time-step, it is clear that the evaluation of this stress-increment using this prestored stiffness can be significantly cheaper than the procedure above (i.e. first determining the Voigt-Taylor stress predictions, then solving for $\boldsymbol{a}$ and finally correcting $\boldsymbol{\sigma}_{VT}$). ◇

Remark 126. For isotropic materials, the formula above can actually be evaluated analytically. Based on $\bar{\boldsymbol{\mathcal{C}}}^{12} = \bar{\lambda}^{12}\boldsymbol{I}\otimes\boldsymbol{I} + 2\bar{\mu}^{12}\boldsymbol{I}^{(s)}$ where $\boldsymbol{I}^{(s)} = \frac{1}{2}\big(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\big)\boldsymbol{e}_i\otimes\boldsymbol{e}_j\otimes\boldsymbol{e}_k\otimes\boldsymbol{e}_l$ is the symmetric fourth order unit tensor and $\bar{\lambda}^{12}$ and $\bar{\mu}^{12}$ the "anti-arithmetic" averages of λ and μ, one has $\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n} = \bar{\lambda}^{12}\boldsymbol{n}\otimes\boldsymbol{n} + 2\bar{\mu}^{12}n_i(\boldsymbol{I}^{(s)})_{ijkl}n_l\boldsymbol{e}_j\otimes\boldsymbol{e}_k$, or, as

$$n_i\big(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\big)n_l\boldsymbol{e}_j\otimes\boldsymbol{e}_k = \big(n_in_j\boldsymbol{e}_j\otimes\boldsymbol{e}_i + n_in_i\boldsymbol{e}_j\otimes\boldsymbol{e}_j\big) = \big(\boldsymbol{n}\otimes\boldsymbol{n} + \|\boldsymbol{n}\|^2\boldsymbol{I}\big) = \big(\boldsymbol{n}\otimes\boldsymbol{n}+\boldsymbol{I}\big),$$

the simple expression

$$\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n} = (\bar{\lambda}^{12}+\bar{\mu}^{12})\boldsymbol{n}\otimes\boldsymbol{n} + \bar{\mu}^{12}\boldsymbol{I}, \quad \text{where} \quad \bar{\lambda}^{12} = h^2\lambda^1 + h^1\lambda^2,\ \bar{\mu}^{12} = h^2\mu^1 + h^1\mu^2$$

for $\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n}$ in terms of the Lamé-parameters λ and μ. Since this is a rank-one perturbation of the (scaled) identity, the inverse of this tensor is easily determined using the Sherman-Morisson-Woodsbury-formula to be given by[62]

$$(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n})^{-1} = \big(\bar{\mu}^{12}\big)^{-1}\left(\boldsymbol{I} - \frac{\frac{\bar{\lambda}^{12}+\bar{\mu}^{12}}{\bar{\mu}^{12}}}{1+\frac{\bar{\lambda}^{12}+\bar{\mu}^{12}}{\bar{\mu}^{12}}}\boldsymbol{n}\otimes\boldsymbol{n}\right) = \big(\bar{\mu}^{12}\big)^{-1}\Big(\boldsymbol{I} - \frac{\bar{\lambda}^{12}+\bar{\mu}^{12}}{\bar{\lambda}^{12}+2\bar{\mu}^{12}}\boldsymbol{n}\otimes\boldsymbol{n}\Big). \tag{7.95}$$

◇

The stress-strain relationship (7.92) above can alternatively - instead of defining it based on the weighted average of the two $\boldsymbol{\sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$, $\alpha = 1,2$ - also be obtained as the derivative $\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial\boldsymbol{\epsilon}}$ of the elastic free energy density

$$\begin{aligned} f_{el} = \quad & \frac{1}{2}\Big(\boldsymbol{\epsilon} - h^2(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^1\Big) : \boldsymbol{\mathcal{C}}^1 : \Big(\boldsymbol{\epsilon} - h^2(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^1\Big)h^1(\phi) \\ & + \frac{1}{2}\Big(\boldsymbol{\epsilon} + h^1(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^2\Big) : \boldsymbol{\mathcal{C}}^2 : \Big(\boldsymbol{\epsilon} + h^1(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^2\Big)h^2(\phi). \end{aligned} \tag{7.96}$$

This correspond to Equation (7.71) with $\boldsymbol{\epsilon}^\alpha$, $\alpha = 1,2$ chosen as in Equation (7.88) provided the vector $\boldsymbol{a}$ is, as above, fixed by the continuity condition $\boldsymbol{\sigma}^1\cdot\boldsymbol{n} = \boldsymbol{\sigma}^2\cdot\boldsymbol{n}$ on the normal stress components. In fact, even though $\boldsymbol{a}$ depends on $\boldsymbol{\epsilon}$ through the defining normal continuity condition,

[61] Two such examples will quickly be considered below.

[62] Alternatively, as in [74], one can also rewrite the formula for $\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n}$ as

$$\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n} = (\bar{\lambda}^{12}+2\bar{\mu}^{12})\boldsymbol{n}\otimes\boldsymbol{n} + \bar{\mu}^{12}\big(\boldsymbol{I} - \boldsymbol{n}\otimes\boldsymbol{n}\big)$$

where $\boldsymbol{n}\otimes\boldsymbol{n}$ and $\boldsymbol{I}-\boldsymbol{n}\otimes\boldsymbol{n}$ are the two orthogonal projection operators onto the normal subspace $\mathrm{Span}(\{\boldsymbol{n}\})$ and the tangential one, $\mathrm{Ker}(\boldsymbol{n})$. By orthogonality both terms can be inverted separately leading to

$$\big(\boldsymbol{n}\cdot\bar{\boldsymbol{\mathcal{C}}}^{12}\cdot\boldsymbol{n}\big)^{-1} = \frac{1}{\bar{\lambda}^{12}+2\bar{\mu}^{12}}\boldsymbol{n}\otimes\boldsymbol{n} + \frac{1}{\bar{\mu}^{12}}\big(\boldsymbol{I}-\boldsymbol{n}\otimes\boldsymbol{n}\big)$$

which is easily seen to be equivalent to the expression in Equation (7.95).

differentiating the free energy density w.r.t. $\boldsymbol{\epsilon}$ and making use of the right subsymmetries of the $\boldsymbol{\mathcal{C}}^\alpha$ leads to

$$\begin{aligned}\frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}} =& h^1 \boldsymbol{\mathcal{C}}^1 (\boldsymbol{\epsilon} - h^2 (\boldsymbol{a} \otimes \boldsymbol{n})_S) : (\boldsymbol{I}^{(s)} - h^2 \boldsymbol{n} \otimes \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}) + h^2 \boldsymbol{\mathcal{C}}^2 (\boldsymbol{\epsilon} + h^1 (\boldsymbol{a} \otimes \boldsymbol{n})_S) : (\boldsymbol{I}^{(s)} + h^1 \boldsymbol{n} \otimes \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}) \\ =& h^1 \boldsymbol{\sigma}^1 : (\boldsymbol{I}^{(s)} - h^2 \boldsymbol{n} \otimes \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}) + h^2 : \boldsymbol{\sigma}^2 : (\boldsymbol{I}^{(s)} + h^1 \boldsymbol{n} \otimes \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}).\end{aligned}$$

While the two terms $h^1 \boldsymbol{\sigma}^1 + h^2 \boldsymbol{\sigma}^2$ give rise to the desired effective stress $\boldsymbol{\sigma}$, the remaining two terms can be combined to $h^1 h^2 \big((\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1) \cdot \boldsymbol{n} \big) \cdot \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}$ and thus simply drop out by the imposed continuity of $\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n}$. The two-phase model above therefore indeed does satisfy the important relation $\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}$ in Equation (7.70).

Remark 127. The crucial point here is that there is no additional contribution due to $\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\epsilon}}$. The same will also be seen to hold with respect to the dependence of $\boldsymbol{a}$ on the other parameters ϕ and $\boldsymbol{n}$ and is in fact to be expected from the variational characterization by [51] to be recalled below. As will be seen in Section 7.2.4, the same property does unfortunately not hold for all of the multiphase generalizations which have been proposed, and then leads to significant complications as compared to the very favorable two-phase setting. ◇

The Mechanical Driving Force

In contrast to the simpler models using an interpolation in terms of the local values of $\boldsymbol{\phi}$ alone, the mechanical free energy density f_{el} now depends not only on the values of ϕ and $\boldsymbol{\epsilon}$ alone, but has an additional dependence on the gradients of $\boldsymbol{\phi}$ due to the heavy use of the normal vector $\boldsymbol{n}$ for constructing the phase-specific stress- and strain-fields. Due to this additional dependence, there are now two different contributions to the mechanical driving for the phasefield model, namely the one due to the derivative w.r.t. ϕ itself and an additional divergence-type contribution $-\nabla \cdot \big(\frac{\partial f_{el}}{\partial \nabla \phi} \big)$ due to the dependence on the orientation of the interface trough $\boldsymbol{n} = -\nabla \phi$[63].

Remark 128. The appearence of an additional contribution to the phasefield equation due to the dependence on $\boldsymbol{n}$ was already clearly stated in [51] but is in contrast (erroneously) claimed in [64] (and later corrected e.g. in [74]) not to be the case[64]. ◇

By the formula (7.89) for $\boldsymbol{a}$, this vector obviously depends both on ϕ and, through $\boldsymbol{n}$, on $\nabla \phi$, $\boldsymbol{a} = \boldsymbol{a}(\phi, \nabla \phi, \boldsymbol{\epsilon})$. Making use of $\frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}^\alpha} = \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha \big) = \boldsymbol{\sigma}^\alpha$, differentiating the free energy density with respect to ϕ leads to

$$\begin{aligned}\frac{\partial f_{el}}{\partial \phi} =& f_{el}^1 \frac{\partial h^1}{\partial \phi} + f_{el}^2 \frac{\partial h^2}{\partial \phi} + \boldsymbol{\sigma}^1 : \frac{\partial \boldsymbol{\epsilon}^1}{\partial \phi} h^1 + \boldsymbol{\sigma}^2 : \frac{\partial \boldsymbol{\epsilon}^2}{\partial \phi} h^2 \\ =& f_{el}^1 \frac{\partial h^1}{\partial \phi} + f_{el}^2 \frac{\partial h^2}{\partial \phi} + \boldsymbol{\sigma}^1 : \Big(- \frac{\partial h^2}{\partial \phi} \big(\boldsymbol{a} \otimes \boldsymbol{n} \big)_S - h^2 \big(\frac{\partial \boldsymbol{a}}{\partial \phi} \otimes \boldsymbol{n} \big)_S \Big) h^1 + \boldsymbol{\sigma}^2 : \Big(\frac{\partial h^1}{\partial \phi} \big(\boldsymbol{a} \otimes \boldsymbol{n} \big)_S + h^1 \big(\frac{\partial \boldsymbol{a}}{\partial \phi} \otimes \boldsymbol{n} \big)_S \Big) h^2.\end{aligned}$$

As indicated in Remark 127, the dependence of $\boldsymbol{a}$ on ϕ actually again drops out since the two terms containing $\frac{\partial a}{\partial \phi}$ can be combined to give $h^1 h^2 \big(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1 \big) \big(\frac{\partial \boldsymbol{a}}{\partial \phi} \otimes \boldsymbol{n} \big)_S = h^1 h^2 \Big(\big(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1 \big) \cdot \boldsymbol{n} \Big) \cdot \frac{\partial \boldsymbol{a}}{\partial \phi} = \boldsymbol{0}$ by the symmetry of $\boldsymbol{\sigma}$, and the normal stress-jump $[\![\boldsymbol{\sigma}]\!] \cdot \boldsymbol{n}$ vanishes by construction. The remaining terms can then, using $\frac{\partial h^2}{\partial \phi} = -\frac{\partial h^1}{\partial \phi}$ and $-h^1 \boldsymbol{\sigma}^1 \frac{\partial h^2}{\partial \phi} + h^2 \boldsymbol{\sigma}^2 \frac{\partial h^1}{\partial \phi} = \big(h^1 \boldsymbol{\sigma}^1 + h^2 \boldsymbol{\sigma}^2 \big) \frac{\partial h^1}{\partial \phi} = \boldsymbol{\sigma} \frac{\partial h^1}{\partial \phi}$, be summarized to

$$\frac{\partial f_{el}}{\partial \phi} = \big(f_{el}^1 - f_{el}^2 \big) \frac{\partial h^1}{\partial \phi} + \boldsymbol{\sigma} : \big(\boldsymbol{a} \otimes \boldsymbol{n} \big)_S \frac{\partial h^1}{\partial \phi} = \big(f_{el}^1 - f_{el}^2 \big) \frac{\partial h^1}{\partial \phi} + \big(\boldsymbol{\sigma} \cdot \boldsymbol{n} \big) \cdot \boldsymbol{a} \frac{\partial h^1}{\partial \phi}, \tag{7.97}$$

[63] Note that in the two-phase setting, the assumption that $\boldsymbol{n}$ is a function of $\nabla \phi$ alone is very natural. This can potentially be different in the multi-phase setting as when e.g. defining the normal(s) in terms of the $\boldsymbol{q}^{\alpha\beta}$ from Section 6.1 (making the normal(s) a function of both $\nabla \boldsymbol{\phi}$ and $\boldsymbol{\phi}$) and thus leading to an additional contribution to $\frac{\partial f_{el}}{\partial \boldsymbol{\phi}^\alpha}$ itself as well.

[64] The error in [64] is hidden in the claim that "the partial derivatives are evaluated at constant $\boldsymbol{\epsilon}_{\boldsymbol{B}}^\alpha = (\boldsymbol{\epsilon}_{\boldsymbol{n}}, \boldsymbol{\epsilon}_{\boldsymbol{t}})$", where $\boldsymbol{\epsilon}_{\boldsymbol{n}}$ and $\boldsymbol{\epsilon}_{\boldsymbol{t}}$ are the normal and tangential strain components.

or, with $\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S = \boldsymbol{\epsilon}^2 - \boldsymbol{\epsilon}^1$, to the alternative expression

$$\frac{\partial f_{el}}{\partial\phi} = \Big(\big(f_{el}^1 - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^1\big) - \big(f_{el}^2 - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^2\big)\Big)\frac{\partial h^1}{\partial\phi} \tag{7.98}$$

using only the phase-specific strains.
Proceeding similarly for the derivative with respect to $\nabla\phi$, one has[65]

$$\begin{aligned}\frac{\partial f_{el}}{\partial\nabla\phi} =& \frac{\partial}{\partial\nabla\phi}\Big(-h^1h^2\big(\boldsymbol{\sigma}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S + h^1h^2\boldsymbol{\sigma}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\big)\Big)\\ =& h^1h^2\frac{\partial}{\partial\boldsymbol{n}}\left(\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{a}\Big)\cdot\boldsymbol{n}\right)\cdot\frac{\partial\boldsymbol{n}}{\partial\nabla\phi} + h^1h^2\frac{\partial}{\partial\boldsymbol{a}}\left(\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{n}\Big)\cdot\boldsymbol{a}\right)\cdot\frac{\partial\boldsymbol{a}}{\partial\nabla\phi}\\ =& h^1h^2\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{a}\Big)\cdot\frac{\partial\boldsymbol{n}}{\partial\nabla\phi},\end{aligned} \tag{7.99}$$

where the term $\frac{\partial a}{\partial\nabla\phi}$ drops out by the continuity of the normal stresses. Using

$$\frac{\partial\boldsymbol{n}}{\partial\nabla\phi} = \frac{\partial}{\partial\nabla\phi}\Big(-\frac{\nabla\phi}{|\nabla\phi|}\Big) = -\frac{|\nabla\phi|\boldsymbol{I}-\nabla\phi\otimes\frac{\nabla\phi}{|\nabla\phi|}}{|\nabla\phi|^2} = -\frac{1}{|\nabla\phi|}\Big(\boldsymbol{I}-\boldsymbol{n}\otimes\boldsymbol{n}\Big)$$

and, by symmetry of $\boldsymbol{\sigma}$, $h^1h^2\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{a}\Big)\cdot\big(\boldsymbol{n}\otimes\boldsymbol{n}\big) = h^1h^2\Big(\boldsymbol{a}\cdot\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{n}\Big)\boldsymbol{n} = \boldsymbol{0}$, one thus finally obtains

$$\frac{\partial f_{el}}{\partial\nabla\phi} = -\frac{1}{|\nabla\phi|}h^1h^2\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{a}\Big). \tag{7.100}$$

Remark 129. The appearence of the prefactor $\frac{1}{|\nabla\phi|}$ in the derivative (7.100) is clearly somewhat problematic in regions where $|\nabla\phi|$ approaches 0 (or even worse, actually is zero). For "realistic" phase-field profiles in the obstacle case, this essentially concerns the transition region from the interface to the neighboring bulk values. Assuming a phasefield profile corresponding roughly to the standard sinusoidal shape, the degree to which this contribution can cause numerical difficulties in these regions depends quite strongly on the interpolation function used. When using any of the higher-order interpolation functions, one does not expect any actual practical difficulties since the product h^1h^2 in this case converges superlinearly to zero (one factor converging to one, the other one superlinearly to zero) as one approaches the outer interface regions. In contrast, $|\nabla\phi|$ does so only linearly and one therefore expects an "extension by zero" to be a legitimate choice.
For the simplest interpolation function $h_0(\phi) = \phi$, the situation is somewhat more difficult though as $\frac{h^1h^2}{|\nabla\phi|}$ would, for the basic one-dimensional profile, converge to a finite but non-zero value. In the discrete setting, a simple extension by zero[66] to the region just outside the interface can thus lead, in combination with the discrete divergence-operator, to a discrete Dirac-type source-term. ◇

7.2.3 Comparison of the Different Formulations in the Two-Phase Case

Before discussing some issues related with the (inherently difficult) extension of the jump-condition based model(s) to the multiphase case, the following paragraphs will proceed to a

[65] Here use is made of $\boldsymbol{\sigma}^\alpha:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S = \big(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{a}\big)\cdot\boldsymbol{n} = \big(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}\big)\cdot\boldsymbol{a}$ in order to avoid the (purely notational) difficulty of differentating the symmetrized form with respect to a vector.

[66] This may in particular happen "automatically" depending on the discretization. Using e.g. a standard cell-centered finite difference discretization based on a central gradient operator, the first cell with $\phi = 0$ would satisfy $h^1h^2 = 0$ even though the discrete gradient would not yet vanish due to its broader stencil. This should not be interpreted as there not being a problem anymore, but simply shows that the non-zero limit $\frac{h^1h^2}{|\nabla\phi|}$ is not properly approximated in this case.

detailed comparison of the formulations in [51], [24], [64] and [74]. Even though this is a somewhat lengthy endeavor as some of them a priori differ quite substantially in their formulation, this effort seems worthwile for several reasons.
Firstly, this will serve to verify - as already indicated above - that the models proposed in [51] (in its small-deformation analogon), [64] (except for a missing contribution to the phasefield equation) and [74] are in fact one and the same model and just differ in their representation. While this is intuitively to be expected, this is in some cases not at all obvious based on the actual expression for the various quantities such as the elastic free energy density, the effective stress and the contributions to the phasefield equation. All three in turn can be considered to improve upon the one in [24], which, while in the same spirit, is in parts based on a somewhat different assumption.
Secondly, even though discussing different representations of an equivalent model may seem like a purely superficial effort, their differences can in fact be quite important. On the one hand, different descriptions will give rise to different implementations which can differ quite significantly in their respective computation effort. On the other hand, the various representations are suggestive of different **non-equivalent** ways of generalizing the two-phase models to the multiphase case. While some of the differences in the resulting models are fairly obvious, others are much harder to understand without properly understanding the differences in the two-phase case.
Last but not least, as each multiphase model generalization has its advantages and disadvantages, it may be quite useful to dispose of an implementation for several ones (such as is the case in the **Pace3D**-framework). As will be seen in the following discussion, the models considered do, despite their inherent complexity and the differences in the precise formulation, for the most part rely on a fairly reduced set of "central" opertations. From a practical point of view, understanding which these are and making them accessible in a unified manner for the different model implementations can significantly reduce code overhead, increase readibility and reduce the risk of implementation errors.

The Jump-Vector Based Approach by Mosler et al. [51]

The model in [51] is, just as the one outlined in Section 7.2.2, also based upon the use of a jump-vector $\boldsymbol{a}$ and the latter can in fact be considered to be a simple consequence of the small deformation analogue of the former. Even though the details with respect to the mechanical model and the drivingforce are already worked out in Section 7.2.2, it is, in particular with respect to Remark 127, worthwhile to recall the slightly different starting point in [51] based on a variational argument instead of the (from a mechanical point of view) sligthly more direct approach above.
Their approach, termed **partial rank one relaxation**, is based upon introducing a jump vector $\boldsymbol{a}$ such that $[\![\boldsymbol{F}]\!] = \boldsymbol{F}^{(2)} - \boldsymbol{F}^{(1)} = \boldsymbol{a} \otimes \boldsymbol{N}$ and, consistent with the averaging condition, $\boldsymbol{F}^1 = \boldsymbol{F} - \phi \boldsymbol{a} \otimes \boldsymbol{N}$ and $\boldsymbol{F}^2 = \boldsymbol{F} + (1-\phi)\boldsymbol{a} \otimes \boldsymbol{N}$. The jump vector $\boldsymbol{a}$ is then chosen such that it minimizes the total free energy for a given deformation gradient $\boldsymbol{F}$.
Translating this to the small deformation setting (i.e. in particular identifying the normal vector $\boldsymbol{N}$ in the reference configuration with the one in the current one, $\boldsymbol{N} \approx \boldsymbol{n}$) and replacing the weighting by ϕ with the one in terms of the weighting-functions $h^1(\phi)$ and $h^2(\phi)$, this corresponds to defining $\boldsymbol{F}^\alpha$ and thus $\boldsymbol{\epsilon}^\alpha$, $\alpha = 1, 2$, as in Equation (7.87), where $\boldsymbol{a}$ is defined in terms of the minimization problem

$$\min_{\boldsymbol{a}} \left\{ f_{el}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}, \boldsymbol{\epsilon}) \right\} = \min_{\boldsymbol{a}} \left\{ \sum_{\alpha=1}^{2} h^\alpha(\phi) \frac{1}{2} (\boldsymbol{\epsilon}^\alpha(\phi, \boldsymbol{a}, \boldsymbol{n}) - \tilde{\boldsymbol{\epsilon}}^\alpha) : \boldsymbol{\mathcal{C}}^\alpha : (\boldsymbol{\epsilon}^\alpha(\phi, \boldsymbol{a}, \boldsymbol{n}) - \tilde{\boldsymbol{\epsilon}}^\alpha) \right\}. \qquad (7.101)$$

Using $\boldsymbol{\epsilon}^1 = \boldsymbol{\epsilon} - h^2\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$ and $\boldsymbol{\epsilon}^2 = \boldsymbol{\epsilon} + h^1\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$ and differentiating with respect to $\boldsymbol{a}$ leads to the FONC

$$\begin{aligned}\frac{\partial f_{el}}{\partial \boldsymbol{a}}(\boldsymbol{\epsilon},\phi,\boldsymbol{a},\boldsymbol{n})\cdot \mathrm{d}\boldsymbol{a} &= \sum_{\alpha=1}^{2} h^\alpha(\phi)\big(\boldsymbol{\mathcal{C}}^\alpha : (\boldsymbol{\epsilon}^\alpha(\phi,\boldsymbol{a},\boldsymbol{n}) - \tilde{\boldsymbol{\epsilon}}^\alpha)\big) : \big(\frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{a}}\cdot \mathrm{d}\boldsymbol{a}\big) = \sum_{\alpha=1}^{2} h^\alpha(\phi)\boldsymbol{\sigma}^\alpha : \frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{a}}\\ &= h^1(\phi)\boldsymbol{\sigma}^1 : \big(-h^2(\mathrm{d}\boldsymbol{a}\otimes\boldsymbol{n})_S\big) + h^2(\phi)\boldsymbol{\sigma}^2 : \big(h^1(\mathrm{d}\boldsymbol{a}\otimes\boldsymbol{n})_S\big)\\ &= h^1(\phi)h^2(\phi)\big(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\big) : (\mathrm{d}\boldsymbol{a}\otimes\boldsymbol{n})_S = h^1(\phi)h^2(\phi)\Big(\big(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\big)\cdot\boldsymbol{n}\Big)\cdot \mathrm{d}\boldsymbol{a} \overset{!}{=} \boldsymbol{0},\end{aligned}$$

and thus, within the interface region (where $h^1h^2 \neq 0$), to the continuity condition $\boldsymbol{\sigma}^1\cdot\boldsymbol{n} = \boldsymbol{\sigma}^2\cdot\boldsymbol{n}$. From this, it is on the one hand clear that the variational characterization by [51] is (at least whenever $h^1(\phi)h^2(\phi) \neq 0$) equivalent to imposing the jump conditions on both the strains and stresses as above, and on the other hand that the dependence of $\boldsymbol{a}$ on ϕ and $\nabla\phi$ has, as already seen above, no bearing on the derivative of f_{el} with respect to ϕ and $\nabla\phi$ as these conditions characterize precisely the points where $\frac{\partial f_{el}}{\partial \boldsymbol{a}} = \boldsymbol{0}$.

An Alternative Variational Characterization of the Model by Schneider et al. [64]

The same two jump conditions are also the ones which the model by [64] and the mechanical part of the model in [24] are based upon. In contrast to the use of the jump vector $\boldsymbol{a}$ above though, the description in [64] uses a transformation matrix $\boldsymbol{Q}$ constructed as $\boldsymbol{Q}^T = \begin{pmatrix}\boldsymbol{n} & \boldsymbol{t} & \boldsymbol{s}\end{pmatrix}$ from an orthonormal set $\big(\boldsymbol{n},\boldsymbol{t},\boldsymbol{s}\big)$ of vectors consisting of the normal $\boldsymbol{n}$ and the two tangential vectors $\boldsymbol{t}$ and $\boldsymbol{s}$. $\boldsymbol{Q}$ is thus a unitary matrix satisfying $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$, and can be used to transform the original entries of the stress-tensor and (analogously for the strain tensor) in their Cartesian representation to a new orthonormal coordinate system by setting

$$\boldsymbol{\sigma}_{\mathcal{B}} \coloneqq \boldsymbol{Q}\boldsymbol{\sigma}\boldsymbol{Q}^T = \begin{pmatrix}\boldsymbol{n}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{n} & \boldsymbol{n}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{t} & \boldsymbol{n}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{s}\\ \boldsymbol{t}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{n} & \boldsymbol{t}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{t} & \boldsymbol{t}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{s}\\ \boldsymbol{s}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{n} & \boldsymbol{s}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{t} & \boldsymbol{s}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{s}\end{pmatrix} =: \begin{pmatrix}\sigma_{nn} & \sigma_{nt} & \sigma_{ns}\\ \sigma_{tn} & \sigma_{tt} & \sigma_{ts}\\ \sigma_{sn} & \sigma_{st} & \sigma_{ss}\end{pmatrix}$$

and analogously for $\boldsymbol{\epsilon}_B \coloneqq \boldsymbol{Q}\boldsymbol{\epsilon}\boldsymbol{Q}^T$.
The advantage of using this coordinate system is that the matrices representing the continuity condition on the phase-specific stresses and strains therefore have a particularly simple algebraic structure as the coordinates of the normal vector $\boldsymbol{n}$ in this new coordinate system are given by $\boldsymbol{Q}^T\boldsymbol{n} = \begin{pmatrix}\boldsymbol{n}\cdot\boldsymbol{n} & \boldsymbol{t}\cdot\boldsymbol{n} & \boldsymbol{s}\cdot\boldsymbol{n}\end{pmatrix}^T = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^T$. More precisely, as $[\![\boldsymbol{\epsilon}]\!]$ is assumed to be of the form $\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$, premultiplying this equality by $\boldsymbol{Q}$ and postmultiplying the result with $\boldsymbol{Q}^T$, it is easily seen that, in this new coordinate system, $[\![\boldsymbol{\epsilon}_B]\!]$ is of the form

$$[\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!] = \frac{1}{2}\begin{pmatrix}\boldsymbol{n}^T\\ \boldsymbol{t}^T\\ \boldsymbol{s}^T\end{pmatrix}\big(\boldsymbol{a}\otimes\boldsymbol{n} + \boldsymbol{n}\otimes\boldsymbol{a}\big)\begin{pmatrix}\boldsymbol{n} & \boldsymbol{t} & \boldsymbol{s}\end{pmatrix} = \frac{1}{2}\begin{pmatrix}2\boldsymbol{a}\cdot\boldsymbol{n} & \boldsymbol{a}\cdot\boldsymbol{t} & \boldsymbol{a}\cdot\boldsymbol{s}\\ \boldsymbol{t}\cdot\boldsymbol{a} & 0 & 0\\ \boldsymbol{s}\cdot\boldsymbol{a} & 0 & 0\end{pmatrix},$$

and, by the symmetry of the scalar product, therefore satisfies

$$[\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!] = \begin{pmatrix}[\![\boldsymbol{\epsilon}_B]\!]_{\boldsymbol{nn}} & [\![\boldsymbol{\epsilon}_B]\!]_{\boldsymbol{nt}} & [\![\boldsymbol{\epsilon}_B]\!]_{\boldsymbol{ns}}\\ [\![\boldsymbol{\epsilon}_B]\!]_{\boldsymbol{nt}} & 0 & 0\\ [\![\boldsymbol{\epsilon}_B]\!]_{\boldsymbol{ns}} & 0 & 0\end{pmatrix}. \tag{7.102}$$

Remark 130. The reverse conclusion also holds, i.e. if $\boldsymbol{\epsilon}_B = \boldsymbol{Q}\boldsymbol{\epsilon}\boldsymbol{Q}^T$ is of the form as in Equation (7.102), there is some $\boldsymbol{a}$ such that $[\![\boldsymbol{\epsilon}]\!]$ is of the form $\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$. In fact, due to the symmetry of $\boldsymbol{\epsilon}_B$, obtaining such an equality requires satisfying the three equations $\boldsymbol{a}\cdot\boldsymbol{n} = [\![\epsilon_{nn}]\!]$, $\boldsymbol{a}\cdot\boldsymbol{t} = 2[\![\epsilon_{nt}]\!]$

and $\boldsymbol{a}\cdot\boldsymbol{s} = 2[\![\epsilon_{ns}]\!]$, or, in matrix-vector form, $\boldsymbol{Q}\boldsymbol{a} = \begin{pmatrix} [\![\epsilon_{nn}]\!] & 2[\![\epsilon_{nt}]\!] & 2[\![\epsilon_{ns}]\!] \end{pmatrix}^T$. As $\boldsymbol{Q}$ is a unitary matrix (with $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$), the vector

$$\boldsymbol{a} = \boldsymbol{Q}^T \begin{pmatrix} [\![\epsilon_{nn}]\!] \\ 2[\![\epsilon_{nt}]\!] \\ 2[\![\epsilon_{ns}]\!] \end{pmatrix} \tag{7.103}$$

clearly satisfies this requirement. ◇

Similarly, pre- and postmultiplying $\boldsymbol{\sigma}$ with $\boldsymbol{Q}$ and $\boldsymbol{Q}^T$ and making use of $\boldsymbol{n}_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$ and the symmetry of $\boldsymbol{\sigma}_{\mathcal{B}}$, it is easy to see that the continuity condition on the normal stresses implies that the stress jump is on the complementary entries and therefore of the form

$$[\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & [\![\boldsymbol{\sigma}_{\mathcal{B}}]\!]_{\boldsymbol{tt}} & [\![\boldsymbol{\sigma}_{\mathcal{B}}]\!]_{\boldsymbol{ts}} \\ 0 & [\![\boldsymbol{\sigma}_{\mathcal{B}}]\!]_{\boldsymbol{st}} & [\![\boldsymbol{\sigma}_{\mathcal{B}}]\!]_{\boldsymbol{ss}} \end{pmatrix} \tag{7.104}$$

whereas the continuous components of $\boldsymbol{\sigma}$ are given by $\boldsymbol{\sigma}_{\boldsymbol{nn}}$, $\boldsymbol{\sigma}_{\boldsymbol{tn}} = \boldsymbol{\sigma}_{\boldsymbol{nt}}$ and $\boldsymbol{\sigma}_{\boldsymbol{sn}} = \boldsymbol{\sigma}_{\boldsymbol{ns}}$.

Remark 131. The following discussion (in particular the retransformation to the original coordinate system in the next section) will, for shortening the notation, only be performed in a two-dimensional setting, in which, with the single tangential vector $\boldsymbol{t}$ the jump in $\boldsymbol{\epsilon}$ takes the form

$$[\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!] = \begin{pmatrix} [\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!]_{\boldsymbol{nn}} & [\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!]_{\boldsymbol{nt}} \\ [\![\boldsymbol{\epsilon}_{\mathcal{B}}]\!]_{\boldsymbol{nt}} & 0 \end{pmatrix}$$

whereas the continuity condition on $\boldsymbol{\sigma}_{\mathcal{B}}$ carries on $\boldsymbol{\sigma}_{\boldsymbol{nn}}$ and $\boldsymbol{\sigma}_{\boldsymbol{nt}} = \boldsymbol{\sigma}_{\boldsymbol{tn}}$. Most of it is in fact independent of whether one works in two or three dimensions though, and the rest easily carries over to the three-dimensional case as indicated in Remark 146, simply requiring somewhat more lenghty formulae.

◇

Remark 132. The following discussion will (as in [64]) primarily make use of the in practice more common Voigt notation. Since this formalism is for the most part based upon matrix-vector and matrix-matrix multiplications only and there in particular being no point in "visually" differentiating between single- and double-contractions, single contractions will not be marked by a $(\cdot)$-symbol except for actual scalar products between vectors or second-order tensors in Voigt-notation.
In contrast, the contraction-symbols will continue to be used in all expressions which are to be interpreted as using the "non-Voigt"-notation. In particular, several formulae contain both Voigt-notation and the normal traction vector $\boldsymbol{\sigma}\cdot\boldsymbol{n}$, a notation which seems much more natural then its more cumbersome[67] counterpart $\boldsymbol{B}\boldsymbol{\sigma}^v$ in Voigt-notation (see below). ◇

Using Voigt-notation, the transformations above can be reexpressed as (and similarly for the eigenstrains)

$$\boldsymbol{\sigma}^v_{\mathcal{B}} = \begin{pmatrix} \sigma_{nn} \\ \sigma_{tt} \\ \sigma_{nt} \end{pmatrix} = \boldsymbol{M}_{\boldsymbol{\sigma}} \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\epsilon}^v_{\mathcal{B}} = \begin{pmatrix} \epsilon_{nn} \\ \epsilon_{tt} \\ 2\epsilon_{nt} \end{pmatrix} = \boldsymbol{M}_{\boldsymbol{\epsilon}} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ 2\epsilon_{xy} \end{pmatrix}$$

using the two matrices

$$\boldsymbol{M}_{\boldsymbol{\sigma}} = \begin{pmatrix} n_x^2 & n_y^2 & 2n_xn_y \\ t_x^2 & t_y^2 & 2t_xt_y \\ n_xt_x & n_yt_y & n_xt_y + n_yt_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{M}_{\boldsymbol{\epsilon}} = \begin{pmatrix} n_x^2 & n_y^2 & n_xn_y \\ t_x^2 & t_y^2 & t_xt_y \\ 2n_xt_x & 2n_yt_y & n_xt_y + n_yt_x \end{pmatrix}.$$

[67] Note that the multiplication of the stress-tensor by $\boldsymbol{n}$ in Voigt-notation has to map the "vector" $\boldsymbol{\sigma}^v$ to the normal traction vector, and thus (by linearity) is represented in terms of a matrix. Even though mapping the second-order tensor $\boldsymbol{\sigma}$ to the same vector in fact involves an even higher-dimensional construct, its representation in terms of the contraction with $\boldsymbol{n}$ is much more compact and readable than the one in the Voigt-case.

Using this new coordinate system, $\left(\boldsymbol{\epsilon}^v\right)^\alpha = \boldsymbol{M}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\epsilon}_{\mathcal{B}}^v$, and, with $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{-1} = \boldsymbol{M}_{\boldsymbol{\sigma}}^T$, this allows rewriting the elastic free energy density of each phase as

$$\begin{aligned} f_{el}^\alpha\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha\Big) =& \frac{1}{2}\left(\boldsymbol{M}_{\boldsymbol{\sigma}}^T\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha\Big)\right)\cdot\left(\boldsymbol{\mathcal{C}}^v\right)^\alpha\left(\boldsymbol{M}_{\boldsymbol{\sigma}}^T\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha\Big)\right) \\ =& \frac{1}{2}\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha\Big)\cdot\Big(\boldsymbol{M}_{\boldsymbol{\sigma}}\left(\boldsymbol{\mathcal{C}}^v\right)^\alpha\boldsymbol{M}_{\boldsymbol{\sigma}}^T\Big)\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha\Big). \end{aligned}$$

Imposing in accordance with the Hadamard condition the equality $\epsilon_{tt}^\alpha = \epsilon_{tt}$, $\alpha = 1,2$ for the tt-component of the given strain and decomposing $\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha$ as[68]

$$\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha - \left(\tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^v\right)^\alpha = \begin{pmatrix} \epsilon_{nn}^\alpha - \tilde{\epsilon}_{nn}^\alpha \\ \epsilon_{tt} - \tilde{\epsilon}_{tt}^\alpha \\ 2\big(\epsilon_{nt}^\alpha - \tilde{\epsilon}_{nt}^\alpha\big) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}\Bigg(\underbrace{\begin{pmatrix} \epsilon_{nn}^\alpha \\ 2\epsilon_{nt}^\alpha \end{pmatrix}}_{=:\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha} - \underbrace{\begin{pmatrix} \tilde{\epsilon}_{nn}^\alpha \\ 2\tilde{\epsilon}_{nt}^\alpha \end{pmatrix}}_{=:\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha}\Bigg) + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\big(\underbrace{\epsilon_{tt}}_{=:\boldsymbol{\epsilon}_{\boldsymbol{t}}} - \underbrace{\tilde{\epsilon}_{tt}^\alpha}_{=:\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha}\big)$$

and using the abbreviating $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha := \Big(\boldsymbol{M}_{\boldsymbol{\sigma}}\left(\boldsymbol{\mathcal{C}}^v\right)^\alpha\boldsymbol{M}_{\boldsymbol{\sigma}}^T\Big)$ together with the symmetry of $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha$, this density can be rewritten, similarly to [64], as

$$\begin{aligned} f_{el}^\alpha\Big(\left(\boldsymbol{\epsilon}_{\mathcal{B}}^v\right)^\alpha\Big) =& \frac{1}{2}\begin{pmatrix} \epsilon_{nn}^\alpha - \tilde{\epsilon}_{nn}^\alpha \\ \epsilon_{tt} - \tilde{\epsilon}_{tt}^\alpha \\ 2\big(\epsilon_{nt}^\alpha - \tilde{\epsilon}_{nt}^\alpha\big) \end{pmatrix}\cdot\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha\begin{pmatrix} \epsilon_{nn}^\alpha - \tilde{\epsilon}_{nn}^\alpha \\ \epsilon_{tt} - \tilde{\epsilon}_{tt}^\alpha \\ 2\big(\epsilon_{nt}^\alpha - \tilde{\epsilon}_{nt}^\alpha\big) \end{pmatrix} \\ =& \frac{1}{2}\Big(\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\big)\cdot\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\big) + 2\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\big)\cdot\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) + \big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big)\cdot\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big)\Big), \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha &= \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \end{aligned} \tag{7.105}$$

Remark 133. Note that one could also reorder the stress- and strain-entries as in [64] in order to have these auxiliary matrices correspond to contiguous subblocks of $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^\alpha$. While this does not affect the theoretical discussion, it is clearly advantageous for a practical implementation of the scheme above as this simplifies the access to the relevant part of the stiffness tensor and can be performed essentially free of overhead by appropriately permuting the rows of $\boldsymbol{M}_{\boldsymbol{\epsilon}}$ and $\boldsymbol{M}_{\boldsymbol{\sigma}}$ during their construction.
As will be discussed below though, the primary advantage of the description in [64] (and in a somewhat different form in [24]) is that they lead (for the most part[69]) to a pleasent theoretical representation. In practice, choosing such a representation in the particular coordinate system described through $\boldsymbol{Q}$ above leads to a significant amount of computational overhead due to the basis transformations involved such that for actual computations, other descriptions may be preferrable. ◇

[68] Even though the definition of $\boldsymbol{\epsilon}_{\boldsymbol{t}}$ below is clearly superfluous in the two-dimensional setting considered here, it leads to a notation which is on the one hand more consistent with the one in [64] and on the other hand leads to calculations which are equally vaild in the three-dimensional case where both $\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha$ and $\boldsymbol{\epsilon}_{\boldsymbol{t}}$ consist of three entries.

[69] An important exception here is performing the derivative with respect to $\nabla\phi$ as this, in contrast to the descriptions in [51] and [74], involves the normal and tangential vectors. In addition, a large part of these dependencies is essentially "artificial" and would, through a likely lenghty calculation, cancel, as their only purpose is to obtain the component representation of the vectors and matrices in this particular coordinate system, which does not affect these quantities themselves.

Fixing $\boldsymbol{\epsilon_t}$ in terms of the given strain still leaves open the definition of the $\boldsymbol{\epsilon_n^\alpha}$ in terms of the given average normal strain components $\boldsymbol{\epsilon_n}$. As a first condition, one can impose that $\boldsymbol{\epsilon_n}$ should be given by the weighted average of the phase-specific normal strains,

$$\boldsymbol{\epsilon_n} = h^1(\phi)\boldsymbol{\epsilon_n^1} + h^2(\phi)\boldsymbol{\epsilon_n^2}. \tag{7.106}$$

As a second condition, one can either, as in [64], directly impose the continuity condition $\boldsymbol{\sigma_n^1} = \boldsymbol{\sigma_n^2} = \boldsymbol{\sigma_n}$, or, in a manner which underlines some of the similarities with the chemical model from Section 7.1, instead rely on an energetic argument in combination with the constraint (7.106). In fact, augmenting this constraint in the Lagrange function

$$L\big(\boldsymbol{\epsilon_n^1}, \boldsymbol{\epsilon_n^2}, \boldsymbol{\epsilon_t}, \boldsymbol{\sigma_n}\big) = f^1\big(\boldsymbol{\epsilon_n^1}, \boldsymbol{\epsilon_n}, \boldsymbol{\epsilon_t}\big)h^1(\phi) + f^2\big(\boldsymbol{\epsilon_n^2}, \boldsymbol{\epsilon_t}\big)h^2(\phi) - \boldsymbol{\sigma_n}\cdot\big(h^1(\phi)\boldsymbol{\epsilon_n^1} + h^2(\phi)\boldsymbol{\epsilon_n^2} - \boldsymbol{\epsilon_n}\big) \tag{7.107}$$

and using the symmetry of $\boldsymbol{\mathcal{C}}^\alpha$, it is easily seen that the minimizers $\boldsymbol{\epsilon_n^1}$ and $\boldsymbol{\epsilon_n^2}$ for the problem

$$\min_{(\boldsymbol{\epsilon_n^1}, \boldsymbol{\epsilon_n^2}) \,:\, h^1(\phi)\boldsymbol{\epsilon_n^1} + h^2(\phi)\boldsymbol{\epsilon_n^2} = \boldsymbol{\epsilon_n}} \Big\{ f^1\big(\boldsymbol{\epsilon_n^1}, \boldsymbol{\epsilon_t}\big)h^1(\phi) + f^2\big(\boldsymbol{\epsilon_n^2}, \boldsymbol{\epsilon_t}\big)h^2(\phi) \Big\}$$

are characterized by

$$\frac{\partial L}{\partial \boldsymbol{\epsilon_n^\alpha}} = h^\alpha(\phi)\Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha\big(\boldsymbol{\epsilon_n^\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) - \boldsymbol{\sigma_n}\Big) = \boldsymbol{0}, \quad \alpha = 1, 2,$$

and thus that, whenever $h^\alpha(\phi) > 0$, the decomposition of the strains with the minimal energy is characterized by the equality of the phase-specific normal stresses $\boldsymbol{\sigma_n^\alpha} = \begin{pmatrix} \sigma_{\boldsymbol{nn}}^\alpha & \sigma_{\boldsymbol{nt}}^\alpha \end{pmatrix}^T$.
As in [64], and with the abbreviation[70] $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha = \big(\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha\big)^{-1}$, one can then first solve for $\big(\boldsymbol{\epsilon_n^1}, \boldsymbol{\epsilon_n^2}\big)$ in terms of the (yet unknown) $\boldsymbol{\sigma_n}$ as $\boldsymbol{\epsilon_n^\alpha} = \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha + \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha\Big(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big)\Big)$ and then determine $\boldsymbol{\sigma_n}$ from the constraint (7.106) through

$$\sum_{\alpha=1}^2 \boldsymbol{\epsilon_n^\alpha} h^\alpha(\phi) = \sum_{\alpha=1}^2 \left(\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha + \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha\Big(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big)\Big)\right) h^\alpha(\phi) \overset{!}{=} \boldsymbol{\epsilon_n}$$

to be given by

$$\boldsymbol{\sigma_n} = \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\boldsymbol{\phi})\Big)^{-1} \left(\boldsymbol{\epsilon_n} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}(\phi) + \sum_{\alpha=1}^2 \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) h^\alpha(\phi) \right) \tag{7.108}$$

with $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}(\phi) \coloneqq \sum_{\alpha=1}^2 \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha h^\alpha(\phi)$.
Reinserting this normal stress into the expressions for the $\boldsymbol{\epsilon_n^\alpha}$ and combining these with the material law then shows that the tangential stress is, as the weighted average of the $\boldsymbol{\sigma_t^\alpha}$, given by

$$\begin{aligned}
\boldsymbol{\sigma_t} &= \sum_{\alpha=1}^2 \left(\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha \Big(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big)\Big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) \right) h^\alpha(\phi) \\
&= \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\boldsymbol{\phi}) \Big) \boldsymbol{\sigma_n} + \sum_{\alpha=1}^2 \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha - \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha \Big) \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) h^\alpha(\phi) \\
&= \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\phi) \Big) \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\phi) \Big)^{-1} \big(\boldsymbol{\epsilon_n} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}(\phi)\big) \\
&\quad + \sum_{\alpha=1}^2 \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha - \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha \Big) \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) h^\alpha(\phi) \\
&\quad + \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\phi) \Big) \Big(\sum_{\alpha=1}^2 \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha h^\alpha(\phi) \Big)^{-1} \sum_{\alpha=1}^2 \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\big) h^\alpha(\phi).
\end{aligned} \tag{7.109}$$

[70] Even though this notation is somewhat inconsistent as $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha$ is in fact the $\boldsymbol{n}-\boldsymbol{n}$-subblock of $\boldsymbol{\mathcal{C}}^\alpha$, whereas $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha$ is not the $\boldsymbol{n}-\boldsymbol{n}$-subblock of $\boldsymbol{\mathcal{S}}^\alpha$, it will nevertheless be used here as it significantly shortens the notation and there is no actual risk of confusing the meaning here as $\boldsymbol{\mathcal{S}}^\alpha$ itself is never used.

Using the same abbreviations

$$\begin{aligned}
&\bar{\mathcal{T}}_{nn} = -\sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} h^{\alpha}(\phi), && \bar{\mathcal{T}}_{nt} = \sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nt}^{\alpha} h^{\alpha}(\phi), \\
&\bar{\mathcal{T}}_{tn} = \sum_{\alpha=1}^{2} \mathcal{C}_{tn}^{\alpha} \mathcal{S}_{nn}^{\alpha} h^{\alpha}(\phi), && \bar{\mathcal{T}}_{tt} = \sum_{\alpha=1}^{2} \left(\mathcal{C}_{tt}^{\alpha} - \mathcal{C}_{tn}^{\alpha} \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nt}^{\alpha} \right) h^{\alpha}(\phi), \\
&\tilde{\chi}_{n} = \sum_{\alpha=1}^{2} \left(\tilde{\epsilon}_{n}^{\alpha} + \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nt}^{\alpha} \tilde{\epsilon}_{t}^{\alpha} \right) h^{\alpha}(\phi), && \tilde{\chi}_{t} = \sum_{\alpha=1}^{2} \left(\mathcal{C}_{tt}^{\alpha} - \mathcal{C}_{tn}^{\alpha} \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nt}^{\alpha} \right) \tilde{\epsilon}_{t}^{\alpha}
\end{aligned} \tag{7.110}$$

as in [64], it is easy to see that these stresses correspond precisely to the expressions

$$\sigma_{n} = -\bar{\mathcal{T}}_{nn}^{-1} \left(\epsilon_{n} + \bar{\mathcal{T}}_{nt} \epsilon_{t} - \tilde{\chi}_{n} \right) \tag{7.111}$$

and

$$\sigma_{t} = \bar{\mathcal{T}}_{tn} \sigma_{n} + \bar{\mathcal{T}}_{tt} \epsilon_{t} - \tilde{\chi}_{t} = -\bar{\mathcal{T}}_{tn} \bar{\mathcal{T}}_{nn}^{-1} \left(\epsilon_{n} + \bar{\mathcal{T}}_{nt} \epsilon_{t} - \tilde{\chi}_{n} \right) + \bar{\mathcal{T}}_{tt} \epsilon_{t} - \tilde{\chi}_{t} \tag{7.112}$$

given there. In the form above, the calculation of the effective stress therefore consists in the following four-step procedure:

1. transform the given strain ϵ^{v} and eigenstrains $\left(\tilde{\epsilon}^{v}\right)^{\alpha}$ to the $\mathcal{B}$-coordinate system;
2. transform the stiffness-tensor $\mathcal{C}$ to the $\mathcal{B}$-coordinate system and calculate the submatrices in Equation (7.110);
3. calculate the stresses $\sigma_{\mathcal{B}}$ as in equations (7.111) and (7.112);
4. retransform $\sigma_{\mathcal{B}}^{v}$ to the standard coordinate system.

Retransformation of the Mechanical Model to the Original Coordinate System

Even though this is a theoretically pleasing description, it involves (as indicated above) a quite notable overhead due to the heavy use of the components of the various tensors in the transformed coordinate system. For actual computational purposes, it is thus useful to reexpress this model based on quantities in the original coordinate system only. Besides the computational point of view, this retransformation is also instructive in its own right, as it relates various operations expressed in term of a - as well as for the model in [74] to be discussed below - with the more transparent "direct" calculations in the aligned coordinate-system $\mathcal{B}$.

For this it is more convenient to take the expression for σ_{n} in Equation (7.108) instead of Equation (7.111) as a starting point. Artificially expanding $\epsilon_{n} - \tilde{\epsilon}_{n}(\phi)$ as $\sum_{\alpha=1}^{2} \left(\epsilon_{n} - \tilde{\epsilon}_{n}^{\alpha}\right) h^{\alpha}(\phi) = \sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nn}^{\alpha} \left(\epsilon_{n} - \tilde{\epsilon}_{n}^{\alpha}\right) h^{\alpha}(\phi)$, σ_{n} can firstly be rewritten as

$$\sigma_{n} = \Big(\sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} h^{\alpha}(\phi) \Big)^{-1} \sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} \Big(\mathcal{C}_{nn}^{\alpha} \left(\epsilon_{n} - \tilde{\epsilon}_{n}^{\alpha} \right) + \mathcal{C}_{nt}^{\alpha} \left(\epsilon_{t} - \tilde{\epsilon}_{t}^{\alpha} \right) \Big) h^{\alpha}(\phi), \tag{7.113}$$

where the term in the latter parenthesis is precisely the normal phase-specific stress component corresponding to the Voigt-Taylor model, i.e. denoting these stresses by

$$\Sigma_{n}^{\alpha} := \mathcal{C}_{nn}^{\alpha} \left(\epsilon_{n} - \tilde{\epsilon}_{n}^{\alpha} \right) + \mathcal{C}_{nt}^{\alpha} \left(\epsilon_{t} - \tilde{\epsilon}_{t}^{\alpha} \right),$$

one has

$$\sigma_{n} = \Big(\sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} h^{\alpha}(\phi) \Big)^{-1} \sum_{\alpha=1}^{2} \mathcal{S}_{nn}^{\alpha} \Sigma_{n}^{\alpha} h^{\alpha}(\phi). \tag{7.114}$$

Based upon the middle expression for σ_{t},

$$\sigma_{t} = \Big(\sum_{\alpha=1}^{2} \mathcal{C}_{tn}^{\alpha} \mathcal{S}_{nn}^{\alpha} h^{\alpha}(\phi) \Big) \sigma_{n} + \sum_{\alpha=1}^{2} \left(\mathcal{C}_{tt}^{\alpha} - \mathcal{C}_{tn}^{\alpha} \mathcal{S}_{nn}^{\alpha} \mathcal{C}_{nt}^{\alpha} \right) \left(\epsilon_{t} - \tilde{\epsilon}_{t}^{\alpha} \right) h^{\alpha}(\phi)$$

in Equation (7.109) (or from the first one in (7.112) and the definitions in Equation (7.110)), one can derive an expression of similar nature for the tangential stress components. Adding and subtracting $\sum_{\alpha=1}^{2} \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)$, the second term (corresponding to $\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{tt}}\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\chi}}_{\boldsymbol{t}}$ in the notation of [64]) can be expanded to

$$\begin{aligned}
&\sum_{\alpha=1}^{2} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)\Big) h^{\alpha}(\phi) - \sum_{\alpha=1}^{2} \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\Big(\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big) h^{\alpha}(\phi) \\
&= \sum_{\alpha=1}^{2} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)\Big) h^{\alpha}(\phi) - \sum_{\alpha=1}^{2} \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big) h^{\alpha}(\phi).
\end{aligned}$$

The term in the last parenthesis is just $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$, whereas the first one is nothing but the tangential component of the stress predicted by the Voigt-Talyor model, i.e. with the analogous abbreviation $\boldsymbol{\Sigma}_{\boldsymbol{t}}^{\alpha} \coloneqq \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)$ as for the normal components, one has

$$\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{tt}}\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\chi}}_{\boldsymbol{t}} = \sum_{\alpha=1}^{2} \big(\boldsymbol{\Sigma}_{\boldsymbol{t}}^{\alpha} - \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big) h^{\alpha}(\phi).$$

The prefactors $\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}$ in the last sum are the same as the ones appearing in the prefactor of $\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{tn}}\boldsymbol{\sigma}_{\boldsymbol{n}}$, allowing to summarize the formula for the tangential stress components to

$$\boldsymbol{\sigma}_{\boldsymbol{t}} = \sum_{\alpha=1}^{2} \boldsymbol{\Sigma}_{\boldsymbol{t}}^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big) h^{\alpha}(\phi), \tag{7.115}$$

i.e. a first Voigt-Taylor-type contribution corrected by a term based on the differences of the real normal stress from the one predicted by the Voigt-Taylor model.

Remark 134. Even though this may not seem particularly convenient at this point, one can also artifically rewrite $\boldsymbol{\sigma}_{\boldsymbol{n}}$ in a similar form. In fact, simply expanding expressions by making use of $\sum_{\alpha=1}^{2} h^{\alpha}(\phi) = 1$ and $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} = \boldsymbol{I}$, one has

$$\boldsymbol{\sigma}_{\boldsymbol{n}} = \sum_{\alpha=1}^{2} \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \big(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big) h^{\alpha}(\phi) = \sum_{\alpha=1}^{2} \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big) h^{\alpha}(\phi). \tag{7.116}$$

This obviously does not actually contain any additional information (and in particular does not allow for the determination of $\boldsymbol{\sigma}_{\boldsymbol{n}}$ itself, for which one has to rely on Equation (7.114)), but shows that both the tangential and normal stress components can be expressed in a common form once the actual normal stress $\boldsymbol{\sigma}_{\boldsymbol{n}}$ is known, namely a Voigt-Taylor-type prediction together with a correction based on the $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big)$, $\alpha = 1, 2$ and $\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}$ resp. $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}$. ◇

Despite of the relative simplicity of the formulae for $\boldsymbol{\sigma}_{\boldsymbol{n}}$ and $\boldsymbol{\sigma}_{\boldsymbol{t}}$ when expressed in terms of the auxiliary quantities $\boldsymbol{\Sigma}_{\boldsymbol{n/t}}^{\alpha}$, both equations (7.114) and (7.115) still rely heavily upon the transformed quantities (and in particular the transformed stiffness-tensors $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^{\alpha}$) in the $\mathcal{B}$-system. A large part of this dependence can in fact be eliminated in a straightforward manner though as a closer look at the interplay between the extraction of the normal components and the transformation matrices will reveal.

A convenient starting point here are the $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}$-matrices, which, by Equation (7.105), are determined by

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{\mathcal{C}}_{\mathcal{B}}^{\alpha} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Big(\boldsymbol{M}_{\boldsymbol{\sigma}} \big(\boldsymbol{\mathcal{C}}^{v}\big)^{\alpha} \boldsymbol{M}_{\boldsymbol{\sigma}}^{T}\Big) \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Making use of

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{M}_{\boldsymbol{\sigma}} = \begin{pmatrix} n_x^2 & n_y^2 & 2n_x n_y \\ n_x t_x & n_y t_y & n_x t_y + n_y t_x \end{pmatrix} = \underbrace{\begin{pmatrix} n_x & n_y \\ t_x & t_y \end{pmatrix}}_{=:\boldsymbol{Q}} \begin{pmatrix} n_x & 0 & n_y \\ 0 & n_y & n_x \end{pmatrix}$$

and similarly

$$\boldsymbol{M}_{\boldsymbol{\sigma}}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} n_x & 0 \\ 0 & n_y \\ n_y & n_x \end{pmatrix} \boldsymbol{Q}^T$$

this can also be rewritten as

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha} = \boldsymbol{Q} \left(\begin{pmatrix} n_x & 0 & n_y \\ 0 & n_y & n_x \end{pmatrix} \left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha} \begin{pmatrix} n_x & 0 \\ 0 & n_y \\ n_y & n_x \end{pmatrix} \right) \boldsymbol{Q}^T.$$

Adopting the same notation $\boldsymbol{B}$ resp. $\boldsymbol{B}^T$ as in [74] (not to be confused with the caligraphic indicator $\mathcal{B}$ of the basis),

$$\boldsymbol{B} := \begin{pmatrix} n_x & 0 & n_y \\ 0 & n_y & n_x \end{pmatrix}, \tag{7.117}$$

these relations can be written more succinctly as

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{M}_{\boldsymbol{\sigma}} = \boldsymbol{Q}\boldsymbol{B}, \quad \boldsymbol{M}_{\boldsymbol{\sigma}}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{B}^T\boldsymbol{Q}^T \quad \text{and} \quad \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha} = \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}\boldsymbol{B}^T\boldsymbol{Q}^T. \tag{7.118}$$

Remark 135. It is easy to verify that $\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}\boldsymbol{B}^T$ corresponds, in non-Voigt notation, to the matrix $\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}$. In addition, while the use of $\boldsymbol{B}$ is convenient for notational purposes, this matrix need not actually be constructed in practice. Instead, it is both more "readable" and more efficient to evaluate the resulting expressions directly (i.e. the first entry of $\boldsymbol{\mathcal{C}}^v\boldsymbol{B}^T$ simply corresponds to $C^{xxxx}n_x + C^{xxxy}n_y$ in 2D resp. $C^{xxxx}n_x + C^{xxxy}n_y + C^{xxxz}n_z$ in 3-D, which is easily translated to Voigt-notation). ◇

Remark 136. The role of the outer factors $\boldsymbol{Q}$ and $\boldsymbol{Q}^T$ here is precisely the same as in [64], namely to transform the inner quantity from the Cartesian coordinate system to the system $\mathcal{B}$. In contrast to the original procedure of first transforming the fourth-order tensors $\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}$ in Voigt-notation using $\boldsymbol{M}_{\boldsymbol{\sigma}}$ and $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{-1} = \boldsymbol{M}_{\boldsymbol{\sigma}}^T$ and then extracting the normal subblocks, the major advantage of first contracting $\boldsymbol{\mathcal{C}}^{\alpha}$ with $\boldsymbol{n}$ from the left and right (resp. left-right-multiplying $\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}$ with $\boldsymbol{B}$ and $\boldsymbol{B}^T$) is that $\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}$ is a "standard" second-order tensor, which, even though it is symmetric, is not subject to the pitfalls of the Voigt-notation. In particular, it transforms based on the standard rules using $\boldsymbol{Q}$ and $\boldsymbol{Q}^T$, instead of the more complex original transformation. Being composed of the orthonormal vectors $\boldsymbol{n}$ and $\boldsymbol{t}$, it is also clear that $\boldsymbol{Q}$ is unitary, i.e. satisfies $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$. ◇

With the outer matrices in the expression for $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}$ being inverses of each other, this leads to the alternative expression

$$\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} = \boldsymbol{Q}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}\boldsymbol{B}^T\right)^{-1}\boldsymbol{Q}^T \tag{7.119}$$

for $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}$ and thus $\boldsymbol{\sigma}_{\boldsymbol{n}}$ from Equation (7.114) can, making use of $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, be rewritten as

$$\boldsymbol{\sigma}_{\boldsymbol{n}} = \boldsymbol{Q}\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}\boldsymbol{B}^T\right)^{-1} h^{\alpha}(\phi)\Big)^{-1} \sum_{\alpha=1}^{2} \left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^{\alpha}\boldsymbol{B}^T\right)^{-1} \boldsymbol{Q}^T\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\phi).$$

While the prefactor itself does not now make explicit use of the transformed stiffess-tensors $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^{\alpha}$, the evaluation of the $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$ itself still does. In accordance with the physical meaning of $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$, it is not surprising that one is also able to reexpress $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$ based on simple quantities from the original coordinate system. In fact, from the original expression $\boldsymbol{\Sigma}_{\boldsymbol{n}} = \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\left(\boldsymbol{\epsilon}_{\boldsymbol{n}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\left(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\right)$, the

definitions of $\boldsymbol{\epsilon_n} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}$ and $\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}$ and again making use of Equation (7.118) together with (7.105), $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$ may be rewritten as

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} = &\left(\boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{M}_{\boldsymbol{\sigma}}^{T}\begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}\right)\left(\begin{pmatrix}1&0&0\\0&0&1\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right)\right)\\ &+\left(\boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{M}_{\boldsymbol{\sigma}}^{T}\begin{pmatrix}0\\1\\0\end{pmatrix}\right)\left(\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right)\right).\end{aligned}$$

By extracting both the common pre- and postfactors $\boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{M}_{\boldsymbol{\sigma}}^{T}$ and $\boldsymbol{M}_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right)$ and since the sum of the two inner dyads is clearly the (3×3) identity matrix, this simplifies first to $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{M}_{\boldsymbol{\sigma}}^{T}\boldsymbol{I}\boldsymbol{M}_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right)$ and then, with $\boldsymbol{M}_{\boldsymbol{\sigma}}^{T} = \boldsymbol{M}_{\boldsymbol{\epsilon}}^{-1}$, to

$$\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right) = \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\Sigma}^{v}\right)^{\alpha} \tag{7.120}$$

with $\left(\boldsymbol{\Sigma}^{v}\right)^{\alpha} := \left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\left(\boldsymbol{\epsilon}^{v}-\left(\tilde{\boldsymbol{\epsilon}}^{v}\right)^{\alpha}\right)$. Reinserting this into the expression for $\boldsymbol{\sigma_n}$ and again cancelling $\boldsymbol{Q}^{T}$ and $\boldsymbol{Q}$, one finally has

$$\boldsymbol{\sigma_n} = \boldsymbol{Q}\left(\sum_{\alpha=1}^{2}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{B}^{T}\right)^{-1}h^{\alpha}(\boldsymbol{\phi})\right)^{-1}\sum_{\alpha=1}^{2}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{B}^{T}\right)^{-1}\boldsymbol{B}\left(\boldsymbol{\Sigma}^{v}\right)^{\alpha}h^{\alpha}(\phi), \tag{7.121}$$

i.e. an expression which, except for the appearence of the $\boldsymbol{Q}$-prefactor in front, is entirely based upon non-transformed quantities.

Remark 137. Instead of interpreting $\boldsymbol{\sigma_n}$ as the sub-vector of normal entries of the stress-tensor in Voigt-notation, one can also interpret it as the components of the normal stress $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ in the system $\mathcal{B}$. In fact, as

$$\boldsymbol{\sigma_n} = \begin{pmatrix}\sigma_{nn}\\ \sigma_{tn}\end{pmatrix} = \begin{pmatrix}\boldsymbol{n}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{n}\\ \boldsymbol{t}\cdot\boldsymbol{\sigma}\cdot\boldsymbol{n}\end{pmatrix} = \boldsymbol{Q}\left(\boldsymbol{\sigma}\cdot\boldsymbol{n}\right),$$

it is clear that the role of $\boldsymbol{Q}$ (being composed row-wise of the two orthonormal vectors $\boldsymbol{n}$ and $\boldsymbol{t}$) in Equation (7.121) is simply to "read out" the coordinates of $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ in the basis formed by $(\boldsymbol{n},\boldsymbol{t})$. Applying $\boldsymbol{Q}^{T}$ from the left (i.e. evaluating $\sigma_{nn}\boldsymbol{n}+\sigma_{tn}\boldsymbol{t}$), it follows that the actual normal stress is given by

$$\begin{aligned}\boldsymbol{\sigma}\cdot\boldsymbol{n} =&\left(\sum_{\alpha=1}^{2}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{B}^{T}\right)^{-1}h^{\alpha}(\boldsymbol{\phi})\right)^{-1}\sum_{\alpha=1}^{2}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\alpha}\boldsymbol{B}^{T}\right)^{-1}\boldsymbol{B}\left(\boldsymbol{\Sigma}^{v}\right)^{\alpha}h^{\alpha}(\phi)\\ =&\left(\sum_{\alpha=1}^{2}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\right)^{-1}h^{\alpha}(\phi)\right)^{-1}\cdot\sum_{\alpha=1}^{2}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{\Sigma}^{\alpha}\cdot\boldsymbol{n}\right)h^{\alpha}(\phi)\end{aligned} \tag{7.122}$$

where $\boldsymbol{\Sigma}^{\alpha} := \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)$ and can in particular be evaluated completely without any reference to the transformed system $\mathcal{B}$. ◇

Remark 138. While it is tempting, based on the role of $\boldsymbol{Q}$ and $\boldsymbol{Q}^{T}$, to directly try to retransform Equation (7.121) back into the corresponding part of the stress in the standard coordinate system, this is not as straightforward as one might expect at first sight. In fact, all that would be required in order to do so would be to "reinsert" the entries of $\boldsymbol{\sigma_n}$ into the full vector $\boldsymbol{\sigma}_{\mathcal{B}}$ and then, as the inverse of $\boldsymbol{M}_{\boldsymbol{\sigma}}$ is given by $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}$, apply $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}$ from the left. In matrix-vector form,

this corresponds to applying

$$\begin{aligned}\boldsymbol{M}_{\boldsymbol{\epsilon}}^T\begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}&=\begin{pmatrix}n_x^2&t_x^2&2n_xt_x\\n_y^2&t_y^2&2n_yt_y\\n_xn_y&t_xt_y&n_xt_y+n_yt_x\end{pmatrix}\begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}=\begin{pmatrix}n_x^2&2n_xt_x\\n_y^2&2n_yt_y\\n_xn_y&n_xt_y+n_yt_x\end{pmatrix}\\&=\begin{pmatrix}n_x&0\\0&n_y\\\frac{1}{2}n_y&\frac{1}{2}n_x\end{pmatrix}\begin{pmatrix}n_x&2t_x\\n_y&2t_y\end{pmatrix}=\begin{pmatrix}n_x&0\\0&n_y\\\frac{1}{2}n_y&\frac{1}{2}n_x\end{pmatrix}\begin{pmatrix}n_x&t_x\\n_y&t_y\end{pmatrix}\begin{pmatrix}1&0\\0&2\end{pmatrix}\end{aligned}\tag{7.123}$$

to the expression for $\boldsymbol{\sigma_n}$. Even though the middle matrix is just $\boldsymbol{Q}^T$, there is an additional "weighting" by the right-most matrix which prohibits cancelling $\boldsymbol{Q}^T$ and the left-most factor $\boldsymbol{Q}$ in Equation (7.121).
One might expect at first sight that the factors 2 in the right-most matrix and the $\frac{1}{2}$-prefactors in the left-most matrix are due to the Voigt-representation, but this is actually not the case (note that the transformation above is between two stress-type quantities). In fact (by the previous remark) $\boldsymbol{\sigma_n}$ in Equation (7.121) corresponds to the correct components of the normal traction vector $\boldsymbol{\sigma_n}$ expressed in the $\mathcal{B}$-system, and retransforming this **vector** back to the Cartesian system could therefore indeed be done by applying $\boldsymbol{Q}^T$ from the left. Interpreted as the components of a symmetric second-order tensor written in Voigt-notation, the transformation is somewhat different though, a point which will be discussed in a little more detail in Remarks 140 and 141 below. ◇

Using this expression for $\boldsymbol{\sigma_n}$, one can now evaluate the tangential stress components $\boldsymbol{\sigma_t}$ in Equation (7.115). On the one hand, arguing analogously as for $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}$ in Equation (7.120), it is easy to verify that $\boldsymbol{\Sigma}_{\boldsymbol{t}}^{\alpha}=\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}$. On the other hand, combining the expression for $\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}$ from Equation (7.105) with the definition of $\boldsymbol{\mathcal{C}}_{\mathcal{B}}^{\alpha}$ and using (7.118), one has

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}=\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}\boldsymbol{Q}^T.$$

and thus with Equation (7.119)

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}=\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}\Big(\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}^T\Big)^{-1}\boldsymbol{Q}^T.$$

Inserting these expressions into the formula (7.115) for $\boldsymbol{\sigma_t}$, one obtains

$$\begin{aligned}\boldsymbol{\sigma_t}&=\sum_{\alpha=1}^{2}\boldsymbol{\Sigma}_{\boldsymbol{t}}^{\alpha}h^{\alpha}(\phi)+\sum_{\alpha=1}^{2}\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma_n}-\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big)h^{\alpha}(\phi)\\&=\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}\left(\sum_{\alpha=1}^{2}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}h^{\alpha}(\phi)+\sum_{\alpha=1}^{2}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}\Big(\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}^T\Big)^{-1}\boldsymbol{Q}^T\big(\boldsymbol{\sigma_n}-\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big)h^{\alpha}(\phi)\right).\end{aligned}$$

By the previous Remark 138, $\boldsymbol{Q}^T\big(\boldsymbol{\sigma_n}-\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big)$ is nothing but $\boldsymbol{\sigma}\cdot\boldsymbol{n}-\boldsymbol{\Sigma}^{\alpha}\cdot\boldsymbol{n}$, i.e. an expression which can be evaluated fully without recourse to the system $\mathcal{B}$. As the other terms in the last factor do not rely on $\mathcal{B}$ either, the only remaining fragment of the transformation consists in the prefactor $\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}$, whose role is simply reading out the tt-component of the expression on the right[71].
It turns out that, with respect to the construction of the actual stress $\boldsymbol{\sigma}^v$, this is now completely

[71] This is clear by construction, but can also be verified directly. In fact,

$$\begin{pmatrix}0&1&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}=\begin{pmatrix}t_x^2&t_y^2&2t_xt_y\end{pmatrix}=\begin{pmatrix}t_x&t_y\end{pmatrix}\begin{pmatrix}t_x&0&t_y\\0&t_y&t_x\end{pmatrix}.$$

It is, as for $\boldsymbol{B}$, easily verified that the pre-multiplication of a stress-type tensor $\hat{\boldsymbol{\sigma}}^v$ in Voigt-notation by second factor simply corresponds to the evaluation of $\hat{\boldsymbol{\sigma}}\cdot\boldsymbol{t}$. Further applying $\begin{pmatrix}t_x&t_y\end{pmatrix}$ then leads to $\boldsymbol{t}\cdot\hat{\boldsymbol{\sigma}}\cdot\boldsymbol{t}=\hat{\boldsymbol{\sigma}}_{tt}$.

superfluous as the right factor in fact already represents the (full) correct stress. Eliminating $\boldsymbol{\sigma_n}$ and $\boldsymbol{\Sigma}^{\alpha}_{\boldsymbol{n}}$ (recall that these quantities are still based upon the system $\mathcal{B}$) with $\sigma_{\boldsymbol{n}} = \boldsymbol{QB\sigma}^v$ (see Remark 137 and recall that a multiplication by $\boldsymbol{B}$ in Voigt-notation corresponds to the multiplication by $\boldsymbol{n}$) and $\boldsymbol{\Sigma}^{\alpha}_{\boldsymbol{n}} = \boldsymbol{QB}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}$, one therefore has the formula

$$\boldsymbol{\sigma}^v = \sum_{\alpha=1}^{2} \big(\boldsymbol{\Sigma}^v\big)^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha} \boldsymbol{B}\Big(\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}^T\Big)^{-1}\big((\boldsymbol{\sigma}\cdot\boldsymbol{n}) - \boldsymbol{B}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}\big)h^{\alpha}(\phi), \tag{7.124}$$

resp., in non-Voigt notation,

$$\boldsymbol{\sigma} = \sum_{\alpha=1}^{2} \boldsymbol{\Sigma}^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \big(\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)\cdot\Big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\Big)^{-1}\cdot\Big(\big(\boldsymbol{\sigma} - \boldsymbol{\Sigma}^{\alpha}\big)\cdot\boldsymbol{n}\Big)h^{\alpha}(\phi), \tag{7.125}$$

with the normal traction vector $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ given by Equation (7.122).
That this is true for the tangential part of the stress is clear from the above. For the normal part, this can either be seen from the "artificial" reexpression of $\boldsymbol{\sigma_n}$ in Equation (7.116) and inserting the expressions for $\boldsymbol{\mathcal{C}}^{\alpha}_{\boldsymbol{nn}}$ and $\boldsymbol{\mathcal{S}}^{\alpha}_{\boldsymbol{nn}}$, or, more succinctly, by verifying that the expression (7.124) reproduces the correct remaining two (normal) components of the stress. This is easily seen to be the case, as extracting the normal componts in the $\mathcal{B}$-system corresponds, by Equation (7.118), to a premultiplication by $\boldsymbol{QB}$, from which one obtains

$$\boldsymbol{QB\sigma}^v = \boldsymbol{QB}\sum_{\alpha=1}^{2} \big(\boldsymbol{\Sigma}^v\big)^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \boldsymbol{QB}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha} \boldsymbol{B}\Big(\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}^T\Big)^{-1}\big((\boldsymbol{\sigma}\cdot\boldsymbol{n}) - \boldsymbol{B}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}\big)h^{\alpha}(\phi).$$

As $\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^v\big)^{\alpha}\boldsymbol{B}$ and its inverse cancel in the second term, this leaves

$$\boldsymbol{QB\sigma}^v = \sum_{\alpha=1}^{2} \boldsymbol{QB}\big(\boldsymbol{\Sigma}^v\big)^{\alpha} h^{\alpha}(\phi) + \sum_{\alpha=1}^{2} \boldsymbol{Q}\big((\boldsymbol{\sigma}\cdot\boldsymbol{n}) - \boldsymbol{B}\big(\boldsymbol{\Sigma}^v\big)^{\alpha}\big)h^{\alpha}(\phi) = \sum_{\alpha=1}^{2} \boldsymbol{Q}(\boldsymbol{\sigma}\cdot\boldsymbol{n})h^{\alpha}(\phi) = \boldsymbol{Q}(\boldsymbol{\sigma}\cdot\boldsymbol{n}).$$

According to Remark 137, $\boldsymbol{Q}(\boldsymbol{\sigma}\cdot\boldsymbol{n}) = \boldsymbol{\sigma_n}$, from which it follows that Equation (7.124) (resp. Equation (7.125)) reproduces the correct values of all three entries of $\boldsymbol{\sigma}^v_{\mathcal{B}}$ in the $\mathcal{B}$-system, and therefore also, after a retransformation to the original system, of $\boldsymbol{\sigma}^v$.

Remark 139. Note that the expressions (7.124) and (7.125) are now entirely based upon quantities in the non-transformed system. While this is of course still the same model, it has the major advantage of avoiding the construction and repeated application of the transformation matrices $\boldsymbol{M_\sigma}$ and $\boldsymbol{M_\epsilon}$ (resp. their transposes). Besides a reduction in computational cost and complexity, this in particular also avoids the construction of a (resp. two in 3D) tangential vector. This is particularly helpful when trying to evaluate $\frac{\partial f_{el}}{\partial \nabla\phi}$, as this contribution now only enters (in an explicit form) in terms of $\boldsymbol{n}$, instead of in a quite implicit form through both the normal and the tangent vector(s) in the original formulation. ◇

Remark 140. As seen above, the expressions in equations (7.124) resp. (7.125) - a priori only necessary for the tangential part of the stress-tensor, as the normal stresses are already known - allow reconstructing the full stress tensor without an explicit separate treatment of the normal and tangential components. It is nevertheless interesting to understand the meaning of the stress-tensor (resp. vector in Voigt-notation) obtained by retransforming the expressions for $\boldsymbol{\sigma_n}$ in Equations (7.120) resp. (7.121) as indicated in Remark 138.
The central observation here is that the combined operation of reinserting the normal stress-components into $\boldsymbol{\sigma}^v_{\mathcal{B}}$ and then retransforming this vector can also be rewritten as

$$\boldsymbol{\sigma}^v_{\boldsymbol{n}} = \boldsymbol{M}^T_{\boldsymbol{\epsilon}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}\boldsymbol{\sigma_n} = \sigma_{nn}\begin{pmatrix} n_x^2 \\ n_y^2 \\ n_x n_y \end{pmatrix} + \sigma_{nt}\begin{pmatrix} 2n_x t_x \\ 2n_y t_y \\ n_x t_y + n_y t_x \end{pmatrix},$$

corresponding to the Voigt-representation of $\boldsymbol{\sigma}_{nn}\boldsymbol{n}\otimes\boldsymbol{n}+\boldsymbol{\sigma}_{nt}\boldsymbol{n}\otimes\boldsymbol{t}+\boldsymbol{\sigma}_{nt}\boldsymbol{t}\otimes\boldsymbol{n}$, i.e. the symmetric second-order tensor with the same normal components as $\boldsymbol{\sigma}_{\boldsymbol{n}}$ and a vanishing tangential one. From this, it is then also clear that extracting the normal components from $\boldsymbol{\sigma}_{\boldsymbol{n}}^{v}$ (in contrast to $\boldsymbol{\sigma}_{\boldsymbol{n}}$ which is a full tensor in Voigt-notation in the standard coordinate system) and reinserting them in the same manner will again lead to the same tensor $\boldsymbol{\sigma}_{\boldsymbol{n}}^{v}$. This implies that, given any symmetric second-oder tress-type tensor $\boldsymbol{\sigma}^{v}$ (in Voigt-notation), the matrix

$$\boldsymbol{N}_{\boldsymbol{\sigma}}^{v} := \boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}\begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&0&1\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}} = \boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}$$

describing this combined operation corresponds to a projection operator onto the "normal subspace" of the stress-type tensors, while the matrix

$$\boldsymbol{T}_{\boldsymbol{\sigma}}^{v} := \boldsymbol{I}-\boldsymbol{N}_{\boldsymbol{\sigma}}^{v} = \boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}\boldsymbol{I}\boldsymbol{M}_{\boldsymbol{\sigma}}-\boldsymbol{N}_{\boldsymbol{\sigma}}^{v} = \boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}\boldsymbol{M}_{\boldsymbol{\sigma}}$$

corresponds to the projection operator onto the "tangent subspace" of the stress-type tensors[72]. The stress obtained by retransforming $\boldsymbol{\sigma}_{\boldsymbol{n}}$ as above is therefore simply the "minimal" one with the prescribed normal component $\boldsymbol{\sigma}\cdot\boldsymbol{n}$. ◇

Besides the calculation rule for $\boldsymbol{\sigma}^{v}$, both the driving force calculation and dealing with more complex models - such as visco-elastic or plastic settings - will also require access to the phase-specific strains and stresses. Their calculation is again straightforward in the $\mathcal{B}$-system as, by assumption, the tangential ones are all equal to the average ones, $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha}=\boldsymbol{\epsilon}_{\boldsymbol{t}}$, while the normal ones can be recovered based upon

$$\boldsymbol{\sigma}_{\boldsymbol{n}} = \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)+\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) = \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}-\boldsymbol{\epsilon}_{\boldsymbol{n}}\big)+\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)+\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)$$

and thus

$$\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}-\boldsymbol{\epsilon}_{\boldsymbol{n}} = \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\Big(\boldsymbol{\sigma}_{\boldsymbol{n}}-\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)-\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}-\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big)\Big) = \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma}_{\boldsymbol{n}}-\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big). \tag{7.126}$$

Combined with the equality of the tangential stresses, $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha}-\boldsymbol{\epsilon}_{\boldsymbol{t}}=\boldsymbol{0}$, one therefore has, in the $\mathcal{B}$-system (again using Equation (7.118)),

$$\boldsymbol{\epsilon}_{\mathcal{B}}^{\alpha}-\boldsymbol{\epsilon}_{\mathcal{B}} = \begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\sigma}_{\boldsymbol{n}}-\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\big) = \begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix}\boldsymbol{Q}\Big(\boldsymbol{B}\big(\boldsymbol{\mathcal{C}}^{v}\big)^{\alpha}\boldsymbol{B}^{T}\Big)^{-1}\Big(\boldsymbol{Q}^{T}\boldsymbol{\sigma}_{\boldsymbol{n}}-\boldsymbol{B}\big(\boldsymbol{\Sigma}^{v}\big)^{\alpha}\Big).$$

In contrast to the recovery of the normal stresses in the original coordinate system, the recovery of this purely normal jump in the strains results in a simple formula. In fact, using $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{-1}=\boldsymbol{M}_{\boldsymbol{\sigma}}^{T}$

[72] One way to see this is by realizing that the inner matrices $\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}$ and $\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}$ in the definitions of $\boldsymbol{N}_{\boldsymbol{\sigma}}^{v}$ and $\boldsymbol{T}_{\boldsymbol{\sigma}}^{v}$ are clearly the corresponding projectors in the $\mathcal{B}$-system with the $\boldsymbol{M}_{\boldsymbol{\sigma}}$- and $\boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}$-matrices just transforming to and back from that system.
Alternatively, it can be verified in a more explicit but also more tedious manner in the original system. A quite useful property in this respect is that, as one can verify making use of the orthogonality properties of $\boldsymbol{n}$ and $\boldsymbol{t}$,

$$\boldsymbol{B}\boldsymbol{M}_{\boldsymbol{\epsilon}}^{T}\begin{pmatrix}1&0\\0&0\\0&1\end{pmatrix} = \boldsymbol{Q}^{T},$$

which in particular ensures that, in combination with Equation (7.118), $\boldsymbol{B}\boldsymbol{N}_{\boldsymbol{\sigma}}^{v}=\boldsymbol{Q}^{T}\boldsymbol{Q}\boldsymbol{B}=\boldsymbol{B}$, and thus the equality of the normal stress vectors $\boldsymbol{B}\boldsymbol{\sigma}^{v}$ and $\boldsymbol{B}\boldsymbol{N}_{\boldsymbol{\sigma}}^{v}\boldsymbol{\sigma}^{v}$ before and after the projection.

and combining $\boldsymbol{M}_{\boldsymbol{\sigma}}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{B}^T\boldsymbol{Q}^T$ (see Equation (7.118)) with $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ and $\boldsymbol{Q}^T\boldsymbol{\sigma}_{\boldsymbol{n}} = \boldsymbol{\sigma}\cdot\boldsymbol{n}$, the retransformation leads to

$$\left(\boldsymbol{\epsilon}^v\right)^\alpha - \boldsymbol{\epsilon}^v = \boldsymbol{B}^T\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^\alpha\boldsymbol{B}^T\right)^{-1}\left(\left(\boldsymbol{\sigma}\cdot\boldsymbol{n}\right) - \boldsymbol{B}\left(\boldsymbol{\Sigma}^v\right)^\alpha\right). \tag{7.127}$$

Defining

$$\boldsymbol{a}^\alpha := \left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^v\right)^\alpha\boldsymbol{B}^T\right)^{-1}\left(\left(\boldsymbol{\sigma}\cdot\boldsymbol{n}\right) - \boldsymbol{B}\left(\boldsymbol{\Sigma}^v\right)^\alpha\right) \text{ resp. } \boldsymbol{a}^\alpha := \left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\left(\boldsymbol{\sigma}-\boldsymbol{\Sigma}^\alpha\right)\cdot\boldsymbol{n}\right) \tag{7.128}$$

this can also, in a form coinciding with the jump vector based formulation outlined above, be written as

$$\left(\boldsymbol{\epsilon}^v\right)^\alpha - \boldsymbol{\epsilon}^v = \boldsymbol{B}^T\boldsymbol{a} = \begin{pmatrix} n_x & 0 \\ 0 & n_y \\ n_y & n_x \end{pmatrix}\boldsymbol{a} = \begin{pmatrix} a_x n_x \\ a_y n_y \\ a_x n_y + a_y n_x \end{pmatrix}, \tag{7.129}$$

with is easily recognized as the Voigt-representation of the strain-type (i.e. with the shear components doubled) tensor $\boldsymbol{\epsilon}^\alpha - \boldsymbol{\epsilon} = \left(\boldsymbol{a}^\alpha \otimes \boldsymbol{n}\right)_S$.

Unlike for the strains, the recovery of the phase-specific stresses is most easily done along the same lines as for the total stress. Starting again in the $\mathcal{B}$-system, the normal stress components are given by $\boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha = \boldsymbol{\sigma}_{\boldsymbol{n}}$, whereas the tangential ones can either be obtained directly from the phase-specific strains calculated above, or, more in line with the previous calculations, through

$$\begin{aligned}\boldsymbol{\sigma}_{\boldsymbol{t}}^\alpha =& \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\right) = \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{n}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \boldsymbol{\epsilon}_{\boldsymbol{n}}\right) \\ =& \boldsymbol{\Sigma}_{\boldsymbol{t}}^\alpha + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha\left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \boldsymbol{\epsilon}_{\boldsymbol{n}}\right) = \boldsymbol{\Sigma}_{\boldsymbol{t}}^\alpha + \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha\left(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^\alpha\right).\end{aligned}$$

The equality $\boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha = \boldsymbol{\sigma}_{\boldsymbol{n}}$ can again, similar to its phase-averaged version in Equation (7.116), artificially be expanded to $\boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha = \boldsymbol{\Sigma}_{\boldsymbol{n}}^\alpha + \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^\alpha\left(\boldsymbol{\sigma}_{\boldsymbol{n}} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^\alpha\right)$. Based on this common representation of the full stress tensor, the same arguments as for the retransformation of $\boldsymbol{\sigma}_{\mathcal{B}}$ can be applied, thus allowing to express the entire phase-specific stress tensor in the original system through

$$\boldsymbol{\sigma}^\alpha = \boldsymbol{\Sigma}^\alpha + \left(\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\right)\cdot\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\right)^{-1}\left(\left(\boldsymbol{\sigma}-\boldsymbol{\Sigma}^\alpha\right)\cdot\boldsymbol{n}\right). \tag{7.130}$$

Remark 141. In relation with Remark 138, it should be noted that the quite straightforward retransformation of the normal strain-jumps in contrast to the more difficult "retransformation" of the normal stress components is due to the interplay of the additional factor 2 on the shear-components in the Voigt-representation of the strain-type tensors in combination with Equation (7.103). In fact, as the strain-jump and the "extension by zero" $\begin{pmatrix} \sigma_{nn} & \sigma_{nt} \\ \sigma_{nt} & 0 \end{pmatrix}$ of the normal stress have the same structure, it is clear that this extension could also be written in terms of a symmetrized dyad $(\boldsymbol{b}\otimes\boldsymbol{n})_S$ with some vector $\boldsymbol{b}$. By Equation (7.103), this vector $\boldsymbol{b}$ is given by $\boldsymbol{b} = (\sigma_{nn}, 2\sigma_{nt})^T$ (this multiplication of the shear-component by 2 being the role of the rightmost-matrix in Equation (7.123)). The middle factor $\boldsymbol{Q}^T$ then simply serves to retransform this vector back to the normal coordinate system. Finally, the left-most matrix - corresponding to the $\boldsymbol{B}$-matrix with the "shear-entries halved" - is the Voigt-representation of a symmetrized dyad with $\boldsymbol{n}$ for stress-type tensors in Voigt notation, i.e. one where the shear-components are not stored in doubled form.

That the retransformation of the normal strain jump is so easily expressed in term of $\boldsymbol{B}^T$ and $\boldsymbol{Q}^T$ is therefore somewhat fortuitous. On the one hand, given that the shear-components of the strain are stored in doubled form, the 2-factor in the calculation of the jump-vector $\boldsymbol{a}$ according to Equation (7.103) is already included. On the other hand, storing the symmetrized dyad $(\boldsymbol{a}\otimes\boldsymbol{n})_S$ with doubled shear-components cancels the $\frac{1}{2}$-prefactor in the left-most matrix in Equation (7.123), leading to $\boldsymbol{B}^T$. ◇

Remark 142. From an implementation point of view, at least in the current setting, the (obviously equivalent) alternative expression

$$\boldsymbol{\sigma} = \sum_{\alpha=1}^{N} \Big(\boldsymbol{\Sigma}^{\alpha} - \big(\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)\cdot\big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1}\cdot\big(\boldsymbol{\Sigma}^{\alpha}\cdot\boldsymbol{n}\big)\Big) h^{\alpha}(\phi) + \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)\cdot\big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1} h^{\alpha}(\phi)\Big)\cdot(\boldsymbol{\sigma}\cdot\boldsymbol{n}) \quad (7.131)$$

for the stress is in principle more convenient than the one in Equation (7.125). The advantage of this form in combination with the expression (repeated from Equation (7.122) for convenience)

$$\boldsymbol{\sigma}\cdot\boldsymbol{n} = \Big(\sum_{\alpha=1}^{2} \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1} h^{\alpha}(\phi)\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \Big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\Big)^{-1} \cdot \big(\boldsymbol{\Sigma}^{\alpha}\cdot\boldsymbol{n}\big) h^{\alpha}(\phi)$$

for the normal stress is that the first term contains only initially known quantities, whereas the second part consist out of a **single** correction operation based on an averaged $\boldsymbol{\mathcal{C}}^{\alpha}$-$\boldsymbol{n}$-combination in the end once $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ has been calculated. In addition, all phase-specific quantities required for both calculating $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ as well as the correction are essentially the same as the ones already required for the evaluation of the first term. It is therefore not necessary to evalute these two times - once for the determination of $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ and then for the evaluation of $\boldsymbol{\sigma}$ itself once the normal stress is known - as one can simply accumulate the required quantities while running over the phases in the evaluation of the $\boldsymbol{\Sigma}^{\alpha} - \big(\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)\cdot\big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1}\cdot\big(\boldsymbol{\Sigma}^{\alpha}\cdot\boldsymbol{n}\big)$.
For more complex models where one requires access to the phase-specific stresses and/or strains for the stress-calculation itself (such as for visco-elastic or plastic problems), an obvious alternative to both equations (7.125) and (7.131) is to simply calculate the total stress as the average of the phase-specific ones (presumably already calculated based upon e.g. Equation (7.130)), $\boldsymbol{\sigma} = \sum_{\alpha} \boldsymbol{\sigma}^{\alpha} h^{\alpha}(\phi)$.

◇

Remark 143. In either case, it should be kept in mind that regardless of the particular formulation and implementation used, the jump-condition based models are fairly expensive to evaluate within the interface region (this being even more problematic for potential multiphase extensions). Whenever - such as in many implicit or linearized schemes - a larger amount of evaluations of stress-increments have to be performed based on fixed phasefield values (an potentially other parameters), it may be significantly cheaper from a computational point of view to precalculate the "effective" or algorithmically consistent stiffness, linking the increments in the total stress to those of the total strain. For the linearly elastic case above, it is straightforward to verify that, given fixed eigenstrains $\tilde{\boldsymbol{\epsilon}}^{\alpha}$, an increment in the average strain $\boldsymbol{\epsilon}$ leads to the increment

$$\delta\big(\boldsymbol{\sigma}\cdot\boldsymbol{n}\big) = \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1} h^{\alpha}(\phi)\Big)^{-1} \cdot \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1} \cdot \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\big) h^{\alpha}(\phi)\Big) : \delta\boldsymbol{\epsilon}.$$

Combined with the increments $\delta\boldsymbol{\Sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha} : \delta\boldsymbol{\epsilon}$ for the Voigt-Taylor stress prediction, it then follows from Equation (7.125) that the increment in the effective stress $\delta\boldsymbol{\sigma}$ is related to the one in the total strain by $\delta\boldsymbol{\sigma} = \boldsymbol{\mathcal{C}}_{eff} : \delta\boldsymbol{\epsilon}$ with $\boldsymbol{\mathcal{C}}_{eff}$ given by

$$\begin{aligned} \boldsymbol{\mathcal{C}}_{eff} = \sum_{\alpha=1}^{2} \Bigg[& \boldsymbol{\mathcal{C}}^{\alpha} + \big(\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)\cdot\big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha}\cdot\boldsymbol{n}\big)^{-1} \\ & \cdot \Bigg(\Big(\sum_{\beta=1}^{2} \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}\big)^{-1} h^{\beta}\Big)^{-1} \cdot \Big(\sum_{\beta=1}^{2} \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}\big)^{-1} \cdot \big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\beta}\big) h^{\beta}\Big) - \boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^{\alpha} \Bigg) \Bigg] h^{\alpha}. \end{aligned} \quad (7.132)$$

◇

As a final step, it remains to derive an expression for the driving force in the original coordinate system. The simplest way of doing so is simply reusing the expression

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\beta=1}^{2} \left(f_{el}^{\beta} - \boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\beta} \right) \frac{\partial h^{\beta}}{\partial \phi} = \sum_{\beta=1}^{2} \left(\frac{1}{2} \left(\boldsymbol{\epsilon}_{\mathcal{B}}^{\beta} - \tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^{\beta} \right) \cdot \boldsymbol{\mathcal{C}}_{\mathcal{B}}^{\beta} \left(\boldsymbol{\epsilon}_{\mathcal{B}}^{\beta} - \tilde{\boldsymbol{\epsilon}}_{\mathcal{B}}^{\beta} \right) - \boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\beta} \right) \frac{\partial h^{\beta}}{\partial \phi}$$

derived in [64]. The first part f_{el}^{β} can directly be reexpressed (the transformation matrices for the strains and the stiffness tensor simply cancel) using either the expressions for $\left(\boldsymbol{\epsilon}^{v}\right)^{\alpha}$ in Equation (7.127) together with the original stiffness-tensor, or by combining the expression for $\boldsymbol{\epsilon}^{\alpha}$ with the one for the phase-specific stresses $\left(\boldsymbol{\sigma}^{v}\right)^{\alpha}$ from Equation (7.130).
Due to its particular form, the additional contribution by the normal stress can also easily be rewritten in terms of quantities in the original system. In fact, firstly using

$$\sum_{\beta=1}^{2} \boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\beta} \frac{\partial h^{\beta}}{\partial \phi} = \sum_{\beta=1}^{2} \boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\beta} - \boldsymbol{\epsilon}_{\boldsymbol{n}} \right) \frac{\partial h^{\beta}}{\partial \phi} + \boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \boldsymbol{\epsilon}_{\boldsymbol{n}} \underbrace{\sum_{\beta=1}^{2} \frac{\partial h^{\beta}}{\partial \phi}}_{=0}$$

and combining this with $\boldsymbol{\sigma}_{\boldsymbol{n}} = \boldsymbol{Q}\boldsymbol{B}\boldsymbol{\sigma}^{v}$ and the expression (7.126) for the phase-specific strains and the definition $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\Sigma}^{v}\right)^{\alpha}$ leads to

$$\boldsymbol{\sigma}_{\boldsymbol{n}} \cdot \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\beta} = \boldsymbol{Q}\boldsymbol{B}\boldsymbol{\sigma}^{v} \cdot \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\sigma}^{v} - \boldsymbol{\Sigma}_{\boldsymbol{n}}^{\alpha}\right) = \boldsymbol{B}\boldsymbol{\sigma}^{v} \cdot \boldsymbol{Q}^{T} \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} \boldsymbol{Q}\boldsymbol{B}\left(\boldsymbol{\sigma}^{v} - \boldsymbol{\Sigma}^{\alpha}\right).$$

The $\boldsymbol{Q}$-factors cancel with the ones in the expression $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\beta} = \boldsymbol{Q}\left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\beta}\boldsymbol{B}^{T}\right)^{-1}\boldsymbol{Q}^{T}$ from Equation (7.119) though, thus leaving

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\beta=1}^{2} \left(f_{el}^{\beta} - \left(\boldsymbol{B}\boldsymbol{\sigma}^{v}\right) \cdot \left(\boldsymbol{B}\left(\boldsymbol{\mathcal{C}}^{v}\right)^{\beta} \boldsymbol{B}^{T} \boldsymbol{B}\left(\boldsymbol{\sigma}^{v} - \boldsymbol{\Sigma}^{\alpha}\right) \right) \frac{\partial h^{\beta}}{\partial \phi}, \tag{7.133}$$

or, in non-Voigt notation,

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\beta=1}^{2} \left(f_{el}^{\beta} - \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right) \cdot \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\beta} \cdot \boldsymbol{n} \right) \cdot \left(\left(\boldsymbol{\sigma} - \boldsymbol{\Sigma}^{\alpha}\right) \cdot \boldsymbol{n} \right) \frac{\partial h^{\beta}}{\partial \phi}. \tag{7.134}$$

Remark 144. Note that, with the definition (7.128) of $\boldsymbol{a}^{\alpha}$, these expressions can also be rewritten in the more pleasant form

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\beta=1}^{2} \left(f_{el}^{\beta} - \left(\boldsymbol{B}\boldsymbol{\sigma}^{v}\right) \cdot \boldsymbol{a}^{\alpha} \right) \frac{\partial h^{\beta}}{\partial \phi} \quad \text{resp.} \quad \frac{\partial f_{el}}{\partial \phi} = \sum_{\beta=1}^{2} \left(f_{el}^{\beta} - \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right) \cdot \boldsymbol{a}^{\alpha} \right) \frac{\partial h^{\beta}}{\partial \phi}. \tag{7.135}$$

With $\boldsymbol{B}\boldsymbol{\sigma}^{v} \cdot \boldsymbol{a}^{\alpha} = \boldsymbol{\sigma}^{v} \cdot \left(\boldsymbol{B}^{T}\boldsymbol{a}^{\alpha}\right)$ (resp. $\left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right) \cdot \boldsymbol{a}^{\alpha} = \boldsymbol{\sigma} : \left(\boldsymbol{n} \otimes \boldsymbol{a}\right)_{S}$ using the symmetry of $\boldsymbol{\sigma}$)[73], it follows that yet another form of the driving force is given by

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\alpha=1}^{2} \left(f_{el}^{\alpha} - \boldsymbol{\sigma} : \left(\boldsymbol{a}^{\alpha} \otimes \boldsymbol{n}\right)_{S} \right) \frac{\partial h^{\alpha}}{\partial \phi}, \tag{7.136}$$

which, as $\left(\boldsymbol{a}^{\alpha} \otimes \boldsymbol{n}\right)_{S} = \boldsymbol{\epsilon}^{\alpha} - \boldsymbol{\epsilon}$ and $\sum_{\alpha=1}^{2} \boldsymbol{\sigma} : \boldsymbol{\epsilon} \frac{\partial h^{\alpha}}{\partial \phi} = \boldsymbol{\sigma} : \boldsymbol{\epsilon} \left(\sum_{\alpha=1}^{2} \frac{\partial h^{\alpha}}{\partial \phi} \right) = 0$, can finally also be replaced by

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\alpha=1}^{2} \left(f_{el}^{\alpha} - \boldsymbol{\sigma} : \boldsymbol{\epsilon}^{\alpha} \right) \frac{\partial h^{\alpha}}{\partial \phi}. \tag{7.137}$$

◇

[73] As in Equation (7.129), $\boldsymbol{B}^{T}\boldsymbol{a}^{\alpha}$ is precisely the Voigt-represenation of the (strain-type) tensor $\left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_{S}$.

Remark 145. It is also possible to directly derive the (same) driving force directly based upon expression in the original coordinate system only. In fact, from $\boldsymbol{\sigma}^\alpha = \frac{\partial f^\alpha}{\partial \boldsymbol{\epsilon}^\alpha}$, one has

$$\frac{\partial f_{el}}{\partial \phi} = \frac{\partial}{\partial \phi}\sum_{\alpha=1}^{2} f^\alpha h^\alpha = \sum_{\alpha=1}^{2} f^\alpha \frac{\partial h^\alpha}{\partial \phi} + \boldsymbol{\mathcal{C}}^\alpha : \left(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\right) : \frac{\partial \boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha}{\partial \phi} h^\alpha = \sum_{\alpha=1}^{2} f^\alpha \frac{\partial h^\alpha}{\partial \phi} + \boldsymbol{\sigma}^\alpha : \frac{\partial \boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha}{\partial \phi} h^\alpha. \tag{7.138}$$

As neither $\tilde{\boldsymbol{\epsilon}}^\alpha$ nor $\boldsymbol{\epsilon}$ depend on ϕ and $\boldsymbol{\epsilon}^\alpha - \boldsymbol{\epsilon} = \left(\boldsymbol{a}^\alpha \otimes \boldsymbol{n}\right)_S$ and $\boldsymbol{\sigma}^\alpha$ is symmetric, the second contribution can be rewritten as

$$\sum_{\alpha=1}^{2} \boldsymbol{\sigma}^\alpha : \frac{\partial \boldsymbol{\epsilon}^\alpha - \boldsymbol{\epsilon}}{\partial \phi} h^\alpha = \sum_{\alpha=1}^{2} \boldsymbol{\sigma}^\alpha : \left(\frac{\partial \boldsymbol{a}^\alpha}{\partial \phi^\alpha} \otimes \boldsymbol{n}\right)_S h^\alpha = \sum_{\alpha=1}^{2} \left(\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n}\right) \cdot \frac{\partial \boldsymbol{a}^\alpha}{\partial \phi^\alpha} h^\alpha.$$

By construction, $\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n} = \boldsymbol{\sigma} \cdot \boldsymbol{n}$, such that it only remains to evaluate $\frac{\partial \boldsymbol{a}^\alpha}{\partial \phi}$. By Equation (7.128), if further follows that

$$\frac{\partial \boldsymbol{a}^\alpha}{\partial \phi^\alpha} = \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi}$$

and thus that

$$\sum_{\alpha=1}^{2} \boldsymbol{\sigma}^\alpha : \frac{\partial \boldsymbol{\epsilon}^\alpha - \boldsymbol{\epsilon}}{\partial \phi} h^\alpha = \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right)\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big) \cdot \frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi}. \tag{7.139}$$

From Equation (7.122), one has

$$\begin{aligned}\frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi} =& \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \left(\boldsymbol{\Sigma}^\alpha \cdot \boldsymbol{n}\right) \frac{\partial h^\alpha}{\partial \phi} \\ &+ \left(\frac{\partial}{\partial \phi}\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1}\right) \cdot \sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \left(\boldsymbol{\Sigma}^\alpha \cdot \boldsymbol{n}\right) h^\alpha,\end{aligned}$$

where the derivative in the second term can be evaluated as

$$-\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \frac{\partial h^\alpha}{\partial \phi}\Big) \cdot \underbrace{\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \left(\boldsymbol{\Sigma}^\alpha \cdot \boldsymbol{n}\right) h^\alpha}_{= \boldsymbol{\sigma} \cdot \boldsymbol{n}}.$$

Reinserting this expression into the derivative of $\boldsymbol{\sigma} \cdot \boldsymbol{n}$ then leads to

$$\begin{aligned}\frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi} =& \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \left(\boldsymbol{\Sigma}^\alpha \cdot \boldsymbol{n}\right) \frac{\partial h^\alpha}{\partial \phi} \\ &- \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \frac{\partial h^\alpha}{\partial \phi}\Big) \cdot \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right) \\ =& \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} \cdot \Big(\left(\boldsymbol{\Sigma}^\alpha \cdot \boldsymbol{n}\right) - \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right)\Big) \frac{\partial h^\alpha}{\partial \phi}.\end{aligned}$$

Recognizing that the summands in the second sum correspond to $\boldsymbol{a}^\alpha$, this further simplifies to

$$\frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi} = -\Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big)^{-1} \cdot \sum_{\alpha=1}^{2} \boldsymbol{a}^\alpha \frac{\partial h^\alpha}{\partial \phi}, \tag{7.140}$$

and finally, combined with Equation (7.138) and (7.139), to

$$\frac{\partial f_{el}}{\partial \phi} = \sum_{\alpha=1}^{2} f^\alpha \frac{\partial h^\alpha}{\partial \phi} + \left(\boldsymbol{\sigma} \cdot \boldsymbol{n}\right) \cdot \Big(\sum_{\alpha=1}^{2} \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^\alpha \cdot \boldsymbol{n}\right)^{-1} h^\alpha\Big) \cdot \frac{\partial (\boldsymbol{\sigma} \cdot \boldsymbol{n})}{\partial \phi} = \sum_{\alpha=1}^{2} \left(f^\alpha - (\boldsymbol{\sigma} \cdot \boldsymbol{n}) \cdot \boldsymbol{a}^\alpha\right) \frac{\partial h^\alpha}{\partial \phi}.$$

This is precisely the same expression as the second one in Equation (7.135). ◇

In contrast to the (correct) derivative with respect to ϕ, there is no expression for the derivative with respect to $\nabla\phi$ in [64], as the paper (falsely) claims that this derivative vanishes. In addition, given that the transformation matrices $\boldsymbol{M}_{\boldsymbol{\sigma}}$ and $\boldsymbol{M}_{\boldsymbol{\epsilon}}$ depend in an intricate fashion on the normal and (in particular in the three-dimensional case) how the tangent vector(s) are constructed based upon $\boldsymbol{n}$, performing this derivative based upon the formulation in [64] is likely to be a very arduous task. In contrast, it is straightforward based upon the reformulation above and runs essentially along the same lines - but actually in a simpler form - as the derivation of the expression for $\frac{\partial f_{el}}{\partial \phi}$ in the previous remark.
Again starting from $f_{el} = \sum_{\alpha=1}^{2} \frac{1}{2}\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\big) : \boldsymbol{\mathcal{C}}^{\alpha} : \big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\big) h^{\alpha}(\phi)$, a differentiation w.r.t. $\boldsymbol{n}$ leads to

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = \sum_{\alpha=1}^{2} \boldsymbol{\sigma}^{\alpha} : \frac{\partial\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\big)}{\partial \boldsymbol{n}} h^{\alpha} = \sum_{\alpha=1}^{2} \boldsymbol{\sigma}^{\alpha} : \frac{\partial}{\partial \boldsymbol{n}}\Big(\boldsymbol{a}^{\alpha} \otimes \boldsymbol{n}\Big)_{S} h^{\alpha}$$

since, similar to above, $\frac{\partial(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha})}{\partial \boldsymbol{n}} = \frac{\partial(\boldsymbol{\epsilon}^{\alpha} - \boldsymbol{\epsilon})}{\partial \boldsymbol{n}}$ as both $\tilde{\boldsymbol{\epsilon}}^{\alpha}$ and $\boldsymbol{\epsilon}$ are independent of $\boldsymbol{n}$. By the symmetry of $\boldsymbol{\sigma}$, this can be simplified to

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = \sum_{\alpha=1}^{2} \Big(\big(\boldsymbol{\sigma}^{\alpha} \cdot \boldsymbol{n}\big) \cdot \frac{\partial \boldsymbol{a}^{\alpha}}{\partial \boldsymbol{n}} + \big(\boldsymbol{\sigma}^{\alpha} \cdot \boldsymbol{a}^{\alpha}\big) \cdot \frac{\partial \boldsymbol{n}}{\partial \boldsymbol{n}}\Big) h^{\alpha} = \sum_{\alpha=1}^{2} \Big(\big(\boldsymbol{\sigma} \cdot \boldsymbol{n}\big) \cdot \frac{\partial \boldsymbol{a}^{\alpha}}{\partial \boldsymbol{n}} + \big(\boldsymbol{\sigma}^{\alpha} \cdot \boldsymbol{a}^{\alpha}\big)\Big) h^{\alpha},$$

where use was made of the equality of the phase-specific normal stresses with the total one. As $\boldsymbol{\sigma}\cdot\boldsymbol{n}$ does not depend on α, the first contribution can be summarized to $\big(\boldsymbol{\sigma}\cdot\boldsymbol{n}\big)\cdot\sum_{\alpha=1}^{2} \frac{\partial(\sum_{\alpha=1}^{2} \boldsymbol{a}^{\alpha} h^{\alpha})}{\partial \boldsymbol{n}}$, and since $\sum_{\alpha=1}^{2} \boldsymbol{a}^{\alpha} h^{\alpha} = \boldsymbol{0}$[74], actually drops out, leaving the very simple expression

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = \sum_{\alpha=1}^{2} \big(\boldsymbol{\sigma}^{\alpha} \cdot \boldsymbol{a}^{\alpha}\big) h^{\alpha} \tag{7.141}$$

for $\frac{\partial f_{el}}{\partial \boldsymbol{n}}$.
Remark 146. An analogous calculation to the one above could also be performed in the three-dimensional case. The only major difference to the two-dimensional one is the larger dimensionality of the vectors and subvectors as the Voigt-representation then consists of six-dimensional vectors and both the normal and tangential components are made up of three entries. For example, the $\boldsymbol{M}_{\boldsymbol{\sigma}}$-matrix for performing the transformation to the $\mathcal{B}$-system becomes [64]

$$\boldsymbol{M}_{\boldsymbol{\sigma}} = \begin{pmatrix} n_x^2 & n_y^2 & n_z^2 & 2n_y n_z & 2n_x n_z & 2n_x n_y \\ t_x^2 & t_y^2 & t_z^2 & 2t_y t_z & 2t_x t_z & 2t_x t_y \\ s_x^2 & s_y^2 & s_z^2 & 2s_y s_z & 2s_x s_z & 2s_x s_y \\ t_x s_x & t_y s_y & t_z s_z & t_y s_z + t_z s_y & t_x s_z + t_z s_x & t_x s_y + t_y s_x \\ n_x s_x & n_y s_y & n_z s_z & n_y s_z + n_z s_y & n_x s_z + n_z s_x & n_x s_y + n_y s_x \\ n_x t_x & n_y t_y & n_z t_z & n_y t_z + n_z t_y & n_x t_z + n_z t_x & n_x t_y + n_y t_x \end{pmatrix}.$$

With this transformation matrix, reading out the normal components in the $\mathcal{B}$-system (in the order $(\boldsymbol{\sigma}_{nn}, \boldsymbol{\sigma}_{nt}, \boldsymbol{\sigma}_{ns})^{T}$) is represented by the combined matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \boldsymbol{M}_{\boldsymbol{\sigma}} = \begin{pmatrix} n_x^2 & n_y^2 & n_z^2 & 2n_y n_z & 2n_x n_z & 2n_x n_y \\ n_x t_x & n_y t_y & n_z t_z & n_y t_z + n_z t_y & n_x t_z + n_z t_x & n_x t_y + n_y t_x \\ n_x s_x & n_y s_y & n_z s_z & n_y s_z + n_z s_y & n_x s_z + n_z s_x & n_x s_y + n_y s_x \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} n_x & n_y & n_z \\ t_x & t_y & t_z \\ s_x & s_y & s_z \end{pmatrix}}_{=:\boldsymbol{Q}} \underbrace{\begin{pmatrix} n_x & 0 & 0 & 0 & n_z & n_y \\ 0 & n_y & 0 & n_z & 0 & n_x \\ 0 & 0 & n_z & n_y & n_x & 0 \end{pmatrix}}_{=:\boldsymbol{B}},$$

[74] This can be verified "manually" based on Equation (7.128) combined with the definition of $\boldsymbol{\sigma}\cdot\boldsymbol{n}$, but is also easily seen from the imposed averaging relation $\boldsymbol{0} = \sum_{\alpha=1}^{2} \boldsymbol{\epsilon}^{\alpha} h^{\alpha} - \boldsymbol{\epsilon} = \sum_{\alpha=1}^{2} \big(\boldsymbol{\epsilon}^{\alpha} - \boldsymbol{\epsilon}\big) h^{\alpha} = \sum_{\alpha=1}^{2} \boldsymbol{a}^{\alpha} h^{\alpha}$.

where $\boldsymbol{Q}$ is again a unitary matrix corresponding to the transformation of "standard" vectors, and $\boldsymbol{B}$ encodes the multiplication of a Voigt-type quantity with $\boldsymbol{n}$. Adjusting the remaining formulae above making explicit use of the two-dimensional nature in an analogous manner, one can then verify that the rest of the calculation can be performed along exactly the same lines, just requiring more "bulky" matrices. ◇

Comparison with the Model by Durga et al. [24]

As already indicated in Section 7.2.2, Durga el al. propose a slightly different model in [24] (generalizing their previous work [23]). The formulation is in fact a priori very similar to the one in [64] in the sense that both approaches rely upon a transformation into a coordinate system where the normal vector $\boldsymbol{n}$ corresponds to the first unit vector $\boldsymbol{e}_1$ and that, within this transformed system and using the notation from [64], one assumes $\boldsymbol{\sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{\sigma}_{\boldsymbol{n}}$, $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon}_{\boldsymbol{t}}$ and $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)$[75]. In contrast to the model in [64], the (common) normal stress is a priori not (but see the discussion below) calculated from $\boldsymbol{\sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)$, but instead simply assumed to be equal to the corresponding components of the one calculated using the Reuss-Sachs model. In non-Voigt notation, one therefore has

$$\boldsymbol{\sigma}\cdot\boldsymbol{n} \coloneqq \boldsymbol{\sigma}_{RS}\cdot\boldsymbol{n} = \Big(\big(\sum_{\alpha=1}^{2}\boldsymbol{\mathcal{S}}^{\alpha}h^{\alpha}(\phi)\big)^{-1} : \big(\boldsymbol{\epsilon} - \sum_{\alpha=1}^{2}\tilde{\boldsymbol{\epsilon}}^{\alpha}h^{\alpha}(\phi)\big)\Big)\cdot\boldsymbol{n}.$$

In order to complete the model, it remains to specify a rule for determining the normal strains $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}$. This is achieved here - in line with the argument underlying the Reuss-Sachs model - by defining $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}$ to be given by the normal components of $\boldsymbol{\mathcal{S}}^{\alpha}:\boldsymbol{\sigma}^{\alpha}$, i.e. $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} = \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\sigma}_{\boldsymbol{n}} + \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\boldsymbol{\sigma}_{\boldsymbol{t}}^{\alpha}$, where $\boldsymbol{\sigma}^{\alpha}$ is the (full) stress reconstructed based upon the normal components from the Reuss-Sachs model and $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\alpha}$ based upon the phase-specific material law[76]. Combined with the expression for $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\alpha}$ in terms of $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}$ leads to

$$\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} = \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\sigma}_{\boldsymbol{n}} + \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big),$$

resp.

$$\Big(\boldsymbol{I} - \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha}\Big)\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) = \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\sigma}_{\boldsymbol{n}} + \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) \tag{7.142}$$

and thus to an equation which can be solved for $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}$ given the prescribed normal stress and tangential strains.
This equality can actually be modified into a more instructive form as $\boldsymbol{\mathcal{S}}^{\alpha}$ and $\boldsymbol{\mathcal{C}}^{\alpha}$ are inverses (in Voigt notation) of each other, i.e. one has

$$\begin{pmatrix} (\boldsymbol{\mathcal{S}}^{\alpha})_{nn} & (\boldsymbol{\mathcal{S}}^{\alpha})_{nt} \\ (\boldsymbol{\mathcal{S}}^{\alpha})_{tn} & (\boldsymbol{\mathcal{S}}^{\alpha})_{tt} \end{pmatrix}\begin{pmatrix} \boldsymbol{\mathcal{C}}_{nn}^{\alpha} & \boldsymbol{\mathcal{C}}_{nt}^{\alpha} \\ \boldsymbol{\mathcal{C}}_{tn}^{\alpha} & \boldsymbol{\mathcal{C}}_{tt}^{\alpha} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\mathcal{S}}^{\alpha})_{nn}\boldsymbol{\mathcal{C}}_{nn}^{\alpha} + (\boldsymbol{\mathcal{S}}^{\alpha})_{nt}\boldsymbol{\mathcal{C}}_{tn}^{\alpha} & (\boldsymbol{\mathcal{S}}^{\alpha})_{nn}\boldsymbol{\mathcal{C}}_{nt}^{\alpha} + (\boldsymbol{\mathcal{S}}^{\alpha})_{nt}\boldsymbol{\mathcal{C}}_{tt}^{\alpha} \\ (\boldsymbol{\mathcal{S}}^{\alpha})_{tn}\boldsymbol{\mathcal{C}}_{nn}^{\alpha} + (\boldsymbol{\mathcal{S}}^{\alpha})_{tt}\boldsymbol{\mathcal{C}}_{tn}^{\alpha} & (\boldsymbol{\mathcal{S}}^{\alpha})_{tn}\boldsymbol{\mathcal{C}}_{nt}^{\alpha} + (\boldsymbol{\mathcal{S}}^{\alpha})_{tt}\boldsymbol{\mathcal{C}}_{tt}^{\alpha} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}.$$

From the upper-left block of this equality, one has $\boldsymbol{I} - \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha} = \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}$, while the upper-right block shows that $\big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nt}}\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha} = -\big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}$. Inserting these expressions into Equation (7.142), the phase-specific normal strains can alternatively be characterized by

$$\big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) = \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\sigma}_{\boldsymbol{n}} - \big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big),$$

and thus, $\big(\boldsymbol{\mathcal{S}}^{\alpha}\big)_{\boldsymbol{nn}}$ being invertible as a diagonal subblock of a positive definite matrix, in fact

$$\boldsymbol{\sigma}_{\boldsymbol{n}} = \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon}_{\boldsymbol{t}} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big). \tag{7.143}$$

[75] In the Voigt-index based notation of the appendix A in [24], the normal and tangential stress components are given by $\boldsymbol{\sigma}_{\boldsymbol{n}} = (\sigma_1, \sigma_5, \sigma_6)^T$ and $\boldsymbol{\sigma}_{\boldsymbol{t}} = (\sigma_2, \sigma_3, \sigma_4)^T$ (and similarly for the strains).

[76] Note the difference in notation here with $(\boldsymbol{\mathcal{S}}^{\alpha})_{\boldsymbol{nn}}$ designating the actual $\boldsymbol{n}-\boldsymbol{n}$-subblock of $\boldsymbol{\mathcal{S}}^{\alpha}$, in contrast with $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha}$ defined above.

From this, it is obvious that the only difference between the model in [24] and the one discussed above[77] is that in the former, the normal stresses are prescribed directly based on the Reuss-Sachs/Steinbach-Apel scheme, whereas in the latter case, $\boldsymbol{\sigma_n}$ is obtained by enforcing the averaging condition

$$\boldsymbol{\epsilon_n} \overset{!}{=} \sum_{\alpha=1}^{2} \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\phi) = \sum_{\alpha=1}^{2} \left(\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\right)^{-1} \Big(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha}\big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big) h^{\alpha}(\phi)$$

based upon Equation (7.143).
In contrast to the model by [64], the model in [24] therefore does not generally satisfy the natural relation $\sum_{\alpha=1}^{2} \boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\phi) = \boldsymbol{\epsilon_n}$ unless the normal stresses in e.g. Equation (7.108) or Equation (7.122) happen to coincide with[78] $\boldsymbol{\sigma}_{RS} \cdot \boldsymbol{n}$.

Remark 147. As the Reuss-Sachs/Steinbach-Apel model is based on the assumption of (fully) equal stresses between all phases, it is, as used in [23] and [24], an intuitively much more reasonable choice for a common normal stress than the one from e.g. the Voigt-Talyor model, which is based on a completely different assumption. In addition, as shown in these papers, combining this assumption with the subsequent imposition of common tangential strains does in fact lead to a marked improvement in comparison with both the Reuss-Sachs- and Voigt-Taylor-model themselves.
Nevertheless, the stress-prediction by the Reuss-Sachs model is (implicitly) based on phase-specific strains where the tangential components need not coincide. In combination with Equation (7.143), the subsequent choice of the phase-specific normal strains based on $\boldsymbol{\sigma}_{RS} \cdot \boldsymbol{n}$ can then be interpreted as adjusting the normal strains such that, now under the different (and more physical assumption) $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon_t}$ and with the phase-specific material law, one obtains this same "predicted" normal stress for both phases. Even though this model presents a major step forward in terms of modeling the stress-strain relationship within a diffuse interface, it does therefore not achieve the same degree of internal consistency as the model(s) by [51] and [64], where the calculation of $\boldsymbol{\sigma} \cdot \boldsymbol{n}$ itself is already based on the same assumptions as the evaluation of the final stresses[79]. ◇

Equivalence of the Two-Phase Model to the One by Tschukin

Yet another formulation for the jump-condition based mechanical model is presented by Tschukin in [74]. As the primary motivations of this alternative formulation to the model in [64] (a collaboration between, among others, Schneider with Tschukin), Tschukin also mentions the computational overhead involved in the transformations to the $\mathcal{B}$-system (including the necessity of generating two tangent vectors instead of relying solely on $\boldsymbol{n}$) as well as correcting the wrong claim of the vanishing contribution $\frac{\partial f_{el}}{\partial \nabla \phi}$ in the variational derivative.
Even though the description in [74] is in many ways formally similar to the one in [64] in the sense that it continues to decompose both the strain- and stress-tensor (and similarly the stiffness tensor) into normal and tangential "components", the crucial difference is that these are not defined in terms of a collection of specific entries within the transformed coordinate system $\mathcal{B}$ anymore. Instead, Tschukin defines these quantities in terms of two (forth-order) projection operators

$$\mathcal{N} = \boldsymbol{I} \square \boldsymbol{\Phi} + \boldsymbol{\Phi} \square \boldsymbol{I} - \boldsymbol{\Phi} \square \boldsymbol{\Phi} \quad \text{and} \quad \mathcal{T} = (\boldsymbol{I} - \boldsymbol{\Phi}) \square (\boldsymbol{I} - \boldsymbol{\Phi}) \tag{7.144}$$

[77] As one explicitly assumes $\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon_t}$, the averaging condition $\boldsymbol{\epsilon_t} = \sum_{\alpha=1}^{2} \boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} h^{\alpha} = \boldsymbol{\epsilon_t}$ is trivially satisfied, and from Equation (7.143) combined with the analogous (explicit) assumption for the tangential stress, the model also satisfies the phase-specific material law $\boldsymbol{\sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha} : (\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha})$, from which the total stress is obtained by a weigted interpolation (trivially reducing to $\boldsymbol{\sigma_n}$ for the normal components).

[78] One (quite particular) case where this actually happens is when both materials are taken as isotropic, with the Lamé-parameters of one phase being chosen as (equal) multiples of those of the other phase, i.e. $\boldsymbol{\mathcal{C}}^2 = c\boldsymbol{\mathcal{C}}^1$ with $\boldsymbol{\mathcal{C}}^1 = \lambda^1 \boldsymbol{I} \otimes \boldsymbol{I} + 2\mu^1 \boldsymbol{I}^{(s)}$.

[79] Note also, that as in [64], the dependence on $\boldsymbol{n}$ is disregarded in the phasefield evolution equation in [23] and [24].

onto the normal and tangential components, where $\boldsymbol{\Phi} = \boldsymbol{n} \otimes \boldsymbol{n}$ and the "box-product" is defined by $(\boldsymbol{A} \boxdot \boldsymbol{B}) : \boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{C} \cdot \boldsymbol{B}$. In terms of these, the normal and tangential components of the stress- and strain-tensor are defined by

$$\boldsymbol{\sigma_n} \coloneqq \mathcal{N} : \boldsymbol{\sigma}, \quad \boldsymbol{\sigma_t} \coloneqq \mathcal{T} : \boldsymbol{\sigma}, \quad \text{and} \quad \boldsymbol{\epsilon_n} \coloneqq \mathcal{N} : \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon_t} \coloneqq \mathcal{T} : \boldsymbol{\epsilon}. \tag{7.145}$$

Using $\boldsymbol{\sigma}^\alpha = \mathcal{N} : \boldsymbol{\sigma}^\alpha + \mathcal{T} : \boldsymbol{\sigma}^\alpha$ and $\boldsymbol{\epsilon}^\alpha = \mathcal{N} : \boldsymbol{\epsilon}^\alpha + \mathcal{T} : \boldsymbol{\epsilon}^\alpha$ together with the idempotence and orthogonality of these projectors, this allows rewriting the normal "part" of the phase-specific material law as

$$\begin{aligned}\boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha =& \mathcal{N} : \boldsymbol{\sigma}^\alpha = \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \left(\mathcal{N} : \boldsymbol{\epsilon}^\alpha + \mathcal{T} : \boldsymbol{\epsilon}^\alpha\right) = \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{N} : \boldsymbol{\epsilon}^\alpha + \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{T} : \boldsymbol{\epsilon}^\alpha \\ =& \left(\mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{N}\right) : \left(\mathcal{N} : \boldsymbol{\epsilon}^\alpha\right) + \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{T} : \left(\mathcal{T} : \boldsymbol{\epsilon}^\alpha\right) = \left(\mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{N}\right) : \boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha + \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{T} : \boldsymbol{\epsilon}_{\boldsymbol{t}}^\alpha.\end{aligned}$$

Proceding similarly with the tangential part and adding, by the same argument, the eigenstrains, the material behavior can be fully characterized by $\boldsymbol{\sigma}^\alpha = \boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha + \boldsymbol{\sigma}_{\boldsymbol{t}}^\alpha$ through

$$\begin{aligned}\boldsymbol{\sigma}_{\boldsymbol{n}}^\alpha =& \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha : \left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha : \left(\boldsymbol{\epsilon}_{\boldsymbol{t}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\right) \text{ and} \\ \boldsymbol{\sigma}_{\boldsymbol{t}}^\alpha =& \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha : \left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^\alpha\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha : \left(\boldsymbol{\epsilon}_{\boldsymbol{t}}^\alpha - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^\alpha\right),\end{aligned} \tag{7.146}$$

where the normal and tangential components of the respective second-order tensors are defined as in Equation (7.145), and

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha \coloneqq \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{N},\ \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^\alpha \coloneqq \mathcal{N} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{T},\ \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^\alpha \coloneqq \mathcal{T} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{N},\ \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^\alpha \coloneqq \mathcal{T} : \boldsymbol{\mathcal{C}}^\alpha : \mathcal{T}. \tag{7.147}$$

Remark 148. Note that Equation (7.146) is formally completely identical with the stress-calculation as it would be performed in the $\mathcal{B}$-system based on the subvectors $\boldsymbol{\epsilon}_{\boldsymbol{n}}^\alpha$ and $\boldsymbol{\epsilon_t}$ and subblocks of the stiffness-tensor defined in (7.105). The interpretation of the varvious quantities is quite different though, as, due to the use of the projection operators, these quantities are now not, as in [64], elements of lower-dimensional subspaces, but instead elements of the original full-dimensional space (but which do of course each actually lie in a lower-dimensional subspace).
While this also induces some technical difficulties, this representation has the advantage of on the one hand not requiring a basis transformation and on the other hand - as seen from the definitions of $\mathcal{N}$ and $\mathcal{T}$ - being constructed based solely on the normal vector $\boldsymbol{n}$. ◇

Due to the strong formal similarities, it is possible to closely mimick the calculations in [64] within this new formulation. The only technical difficulty is that, as in particular $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha$ now has a non-trivial kernel even in Voigt-notation as it is based upon a pre- and postmultiplication by a (non-trivial) projector, one cannot rely on actual inverses for systems involving $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^\alpha$. One can, as in Tschukin, replace the (non-existent) inverses with respect to the normal subspace by Moore-Penrose pseudo-inverses though[80]. At least in theory, the lack of actual invertibility does then not cause any significant issues. In particular, using the corresponding pseudo-inverse $\bar{\boldsymbol{\mathcal{S}}}_{\boldsymbol{nn}}^{12}$ of

$$\bar{\boldsymbol{\mathcal{C}}}_{\boldsymbol{nn}}^{12} \coloneqq h^1(\phi)\mathcal{N} : \boldsymbol{\mathcal{C}}^1 : \mathcal{N} + h^2(\phi)\mathcal{N} : \boldsymbol{\mathcal{C}}^1 : \mathcal{N} = \mathcal{N} : \bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) : \mathcal{N}$$

with the anti-arithmetic average $\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) = h^1(\phi)\boldsymbol{\mathcal{C}}^1 + h^2(\phi)\boldsymbol{\mathcal{C}}^2$ as in Equation (7.90), he derives the expression[81]

$$f_{el} = \frac{1}{2}\Big(\boldsymbol{\Sigma}^1 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^1\right)h^1(\phi) + \boldsymbol{\Sigma}^2 : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^2\right)h^2(\phi) - \left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right) : \bar{\boldsymbol{\mathcal{S}}}_{\boldsymbol{nn}}^{12} : \left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right)h^1(\phi)h^2(\phi)\Big) \tag{7.148}$$

[80] Constructed such as to act as the corresponding inverses when restricted to the normal subspace only, these satisfy in particular $\left(\cdot\right)^\dagger : \left(\cdot\right) = \mathcal{N}$ as well as $\left(\cdot\right)^\dagger : \mathcal{N} = \mathcal{N} : \left(\cdot\right)^\dagger = \left(\cdot\right)^\dagger$, $\left(\cdot\right)^\dagger : \mathcal{T} = \mathcal{T} : \left(\cdot\right)^\dagger = \mathbf{0}$, see [74] and the discussion below.

[81] Note that, while Tschukin's starting point is - similarly to the work in [64] - an expression for the elastic free energy in terms of the "homogeneous" variables $\boldsymbol{\sigma_n}$ and $\boldsymbol{\epsilon_t}$, he then proceeds a somewhat different line as a preparation to a quite different generalization to the multiphase case from the one in [62]. This will further be discussed in Section 7.2.4.

for the elastic free energy density, where $\boldsymbol{\Sigma}^\alpha = \boldsymbol{\mathcal{C}}^\alpha : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\alpha\right)$, $\alpha = 1, 2$, corresponds to the stress obtained under the Voigt-Taylor assumption $\boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon}$[82]. The effective stress is then, in accordance with Equation (7.70), determined by differentiating f_{el} with respect to the average strain $\boldsymbol{\epsilon}$ to be given by

$$\boldsymbol{\sigma} = \boldsymbol{\Sigma}^1 h^1(\phi) + \boldsymbol{\Sigma}^2 h^2(\phi) - h^1(\phi) h^2(\phi) \left(\boldsymbol{\mathcal{C}}^1 - \boldsymbol{\mathcal{C}}^2\right) : \bar{\boldsymbol{\mathcal{S}}}^{12}_{\boldsymbol{nn}} : \left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right). \tag{7.149}$$

Even though the model in [74] starts from a similar argument as the one in [64], as one might guess by comparing equations (7.149) and (7.91), it is in fact much easier to relate to the jump-vector based formulation in Section 7.2.2 than to the one in [64]. The key observation to this is contained in the following lemma:

Lemma 9. *Let $\boldsymbol{\mathcal{C}}$ be a symmetric (in the same sense as the standard stiffness-tensor with the major and minor subsymmetries) fourth-order such that $\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}} \cdot \boldsymbol{n}$ is invertible. Then the pseudo-inverse $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} := \left(\mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N}\right)^\dagger$ satisfies*

$$\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \boldsymbol{\Sigma} = \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S \quad \textit{where} \quad \boldsymbol{a} = \left(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}} \cdot \boldsymbol{n}\right)^{-1} \cdot \left(\boldsymbol{\Sigma} \cdot \boldsymbol{n}\right) \tag{7.150}$$

for any symmetric second-order tensor $\boldsymbol{\Sigma}$.

Proof. Even though it would be possible to verify this in Voigt-notation by retransforming to the $\mathcal{B}$-system and using Remark 130 together with the discussions in Remark 140 and 141, it is actually simpler to directly verify that the construction in Equation (7.150) satisfies all the required properties.
Based on the definitions of the projectors in Equation (7.144), it is easy to verify that $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$ satisfies $\mathcal{T} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} = \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \mathcal{T} = \boldsymbol{0}$ and $\mathcal{N} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} = \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \mathcal{N} = \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$. With the definition in the left-most equality in Equation (7.150), the equalities for $\mathcal{T} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$ and $\mathcal{N} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$ are in fact independent of the particular choice of $\boldsymbol{a}$ but simply rely on the fact that $\left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S$ is an element of the normal subspace[83] and thus satisfies $\mathcal{T} : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S = \boldsymbol{0}$ and $\mathcal{N} : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S = \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S$. Applying $\mathcal{T}$ resp. $\mathcal{N}$ to $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \boldsymbol{\Sigma} = \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S$ then leads to the desired conclusion.
The corresponding equalities with $\mathcal{T}$ and $\mathcal{N}$ applied from the right in turn depend upon $\boldsymbol{\Sigma}$ entering the formula for $\boldsymbol{a}$ in terms of its normal component $\boldsymbol{\Sigma} \cdot \boldsymbol{n}$ only. Since $(\mathcal{T} : \boldsymbol{\Sigma}) \cdot \boldsymbol{n} = \boldsymbol{0}$ and $(\mathcal{N} : \boldsymbol{\Sigma}) \cdot \boldsymbol{n} = \boldsymbol{\Sigma} \cdot \boldsymbol{n}$[84], the definition of $\boldsymbol{a}$ shows that $\mathcal{S}_{\boldsymbol{nn}} : \mathcal{T} : \boldsymbol{\Sigma} = \boldsymbol{0}$ in the first case, whereas $\boldsymbol{a}$ remains unchanged in the second case and thus $\mathcal{S}_{\boldsymbol{nn}} : \mathcal{N} : \boldsymbol{\Sigma} = \mathcal{S}_{\boldsymbol{nn}} : \boldsymbol{\Sigma}$.
As $\mathcal{S}_{\boldsymbol{nn}}$ therefore maps into the normal subspace and its kernel contains (at least) the tangent subspace, all that remains to verify is that $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$ acts as an inverse to $\mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N}$ on the normal subspace (of symmetric tensors), i.e. $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N} : \boldsymbol{\Sigma} = \mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \boldsymbol{\Sigma} = \mathcal{N} : \boldsymbol{\Sigma}$ for all symmetric $\boldsymbol{\Sigma}$. Applying $\mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N}$ from the left to the left equality in equaion (7.150), it first follows from $\mathcal{N} : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S$ and the right subsymmetry of $\boldsymbol{\mathcal{C}}$ that

$$\left(\mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N}\right) : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}} : \boldsymbol{\Sigma} = \left(\mathcal{N} : \boldsymbol{\mathcal{C}} : \mathcal{N}\right) : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S = \mathcal{N} : \boldsymbol{\mathcal{C}} : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right)_S = \mathcal{N} : \left((\boldsymbol{\mathcal{C}} \cdot \boldsymbol{n}) \cdot \boldsymbol{a}\right).$$

[82] Note that Tschukin also uses the notation $\boldsymbol{E}$ instead of $\boldsymbol{\epsilon}$ for the strain do distinguish it from its representation in Voigt-notation.

[83] While this has already been used a number of times, it can also be verified directly based on $\boldsymbol{\Phi} = \boldsymbol{n} \otimes \boldsymbol{n}$ and the definition of $\mathcal{T}$. One has

$$\mathcal{T} : \left(\boldsymbol{a} \otimes \boldsymbol{n}\right) = \left(\boldsymbol{I} - \boldsymbol{n} \otimes \boldsymbol{n}\right) \cdot \left(\boldsymbol{a} \otimes \boldsymbol{n}\right) \cdot \left(\boldsymbol{I} - \boldsymbol{n} \otimes \boldsymbol{n}\right) = \left(\boldsymbol{I} - \boldsymbol{n} \otimes \boldsymbol{n}\right) \cdot \left(\boldsymbol{a} \otimes \left(\boldsymbol{I} \cdot \boldsymbol{n}\right) - \|\boldsymbol{n}\|^2 \boldsymbol{a} \otimes \boldsymbol{n}\right) = \boldsymbol{0}$$

and similarly $\mathcal{T} : \left(\boldsymbol{n} \otimes \boldsymbol{a}\right) = \boldsymbol{0}$. As $\mathcal{N} = \boldsymbol{I} - \mathcal{T}$, $\mathcal{N} : \left(\boldsymbol{n} \otimes \boldsymbol{a}\right)_S = \left(\boldsymbol{n} \otimes \boldsymbol{a}\right)_S - \mathcal{T} : \left(\boldsymbol{n} \otimes \boldsymbol{a}\right)_S = \left(\boldsymbol{n} \otimes \boldsymbol{a}\right)_S$.

[84] This is intuitively clear again as the former is the projection onto the tangential subspace and the latter the one onto the normal subspace. It can also easily be verified explicitly by applying the projectors to $\boldsymbol{\Sigma}$ and contracting the result with $\boldsymbol{n}$.

Insering $\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)$ into the definition of $\mathcal{N}$ and using the left subsymmetry of $\boldsymbol{\mathcal{C}}$, one further has

$$\begin{aligned}\mathcal{N}:\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big) =&\big(\boldsymbol{I}\,\square\,\boldsymbol{\Phi}+\boldsymbol{\Phi}\,\square\,\boldsymbol{I}-\boldsymbol{\Phi}\,\square\,\boldsymbol{\Phi}\big):\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\\ =&\Big(\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\cdot\boldsymbol{n}\Big)\otimes\boldsymbol{n}+\boldsymbol{n}\otimes\Big(\boldsymbol{n}\cdot\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\Big)-\Big(\boldsymbol{n}\cdot\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\cdot\boldsymbol{n}\Big)\boldsymbol{n}\otimes\boldsymbol{n}\\ =&\big((\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\otimes\boldsymbol{n}+\boldsymbol{n}\otimes\big((\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)-\Big(\big((\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)\cdot\boldsymbol{n}\Big)\boldsymbol{n}\otimes\boldsymbol{n}.\end{aligned}$$

Together with the definition of $\boldsymbol{a}$ above, the prefactors $\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n}$ cancel, leaving

$$\mathcal{N}:\big((\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n})\cdot\boldsymbol{a}\big)=(\boldsymbol{\Sigma}\cdot\boldsymbol{n})\otimes\boldsymbol{n}+\boldsymbol{n}\otimes(\boldsymbol{\Sigma}\cdot\boldsymbol{n})-\Big((\boldsymbol{\Sigma}\cdot\boldsymbol{n})\cdot\boldsymbol{n}\Big)\boldsymbol{n}\otimes\boldsymbol{n}.$$

As this is precisely $\mathcal{N}:\boldsymbol{\Sigma}$, it follows that

$$\big(\mathcal{N}:\boldsymbol{\mathcal{C}}:\mathcal{N}\big):\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}:\boldsymbol{\Sigma}=\mathcal{N}:\boldsymbol{\Sigma}$$

for any (symmetric) $\boldsymbol{\Sigma}$, and thus the result. □

Remark 149. Note that the (somewhat tedious) restriction to symmetric tensors in the lemma is due to the fact that it is on the one hand fully sufficient in the current setting and and on the other hand avoids a slight technical difficulty related to the various subsymmetries of fourth-order tensors. More precisely, as $(\boldsymbol{a}\otimes\boldsymbol{n})_S$ is obviously a symmetric tensor, $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}$ as defined above maps into the subspace of symmetric second order tensors and thus enjoys the left subsymmetric $\big(\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}\big)_{ijkl}=\big(\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}\big)_{jikl}$. One could also enforce the right subsymmtry by replacing the definition of $\boldsymbol{a}$ by $\boldsymbol{a}:=\big(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}\cdot\boldsymbol{n}\big)^{-1}\cdot\frac{1}{2}\big(\boldsymbol{\Sigma}\cdot\boldsymbol{n}+\boldsymbol{\Sigma}^T\cdot\boldsymbol{n}\big)$, which, although not really used in the current settting, would not require any major modifications either and would then indeed lead to the required pseudoinverse.
The only somewhat unpleasent issue in the general case is related to the definition of the projectors $\mathcal{N}$ and $\mathcal{T}$ themselves. In fact, even though the projector $\mathcal{N}$ in the form of Equation (7.144) does **not** posess either the left nor right subsymmetry[85], it is well suited to the setting here[86] and avoids introducing unnecessary symmetrization operations. ◇

Based on Lemma 9, it is now easy to see that, within the two-phase setting, the model by [74] is equivalent to the jump-condition based one in Section 7.2.2 and by the previous discussion thus also (except for the missing contribtuion due to $\frac{\partial f_{el}}{\partial\nabla\phi}$ in [64]) with the one in [64].
Firstly, using this lemma for replacing the term $\boldsymbol{\mathcal{S}}^{12}_{\boldsymbol{nn}}(\phi)$ in Equation (7.149) with $(\boldsymbol{a}\otimes\boldsymbol{n})_S$ with $\boldsymbol{a}$ defined by $\bar{\boldsymbol{\mathcal{C}}}^{12}_{\boldsymbol{nn}}(\phi)\cdot\boldsymbol{a}=\big(\boldsymbol{\Sigma}^1-\boldsymbol{\Sigma}^2\big)\cdot\boldsymbol{n}$ with the jump-vector $\boldsymbol{a}$ coinciding with the one in Equation (7.89) and making use of the $\boldsymbol{\mathcal{C}}^\alpha$, the expression for the effective stress in [74] clearly reduces to the one from Equation (7.91) in Section 7.2.2. Secondly, the same replacement in the expression for the elastic free energy density reduces Equation (7.148) to

$$f_{el}=\frac{1}{2}\Big(\boldsymbol{\Sigma}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)h^1(\phi)+\boldsymbol{\Sigma}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)h^2(\phi)-\big(\boldsymbol{\Sigma}^1-\boldsymbol{\Sigma}^2\big):\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S h^1(\phi)h^2(\phi)\Big).\qquad(7.151)$$

For comparison with the jump-condition based approach, note that, by construction, the elastic free energy density from Equation (7.96) in Section 7.2.2 can be rewritten as

$$\begin{aligned}f_{el}=&\frac{1}{2}\Big(\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big)h^1(\phi)+\boldsymbol{\sigma}^2:\big(\boldsymbol{\epsilon}^2-\tilde{\boldsymbol{\epsilon}}^2\big)h^2(\phi)\Big)\\ =&\frac{1}{2}\Big(\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}-h^2(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S-\tilde{\boldsymbol{\epsilon}}^1\big)h^1(\phi)+\boldsymbol{\sigma}^2:\big(\boldsymbol{\epsilon}+h^1(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S-\tilde{\boldsymbol{\epsilon}}^2\big)h^2(\phi)\Big).\end{aligned}$$

[85] This can be seen from $\mathcal{N}:\boldsymbol{\Sigma}=(\boldsymbol{\Sigma}\cdot\boldsymbol{n})\otimes\boldsymbol{n}+\boldsymbol{n}\otimes(\boldsymbol{\Sigma}^T\cdot\boldsymbol{n})-(\boldsymbol{n}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{n})\boldsymbol{n}\otimes\boldsymbol{n}$, which is not in general symmetric unless $\boldsymbol{\Sigma}$ is so, nor depends symmetrically on $\boldsymbol{\Sigma}$.

[86] Both $\mathcal{N}$ and $\mathcal{T}$ are here only applied to quantities with the necessary subsymmetries for this not to matter (either symmetric second-order tensors of fourth-order tensors with the left and right subsymmetries).

As already used in various forms above, the terms in $(\boldsymbol{a}\otimes\boldsymbol{n})_S$ cancel as they can be combined to $\Big(\big(\boldsymbol{\sigma}^2-\boldsymbol{\sigma}^1\big)\cdot\boldsymbol{n}\Big)h^1(\phi)h^2(\phi)$ which vanishes due to the common normal stress, i.e. an equivalent expression is given by

$$f_{el}=\frac{1}{2}\Big(\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)h^1(\phi)+\boldsymbol{\sigma}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)h^2(\phi)\Big).$$

Further inserting

$$\begin{aligned}\boldsymbol{\sigma}^1=&\boldsymbol{\mathcal{C}}^1:\Big(\boldsymbol{\epsilon}-h^2(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S-\tilde{\boldsymbol{\epsilon}}^1\Big)=\boldsymbol{\Sigma}^1-h^2(\phi)\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S,\\ \boldsymbol{\sigma}^2=&\boldsymbol{\mathcal{C}}^2:\Big(\boldsymbol{\epsilon}+h^1(\phi)\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S-\tilde{\boldsymbol{\epsilon}}^2\Big)=\boldsymbol{\Sigma}^2+h^1(\phi)\boldsymbol{\mathcal{C}}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\end{aligned}\tag{7.152}$$

then leads to

$$f_{el}=\frac{1}{2}\Big(\boldsymbol{\Sigma}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)h^1(\phi)+\boldsymbol{\Sigma}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)h^2(\phi)\Big)-\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S:\Big(\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)-\boldsymbol{\mathcal{C}}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)\Big)h^1(\phi)h^2(\phi),$$

which is precisely Equation (7.151), thus showing that both free energy densities also coincide[87].

Based on the equality of f_{el}, it is clear that the contributions $\frac{\partial f_{el}}{\partial\phi}$ and $\frac{\partial f_{el}}{\partial\nabla\phi}=\frac{\partial f_{el}}{\partial\boldsymbol{n}}\cdot\frac{\partial\boldsymbol{n}}{\partial\nabla\phi}$ should also coincide. Tschukin specifies the expression for $\frac{\partial f_{el}}{\partial\phi}$ as[88]

$$\frac{\partial f_{el}}{\partial\phi}=\frac{1}{2}\left(\Big(\boldsymbol{\Sigma}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)-\Delta\bar{\boldsymbol{\epsilon}}^{12}:\boldsymbol{\mathcal{C}}^1:\Delta\bar{\boldsymbol{\epsilon}}^{12}\big(h^2\big)^2\Big)-\Big(\boldsymbol{\Sigma}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)-\Delta\bar{\boldsymbol{\epsilon}}^{21}:\boldsymbol{\mathcal{C}}^2:\Delta\bar{\boldsymbol{\epsilon}}^{21}\big(h^1\big)^2\Big)\right)\frac{\partial h^1}{\partial\phi}\tag{7.153}$$

where $\Delta\bar{\boldsymbol{\epsilon}}^{12}$ and $\Delta\bar{\boldsymbol{\epsilon}}^{21}$ are specified by

$$\Delta\bar{\boldsymbol{\epsilon}}^{12}=\bar{\boldsymbol{\mathcal{S}}}^{12}_{\boldsymbol{nn}}:\big(\boldsymbol{\Sigma}^1-\boldsymbol{\Sigma}^2\big)\quad\text{and}\quad\Delta\bar{\boldsymbol{\epsilon}}^{21}=\bar{\boldsymbol{\mathcal{S}}}^{21}_{\boldsymbol{nn}}:\big(\boldsymbol{\Sigma}^2-\boldsymbol{\Sigma}^1\big).$$

From Lemma 9, and $\bar{\boldsymbol{\mathcal{S}}}^{12}_{\boldsymbol{nn}}=\bar{\boldsymbol{\mathcal{S}}}^{21}_{\boldsymbol{nn}}$, it is easily seen that these quantities in fact correspond to $\Delta\bar{\boldsymbol{\epsilon}}^{12}=\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$ and $\Delta\bar{\boldsymbol{\epsilon}}^{21}=-\Delta\bar{\boldsymbol{\epsilon}}^{12}=-\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$.
In addition, using $\boldsymbol{\epsilon}=\boldsymbol{\epsilon}^1+h^2\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$ and thus $\boldsymbol{\Sigma}^1=\boldsymbol{\sigma}^1+h^2\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$, one can rewrite the first term $\boldsymbol{\Sigma}^1:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^1\big)$ as

$$\begin{aligned}&\boldsymbol{\Sigma}^1:\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big)+h^2\boldsymbol{\Sigma}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\\ =&\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big)+h^2\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big):\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S+h^2\boldsymbol{\sigma}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S+\big(h^2\big)^2\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S:\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\\ =&\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big)+2h^2\boldsymbol{\sigma}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S+\big(h^2\big)^2\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S:\boldsymbol{\mathcal{C}}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\end{aligned}$$

and similarly, with $\boldsymbol{\epsilon}=\boldsymbol{\epsilon}^2-h^1\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$ and thus $\boldsymbol{\Sigma}^2=\boldsymbol{\sigma}^2-h^1\boldsymbol{\mathcal{C}}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S$,

$$\boldsymbol{\Sigma}^2:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^2\big)=\boldsymbol{\sigma}^2:\big(\boldsymbol{\epsilon}^2-\tilde{\boldsymbol{\epsilon}}^2\big)-2h^1\boldsymbol{\sigma}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S+\big(h^1\big)^2\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S:\boldsymbol{\mathcal{C}}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S.$$

Combining these observations in the expression (7.153), the quadratic terms in $\Delta\bar{\boldsymbol{\epsilon}}^{12}$ and $\Delta\bar{\boldsymbol{\epsilon}}^{21}$ cancel, leaving

$$\frac{\partial f_{el}}{\partial\phi}=\frac{1}{2}\left(\Big(\boldsymbol{\sigma}^1:\big(\boldsymbol{\epsilon}^1-\tilde{\boldsymbol{\epsilon}}^1\big)+2h^2\boldsymbol{\sigma}^1:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\Big)-\Big(\boldsymbol{\sigma}^2:\big(\boldsymbol{\epsilon}^2-\tilde{\boldsymbol{\epsilon}}^2\big)-2h^1\boldsymbol{\sigma}^2:\big(\boldsymbol{a}\otimes\boldsymbol{n}\big)_S\Big)\right)\frac{\partial h^1}{\partial\phi}.$$

[87] Note that the equality of the stresses could also have been derived from the equality of the free energy densities, as the effective stress in both cases satisfies $\boldsymbol{\sigma}=\frac{\partial f_{el}}{\partial\boldsymbol{\epsilon}}$. Nevertheless, the direct verification of the equality of the stresses is much simpler.

[88] Note that the expressions are slightly adopted as compared to the ones in [74] since, in contrast to the discussion in Section 7.2.2, Tschukin does not explicitly use a reduced formulation $\phi^2=1-\phi^1$ but instead works with $h^1(\phi^1)$ and $h^2(\phi^2)$ with ϕ^1 and ϕ^2 considered as independent. Once the Lagrange-multiplier for the sum-constraint is taken into account, this reduces, up to the usual factor 2, to the same.

While the terms $\frac{1}{2}\boldsymbol{\sigma}^\alpha : \left(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\right)$ correspond to the phase-specific contributions f_{el}^α to f_{el}, the remaining terms in the jump vector $\boldsymbol{a}$ can, making use of the equality of the normals stresses, be summarized to

$$h^2\boldsymbol{\sigma}^1 : \left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S + h^1\boldsymbol{\sigma}^2 : \left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S = \left(h^2\boldsymbol{\sigma}^1\cdot\boldsymbol{n} + h^1\boldsymbol{\sigma}^2\cdot\boldsymbol{n}\right)\cdot\boldsymbol{a} = \left(h^2+h^1\right)\boldsymbol{\sigma}\cdot\boldsymbol{a} = (\boldsymbol{\sigma}\cdot\boldsymbol{n})\cdot\boldsymbol{a},$$

showing that the expression for $\frac{\partial f_{el}}{\partial\phi}$ coincides with the one in Equation (7.97) (and therefore of course also with the other alternative expressions listed in Section 7.2.2).
Tschukin also specifies the (auxiliary but decisive) quantity $\frac{\partial f_{el}}{\partial n_i}$ in terms of $\Delta\bar{\boldsymbol{\epsilon}}^{12}$ as

$$\frac{\partial f_{el}}{\partial n_i} = -\left(\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2\right) : \left(\mathcal{D}_i - \bar{\boldsymbol{\mathcal{S}}}^{12}_{\boldsymbol{nn}} : \mathcal{D}_i : \bar{\boldsymbol{\mathcal{C}}}^{12}\right) : \Delta\bar{\boldsymbol{\epsilon}}^{12} h^1(\phi) h^2(\phi) \tag{7.154}$$

where $\mathcal{D}_i = \frac{\partial}{\partial n_i}\mathcal{N}$. Noting first that $\bar{\boldsymbol{\mathcal{S}}}^{12}_{\boldsymbol{nn}} : (\boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^2) = \Delta\bar{\boldsymbol{\epsilon}}^{12}$, this can, together with the major symmetry of $\mathcal{D}_i$, be rewritten as[89]

$$\frac{\partial f_{el}}{\partial n_i} = \left(\boldsymbol{\Sigma}^2 - \boldsymbol{\Sigma}^1 - \bar{\boldsymbol{\mathcal{C}}}^{12} : \Delta\bar{\boldsymbol{\epsilon}}^{12}\right) : \mathcal{D}_i : \Delta\bar{\boldsymbol{\epsilon}}^{12} h^1(\phi) h^2(\phi).$$

Replacing $\Delta\bar{\boldsymbol{\epsilon}}^{12}$ as above with $\left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S$ and with $\bar{\boldsymbol{\mathcal{C}}}^{12}(\phi) = h^2(\phi)\boldsymbol{\mathcal{C}}^1 + h^1(\phi)\boldsymbol{\mathcal{C}}^2$, the first term can be rewritten as

$$\boldsymbol{\Sigma}^2 - \boldsymbol{\Sigma}^1 + \bar{\boldsymbol{\mathcal{C}}}^{12} : \left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S = \left(\boldsymbol{\Sigma}^2 + h^1\boldsymbol{\mathcal{C}}^2 : \left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S\right) - \left(\boldsymbol{\Sigma}^1 - h^2\boldsymbol{\mathcal{C}}^1 : \left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S\right) = \boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1,$$

where the final equality follows from Equation (7.152). This further simplifies $\frac{\partial f_{el}}{\partial n_i}$ to $\frac{\partial f_{el}}{\partial n_i} = \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) : \mathcal{D}_i\left(\boldsymbol{a}\otimes\boldsymbol{n}\right)_S$. As a final step for relating the expression (7.154) with the one previously derived in Equation (7.99) for the jump-vector based approach, it remains to eliminate the derivatives $\mathcal{D}_i$ of the projector $\mathcal{N}$. Firstly, by a simple application of the chain-rule,

$$\mathcal{D}_i : (\boldsymbol{a}\otimes\boldsymbol{n})_S = \left(\frac{\partial}{\partial n_i}\mathcal{N}\right) : (\boldsymbol{a}\otimes\boldsymbol{n})_S = \frac{\partial}{\partial n_i}\left(\mathcal{N}(\boldsymbol{a}\otimes\boldsymbol{n})_S\right) - \mathcal{N} : \frac{\partial}{\partial n_i}(\boldsymbol{a}\otimes\boldsymbol{n})_S$$

Secondly, $(\boldsymbol{a}\otimes\boldsymbol{n})_S$ being an element of the normal subspace, $\mathcal{N} : (\boldsymbol{a}\otimes\boldsymbol{n})_S = (\boldsymbol{a}\otimes\boldsymbol{n})_S$, and thus

$$\mathcal{D}_i : (\boldsymbol{a}\otimes\boldsymbol{n})_S = \frac{\partial}{\partial n_i}(\boldsymbol{a}\otimes\boldsymbol{n})_S - \mathcal{N} : \frac{\partial}{\partial n_i}(\boldsymbol{a}\otimes\boldsymbol{n})_S = \mathcal{T} : \frac{\partial}{\partial n_i}(\boldsymbol{a}\otimes\boldsymbol{n})_S.$$

Even tough both $\boldsymbol{a}$ and $\boldsymbol{n}$ depend on $\boldsymbol{n}$, the presence of the tangential projector then ensures that the contribution $\left(\frac{\partial\boldsymbol{a}}{\partial n_i}\otimes\boldsymbol{n}\right)_S$ due to the derivative of $\boldsymbol{a}$ drops out as this is (regardless of the entries of the first vector in this dyad), again an element of the normal space. This further simplifies the expression for $\mathcal{D}_i : (\boldsymbol{a}\otimes\boldsymbol{n})_S$ to

$$\mathcal{D}_i : (\boldsymbol{a}\otimes\boldsymbol{n})_S = \mathcal{T} : \left(\boldsymbol{a}\otimes\frac{\partial n}{\partial n_i}\right)_S = \mathcal{T} : \left(\boldsymbol{a}\otimes\boldsymbol{e}_i\right)_S.$$

Reinserting this into the expression for $\frac{\partial f_{el}}{\partial n_i}$, as the last step, it suffices to realize that by the normal continuity of the stresses, $\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1$ is an element of the tangential subspace and thus satisfies $\mathcal{T} : \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) = \boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1$, from which it then follows that

$$\frac{\partial f_{el}}{\partial n_i} = \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) : \mathcal{T} : \left(\boldsymbol{a}\otimes\boldsymbol{e}_i\right)_S = \left(\mathcal{T} : \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right)\right) : \left(\boldsymbol{a}\otimes\boldsymbol{e}_i\right)_S = \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) : \left(\boldsymbol{a}\otimes\boldsymbol{e}_i\right)_S.$$

[89] This expression corresponds to the non-Voigt version of the alternative expression for $\frac{\partial f_{el}}{\partial n_i}$ provided by equation (4.58) in [74].

Together with the symmetry of the first term, this is the same as $\left(\left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) \cdot \boldsymbol{a}\right) \cdot \boldsymbol{e}_i$, i.e. the derivative of f_{el} with respect to n_i is simply the i'th of the vector $\left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) \cdot \boldsymbol{a}$ and therefore

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = \left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) \cdot \boldsymbol{a}. \tag{7.155}$$

Contracting this expression with $\frac{\partial \boldsymbol{n}}{\partial \nabla \phi}$, this obviously leads to the same expression for $\frac{\partial f_{el}}{\partial \nabla \phi}$ as in Equation (7.100).

Remark 150. Note that Tschukin also provides a description of his model in Voigt-representation, where he includes some additional simplifications as compared to the non-Voigt description above. Some of these are in part similar to the manipulations above, whereas others are more similar to the ones performed when discussing the model in [64].
In particular, he factorizes the Voigt-representation $\boldsymbol{N}_{\boldsymbol{\sigma}}^v$ of the projector onto the normal space as $\boldsymbol{N}_{\boldsymbol{\sigma}} = \boldsymbol{A}\boldsymbol{B}$, where $\boldsymbol{A} \in \boldsymbol{R}^{6\times 3}$ and $\boldsymbol{B} \in \boldsymbol{R}^{3\times 6}$ (as above) is the matrix representing the multiplication of a stress-type tensor in Voigt-notation by the normal vector $\boldsymbol{n}$. Recalling Remark 140, $\boldsymbol{N}_{\boldsymbol{\sigma}}^v$ (in the two-dimensional setting) and making use of Equation (7.118), this projector can also be written as

$$\boldsymbol{N}_{\boldsymbol{\sigma}}^v \coloneqq \boldsymbol{M}_{\boldsymbol{\epsilon}}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{M}_{\boldsymbol{\sigma}} = \boldsymbol{M}_{\boldsymbol{\epsilon}}^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{Q}\boldsymbol{B},$$

from which it follows that $\boldsymbol{A}$ corresponds precisely to the product of the first three matrices[90]. He then makes use of the idempotence of $\boldsymbol{N}_{\boldsymbol{\sigma}}^v$ and the "good guess" for the required pseudo-inverse $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{12}$ in Voigt-nation for finally eliminating $\boldsymbol{A}$ completely, reducing e.g. his stress-calcuation to a form much more similar to the one in Equation (7.124).
A particular consequence of this is that the matrices $\boldsymbol{D}_i$ introduced in this Voigt-setting should not be confused with the derivatives of the projector above as these are actually defined as the derivatives of the matrix $\boldsymbol{B}$ with respect to n_i. As this is just a matrix-form of encoding the multiplication by $\boldsymbol{n}$, it is obvious that $\frac{\partial \boldsymbol{B}}{\partial n_i}$ is simply a matrix-form of encoding a multiplication by the unit vector $\boldsymbol{e}_i$ and that his expression for the derivative in $\frac{\partial f_{el}}{\partial n_i}$ in equation 4.58 thus in fact just corresponds to the i'th component of $\left(\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^1\right) \cdot \boldsymbol{a}$ in Equation (7.155). ◇

[90] One major difference being that Tschukin reexpresses this in terms of the normal vector only.

7.2.4 Some Partial Extension to the Multiphase Case

Given that the jump-condition based model delivers very satisfactory results in the two-phase case, it is natural to try to extend it to the more general multiphase setting. Unfortunately, in constrast to e.g. the simpler Voigt-Talor or Reuss-Sachs approach, the extension of the more elaborate model described in the previous section faces a very fundamental difficulty which is not a priori related to the phasefield modeling approach itself. Rather, it is related to the description of the mechanical behavior at irregular points such as corner points or, within the present context, points where multiple interfaces meet, since it is well known that - unlike for flat or smoothly curved interfaces - such points (and/or edges in the three-dimensional setting) are generally associated with singularities in the behavior of the solution.

In relation with the jump condition approach from the previous section and the following discussion, it is instructive to illustrate this difficulty in a more concrete form in the example in Figure 7.1, which, for notational simplicity, is based on the analogous approach for the linear heat-conduction problem

$$\begin{cases} \nabla \cdot \boldsymbol{q} = f, \quad \boldsymbol{q} = -\kappa \nabla T & \text{in } \Omega \\ T = 0 & \text{on } \partial\Omega \end{cases}$$

In this case, the continuity of the traces (and their tangential derivatives) leads to the jump condition $\nabla T^\beta = \nabla T^\alpha + a^{\alpha\beta}\boldsymbol{n}^{\alpha\beta}$, whereas for f regular (e.g. in $L^2(\Omega)$), the continuity of the normal stress-components needs to be replaced by $[\![\boldsymbol{q}^\alpha - \boldsymbol{q}^\beta]\!] \cdot \boldsymbol{n}^{\alpha\beta} = 0$.
Simply **assuming** that these jump conditions between the two-phase interfaces continue to hold as one approaches the triple point (i.e. there is a well-defined limit, regardless along which interface this point is approached), the jump condition on the 2-3-interface and the combination of the two jump-conditions for ∇T^1 at the 1-2- and 1-3-interface lead to

$$\nabla T^3 = \nabla T^2 + a^{23}\boldsymbol{e}_x \text{ and } \nabla T^2 + a^{21}\boldsymbol{e}_y = \nabla T^1 = \nabla T^3 + a^{31}\boldsymbol{e}_y,$$

implying that $a^{21}\boldsymbol{e}_y = a^{23}\boldsymbol{e}_x + a^{31}\boldsymbol{e}_y$. By linear independence, this shows that $a^{23} = 0$, i.e. (as expected from the continuity of the tangential derivatives along the 1-2- and 1-3-interface) $\frac{\partial T^1}{\partial x} = \frac{\partial T^2}{\partial x} = \frac{\partial T^3}{\partial x}$, and $a^{21} = a^{31}$, i.e. (as expected from the continuity of the tangential derivative along the 2-3-interface) $\frac{\partial T^2}{\partial y} = \frac{\partial T^1}{\partial y} - a^{21}\boldsymbol{e}_y = \frac{\partial T^1}{\partial y} - a^{31}\boldsymbol{e}_y = \frac{\partial T^3}{\partial y}$. Inserting these equalities into the jump conditions on the normal fluxes further leads to the conclusion that

$$\kappa^2 \frac{\partial T^2}{\partial x} = \kappa^3 \frac{\partial T^3}{\partial x} \quad \text{and} \quad \kappa^2 \frac{\partial T^2}{\partial y} = \kappa^2\Big(\frac{\partial T^1}{\partial y} - a^{21}\Big) = \kappa^3\Big(\frac{\partial T^1}{\partial y} - a^{31}\Big) = \kappa^3 \frac{\partial T^3}{\partial y}.$$

As $\nabla T^2 = \nabla T^3$ by the above, these equality is possible iff either $\kappa^2 = \kappa^3$ (i.e. there is no "real" triple-junction) or both gradients happen to vanish at this point. In the latter case, one necessarily also has $\frac{\partial T^1}{\partial x} = 0$ by continuity of the tangential derivatives and $\frac{\partial T^1}{\partial y} = 0$ due to the equality of the normal fluxes. In summary, the jump conditions between all three phases in this case can only be satisfied when $\kappa^2 = \kappa^3$ or when all three gradients happen to vanish at this point[91].

Remark 151. Note that, from a phasefield perspective, the issue at internal corners does not a priori seem to impose any particular difficulties from a pure modeling point of view as such corners are usually "smoothed out" due to the diffuse interface profile. This seeming simplicity is somewhat misleading in the same sense that, even though e.g. a standard (sharp-interface) FEM-simulation does not in principle require any special treatment of corners points, the accuracy of the resulting approximation will in general be significantly reduced as compared to a "smooth"

[91] In contrast, when either $\kappa^2 = \kappa^1$ or $\kappa^3 = \kappa^1$, one would still expect a singular behavior due to the presence of a corner between the two "effective" material regions.

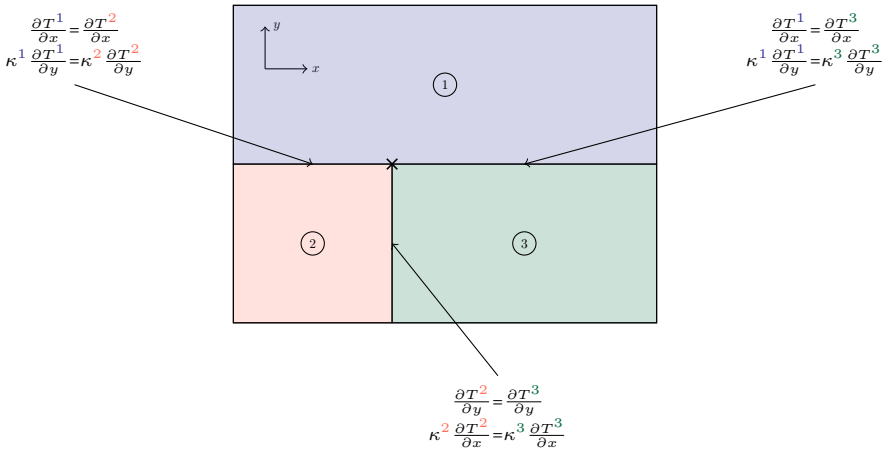


Figure 7.1: Illustration of the contradictory nature of the jump-conditions at a triple point.

problem. Understanding this difficulty is, at least at a qualitative level, quite intuititve, as approximation a solution which has a highly complex behavior within a relatively small region is of course much harder than approximating one which which is highly regular.
In contrast to the case of internal corners, which can in principle be tackled using the jump-condition based model above, the case of intersecting interfaces also poses a "visible" difficulty in the sense that the modeling of such multiple junctions within the phasefield context requires using more than two coexistent phases, and thus requires some extension of the previous model(s) to the multiphase case. ◇

A very similar (though algebraically more challenging) difficulty arises when trying to extend the jump-condition based model from the previous section to the multiphase case. A straightforward generalization of the jump relations on the symmetrized displacement gradient from Equation (7.87) would consist in imposing

$$[\![\boldsymbol{F}]\!]^{\alpha\beta} = \boldsymbol{F}^{\beta} - \boldsymbol{F}^{\alpha} = \boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{N}^{\alpha\beta} \quad \text{resp.} \quad [\![\boldsymbol{\epsilon}]\!]^{\alpha\beta} = \boldsymbol{\epsilon}^{\beta} - \boldsymbol{\epsilon}^{\alpha} = (\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta})_S \tag{7.156}$$

for all $\alpha \neq \beta$ and then trying to determine the $\boldsymbol{a}^{\alpha\beta}$, $\alpha \neq \beta$ from the corresponding continuity conditions

$$[\![\boldsymbol{\sigma}]\!]^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta} = \left(\boldsymbol{\sigma}^{\beta} - \boldsymbol{\sigma}^{\alpha}\right) \cdot \boldsymbol{n}^{\alpha\beta} = \boldsymbol{0} \quad , \alpha \neq \beta \tag{7.157}$$

for the normal stresses.
As above, such a straightforward generalization is unfortunately doomed to failure (except for some special cases). In fact, even though Equation (7.156) would a priori introduces $N(N-1)$ unknown jump vectors $\left(\boldsymbol{a}^{\alpha\beta}\right)_{1\leq\alpha\neq\beta\leq N}$ together with an equal number of continuity conditions, these conditions are not all independendent. Firstly, using $\boldsymbol{n}^{\beta\alpha} = -\boldsymbol{n}^{\alpha\beta}$ and thus from

$$\boldsymbol{a}^{\beta\alpha} \otimes \boldsymbol{n}^{\beta\alpha} = -\boldsymbol{a}^{\beta\alpha} \otimes \boldsymbol{n}^{\alpha\beta} = [\![\boldsymbol{F}]\!]^{\beta\alpha} = -[\![\boldsymbol{F}]\!]^{\alpha\beta} = -\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta}$$

that $\boldsymbol{a}^{\beta\alpha} = \boldsymbol{a}^{\alpha\beta}$, the number of unknowns can be reduced to the $\frac{N(N-1)}{2}$ unknown jump vectors $\left(\boldsymbol{a}^{\alpha\beta}\right)_{1\leq\alpha<\beta\leq N}$. At the same time, as $[\![\boldsymbol{\sigma}]\!]^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta} = \boldsymbol{0}$ directly implies $[\![\boldsymbol{\sigma}]\!]^{\beta\alpha} \cdot \boldsymbol{n}^{\beta\alpha}$, one can simultaneously reduce the number of jump conditions to be satisfied in Equation (7.157) to those for $1 \leq \alpha < \beta \leq N$.
The number of truly independent entries for the jump vectors can be significantly lower though as the algebraic equality

$$[\![\boldsymbol{F}]\!]^{\alpha\beta} = \boldsymbol{F}^{\beta} - \boldsymbol{F}^{\alpha} = \boldsymbol{F}^{\beta} - \boldsymbol{F}^{\delta} + \boldsymbol{F}^{\delta} - \boldsymbol{F}^{\alpha} = [\![\boldsymbol{F}]\!]^{\delta\beta} - [\![\boldsymbol{F}]\!]^{\delta\alpha} = [\![\boldsymbol{F}]\!]^{\delta\beta} - [\![\boldsymbol{F}]\!]^{\delta\alpha}$$

imposes the additional compatibility conditions

$$\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta} = \boldsymbol{a}^{\delta\beta} \otimes \boldsymbol{n}^{\delta\beta} - \boldsymbol{a}^{\delta\alpha} \otimes \boldsymbol{n}^{\delta\alpha} \quad \forall \delta \neq \alpha, \beta \tag{7.158}$$

on the $\boldsymbol{a}^{\alpha\beta}$. Since $\boldsymbol{n}^{\alpha\beta}$ is a unit vector, right-multiplying this equation with $\boldsymbol{n}^{\alpha\beta}$ direclty fixes $\boldsymbol{a}^{\alpha\beta}$ in terms of $\boldsymbol{a}^{\alpha\delta}$ and $\boldsymbol{a}^{\beta\delta}$ as[92]

$$\boldsymbol{a}^{\alpha\beta} = \left(\boldsymbol{n}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}\right)\boldsymbol{a}^{\alpha\beta} = \left(\boldsymbol{n}^{\alpha\beta} \cdot \boldsymbol{n}^{\delta\beta}\right)\boldsymbol{a}^{\delta\beta} - \left(\boldsymbol{n}^{\alpha\beta} \cdot \boldsymbol{n}^{\delta\alpha}\right)\boldsymbol{a}^{\delta\alpha} \quad \forall \delta \neq \alpha.$$

This already implies that **all** jump vectors would immediately follow once they are known with respect to a single phase[93].
Even when all normals happen to be parallel (and "testing" the combatibility conditions (7.158) with the common normal vector $\boldsymbol{n}^{\alpha\beta} = \pm\boldsymbol{n}$, $\alpha \neq \beta$, therefore provides all relevant information contained in the system) already reduces the number of effectively independent jump vectors to at most $N-1$ (e.g. the $\boldsymbol{a}^{1\alpha}$, $\alpha \geq 2$). For every triplet (α, β, δ) of phases for which the normals $\boldsymbol{n}^{\alpha\beta}$, $\boldsymbol{n}^{\delta\alpha}$ and $\boldsymbol{n}^{\delta\beta}$ are not all parallel, Equation (7.158) imposes an additional set of restrictions on the $\boldsymbol{a}^{\alpha\beta}$ though. In fact, if either $\boldsymbol{n}^{\delta\alpha}$ or $\boldsymbol{n}^{\delta\beta}$ are not parallel to $\boldsymbol{n}^{\alpha\beta}$, there is a tangential vector $\boldsymbol{t}^{\alpha\beta}$ such that $\boldsymbol{t}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}$ is zero but at least one of $\boldsymbol{n}^{\delta\beta} \cdot \boldsymbol{n}^{\delta\alpha}$ and $\boldsymbol{t}^{\alpha\beta} \cdot \boldsymbol{n}^{\delta\beta}$ is non-zero. Applying $\boldsymbol{t}^{\alpha\beta}$ from the right to the compatibility conditions,

$$\boldsymbol{0} = \left(\boldsymbol{t}^{\alpha\beta} \cdot \boldsymbol{n}^{\delta\beta}\right)\boldsymbol{a}^{\delta\beta} - \left(\boldsymbol{t}^{\alpha\beta} \cdot \boldsymbol{n}^{\delta\alpha}\right)\boldsymbol{a}^{\delta\alpha}$$

which allows, besides fixing $\boldsymbol{a}^{\alpha\beta}$ in terms of the two respective other jump vectors, fixing one jump vector on the right-hand side in terms of the other one (or directly as zero if one of the other normals also happens to be parallel to $\boldsymbol{n}^{\alpha\beta}$). Having non-parralel "two-phase-normal vectors" - such as is normally the case - thus even further reduces the available degrees of freedom in the ansatz (7.156).

Just as the jumps in the strains are not all independent, neither are of course the jumps in the normal stresses. Nevertheless, this does - similar to the example from Figure 7.1 - not usually entail a sufficient reduction of the number of normal continuity equations to be satified, with one notable exception being the case when there is only a single normal involved. In combination with the continuity condition, the resulting system is therefore generally inconsistent since there are not enough degrees of freedom to enforce all equations simultaneously.

Remark 152. Together with the averaging condition $\boldsymbol{F} = \sum_{\alpha=1}^{N} \boldsymbol{F}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$, it is, similarly to the two-phase setting, possible to reexpress $\boldsymbol{F}^{\alpha}$ in terms of the average displacement gradient $\boldsymbol{F}$ and the jumps with respect to the other phases. If fact, making use of $\boldsymbol{F}^{\beta} = \boldsymbol{F}^{\alpha} + [\![\boldsymbol{F}]\!]^{\alpha\beta}$ and $\sum_{\alpha=1}^{N} h^{\alpha}(\boldsymbol{\phi}) = 1$ allows rewriting the averaging condition as

$$\boldsymbol{F} = \boldsymbol{F}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) + \sum_{\beta \neq \alpha} \boldsymbol{F}^{\beta} h^{\beta}(\boldsymbol{\phi}) = \boldsymbol{F}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) + \sum_{\beta \neq \alpha} [\![\boldsymbol{F}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi}) = \boldsymbol{F} + \sum_{\beta \neq \alpha} [\![\boldsymbol{F}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi})$$

[92] Note that, even though this argument is slightly simpler in the non-symmetrized case, the symmetrized form

$$\left(\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta}\right)_S = \left(\boldsymbol{a}^{\delta\beta} \otimes \boldsymbol{n}^{\delta\beta}\right)_S - \left(\boldsymbol{a}^{\delta\alpha} \otimes \boldsymbol{n}^{\delta\alpha}\right)_S$$

leads to the same conclusion. In fact, a first contraction of the equality $\left(\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta}\right)_S = \boldsymbol{T}$ with known $\boldsymbol{T}$ leads to $\frac{1}{2}\left(\boldsymbol{a}^{\alpha\beta} + \left(\boldsymbol{a}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}\right)\boldsymbol{n}^{\alpha\beta}\right) = \boldsymbol{T} \cdot \boldsymbol{n}^{\alpha\beta}$, which is not yet fully explicit in $\boldsymbol{a}^{\alpha\beta}$. A second contraction then shows $\boldsymbol{a}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta} = \boldsymbol{n}^{\alpha\beta} \cdot \boldsymbol{T} \cdot \boldsymbol{n}^{\alpha\beta}$, which, reinserted into the previous equation shows that $\boldsymbol{a}^{\alpha\beta} = 2\boldsymbol{T} \cdot \boldsymbol{n}^{\alpha\beta} - \left(\boldsymbol{n}^{\alpha\beta} \cdot \boldsymbol{T} \cdot \boldsymbol{n}^{\alpha\beta}\right)\boldsymbol{n}^{\alpha\beta}$ and thus fully specifies $\boldsymbol{a}^{\alpha\beta}$ in terms of $\boldsymbol{T}$.

[93] In fact, assuming for convenience for this to be the first phase, knowing the $\boldsymbol{a}^{1\alpha}$, $2 \leq \alpha \leq N$ and using the appropriate column in Equation (7.158) with $\delta = 1$ and $\beta = \alpha + 1$ one can obtain all jump vectors $\boldsymbol{a}^{\alpha\beta}$ for $\alpha = \beta - 1$. Thus knowing $\boldsymbol{a}^{\alpha(\alpha+1)}$ and $\boldsymbol{a}^{(\alpha+1)(\alpha+2)}$ for all $\alpha \geq 2$, the $\boldsymbol{a}^{\alpha(\alpha+2)}$ follow immediately from $\boldsymbol{a}^{\beta\alpha} = \boldsymbol{a}^{\alpha\beta}$ and $\boldsymbol{a}^{\alpha(\alpha+2)} \otimes \boldsymbol{n}^{\alpha(\alpha+2)} = \boldsymbol{a}^{(\alpha+1)(\alpha+2)} \otimes \boldsymbol{n}^{(\alpha+1)(\alpha+2)} - \boldsymbol{a}^{(\alpha+1)\alpha} \otimes \boldsymbol{n}^{(\alpha+1)\alpha}$ and, repeating this argument, finally all $\boldsymbol{a}^{\alpha\beta}$ for $\beta \neq \alpha$.

and thus leads to[94]

$$\boldsymbol{F}^{\alpha} = \boldsymbol{F} - \sum_{\beta \neq \alpha} [\![\boldsymbol{F}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi}), \forall \alpha, \tag{7.159}$$

respectively, through a completely analogous calculation, the relation

$$\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon} - \sum_{\beta \neq \alpha} [\![\boldsymbol{\epsilon}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi}), \forall \alpha \tag{7.160}$$

for the $\boldsymbol{\epsilon}^{\alpha}$. ◇

Remark 153. Note that, taking the weighted average of Equation (7.159) and Equation (7.160) shows that the jumps in addition satisfy

$$\sum_{\alpha=1}^{N} \sum_{\beta \neq \alpha} [\![\boldsymbol{F}]\!]^{\alpha\beta} h^{\alpha}(\boldsymbol{\phi}) h^{\beta}(\boldsymbol{\phi}) = \boldsymbol{0} \quad \text{and} \quad \sum_{\alpha=1}^{N} \sum_{\beta \neq \alpha} [\![\boldsymbol{\epsilon}]\!]^{\alpha\beta} h^{\alpha}(\boldsymbol{\phi}) h^{\beta}(\boldsymbol{\phi}) = \boldsymbol{0}. \tag{7.161}$$

◇

A Model Based on a Common Normal ([62], §4.5 [61])

A first way of generalizing the two-phase model from the previous section to the multiphase setting is proposed in [62] and [61] and further successfully applied in e.g. [66], [4], [5]. It is based upon generalizing the representation of the normal and tangential stresses in equations (7.111) and (7.112) in terms of a basis-transformation in equation through the use of a common "average" normal vector $\boldsymbol{n}$. Using this normal vector, it assumes the equality of all tangential strains with the (given) average one $\boldsymbol{\epsilon_t}$ and then enforces the equality of all normal phase-specific stresses with the effective (a priori unknown) normal stress $\boldsymbol{\sigma_n}$. Using, instead of the base-transformation as in [62] the projector-based formalism in [74] and setting $\left(\cdot\right)_{\boldsymbol{n}} := \mathcal{N} : \left(\cdot\right)$, $\left(\cdot\right)_{\boldsymbol{t}} = \mathcal{T} : \left(\cdot\right)$, the underlying assumption is thus that $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon_t} \quad \forall \alpha$ and that $\boldsymbol{\sigma}_{\boldsymbol{n}}^{\alpha} = \boldsymbol{\sigma_n} \quad \forall \alpha$. This is combined with the usual averaging conditions $\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$ and $\boldsymbol{\sigma} = \sum_{\alpha=1}^{N} \boldsymbol{\sigma}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$ on the total strains and stresses, where, due to the conditions above, it is clear that the averaging procedure actually only affects the normal resp. tangential components. Together with the phase-specific material law $\boldsymbol{\sigma}^{\alpha} = \boldsymbol{C}^{\alpha} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)$ and making use of the idempotence of the projection operators ($\mathcal{N} = \mathcal{N}^2$ and $\mathcal{T} = \mathcal{T}^2$), the continuity condition on the normal components of the stresses may be rewritten as

$$\begin{aligned}
\boldsymbol{\sigma}_{\boldsymbol{n}}^{\alpha} =& \mathcal{N} : \left(\boldsymbol{\mathcal{C}}^{\alpha} : \left(\mathcal{N} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) + \mathcal{T} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right) = \mathcal{N} : \left(\boldsymbol{\mathcal{C}}^{\alpha} : \left(\mathcal{N} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) + \mathcal{T} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right)\right. \\
=& \boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha} : \left(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\right) + \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} : \left(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\right) \overset{!}{=} \boldsymbol{\sigma_n} := \mathcal{N} : \boldsymbol{\sigma} \quad \forall \alpha,
\end{aligned}$$

where $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha} = \mathcal{N} : \boldsymbol{\mathcal{C}}^{\alpha} : \mathcal{N}$, $\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} = \mathcal{N} : \boldsymbol{\mathcal{C}}^{\alpha} : \mathcal{T}$ and use was made of $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon_t}$. Using, as in [74], an appropriate pseudo-inverse $\boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} = \left(\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}\right)^{\dagger}$ of $\boldsymbol{\mathcal{C}}_{\boldsymbol{nn}}^{\alpha}$, this equation permits reexpressing all unknown normal components of the phase-specific strains in terms of the single common normal stress-component as

$$\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} = \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} + \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} : \left(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right) \tag{7.162}$$

and thus leads, through a weighted average, to the condition

$$\boldsymbol{\epsilon_n} = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) + \sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} : \left(\boldsymbol{\sigma_n} - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right) h^{\alpha}(\boldsymbol{\phi})$$

[94] Note that if $h^{\alpha}(\boldsymbol{\phi}) = 0$, $\boldsymbol{F}^{\alpha}$ does not appear in the averaging condition in terms of the $\left(\boldsymbol{F}^{\beta}\right)_{1 \leq \beta \leq N}$ and $\boldsymbol{F}$ and can therefore in principle be defined arbitrarily regardless of the values of the other $\boldsymbol{F}^{\beta}$, $\beta \neq \alpha$. In contrast, **given** a set of jump vector $[\![\boldsymbol{F}]\!]^{\alpha\beta}$ and $\boldsymbol{F}$, the relation (7.159) still has to hold for $\boldsymbol{F}^{\alpha}$, the only difference to the case $h^{\alpha}(\boldsymbol{\phi}) \neq 0$ being that, given all other $\boldsymbol{F}^{\beta}$ and $\boldsymbol{F}$, one is free to choose an arbitrary $\boldsymbol{F}^{\alpha}$ for defining the jumps $[\![\boldsymbol{F}]\!]^{\alpha\beta} = \boldsymbol{F}^{\beta} - \boldsymbol{F}^{\alpha}$.

relating the $\boldsymbol{\sigma_n}$ with the given average normal strain $\boldsymbol{\epsilon_n}$. Together with the corresponding pseudo-inverse of $\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$, which, for consistency with the description in [62] will be abbreviated as $-\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}}$, i.e. $\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}} := -\Big(\sum_{\alpha=1}^{N} \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} h^{\alpha}(\boldsymbol{\phi})\Big)^{\dagger}$, this allows obtaining the commmon normal stress component as

$$\boldsymbol{\sigma_n} = -\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}} : \left(\boldsymbol{\epsilon_n} - \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) + \sum_{\alpha=1}^{N} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} : \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big) h^{\alpha}(\boldsymbol{\phi}) \right). \tag{7.163}$$

Finally, again using Equation (7.162) with this now known normal stress, one can first obtain the normal phase-specific strain components $\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha}$ given by

$$\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} = \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha} + \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} : \left[-\bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}} : \left(\boldsymbol{\epsilon_n} - \sum_{\beta=1}^{N} \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\beta} h^{\beta}(\boldsymbol{\phi}) + \sum_{\beta=1}^{N} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\beta} : \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\beta}\big)\Big) h^{\beta}(\boldsymbol{\phi}) \right) - \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} : \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) \right]$$

and, combined with the already known tangential strains $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon_t}$ and through another averaging using the phase-specific material law, the tangential stress components as

$$\begin{aligned}
\boldsymbol{\sigma_t} &= \sum_{\alpha=1}^{N} \boldsymbol{\sigma}_{\boldsymbol{t}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha} : \big(\boldsymbol{\epsilon}_{\boldsymbol{n}}^{\alpha} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\alpha}\big) + \boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha} : \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big)\Big) h^{\alpha}(\boldsymbol{\phi}) \\
&= -\sum_{\alpha=1}^{N} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) \Big) : \bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}} : \left(\boldsymbol{\epsilon_n} - \sum_{\beta=1}^{N} \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}^{\beta} h^{\beta}(\boldsymbol{\phi}) \right) \\
&\quad + \sum_{\alpha=1}^{N} \left(\big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tt}}^{\alpha} - \boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\alpha} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\alpha} \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} \big) - \sum_{\beta=1}^{N} \Big(\boldsymbol{\mathcal{C}}_{\boldsymbol{tn}}^{\beta} : \boldsymbol{\mathcal{S}}_{\boldsymbol{nn}}^{\beta} h^{\beta}(\boldsymbol{\phi}) \Big) : \bar{\boldsymbol{\mathcal{T}}}_{\boldsymbol{nn}} : \boldsymbol{\mathcal{C}}_{\boldsymbol{nt}}^{\alpha} \right) : \big(\boldsymbol{\epsilon_t} - \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}}^{\alpha}\big) h^{\alpha}(\boldsymbol{\phi})
\end{aligned} \tag{7.164}$$

As in the two-phase case, one can alternatively choose a description in terms of appropriately chosen jump vectors. As $\mathcal{T} : \boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon}_{\boldsymbol{t}}^{\alpha} = \boldsymbol{\epsilon_t}$ for all α, $\boldsymbol{\epsilon}^{\alpha}$ is necessarily of the form

$$\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon} - \big(\boldsymbol{a}^{\alpha} \otimes \boldsymbol{n}\big)_S. \tag{7.165}$$

Together with the averaging condition $\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^{\alpha} h^{\alpha}(\boldsymbol{\phi})$, this in addition implies the constraint

$$\sum_{\alpha=1}^{N} h^{\alpha}(\boldsymbol{\phi}) \boldsymbol{a}^{\alpha} \stackrel{!}{=} \boldsymbol{0} \tag{7.166}$$

on the jump-vectors $\boldsymbol{a}^{\alpha}$.
In terms of the jump vectors and again using the abbreviation $\boldsymbol{\Sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha} : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\big)$, the condition on the equality of the normal stresses can be written as

$$\boldsymbol{\sigma}^{\alpha} \cdot \boldsymbol{n} = \boldsymbol{n} \cdot \Big(\boldsymbol{\mathcal{C}}^{\alpha} : \big(\boldsymbol{\epsilon} - \big(\boldsymbol{a}^{\alpha} \otimes \boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^{\alpha} \big) \Big) = \boldsymbol{\Sigma}^{\alpha} \cdot \boldsymbol{n} - \big(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n} \big) \cdot \boldsymbol{a}^{\alpha} \stackrel{!}{=} \boldsymbol{\sigma} \cdot \boldsymbol{n}$$

and thus shows that the phase-specific jump vectors $\boldsymbol{a}^{\alpha}$ are given by

$$\boldsymbol{a}^{\alpha} = \big(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n} \big)^{-1} \cdot \Big(\big(\boldsymbol{\Sigma}^{\alpha} - \boldsymbol{\sigma} \big) \cdot \boldsymbol{n} \Big) \tag{7.167}$$

in terms of the common value $\boldsymbol{\sigma} \cdot \boldsymbol{n}$. Together with the zero-average condition on the jump-vectors, it follows that $\boldsymbol{\sigma} \cdot \boldsymbol{n}$ has to be such that $\sum_{\alpha=1}^{N} \boldsymbol{a} h^{\alpha}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N} \Big(\big(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n} \big)^{-1} \cdot \Big(\big(\boldsymbol{\Sigma}^{\alpha} - \boldsymbol{\sigma} \big) \cdot \boldsymbol{n} \Big) h^{\alpha}(\boldsymbol{\phi}) \stackrel{!}{=} \boldsymbol{0}$, i.e.

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = \Big(\sum_{\alpha=1}^{N} \big(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n} \big)^{-1} h^{\alpha}(\boldsymbol{\phi}) \Big)^{-1} \cdot \sum_{\alpha=1}^{N} \big(\boldsymbol{n} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n} \big)^{-1} \cdot \big(\boldsymbol{\Sigma}^{\alpha} \cdot \boldsymbol{n} \big) h^{\alpha}(\boldsymbol{\phi}). \tag{7.168}$$

Similarly to before, recombining the now known normal stress with Equation (7.167) shows that the phase-specific jump vectors are given by

$$\boldsymbol{a}^\alpha = \left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\right)^{-1}\cdot\left[\left(\boldsymbol{\Sigma}^\alpha\cdot\boldsymbol{n} - \Big(\sum_{\beta=1}^{N}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\beta\cdot\boldsymbol{n}\right)^{-1}h^\beta(\boldsymbol{\phi})\Big)^{-1}\cdot\sum_{\beta=1}^{N}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\beta\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{\Sigma}^\beta\cdot\boldsymbol{n}\right)h^\beta(\boldsymbol{\phi})\right],$$

thus leading, with the given value of $\boldsymbol{\epsilon}$ to the alternative expression

$$\begin{aligned}\boldsymbol{\sigma} =& \sum_{\alpha=1}^{N}\boldsymbol{\Sigma}^\alpha h^\alpha(\boldsymbol{\phi}) + \sum_{\alpha=1}^{N}\Big(\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\Big)\cdot\boldsymbol{a}^\alpha h^\alpha(\boldsymbol{\phi})\\ =&\boldsymbol{\Sigma} + \sum_{\alpha=1}^{N}\Big(\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\Big)\cdot\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\alpha\cdot\boldsymbol{n}\right)^{-1}\\ &\cdot\left[\left(\boldsymbol{\Sigma}^\alpha\cdot\boldsymbol{n} - \Big(\sum_{\beta=1}^{N}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\beta\cdot\boldsymbol{n}\right)^{-1}h^\beta(\boldsymbol{\phi})\Big)^{-1}\cdot\sum_{\beta=1}^{N}\left(\boldsymbol{n}\cdot\boldsymbol{\mathcal{C}}^\beta\cdot\boldsymbol{n}\right)^{-1}\cdot\left(\boldsymbol{\Sigma}^\beta\cdot\boldsymbol{n}\right)h^\beta(\boldsymbol{\phi})\right]h^\alpha(\boldsymbol{\phi})\end{aligned} \tag{7.169}$$

for the effective stress.

Based on the use of the jump-vectors above, the driving-force calculation for this model can, alternatively to the procedure in [62] in the transformed coordinate system, be performed in a manner similar to the two-phase case considered previously. Differentiating the expression $f_{el}(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta f_{el}^\beta h^\beta(\boldsymbol{\phi}) = \sum_\beta \frac{1}{2}\left(\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon}) - \tilde{\boldsymbol{\epsilon}}^\beta\right):\boldsymbol{\mathcal{C}}^\beta:\left(\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon}) - \tilde{\boldsymbol{\epsilon}}^\beta\right)h^\beta(\boldsymbol{\phi})$ with respect to ϕ^α for a **fixed** normal vector $\boldsymbol{n}$, if first follows that

$$\frac{\partial}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}} f_{el}(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta f_{el}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_\beta\boldsymbol{\sigma}^\beta:\frac{\partial\boldsymbol{\epsilon}^\beta}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}}h^\beta.$$

In contrast to the chemical case, one can again not directly make use of the product rule and $\frac{\partial\sum_\beta\boldsymbol{\epsilon}^\beta h^\beta}{\partial\phi^\alpha} = \frac{\partial\boldsymbol{\epsilon}}{\partial\phi^\alpha} = \boldsymbol{0}$ for simplifying the last term as the tangential parts of the $\boldsymbol{\sigma}^\beta$ do not reduce to a common value. Nevertheless, since $\boldsymbol{\epsilon}^\beta = \boldsymbol{\epsilon} - (\boldsymbol{a}^\beta\otimes\boldsymbol{n})_S$, it follows that $\frac{\partial\boldsymbol{\epsilon}^\beta}{\partial\phi^\alpha}\big|_{\boldsymbol{n}} = -(\frac{\partial\boldsymbol{a}^\beta}{\partial\phi^\alpha}\otimes\boldsymbol{n})_S$ and thus

$$\sum_\beta\boldsymbol{\sigma}^\beta:\frac{\partial\boldsymbol{\epsilon}^\beta}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}}h^\beta = -\sum_\beta\left(\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n}\right)\cdot\frac{\partial\boldsymbol{a}^\beta}{\partial\phi^\alpha}h^\beta = -(\boldsymbol{\sigma}\cdot\boldsymbol{n})\cdot\sum_\beta\frac{\partial\boldsymbol{a}^\beta}{\partial\phi^\alpha}h^\beta$$

since all normal stresses for the different phases are in fact equal. Further making use of the condition $\sum_\beta\boldsymbol{a}^\beta h^\beta = \boldsymbol{0}$ from Equation (7.166) and using the product rule $\sum_\beta\frac{\partial\boldsymbol{a}^\beta}{\partial\phi^\alpha}h^\beta = \frac{\partial\sum_\beta\boldsymbol{a}^\beta h^\beta}{\partial\phi^\alpha} - \sum_\beta\boldsymbol{a}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}$, one can eliminate the derivatives of the $\boldsymbol{a}^\beta$, finally leading to

$$\frac{\partial}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}} f_{el}(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta f_{el}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha} + (\boldsymbol{\sigma}\cdot\boldsymbol{n})\cdot\sum_\beta\boldsymbol{a}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha} = \sum_\beta\left(f_{el}^\beta + (\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n})\cdot\boldsymbol{a}^\beta\right)\frac{\partial h^\beta}{\partial\phi^\alpha}. \tag{7.170}$$

Alternatively using $\boldsymbol{\epsilon}^\beta = \boldsymbol{\epsilon} - \left(\boldsymbol{a}^\beta\otimes\boldsymbol{n}\right)_S$, one can also rewrite the second contribution to the driving force as

$$\sum_{\beta=1}^{N}\left(\boldsymbol{\sigma}\cdot\boldsymbol{n}\right)\cdot\boldsymbol{a}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha} = \sum_{\beta=1}^{N}\boldsymbol{\sigma}:\left(\boldsymbol{a}^\beta\otimes\boldsymbol{n}\right)_S\frac{\partial h^\beta}{\partial\phi^\alpha} = \sum_{\beta=1}^{N}\boldsymbol{\sigma}:\left(\boldsymbol{\epsilon}-\boldsymbol{\epsilon}^\beta\right)\frac{\partial h^\beta}{\partial\phi^\alpha} = -\sum_{\beta=1}^{N}\boldsymbol{\sigma}:\boldsymbol{\epsilon}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha} + \left(\boldsymbol{\sigma}:\boldsymbol{\epsilon}\right)\frac{\partial\sum_{\beta=1}^{N}h^\beta}{\partial\phi^\alpha}.$$

Since the last term vanishes, this leaves

$$\frac{\partial f_{el}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}} = \sum_{\beta=1}^{N}\left(f_{el}^\beta(\boldsymbol{\epsilon}^\beta) - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^\beta\right)\frac{\partial h^\beta}{\partial\phi^\alpha}. \tag{7.171}$$

The remaining contributions to $\frac{\partial f_{el}}{\partial \phi^\alpha}$ and $\frac{\partial f_{el}}{\partial \nabla \phi^\alpha}$ through the dependence on $\boldsymbol{n}$ follow from a similar reasoning in combination with the definition of $\boldsymbol{n}$ in terms of $\boldsymbol{\phi}$ and $\nabla\boldsymbol{\phi}$. Differentiating f_{el} with respect to $\boldsymbol{n}$ one has

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = \sum_\beta \boldsymbol{\sigma}^\beta : \frac{\partial \boldsymbol{\epsilon}^\beta}{\partial \boldsymbol{n}} h^\beta = -\sum_\beta \Big((\boldsymbol{\sigma}^\beta \cdot \boldsymbol{n}) \cdot \frac{\partial \boldsymbol{a}^\beta}{\partial \boldsymbol{n}} + (\boldsymbol{\sigma}^\beta \cdot \boldsymbol{a}^\beta) \Big) h^\beta .$$

The first contribution can be eliminated by making use of the common normal stress $\boldsymbol{\sigma}^\beta \cdot \boldsymbol{n} = \boldsymbol{\sigma} \cdot \boldsymbol{n}$ for all β and the condition $\sum_\beta \boldsymbol{a}^\beta h^\beta = \boldsymbol{0}$ together with the independence of the h^β on $\boldsymbol{n}$, thus leaving the simple expression

$$\frac{\partial f_{el}}{\partial \boldsymbol{n}} = -\sum_{\beta=1}^{N} \big(\boldsymbol{\sigma}^\beta \cdot \boldsymbol{a}^\beta \big) h^\beta(\boldsymbol{\phi}) \tag{7.172}$$

for the derivative with respect to $\boldsymbol{n}$.

Remark 154. As the similarity of the calculations and the resulting expressions for the driving force contributions to the two-phase case shows, this model is able to maintain the most of the pleasant properties of its two-phase analogon in the multiphase case. In particular, its free energy density is defined in the natural manner as the weigted average of the $f^\alpha = \frac{1}{2}\big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big) : \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$ of the phase-specific ones with the $\boldsymbol{\epsilon}^\alpha$ satisfying the Hadamard jump conditions with respect to every other phase. Furthermore, the normal stresses between all phases are continuous, and the total stress satisfies both $\boldsymbol{\sigma} = \sum_\alpha \boldsymbol{\sigma}^\alpha h^\alpha(\boldsymbol{\phi})$ and the compatibility condition $\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}$. In fact, one has $\frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}} = \sum_\alpha \boldsymbol{\sigma}^\alpha : \frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{\epsilon}} h^\alpha$, which, from $\boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon} - (\boldsymbol{a}^\alpha \otimes \boldsymbol{n})_S$, leads to

$$\frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}} = \sum_\alpha \boldsymbol{\sigma}^\alpha h^\alpha - \sum_\alpha (\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n}) \cdot \frac{\partial \boldsymbol{a}^\alpha}{\partial \boldsymbol{\epsilon}} h^\alpha = \boldsymbol{\sigma} - (\boldsymbol{\sigma} \cdot \boldsymbol{n}) \frac{\partial \sum_\alpha \boldsymbol{a}^\alpha h^\alpha}{\partial \boldsymbol{\epsilon}} = \boldsymbol{\sigma}.$$

Its only drawback - and this is a potentially major one from a physical point of view - is that this is primarily due to the usage of a single common normal vector $\boldsymbol{n}$. While the appropriate definition of a normal vector $\boldsymbol{n}^{\alpha\beta}$ between any phase-pairing is already a topic of some debate, i.e. for example in terms of the $\boldsymbol{q}^{\alpha\beta}$ or the normalized differences $\nabla\phi^\beta - \nabla\phi^\alpha$, doing so for all phases and thus also phase-pairings at once is an even more difficult problem. This is also the primary criticism of the model in [62] by Tschukin in [74].
In addition, even though the contribution to the driving force through the dependence on $\boldsymbol{n}$ is left out in [62], defining $\boldsymbol{n}$ in terms of a fairly complex averaging procedure based on the phasefield and various mechanical properties will, despite of the simplicity of the expression in Equation (7.172), lead to a very complex total driving force in combination with the corresponding expressions for $\frac{\partial \boldsymbol{n}}{\partial \phi^\alpha}$ and $\frac{\partial \boldsymbol{n}}{\partial \nabla \phi^\alpha}$. ◇

The Model by Schneider et al. ([63] and [61])

A quite different generalization of the model from Section 7.2.2 to the multiphase setting was proposed by Schneider et al. in [63] in the finite deformation setting (see also [61]).

Dropping the restrictions imposed by Equation (7.156) except with respect to a given reference phase R, Equation (7.159) applied for this reference phase becomes

$$\boldsymbol{F}^{\mathrm{R}} = \boldsymbol{F} - \sum_{\beta \neq \mathrm{R}} \boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{N}^{\mathrm{R}\beta} h^\beta(\boldsymbol{\phi}) \quad \text{and} \quad \boldsymbol{F}^\alpha = \boldsymbol{F}^{\mathrm{R}} + \boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{N}^{\mathrm{R}\alpha}, \quad \alpha \neq R \tag{7.173}$$

resp.

$$\boldsymbol{\epsilon}^{\mathrm{R}} = \boldsymbol{\epsilon} - \sum_{\beta \neq \mathrm{R}} \big(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta} \big)_S h^\beta(\boldsymbol{\phi}) \quad \text{and} \quad \boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon}^{\mathrm{R}} + \big(\boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{n}^{\mathrm{R}\alpha} \big)_S, \quad \alpha \neq \mathrm{R} \tag{7.174}$$

in the small deformation case. The remaining $3(N-1)$ unknowns in the jump vectors $\boldsymbol{a}^{\mathrm{R}\alpha}$, $\alpha \neq R$, can then again be fixed by imposing the continuity conditions on the normal stresses in Equation (7.157) with respect to this reference phase only, i.e.

$$[\![\boldsymbol{\sigma}]\!]^{\mathrm{R}\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha} = \left(\boldsymbol{\sigma}^{\alpha} - \boldsymbol{\sigma}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\alpha} = \boldsymbol{0} \tag{7.175}$$

for all $\alpha \neq \mathrm{R}$.

Inserting the definition of the $\boldsymbol{\epsilon}^{\alpha}$ in terms of $\boldsymbol{\epsilon}$ and the $\boldsymbol{a}^{\mathrm{R}\alpha}$ into the phase-specific stress strain relationships, $\alpha = 1, ..., N$, one has

$$\boldsymbol{\sigma}^{\alpha} = \begin{cases} \boldsymbol{\mathcal{C}}^{\mathrm{R}} : \left(\boldsymbol{\epsilon} - \sum_{\beta \neq \mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^{\beta}(\phi) - \tilde{\boldsymbol{\epsilon}}^{\mathrm{R}}\right) & \alpha = R, \\ \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon} - \sum_{\beta \neq \mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^{\beta}(\phi) + \left(\boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{n}^{\mathrm{R}\alpha}\right)_S - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) & \text{else.} \end{cases}$$

Together with the right subsymmetry of $\boldsymbol{\mathcal{C}}^{\alpha}$ and defining as before $\boldsymbol{\Sigma}^{\alpha} := \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)$, this simplifies to

$$\boldsymbol{\sigma}^{\alpha} = \begin{cases} \boldsymbol{\Sigma}^{\mathrm{R}} - \sum_{\beta \neq \mathrm{R}} h^{\beta}(\phi) \left(\boldsymbol{\mathcal{C}}^{\mathrm{R}} \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta} & \alpha = R, \\ \boldsymbol{\Sigma}^{\alpha} - \sum_{\beta \neq \mathrm{R}} h^{\beta}(\phi) \left(\boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta} + \left(\boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right) \cdot \boldsymbol{a}^{\mathrm{R}\alpha} & \text{else,} \end{cases}$$

and thus, in combination with the condition on the normal continuity of the stresses leads to

$$\left(\left(\boldsymbol{\Sigma}^{\mathrm{R}} - \sum_{\beta \neq \mathrm{R}} \left(\boldsymbol{\mathcal{C}}^{\mathrm{R}} \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta}\right) - \left(\boldsymbol{\Sigma}^{\alpha} - \left(\sum_{\beta \neq \mathrm{R}} \left(\boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \otimes \boldsymbol{a}^{\mathrm{R}\beta} h^{\beta}\right) + \left(\boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right) \cdot \boldsymbol{a}^{\mathrm{R}\alpha}\right)\right) \cdot \boldsymbol{n}^{\mathrm{R}\alpha} \stackrel{!}{=} \boldsymbol{0}.$$

Regrouping terms, one therefore has to solve the $3(N-1)$-system

$$\begin{aligned} &\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right) \cdot \boldsymbol{a}^{\mathrm{R}\alpha} - \sum_{\beta \neq \mathrm{R}} h^{\beta}\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \left(\boldsymbol{\mathcal{C}}^{\alpha} - \boldsymbol{\mathcal{C}}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta} \\ =&\left(\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \boldsymbol{\mathcal{C}}^{\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right) - h^{\alpha}\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \left(\boldsymbol{\mathcal{C}}^{\alpha} - \boldsymbol{\mathcal{C}}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right)\right) \cdot \boldsymbol{a}^{\mathrm{R}\alpha} \\ &- \sum_{\beta \neq \mathrm{R}, \alpha} h^{\beta}\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \left(\boldsymbol{\mathcal{C}}^{\alpha} - \boldsymbol{\mathcal{C}}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta} \\ =&\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \left((1 - h^{\alpha})\boldsymbol{\mathcal{C}}^{\alpha} + h^{\alpha}\boldsymbol{\mathcal{C}}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\alpha}\right)\right) \cdot \boldsymbol{a}^{\mathrm{R}\alpha} - \sum_{\beta \neq \mathrm{R}, \alpha} h^{\beta}\left(\boldsymbol{n}^{\mathrm{R}\alpha} \cdot \left(\boldsymbol{\mathcal{C}}^{\alpha} - \boldsymbol{\mathcal{C}}^{\mathrm{R}}\right) \cdot \boldsymbol{n}^{\mathrm{R}\beta}\right) \cdot \boldsymbol{a}^{\mathrm{R}\beta} \\ \stackrel{!}{=}&\left(\boldsymbol{\Sigma}^{\mathrm{R}} - \boldsymbol{\Sigma}^{\alpha}\right) \cdot \boldsymbol{n}^{\mathrm{R}\alpha} = -[\![\boldsymbol{\Sigma}]\!]^{\mathrm{R}\alpha} \cdot \boldsymbol{n}^{\mathrm{R}\alpha} =: \boldsymbol{r}^{\mathrm{R}\alpha} \end{aligned} \tag{7.176}$$

with the "residual" $\boldsymbol{r}^{\mathrm{R}\alpha}$ corresponding to the (negative) stress jump as it would arise in the Voigt-Taylor model.

Finally, having determined the $\boldsymbol{a}^{\mathrm{R}\alpha}$, the total stress $\boldsymbol{\sigma}$ is obtained, based on a weighted average of the phase-specific ones, as

$$\begin{aligned} \boldsymbol{\sigma} &= \sum_{\alpha=1}^{N} \boldsymbol{\sigma}^{\alpha} h^{\alpha}(\phi) = \sum_{\alpha=1}^{N} \left(\boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right) h^{\alpha}(\phi) \\ &= \sum_{\alpha=1}^{N} \left(\boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon} - \sum_{\beta \neq \mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^{\beta}(\phi) - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) + \begin{cases} \boldsymbol{0} & \alpha = 1, \\ \left(\boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{n}^{\mathrm{R}\alpha}\right)_S & \text{else} \end{cases}\right) h^{\alpha}(\phi) \\ &= \sum_{\alpha=1}^{N} \left(\boldsymbol{\Sigma}^{\alpha} - \boldsymbol{\mathcal{C}}^{\alpha} : \left(\sum_{\beta \neq 1} \left(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^{\beta}(\phi)\right)\right) h^{\alpha}(\phi) + \sum_{\beta \neq \mathrm{R}} \boldsymbol{\mathcal{C}}^{\beta} : \left(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^{\beta}(\phi). \end{aligned}$$

Using $1 = \sum_{\alpha=1}^{N} h^\alpha(\boldsymbol{\phi})$, this can further be combined to give[95]

$$\boldsymbol{\sigma} = \sum_{\alpha=1}^{N} \left(\boldsymbol{\Sigma}^\alpha - \Big(\sum_{\beta \neq \mathrm{R}} \big(\boldsymbol{\mathcal{C}}^\alpha - \boldsymbol{\mathcal{C}}^\beta\big) : \big(\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\big)_S h^\beta(\boldsymbol{\phi}) \Big) \right) h^\alpha(\boldsymbol{\phi}). \tag{7.177}$$

Remark 155. Similarly to the chemical case considered in Section 7.1, it is not really necessary to solve for all N phases at once. Instead, the only equations which are actually coupled are the ones for which $h^\alpha(\boldsymbol{\phi}) > 0$, whereas the strains and stresses for all remaining phases can, if required, be recovered a posteriori. ◇

In contrast to the model based on a common normal, a clear advantage of this model is that it allows, more in line with the standard phasefield approach, using different normal vector $\boldsymbol{n}^{\alpha\beta}$ for each of the phase-pairings. This comes at a heavy price though, as this is made possible here by enforcing the desired jump-conditions with respect to a single reference phase R only, whereas all other phase-pairings will generally not satisfy the jump conditions on either the strains or the stresses[96].

Remark 156. Due to this, an important question is also how the reference phase can be chosen "appropriately". A simple and quite intuitive possibilty is to use the phase with the highest local volume-fraction ϕ^α. Other choices are of course possible, i.e. one could for example try to take the mechanical properties of the phases into account. Regardless of the actual choice, the necessity of having to choose is on the one hand an unpleasant one, and on the other hand also entails some issues (there being others as well as will be seen below) in the calculation of the driving force as the points where the reference phases change will generally be associated with discontinuities in the mechanical energies due to the sudden change of the subset of jump conditions to be satisfied. ◇

Furthermore - and this is an in principle very problematic issue from a variational point of view - even though one can conveniently define an elastic free energy density f_{el} as in Equation (7.71) based on the $\boldsymbol{\epsilon}^\alpha$ in Equation (7.174), there is no extremum principle with respect to this density underlying the actual determination of the jump vectors $\boldsymbol{a}^{\mathrm{R}\alpha}$.
In fact, a minimization of the free energy density (7.71) as in [51] for the given form of the $\boldsymbol{\epsilon}^\alpha$ would, with

$$\delta\boldsymbol{\epsilon}^\alpha = \begin{cases} -\sum_{\beta\neq\mathrm{R}} h^\beta \big(\delta\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\big)_S & \alpha = \mathrm{R}, \\ -\sum_{\beta\neq\mathrm{R}} h^\beta \big(\delta\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\big)_S + \big(\delta\boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{n}^{\mathrm{R}\alpha}\big)_S & \alpha \neq R \end{cases}$$

result in

$$\begin{aligned} &- h^R\boldsymbol{\sigma}^R : \sum_{\beta\neq\mathrm{R}} h^\beta \big(\delta\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\big)_S + \sum_{\alpha\neq\mathrm{R}} \Big(- h^\alpha\boldsymbol{\sigma}^\alpha : \sum_{\beta\neq\mathrm{R}} h^\beta \big(\delta\boldsymbol{a}^{\mathrm{R}\beta} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\big)_S + h^\alpha\boldsymbol{\sigma}^\alpha : \big(\delta\boldsymbol{a}^{\mathrm{R}\alpha} \otimes \boldsymbol{n}^{\mathrm{R}\alpha}\big)_S \Big) \\ &= \sum_{\alpha\neq\mathrm{R}} h^\alpha \left(\Big(\boldsymbol{\sigma}^\alpha - \sum_{\beta\neq\mathrm{R}} h^\beta\boldsymbol{\sigma}^\beta - h^R\boldsymbol{\sigma}^R \Big) \cdot \boldsymbol{n}^{\mathrm{R}\alpha} \right) \cdot \delta\boldsymbol{a}^{\mathrm{R}\alpha} = \sum_{\alpha\neq\mathrm{R}} h^\alpha \left(\Big(\boldsymbol{\sigma}^\alpha - \sum_{\beta} h^\beta\boldsymbol{\sigma}^\beta \Big) \cdot \boldsymbol{n}^{\mathrm{R}\alpha} \right) \cdot \delta\boldsymbol{a}^{\mathrm{R}\alpha} \overset{!}{=} \mathbf{0}, \end{aligned}$$

and thus the requirement that each $\boldsymbol{a}^{\mathrm{R}\alpha}$ should be such that the normal stress components w.r.t. $\boldsymbol{n}^{\mathrm{R}\alpha}$ coincides with the ones of the total resulting stress. As this differs from the characterization in Equation (7.175), there is no reason for the dependence of the $\boldsymbol{a}^{\mathrm{R}\alpha}$ with respect to the parameters $\boldsymbol{\phi}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{n}$ to drop out when differentiating f_{el}. In particular, the stress defined by $\boldsymbol{\sigma} = \sum_\alpha \boldsymbol{\sigma}^\alpha h^\alpha = \sum_\alpha \boldsymbol{\mathcal{C}}^\alpha : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha\big)$ does **not** generally satisfy the crucial relation $\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}$ except within bulk- or two-phase regions.
This has, at least at a theoretical level, far-reaching consequences in terms of the postulated gradient flow for $\boldsymbol{\phi}$ similar to the discussion in [41]. In fact, one now has to choose between

[95] The summation could clearly also be restricted to $\beta \neq \mathrm{R}, \alpha$.
[96] Unfortunately, the claim in [61] that the remaining equations are redundant is incorrect.

maintaining the simple structure of the algorithm, i.e. ignoring the interdependencies between the different parameters $\boldsymbol{\phi}$ and the displacement respectively $\boldsymbol{\epsilon}$, or, ultimately, the minimization of the functional in Equation (7.67).

Remark 157. From a more practical point of view, this global variational inconsistency can likely be considered as a secondary issue as compared to the many difficult questions related to multiphase regions such as the choice of normal vectors and implications associated to the choice of reference phase and, more importantly, the continued neglection of the dependence on $\boldsymbol{n}$ in [63].
One can in principle derive a driving force consistent with the minimization of $\mathcal{F}_\epsilon$ subject to the momentum-balance based on $\boldsymbol{\sigma}$ as above considered as a side-constraint using e.g. an adjoint equation corresponding to a reduced formulation in terms of the displacement $\boldsymbol{u}$ as a function of $\boldsymbol{\phi}$ and the $\boldsymbol{n}^{\mathrm{R}\alpha}$. This has the drawback that determining the adjoint state gives rise to a second **global** problem similar in nature to the steady-state momentum-balance Equation (7.69) and is therefore potentially relatively expensive. ◇

For the calculation of the driving force, it is again convenient to separate the derivative into two separate contributions, one based on fixed normal vectors $(\boldsymbol{n}^{\alpha\beta})_{\alpha\neq\beta}$ and the other one due to the changes in the normal vectors. A partial differentiation with respect to ϕ^α for fixed normals leads to

$$\frac{\partial}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} f(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta f^\beta \frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_\beta \boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon})}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} h^\beta(\boldsymbol{\phi}).$$

From

$$\boldsymbol{\epsilon}^{\mathrm{R}}(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \boldsymbol{\epsilon} - \sum_{\beta\neq\mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta}(\boldsymbol{\phi},\boldsymbol{\epsilon})\otimes\boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^\beta(\boldsymbol{\phi}), \quad \boldsymbol{\epsilon}^\alpha(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}^{\mathrm{R}}(\boldsymbol{\phi},\boldsymbol{\epsilon}) + \left(\boldsymbol{a}^{\mathrm{R}\alpha}(\boldsymbol{\phi},\boldsymbol{\epsilon})\otimes\boldsymbol{n}^{\mathrm{R}\alpha}\right)_S \quad \alpha\neq\mathrm{R}$$

together with $\boldsymbol{\sigma} = \sum_{\beta=1}^N \boldsymbol{\sigma}^\beta h^\beta(\boldsymbol{\phi})$, the second sum can be rewritten to to

$$\begin{aligned}
\sum_\beta h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon})}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} &= \Big(\sum_\beta \boldsymbol{\sigma}^\beta h^\beta\Big) : \frac{\partial\boldsymbol{\epsilon}^{\mathrm{R}}(\boldsymbol{\phi},\boldsymbol{\epsilon})}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} + \sum_{\beta\neq\mathrm{R}} \boldsymbol{\sigma}^\beta : \Big(\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\Big)_S h^\beta(\boldsymbol{\phi}) \\
&= -\boldsymbol{\sigma} : \sum_{\beta\neq\mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta}\otimes\boldsymbol{n}^{\mathrm{R}\beta}\right)_S \frac{\partial h^\beta}{\partial\phi^\alpha} - \boldsymbol{\sigma} : \sum_{\beta\neq\mathrm{R}} \Big(\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\Big)_S h^\beta \\
&\quad + \sum_{\beta\neq\mathrm{R}} \boldsymbol{\sigma}^\beta : \Big(\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} \otimes \boldsymbol{n}^{\mathrm{R}\beta}\Big)_S h^\beta \\
&= -\boldsymbol{\sigma} : \sum_{\beta\neq\mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta}\otimes\boldsymbol{n}^{\mathrm{R}\beta}\right)_S \frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_{\beta\neq\mathrm{R}} \Big(\big(\boldsymbol{\sigma}^\beta - \boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha} h^\beta.
\end{aligned}$$

Using $\left(\boldsymbol{a}^{\mathrm{R}\beta}(\boldsymbol{\phi},\boldsymbol{\epsilon})\otimes\boldsymbol{n}^{\mathrm{R}\beta}\right)_S = \boldsymbol{\epsilon}^\beta - \boldsymbol{\epsilon}^{\mathrm{R}}$, an alternative expression is given by

$$\sum_{\beta=1}^N h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon})}{\partial\phi^\alpha} = -\boldsymbol{\sigma} : \sum_{\beta\neq\mathrm{R}} \left(\boldsymbol{\epsilon}^\beta - \boldsymbol{\epsilon}^{\mathrm{R}}\right)\frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_{\beta\neq\mathrm{R}} \Big(\big(\boldsymbol{\sigma}^\beta - \boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha} h^\beta$$

which, with

$$\boldsymbol{\sigma} : \sum_{\beta\neq\mathrm{R}} \boldsymbol{\epsilon}^{\mathrm{R}} \frac{\partial h^\beta}{\partial\phi^\alpha} = \boldsymbol{\sigma} : \boldsymbol{\epsilon}^{\mathrm{R}} \frac{\partial\sum_{\beta\neq\mathrm{R}} h^\beta}{\partial\phi^\alpha} = \boldsymbol{\epsilon}^{\mathrm{R}} \frac{\partial\big(1 - h^{\mathrm{R}}(\boldsymbol{\phi})\big)}{\partial\phi^\alpha} = -\boldsymbol{\sigma} : \boldsymbol{\epsilon}^{\mathrm{R}} \frac{\partial h^{\mathrm{R}}}{\partial\phi^\alpha}$$

show that

$$\sum_\beta h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^\beta(\boldsymbol{\phi},\boldsymbol{\epsilon})}{\partial\phi^\alpha} = -\boldsymbol{\sigma} : \sum_{\beta=1}^N \boldsymbol{\epsilon}^\beta \frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_{\beta\neq\mathrm{R}} \Big(\big(\boldsymbol{\sigma}^\beta - \boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha} h^\beta.$$

This makes it possible to rewrite this contribution to the total driving force in a manner more similar to the one for the model with the common normals in Equation (7.171) as

$$\frac{\partial}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} f(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta \Big(f^\beta - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^\beta\Big)\frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_{\beta\neq \mathrm{R}} h^\beta\Big(\big(\boldsymbol{\sigma}^\beta-\boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}. \tag{7.178}$$

In contrast to the former model, there is no reason here for the terms in $\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}$, except of course for two-phase regions where both phases with $h^\beta > 0$ share the only common relevant normal and therefore the normal stress. This is considerably less problematic than the issues caused by $\boldsymbol{\sigma} \neq \frac{\partial f_{el}}{\partial\boldsymbol{\epsilon}}$ as the required expressions for these derivatives can be recovered in a **local** fashion after a differentiation of their defining system in Equation (7.176), leading to

$$\begin{aligned}
&\Big(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\big((1-h^\alpha)\boldsymbol{\mathcal{C}}^\alpha + h^\alpha\boldsymbol{\mathcal{C}}^{\mathrm{R}}\big)\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\big)\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\alpha}}{\partial\phi^\delta}\Big|_{\boldsymbol{n}^{\alpha\beta}} - \sum_{\beta\neq\mathrm{R},\alpha} h^\beta\Big(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\big(\boldsymbol{\mathcal{C}}^\alpha-\boldsymbol{\mathcal{C}}^{\mathrm{R}}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\delta}\Big|_{\boldsymbol{n}^{\alpha\beta}}\\
&= -\Big(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\big((1-\frac{\partial h^\alpha}{\partial\phi^\delta})\boldsymbol{\mathcal{C}}^\alpha + \frac{\partial h^\alpha}{\partial\phi^\delta}\boldsymbol{\mathcal{C}}^{\mathrm{R}}\big)\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\big)\Big)\cdot\boldsymbol{a}^{\mathrm{R}\alpha} + \sum_{\beta\neq\mathrm{R},\alpha}\frac{\partial h^\beta}{\partial\phi^\delta}\Big(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\big(\boldsymbol{\mathcal{C}}^\alpha-\boldsymbol{\mathcal{C}}^{\mathrm{R}}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\boldsymbol{a}^{\mathrm{R}\beta}.
\end{aligned}$$

Remark 158. A pleasant property of this system is that it is based on the same matrix as the original system (7.176) for the determination of the $\boldsymbol{a}^{\mathrm{R}\alpha}$ and differs only in the right-hand sides and one can thus reuse a previous factorization for solving for the $\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}$. Nevertheless, one needs to solve such a system for the derivative with respect to each of the phases.
A very pragmatic alternative is to instead simply neglect these additional contributions. While this has no real mathematical justification, it leads to a simpler driving force, which remains correct within the two-phase regions, while the errors in the multiphase regions are expected to be of a similar order as the ones already induced by the variational inconsistency due to the construction of $\boldsymbol{\sigma}$. ◇

As the calulation of the phase-specific strains relies on the $\boldsymbol{n}^{\mathrm{R}\alpha}$ only, there are no additional contributions to $\frac{\partial f_{el}}{\partial\phi^\alpha}$ and $\frac{\partial f_{el}}{\partial\nabla\phi^\alpha}$ due to the other normal vectors. For the $\boldsymbol{n}^{\mathrm{R}\alpha}$, one has

$$\begin{aligned}
\frac{\partial f_{el}}{\partial\boldsymbol{n}^{R\alpha}} &= \sum_\beta h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^\beta}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}} = \sum_\beta h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial\boldsymbol{\epsilon}^{\mathrm{R}}}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}} + \sum_{\beta\neq\mathrm{R}} h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}\big(\boldsymbol{a}^{\mathrm{R}\beta}\otimes\boldsymbol{n}^{\mathrm{R}\beta}\big)_S\\
&= -\sum_\beta h^\beta\boldsymbol{\sigma}^\beta : \sum_{\delta\neq\mathrm{R}} h^\delta\frac{\partial}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}\big(\boldsymbol{a}^{\mathrm{R}\delta}\otimes\boldsymbol{n}^{\mathrm{R}\delta}\big)_S + \sum_{\beta\neq\mathrm{R}} h^\beta\boldsymbol{\sigma}^\beta : \frac{\partial}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}\big(\boldsymbol{a}^{\mathrm{R}\beta}\otimes\boldsymbol{n}^{\mathrm{R}\beta}\big)_S.
\end{aligned}$$

Whereas the explicit dependence on the normal vectors only entails a single contribution for $\alpha=\delta$ resp. $\alpha=\delta$, the $\boldsymbol{a}^{\mathrm{R}\delta}$ depend on all $\boldsymbol{n}^{\mathrm{R}\alpha}$ simultaneously, such that one has

$$\begin{aligned}
\frac{\partial f_{el}}{\partial\boldsymbol{n}^{R\alpha}} &= -\sum_\beta h^\beta\boldsymbol{\sigma}^\beta\cdot\boldsymbol{a}^{\mathrm{R}\alpha}h^\alpha - \sum_\beta\sum_{\delta\neq\mathrm{R}} h^\delta h^\beta\big(\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n}^{\mathrm{R}\delta}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\delta}}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}\\
&\quad + h^\alpha\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{a}^{\mathrm{R}\alpha} + \sum_{\beta\neq\mathrm{R}} h^\beta\big(\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n}^{\mathrm{R}\beta}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}.
\end{aligned}$$

Using $\boldsymbol{\sigma} = \sum_\beta h^\beta\boldsymbol{\sigma}^\beta$ and exchanging the dummy-indices, this can be rewritten in a more compact form as

$$\frac{\partial f_{el}}{\partial\boldsymbol{n}^{R\alpha}} = h^\alpha\big(\boldsymbol{\sigma}^\alpha-\boldsymbol{\sigma}\big)\cdot\boldsymbol{a}^{\mathrm{R}\alpha} + \sum_{\beta\neq\mathrm{R}} h^\beta\Big(\big(\boldsymbol{\sigma}^\beta-\boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}. \tag{7.179}$$

There is again no reason for the second contribution to drop out. Unless one decides to neglect this contribution for simplicity, the required contributions through $\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{n}^{\mathrm{R}\alpha}}$ can be recovered, similarly to those due to $\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}$ through a linearization of the system (7.176).

The Model by Tschukin ([74])

A quite different extension of the two-phase models from Section 7.2.2 is proposed by Tschukin in [74]. In contrast to the work by Schneider, the model in [74] is a priori not based upon a prefixed expression of the strains and stresses in each phase but instead obtained through a generalization of the expression (7.151) of the elastic free energy density. He suggests replacing the first part in Equation (7.151) through a summation over the analogous expression over all phases and, in analogy to the treatment of the surface energy contributions through a summation over all phase-pairings in Section 6.1, replacing the last term through a summation over all two-phase pairings. More precisely, using the same definitions $\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} = \bar{\boldsymbol{\mathcal{S}}}^{\beta\alpha}_{\boldsymbol{nn}} = \left(\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}_{\boldsymbol{nn}}\right)^{\dagger}$ where $\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}_{\boldsymbol{nn}} = \mathcal{N} : \left(h^{\alpha}\boldsymbol{\mathcal{C}}^{\beta} + h^{\beta}\boldsymbol{\mathcal{C}}^{\alpha}\right) : \mathcal{N}$, he introduces the expression[97]

$$f_{el}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \frac{1}{2}\sum_{\alpha=1}^{N}\left(\boldsymbol{\Sigma}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) - \frac{1}{2}\sum_{\beta\neq\alpha}[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} : \bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} : [\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi})\right) h^{\alpha}(\boldsymbol{\phi}) \tag{7.180}$$

for the elastic free energy density. The average (effective) stress $\boldsymbol{\sigma}$ is then derived from Equation (7.180) in accordance to Equation (7.70) as $\boldsymbol{\sigma} := \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}$, which, together with

$$\frac{\partial}{\partial \boldsymbol{\epsilon}}[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} = \frac{\partial}{\partial \boldsymbol{\epsilon}}\left(\boldsymbol{\mathcal{C}}^{\beta} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\beta}\right) - \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)\right) = \boldsymbol{\mathcal{C}}^{\beta} - \boldsymbol{\mathcal{C}}^{\alpha}$$

leads to the expression

$$\boldsymbol{\sigma} := \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}} = \sum_{\alpha=1}^{N}\left(\boldsymbol{\Sigma}^{\alpha} - \frac{1}{2}\sum_{\beta\neq\alpha}\left(\boldsymbol{\mathcal{C}}^{\beta} - \boldsymbol{\mathcal{C}}^{\alpha}\right) : \bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} : [\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi})\right) h^{\alpha}(\boldsymbol{\phi}) \tag{7.181}$$

generalizing the one in Equation (7.149) from the two-phase case.

Remark 159. A clear advantages as compared to the model by [63], [61] is that it is obvious from equations (7.180) and (7.181) that there is no "preferential" or reference phase. In addition, the stress tensor will, by its very definition in Equation (7.181), satisfy the consistency condition (7.70). ◇

In constrast to the model by Schneider, Tschukin avoids the issue of defining phase-specific strains and stresses. In particular, the free energy density being based upon considering a sum over essentially "independent" two-phase interactions, it is not at all clear a priori how his model can be extended to situations (such as viscoelastic or plastic problems), where one would require a consistent (i.e. independent of the phase-pairing) definition of these phase-inherent quantities. Simply **defining** similar to the discussion in the two-phase case in Section 7.2.3 the jump vector $\boldsymbol{a}^{\alpha\beta}$ for a given phase-pairing through $\left(\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta}\right)_S = \bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} : \left(\boldsymbol{\Sigma}^{\alpha} - \boldsymbol{\Sigma}^{\beta}\right) = -\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} : [\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}$, or, in a simpler form similar to Lemma 9, through

$$\boldsymbol{a}^{\alpha\beta} = \left(\boldsymbol{n}^{\alpha\beta} \cdot \bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}\right)^{-1} \cdot \left(\left(\boldsymbol{\Sigma}^{\alpha} - \boldsymbol{\Sigma}^{\beta}\right) \cdot \boldsymbol{n}^{\alpha\beta}\right) = -\left(\boldsymbol{n}^{\alpha\beta} \cdot \bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}\right)^{-1} \cdot \left([\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} \cdot \boldsymbol{n}^{\alpha\beta}\right), \tag{7.182}$$

with $\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta} = h^{\beta}\boldsymbol{\mathcal{C}}^{\alpha} + h^{\alpha}\boldsymbol{\mathcal{C}}^{\beta}$, it turns out that it is in fact possible to obtain the stress-strain relationship above based upon consistently defined phase-specific stresses $\boldsymbol{\sigma}^{\alpha}$ and strains $\boldsymbol{\epsilon}^{\alpha}$ given by

$$\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon} - \sum_{\beta\neq\alpha} h^{\beta}\left(\boldsymbol{a}^{\alpha\beta} \otimes \boldsymbol{n}^{\alpha\beta}\right)_S \quad \text{and} \quad \boldsymbol{\sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right). \tag{7.183}$$

[97] Note that the additional factor $\frac{1}{2}$ in front of the last term is due to each phase-pairing appearing twice in this generalization and can be eliminated by including each pairing only once by e.g. instead summing over all $1 \leq \alpha < \beta \leq N$.

In fact, using $\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} = \bar{\boldsymbol{\mathcal{S}}}^{\beta\alpha}_{\boldsymbol{nn}}$ and $[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} = -[\![\boldsymbol{\Sigma}]\!]^{\beta\alpha}$, the second term in Equation (7.181) can be rewritten as

$$\frac{1}{2}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\mathcal{C}}^{\alpha}:\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\beta}(\boldsymbol{\phi})h^{\alpha}(\boldsymbol{\phi}) - \frac{1}{2}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\mathcal{C}}^{\beta}:\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\beta}(\boldsymbol{\phi})h^{\alpha}(\boldsymbol{\phi})$$

$$=\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\mathcal{C}}^{\alpha}:\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\beta}(\boldsymbol{\phi})h^{\alpha}(\boldsymbol{\phi}),$$

thus leading to the alternative expression

$$\boldsymbol{\sigma} := \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}} = \sum_{\alpha=1}^{N}\left(\boldsymbol{\Sigma}^{\alpha} + \boldsymbol{\mathcal{C}}^{\alpha}:\Big(\sum_{\beta\neq\alpha}\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\beta}(\boldsymbol{\phi})\Big)\right)h^{\alpha}(\boldsymbol{\phi}). \tag{7.184}$$

for the effective stress. As $\boldsymbol{\Sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha}:\big(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\alpha}\big)$, this is the same as

$$\boldsymbol{\sigma} = \sum_{\alpha=1}^{N}\boldsymbol{\mathcal{C}}^{\alpha}:\left(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\alpha}+\Big(\sum_{\beta\neq\alpha}\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\beta}(\boldsymbol{\phi})\Big)\right)h^{\alpha}(\boldsymbol{\phi})$$

$$=\sum_{\alpha=1}^{N}\boldsymbol{\mathcal{C}}^{\alpha}:\Big(\boldsymbol{\epsilon}-\sum_{\beta\neq\alpha}h^{\beta}\big(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\big)_{S}-\tilde{\boldsymbol{\epsilon}}^{\alpha}\Big)h^{\alpha}(\boldsymbol{\phi}),$$

and thus a weighted average of the phase-specific stresses defined as in Equation (7.183). Furhtermore, $\boldsymbol{\epsilon}$ and the $\boldsymbol{\epsilon}^{\alpha}$ also satisfy the averaging condition $\boldsymbol{\epsilon} = \sum_{\alpha}\boldsymbol{\epsilon}^{\alpha}h^{\alpha}(\boldsymbol{\phi})$, since

$$\sum_{\alpha=1}^{N}\boldsymbol{\epsilon}^{\alpha}h^{\alpha}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N}\boldsymbol{\epsilon}h^{\alpha}(\boldsymbol{\phi}) - \sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\big(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\big)_{S}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

and the last term vanishes due to

$$-\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\big(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\big)_{S}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi}) = \sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:\big(\boldsymbol{\Sigma}^{\beta}-\boldsymbol{\Sigma}^{\alpha}\big)h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

$$=\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:\boldsymbol{\Sigma}^{\beta}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi}) - \sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:\boldsymbol{\Sigma}^{\alpha}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

and thus

$$\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\big(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\big)_{S}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi}) = \boldsymbol{0}.$$

The model in [74] therefore avoids two of the major drawbacks of the model in [63], namely the dependence on a reference phase and the inconsistency between $\boldsymbol{\sigma}$ and $\frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}$. Unfortunately, it has other issues within multiphase regions. Firstly, since the calculation of the jump-vectors $\boldsymbol{a}^{\alpha\beta}$ is such that it would only enforce the normal continuity of the stresses if α and β where the only phases present, there will generally be no phase-pairing for which the phase-specific stresses defined in Equation (7.183) will actually satisfy this continuity condition in multiphase regions. Secondly, and this is probably the more problematic point, the definition of f_{el} in Equation (7.180) is much less natural than it may seem at first sight. More precisely, with

$$-\frac{1}{4}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\big(\boldsymbol{\Sigma}^{\beta}-\boldsymbol{\Sigma}^{\alpha}\big):\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

$$=\frac{1}{4}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\Sigma}^{\alpha}:\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi}) + \frac{1}{4}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\Sigma}^{\beta}:\bar{\boldsymbol{\mathcal{S}}}^{\beta\alpha}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\beta\alpha}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

$$=\frac{1}{2}\sum_{\alpha=1}^{N}\sum_{\beta\neq\alpha}\boldsymbol{\Sigma}^{\alpha}:\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}:[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}h^{\alpha}(\boldsymbol{\phi})h^{\beta}(\boldsymbol{\phi})$$

f_{el} can be written as

$$f_{el}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \frac{1}{2}\sum_{\alpha=1}^{N}\left(\boldsymbol{\Sigma}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) + \boldsymbol{\Sigma}^{\alpha} : \Big(\sum_{\beta\neq\alpha} \bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}} : [\![\boldsymbol{\Sigma}]\!]^{\alpha\beta} h^{\beta}(\boldsymbol{\phi})\Big)\right) h^{\alpha}(\boldsymbol{\phi}).$$

Together the definition of the $\boldsymbol{a}^{\alpha\beta}$ and the $\boldsymbol{\epsilon}^{\alpha}$ in Equation (7.183), this further simplifies to

$$f_{el}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}, \boldsymbol{\epsilon}) = \frac{1}{2}\sum_{\alpha=1}^{N} \boldsymbol{\Sigma}^{\alpha}(\boldsymbol{\epsilon}) : \left(\boldsymbol{\epsilon}^{\alpha}(\boldsymbol{\phi}, \nabla\boldsymbol{\phi}, \boldsymbol{\epsilon}) - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) h^{\alpha}(\boldsymbol{\phi}), \tag{7.185}$$

and thus a free energy density where one half of the expression is based on the supposedly bad approximation $\boldsymbol{\Sigma}^{\alpha} = \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right)$ instead of $\boldsymbol{\sigma}^{\alpha}$ corresponding to the "correct" strains $\boldsymbol{\epsilon}^{\alpha}$.
One particularly problematic point with this definition is that, whereas the weighted average $\frac{1}{2}\sum_{\alpha}\left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) : \boldsymbol{\mathcal{C}}^{\alpha} : \left(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}\right) h^{\alpha}(\boldsymbol{\phi})$ is necessarily non-negative regardless of the precise choice of the $\boldsymbol{\epsilon}^{\alpha}$, this need not be the case for the expression in Equation (7.185), and it can in fact be observed numerically that defining f_{el} as in [74] may indeed lead to locally negative free energy densities.

The last expression in Equation (7.185) also allows for a somewhat simpler derivation of the correct driving force contributions (in [74] based on the $\boldsymbol{n}^{\alpha\beta}$ defined in terms of the gradients of $\boldsymbol{\phi}$ only)

$$\frac{\partial f_{el}}{\partial \phi^{\alpha}} = \frac{1}{2}\sum_{\beta=1}^{N}\left(\boldsymbol{\Sigma}^{\beta} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\beta}\right) - \sum_{\delta\neq\beta}\left(\boldsymbol{a}^{\beta\delta}\otimes\boldsymbol{n}^{\beta\delta}\right)_S : \boldsymbol{\mathcal{C}}^{\beta} : \left(\boldsymbol{a}^{\beta\delta}\otimes\boldsymbol{n}^{\beta\delta}\right)_S \left(h^{\delta}\right)^2\right)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} \tag{7.186}$$

as in [74], which, while correct, requires a somewhat heavy formalism due to the use of the pseudo-inverses in Equation (7.180). A partial differentiation with respect to ϕ^{α} while holding the normal vectors fixed first leads to

$$\frac{\partial f_{el}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}} = \frac{1}{2}\sum_{\beta}\boldsymbol{\Sigma}^{\beta} : \left(\boldsymbol{\epsilon}^{\beta} - \tilde{\boldsymbol{\epsilon}}^{\beta}\right)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} + \frac{1}{2}\sum_{\beta} h^{\beta}\boldsymbol{\Sigma}^{\beta} : \frac{\partial \boldsymbol{\epsilon}^{\beta}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}},$$

where, by Equation (7.183),

$$\frac{\partial \boldsymbol{\epsilon}^{\beta}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}} = -\sum_{\delta\neq\beta}\left(\boldsymbol{a}^{\beta\delta}\otimes\boldsymbol{n}^{\beta\delta}\right)_S \frac{\partial h^{\delta}}{\partial \phi^{\alpha}} - \sum_{\delta\neq\beta} h^{\delta}\Big(\frac{\partial \boldsymbol{a}^{\beta\delta}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}}\otimes\boldsymbol{n}^{\beta\delta}\Big)_S.$$

By a partial differentiation of the defining Equation (7.182) for the jump vectors and the definition of $\boldsymbol{\mathcal{C}}^{\beta\delta}$, it is easily seen that

$$\frac{\partial \boldsymbol{a}^{\beta\delta}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}} = -\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\Big(\boldsymbol{n}^{\beta\delta}\cdot\Big(\frac{\partial h^{\delta}}{\partial \phi^{\alpha}}\boldsymbol{\mathcal{C}}^{\beta} + \frac{\partial h^{\beta}}{\partial \phi^{\alpha}}\boldsymbol{\mathcal{C}}^{\delta}\Big)\cdot\boldsymbol{n}^{\beta\delta}\Big)\cdot\boldsymbol{a}^{\beta\delta}.$$

Further expanding the term $\boldsymbol{\Sigma}^{\beta} : \left(\boldsymbol{\epsilon}^{\beta} - \tilde{\boldsymbol{\epsilon}}\right)$ using $\boldsymbol{\epsilon}^{\beta} = \boldsymbol{\epsilon} - \sum_{\delta\neq\beta}\left(\boldsymbol{a}^{\beta\delta}\otimes\boldsymbol{n}^{\beta\delta}\right)_S h^{\delta}$, it follows that

$$\begin{aligned}
\frac{\partial f_{el}}{\partial \phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}} =& \frac{1}{2}\sum_{\beta}\boldsymbol{\Sigma}^{\beta} : \left(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\beta}\right)\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} \\
&- \frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}h^{\delta}\frac{\partial h^{\beta}}{\partial \phi^{\alpha}} - \frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial \phi^{\alpha}} \\
&+ \frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\Big(\boldsymbol{n}^{\beta\delta}\cdot\Big(\frac{\partial h^{\delta}}{\partial \phi^{\alpha}}\boldsymbol{\mathcal{C}}^{\beta} + \frac{\partial h^{\beta}}{\partial \phi^{\alpha}}\boldsymbol{\mathcal{C}}^{\delta}\Big)\cdot\boldsymbol{n}^{\beta\delta}\Big)\cdot\boldsymbol{a}^{\beta\delta}
\end{aligned}$$

Rewriting the last line as

$$
\begin{aligned}
&\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial\phi^{\alpha}}\\
+&\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}\\
=&\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\left(h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}+h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}-h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}\right)\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial\phi^{\alpha}}\\
+&\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\left(h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}+h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}-h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}\right)\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}\\
=&\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial\phi^{\alpha}}+\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}\\
&-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial\phi^{\alpha}}\\
&-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}
\end{aligned}
$$

allows canceling the middle row, leaving

$$
\begin{aligned}
\frac{\partial f_{el}}{\partial\phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}}=&\frac{1}{2}\sum_{\beta}\boldsymbol{\Sigma}^{\beta}:\left(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\beta}\right)\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}\\
&-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\beta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\beta}\boldsymbol{\mathcal{C}}^{\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\delta}}{\partial\phi^{\alpha}}\\
&-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\delta}\left(\boldsymbol{\Sigma}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}.
\end{aligned}
$$

Finally, exchanging dummy-indices and making use of $\boldsymbol{n}^{\beta\delta}=-\boldsymbol{n}^{\delta\beta}$ and $\boldsymbol{a}^{\beta\delta}=\boldsymbol{a}^{\delta\beta}$, the last two rows can further be summarized to

$$
-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta} h^{\delta}\Big(\left(\boldsymbol{\Sigma}^{\beta}-\boldsymbol{\Sigma}^{\delta}\right)\cdot\boldsymbol{n}^{\beta\delta}\Big)\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot h^{\delta}\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}},
$$

which, since $\boldsymbol{a}^{\beta\delta}=\left(\boldsymbol{n}^{\beta\delta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\beta\delta}\cdot\boldsymbol{n}^{\beta\delta}\right)^{-1}\cdot\Big(\left(\boldsymbol{\Sigma}^{\beta}-\boldsymbol{\Sigma}^{\delta}\right)\cdot\boldsymbol{n}^{\beta\delta}\Big)$, leads to the relatively simple final expresssion

$$
\frac{\partial f_{el}}{\partial\phi^{\alpha}}\Big|_{\boldsymbol{n}^{\alpha\beta}}=\frac{1}{2}\sum_{\beta}\boldsymbol{\Sigma}^{\beta}:\left(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\beta}\right)\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}-\frac{1}{2}\sum_{\beta}\sum_{\delta\neq\beta}\left(h^{\delta}\right)^{2}\boldsymbol{a}^{\beta\delta}\cdot\left(\boldsymbol{n}^{\beta\delta}\cdot\boldsymbol{\mathcal{C}}^{\beta}\cdot\boldsymbol{n}^{\beta\delta}\right)\cdot\boldsymbol{a}^{\beta\delta}\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}
$$

which amounts to the same as the one in Equation (7.186) by the subsymmetries of $\boldsymbol{\mathcal{C}}^{\beta}$.
For the remaining contributions arising from the dependence on the $\boldsymbol{n}^{\alpha\beta}$, it is convenient to explicitly make use of $\boldsymbol{n}^{\alpha\beta}=-\boldsymbol{n}^{\beta\alpha}$ in order to reduce f_{el} to a function of e.g. the $(\boldsymbol{n}^{\alpha\beta})_{\alpha<\beta}$ only. Doing so and with $\boldsymbol{a}^{\alpha\beta}=\boldsymbol{a}^{\beta\alpha}$, it follows that

$$
\begin{aligned}
\frac{\partial f_{el}}{\partial\boldsymbol{n}^{\alpha\beta}}=&\frac{1}{2}\sum_{\delta}h^{\delta}\boldsymbol{\Sigma}^{\delta}:\frac{\partial\boldsymbol{\epsilon}^{\delta}}{\partial\boldsymbol{n}^{\alpha\beta}}=\frac{1}{2}\sum_{\delta}h^{\delta}\boldsymbol{\Sigma}^{\delta}:\frac{\partial}{\partial\boldsymbol{n}^{\alpha\beta}}\Big(\boldsymbol{\epsilon}-\sum_{\eta\neq\delta}h^{\eta}\left(\boldsymbol{a}^{\delta\eta}\otimes\boldsymbol{n}^{\delta\eta}\right)_{S}\Big)\\
=&-\frac{1}{2}h^{\alpha}h^{\beta}\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta}-\frac{1}{2}h^{\alpha}h^{\beta}\Big(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\boldsymbol{n}^{\alpha\beta}\Big)\cdot\frac{\partial\boldsymbol{a}^{\alpha\beta}}{\partial\boldsymbol{n}^{\alpha\beta}}
\end{aligned}
$$

Since $\mathrm{d}\boldsymbol{a}^{\alpha\beta}$ is, for fixed $\boldsymbol{\phi}$, characterized in terms of $\delta\boldsymbol{n}^{\alpha\beta}$ as

$$\left(\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\mathrm{d}\boldsymbol{a}^{\alpha\beta}=\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}-\left(\mathrm{d}\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta}-\left(\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta},$$

one has

$$\begin{aligned}\left(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\mathrm{d}\boldsymbol{a}^{\alpha\beta}=&\left(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\left(\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{n}^{\alpha\beta}\right)^{-1}\\&\cdot\left(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}-\left(\mathrm{d}\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta}-\left(\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta}\right)\\=&\boldsymbol{a}^{\alpha\beta}\cdot\left(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}-2\left(\left(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\right)_{S}:\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{a}^{\alpha\beta}\right)\cdot\mathrm{d}\boldsymbol{n}^{\alpha\beta}\right),\end{aligned}$$

where use was made of the symmetries of $\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}$. Combining this with the previous expression for $\frac{\partial f_{el}}{\partial\boldsymbol{n}^{\alpha\beta}}$, one therefore has

$$\frac{\partial f_{el}}{\partial\boldsymbol{n}^{\alpha\beta}}=-h^{\alpha}h^{\beta}\left(\left(\boldsymbol{\Sigma}^{\alpha}-\boldsymbol{\Sigma}^{\beta}\right)\cdot\boldsymbol{a}^{\alpha\beta}-\left(\boldsymbol{a}^{\alpha\beta}\otimes\boldsymbol{n}^{\alpha\beta}\right)_{S}:\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}\cdot\boldsymbol{a}^{\alpha\beta}\right),\tag{7.187}$$

which is a slightly simpler version of expressions in [74] since it avoids the use of the derivative of the projector $\mathcal{N}$ onto the normal subspace of symmetric second-order tensors.

Remark 160. Summarizing the discussion of the three models considered here, each has its own advantages and disadvantages. Whereas the model in [62] has the mathematically most pleasant properties and in particular satisfies the desired jump conditions between all phase-pairings, this can only be achieved through a very significant geometrical simplification of using a common normal vector for all phase-pairings. The models in [63] and [74] do not make this assumption, but are therefore not generally able to satisfy all jump conditions simultaneously except for in the two-phase regions.

Whereas the model in [63] has the advantage of enforcing the jump conditions on both the strains and stresses at least with respect to the reference phase R, the use of such a preferential phase is clearly an undesirable feature. In constrast, the model in [74] does not rely on a reference phase, but will generally not satisfy any jump conditions on either the stresses or strains within multiphase regions.

From an energetic point of view, both models have some deficiencies when more than two phases are present at a given point. Whereas the small-deformation analogon of the model in [63] is seemingly based on the very natural definition f_{el} as a weighted average of the elastic free energy densities $f_{el}^{\alpha}(\boldsymbol{\epsilon}^{\alpha})$ evaluated in the respective phase-specific strain and thus in particular satisfies $\boldsymbol{\sigma}^{\alpha}=\frac{\partial f_{el}^{\alpha}}{\partial\boldsymbol{\epsilon}^{\alpha}}$, it does generally not satisfy $\boldsymbol{\sigma}=\frac{\partial f_{el}}{\partial\boldsymbol{\epsilon}}$. In contrast, the effective stress in [74] satisfies this relation by construction, but in the form in Equation (7.180) does not allow for such an easy interpretation in terms of phase-specific free energy densities. Even though it was seen in Equation (7.185) that one can artificially rewrite f_{el} in such a form, the resulting expression is quite unsatisfactory as it is based on a mixture of a Voigt-Taylor-type stress prediction and a more quantitative prediction of the strains and does in particular not satisfy $\boldsymbol{\sigma}^{\alpha}=\frac{\partial f_{el}^{\alpha}}{\partial\boldsymbol{\epsilon}^{\alpha}}$.

This should not be misinterpreted as saying that these are bad models, as both achieve a fairly reasonable extension of the very satisfactory two-phase model to within the highly difficult multiphase. Nevertheless, their drawbacks have to be kept in mind when choosing between them for a particular application, and will be inhereted from any extension to more complex physical situtations. ◇

7.2.5 Chemo-Elasticity

A more complex situation arises when there is a coupling between the both chemical and elastic influences and the evolution of a microstructure. Such a situation can arise in a diffusion process under a simultaneous elastic deformations if the elastic free energy density depends not only upon the total strain and a phasedependent eigenstrain, but in addition on the concentration itself, for example if the eigenstrains and/or the stiffness tensors have an additional dependence upon the concentration vector. In the bulk phases, this leads to a total free energy densities depending both on the "purely" chemical part $f_{ch}^{\alpha}(\boldsymbol{c}^{\alpha})$ as well as an additional elastic contribution $f_{el}^{\alpha}(\boldsymbol{c}^{\alpha}, \boldsymbol{\epsilon}^{\alpha})$, i.e.

$$f^{\alpha}(\boldsymbol{c}^{\alpha}, \boldsymbol{\epsilon}^{\alpha}) = f_{ch}^{\alpha}(\boldsymbol{c}^{\alpha}) + f_{el}^{\alpha}(\boldsymbol{\epsilon}^{\alpha}, \boldsymbol{c}^{\alpha}) = f_{ch}^{\alpha}(\boldsymbol{c}^{\alpha}) + \frac{1}{2}\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \boldsymbol{C}^{\alpha}(\boldsymbol{c}^{\alpha}) : \big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{c}^{\alpha})\big). \quad (7.188)$$

For constructing a phasefield model based on these free energy densities, one again has to choose how this definition should be extended to within the diffuse interface region, where several phases coexist. Assuming as before that one wishes to formulate the model purely in terms of the average concentration $\boldsymbol{c}$ and strain $\boldsymbol{\epsilon}$, assumed to satisfy

$$\boldsymbol{c} = \sum_{\alpha=1}^{N} \boldsymbol{c}^{\alpha} h^{\alpha}(\boldsymbol{\phi}) \quad \text{and } \boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^{\alpha} h^{\alpha}(\boldsymbol{\phi}), \quad (7.189)$$

an interpolation of the phase-specific energies $f^{\alpha}(\boldsymbol{c}^{\alpha}, \boldsymbol{\epsilon}^{\alpha})$ in Equation (7.188) of the form

$$f(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}) = \sum_{\alpha=1}^{N} f^{\alpha}(\boldsymbol{c}^{\alpha}, \boldsymbol{\epsilon}^{\alpha}) h^{\alpha}(\boldsymbol{\phi}) \quad (7.190)$$

requires an appropriate additional set of conditions for the determination of the phase-specific quantities $(\boldsymbol{c}^{\alpha}, \boldsymbol{\epsilon}^{\alpha})_{1 \le \alpha \le N}$ in terms of $\boldsymbol{\phi}, \boldsymbol{c}$ and $\boldsymbol{\epsilon}$.

Probably the simplest approach is to combine the assumption $\boldsymbol{c}^{\alpha} = \boldsymbol{c}, \alpha = 1, ..., N$, with one of the classical mechanical models, i.e. the Voigt-Talyor or Reuss-Sachs appraoch. In the former case, the assumption of the equality of all strains $\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\epsilon}, \alpha = 1, ..., N$ leads to the total free energy density

$$f(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}) = f_{ch}(\boldsymbol{\phi}, \boldsymbol{c}) + \sum_{\alpha=1}^{N} \Big(\frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) : \boldsymbol{C}^{\alpha}(\boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) \Big) h^{\alpha}(\boldsymbol{\phi}).$$

The major advantage of this model lies in its particularly simple form, as, assuming the individual parts are explicit in $\boldsymbol{c}$, it results in the formula

$$\boldsymbol{\sigma} = \frac{\partial f}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}) = \sum_{\alpha=1}^{N} \Big(\boldsymbol{C}^{\alpha}(\boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) \Big) h^{\alpha}(\boldsymbol{\phi})$$

for the average stress and

$$\boldsymbol{\mu} := \frac{\partial f}{\partial \boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}) = \underbrace{\frac{\partial f_{ch}}{\partial \boldsymbol{c}}(\boldsymbol{\phi}, \boldsymbol{c})}_{=:\boldsymbol{\mu}_{ch}(\boldsymbol{c})} + \underbrace{\sum_{\alpha=1}^{N} \left(-\Big(\boldsymbol{C}^{\alpha}(\boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) \Big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial \boldsymbol{c}} + \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) : \Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})\big) : \frac{\partial \boldsymbol{C}^{\alpha}}{\partial \boldsymbol{c}} \Big) \right) h^{\alpha}(\boldsymbol{\phi})}_{=:\boldsymbol{\mu}_{el}(\boldsymbol{c}, \boldsymbol{\epsilon})}$$

for the chemical potential, both of which are "fully explicit" in $\boldsymbol{c}$ and $\boldsymbol{\epsilon}$.
In the latter case, the situation is - similarly to before - slightly more complex due to the appearence of the derivative of an inverse matrix. As for Equation (7.82), the assumption of equal stresses $\boldsymbol{\sigma}^{\alpha} = \boldsymbol{\sigma}\ \forall \alpha$ leads to the total free energy density

$$f(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}) = f_{ch}(\boldsymbol{\phi}, \boldsymbol{c}) + \frac{1}{2}\Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi}, \boldsymbol{c})\big) : \boldsymbol{C}_{RS}(\boldsymbol{\phi}, \boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{c})\big) \Big), \quad \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi}, \boldsymbol{c}) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}) h^{\alpha}(\boldsymbol{\phi})$$

with $\boldsymbol{C}_{RS}(\boldsymbol{\phi},\boldsymbol{c}) = \Big(\sum_{\alpha=1}^{N} \boldsymbol{S}^{\alpha}(\boldsymbol{c})h^{\alpha}(\boldsymbol{\phi})\Big)^{-1}$. From this, the stress calculation is in fact even slightly simpler than for the Voigt-Talyer model as it now only involves the averaged eigenstrain $\tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi},\boldsymbol{c}) = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c})h^{\alpha}(\boldsymbol{\phi})$,

$$\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) = \boldsymbol{C}_{RS}(\boldsymbol{\phi},\boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi},\boldsymbol{c})\big).$$

The definition of $\boldsymbol{C}_{RS}(\boldsymbol{\phi},\boldsymbol{c})$ in terms of the inverse of the interpolated compliances gives rise to the same minor complication already encountered in the calculation of the drivingforce for the Reuss-Sachs model as

$$\boldsymbol{\mu}_{el}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) = \frac{\partial f_{el}}{\partial \boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) = -\boldsymbol{C}_{RS}(\boldsymbol{\phi},\boldsymbol{c}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{c})\big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}}{\partial \boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c}) + \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{\phi},\boldsymbol{c})\big) : \left(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}(\boldsymbol{c})\big) : \frac{\partial \boldsymbol{C}_{RS}(\boldsymbol{\phi},\boldsymbol{c})}{\partial \boldsymbol{c}}\right).$$

By the same argument as for Equation (7.84) with $\boldsymbol{c}$ taking the role of $\boldsymbol{\phi}$, this can again be replaced by a fully explicit expression in terms of the stresses and the derivatives of the $\boldsymbol{S}^{\alpha}(\boldsymbol{c})$ leading to

$$\boldsymbol{\mu}_{el}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) = -\boldsymbol{\sigma}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) : \left(\frac{\partial \tilde{\boldsymbol{\epsilon}}}{\partial \boldsymbol{c}}(\boldsymbol{\phi},\boldsymbol{c}) + \frac{1}{2}\boldsymbol{\sigma}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) : \Big(\sum_{\alpha=1}^{N} \frac{\partial \boldsymbol{S}^{\alpha}(\boldsymbol{c})}{\partial \boldsymbol{c}} h^{\alpha}(\boldsymbol{\phi})\Big)\right).$$

It is clear that, while this model is simple, it will inherit the same drawbacks already discussed in the chemical and mechanical case above. As a natural extension of the discussion in Section 7.1, one could instead consider a "grandchem"-type approach with the phase-specific concentrations $\boldsymbol{c}^{\alpha}$ fixed through equality chemical potentials, i.e.

$$f(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon}) = \sum_{\alpha=1}^{N} \left(f_{ch}^{\alpha}(\boldsymbol{c}^{\alpha}) + \frac{1}{2}\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \boldsymbol{C}^{\alpha}(\boldsymbol{c}^{\alpha}) : \big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big)\right) h^{\alpha}(\boldsymbol{\phi})$$

with $(\boldsymbol{c}^{\alpha})_{1\leq\alpha\leq N}$ fixed by either the equality up to a phase-specific multiplier $\lambda^{\alpha}\boldsymbol{e}$ of the chemical potentials

$$\boldsymbol{\mu}^{\alpha} = \frac{\partial f_{ch}^{\alpha}}{\partial \boldsymbol{c}^{\alpha}} - \Big(\boldsymbol{C}^{\alpha}(\boldsymbol{c}^{\alpha}) : \big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big)\Big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial \boldsymbol{c}^{\alpha}} + \frac{1}{2}\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \Big(\big(\boldsymbol{\epsilon}^{\alpha} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \frac{\partial \boldsymbol{C}^{\alpha}}{\partial \boldsymbol{c}^{\alpha}}\Big) \quad (7.191)$$

or the equality of the $\tilde{\boldsymbol{\mu}}^{\alpha}$ when using a reduced formulation, where the first purely chemical distribution will be designated by $\boldsymbol{\mu}_{ch}^{\alpha}$ and the second one by $\boldsymbol{\mu}_{el}^{\alpha}$. The complexity of such a model depends heavily upon the particular mechanical model this is combined with.

The Voigt-Taylor Case

The simplest situation arises if the $\boldsymbol{\epsilon}^{\alpha}$ are, as in the Voigt-Taylor model, all assumed to the be equal to the average strain $\boldsymbol{\epsilon}$. Even though the determination of the mechanical stress does then require the knowledge of the $\tilde{\boldsymbol{c}}^{\alpha}$, the only unknown parameter for the evaluation of the chemical potentials are the $\boldsymbol{c}^{\alpha}$,

$$\boldsymbol{\mu}^{\alpha} = \boldsymbol{\mu}_{ch}^{\alpha} - \Big(\boldsymbol{C}^{\alpha}(\boldsymbol{c}^{\alpha}) : \big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big)\Big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial \boldsymbol{c}^{\alpha}} + \frac{1}{2}\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \Big(\big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big) : \frac{\partial \boldsymbol{C}^{\alpha}}{\partial \boldsymbol{c}^{\alpha}}\Big).$$

One can therefore proceed in a two-step fashion by first solving for the $\tilde{\boldsymbol{c}}^{\alpha}$ as in Section 7.1, the only difference being that the expressions for the phase-specific chemical potentials become somewhat more complex due to the additional - but known as a function of $\boldsymbol{c}^{\alpha}$ - elastic contributions.

More precisely, with

$$\begin{aligned}\frac{\partial(\boldsymbol{\mu}_{el}^{\alpha}\cdot\boldsymbol{\zeta}^{\alpha})}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha} =& \frac{\partial}{\partial\boldsymbol{c}^{\alpha}}\Big(-\boldsymbol{\sigma}^{\alpha}(\boldsymbol{c}^{\alpha},\boldsymbol{\epsilon}_{el}^{\alpha}):\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\big)+\frac{1}{2}\Big(\boldsymbol{\epsilon}_{el}^{\alpha}:\big(\frac{\partial\boldsymbol{\mathcal{C}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\big):\boldsymbol{\epsilon}_{el}^{\alpha}\Big)\Big)\cdot\boldsymbol{\chi}^{\alpha}\\ =&-\boldsymbol{\sigma}^{\alpha}:\big(\frac{\partial^2\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial(\boldsymbol{c}^{\alpha})^2}:(\boldsymbol{\zeta}^{\alpha}\otimes\boldsymbol{\chi}^{\alpha})\big)-\big(\frac{\partial\boldsymbol{\sigma}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha}\big)\cdot\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\\ &+\frac{1}{2}\boldsymbol{\epsilon}_{el}^{\alpha}:\big(\frac{\partial^2\boldsymbol{\mathcal{C}}^{\alpha}}{\partial(\boldsymbol{c}^{\alpha})^2}:(\boldsymbol{\zeta}^{\alpha}\otimes\boldsymbol{\chi}^{\alpha})\big):\boldsymbol{\epsilon}_{el}^{\alpha}-\Big(\big(\frac{\partial\boldsymbol{\mathcal{C}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\big):\boldsymbol{\epsilon}_{el}^{\alpha}\Big):\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha}\big),\end{aligned}$$

or, using

$$\frac{\partial\boldsymbol{\sigma}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\mathrm{d}\boldsymbol{c}^{\alpha}=-\boldsymbol{\mathcal{C}}^{\alpha}:\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\mathrm{d}\boldsymbol{c}^{\alpha}\big)+\Big(\frac{\partial\boldsymbol{\mathcal{C}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\mathrm{d}\boldsymbol{c}^{\alpha}\Big):\boldsymbol{\epsilon}_{el}^{\alpha} \tag{7.192}$$

for replacing the derivative of the stress w.r.t. $\boldsymbol{c}^{\alpha}$, one obtains the formula

$$\begin{aligned}\frac{\partial^2 f_{el}^{\alpha}}{\partial(\boldsymbol{c}^{\alpha})^2}:(\boldsymbol{\zeta}^{\alpha}\otimes\boldsymbol{\chi}^{\alpha}) =& \frac{1}{2}\boldsymbol{\epsilon}_{el}^{\alpha}:\big(\frac{\partial^2\boldsymbol{\mathcal{C}}^{\alpha}}{\partial(\boldsymbol{c}^{\alpha})^2}:(\boldsymbol{\zeta}^{\alpha}\otimes\boldsymbol{\chi}^{\alpha})\big):\boldsymbol{\epsilon}_{el}^{\alpha}-\boldsymbol{\sigma}^{\alpha}:\big(\frac{\partial^2\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial(\boldsymbol{c}^{\alpha})^2}:(\boldsymbol{\zeta}^{\alpha}\otimes\boldsymbol{\chi}^{\alpha})\big)\\ &-\Big(\big(\frac{\partial\boldsymbol{\mathcal{C}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha}\big):\boldsymbol{\epsilon}_{el}^{\alpha}\Big):\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}\big)-\Big(\big(\frac{\partial\boldsymbol{\mathcal{C}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\big):\boldsymbol{\epsilon}_{el}^{\alpha}\Big):\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha}\big)\\ &+\big(\frac{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\chi}^{\alpha}\big):\boldsymbol{\mathcal{C}}^{\alpha}:\big(\frac{\partial\boldsymbol{\epsilon}^{\alpha}}{\partial\boldsymbol{c}^{\alpha}}\cdot\boldsymbol{\zeta}^{\alpha}\big)\end{aligned} \tag{7.193}$$

for characterizing the second derivative of the mechanical contribution to the free energy density with respect to the phase-specific concentration.
Once the $\boldsymbol{c}^{\alpha}$ have been determined, the evaluation of the stress $\boldsymbol{\sigma}$ is then straightforward.

Remark 161. This can nevertheless exacerbate the issues already discussed in Remark 104 when considering the $\boldsymbol{c}^{\alpha}$ as functions of the chemical potential, since any additional contribution further decreases the likelyhood of being able to perform the conversion from $(\boldsymbol{\mu},\boldsymbol{\epsilon})$ to $\boldsymbol{c}^{\alpha}$ explicitly. ◇

In terms of the driving force $\frac{\partial f}{\partial\phi^{\alpha}}$ for the phasefield equation, it is clear that, due to the choice $\boldsymbol{\epsilon}^{\beta}=\boldsymbol{\epsilon}$, this approach entails very little changes with respect to the previous considerations and one has

$$\frac{\partial f}{\partial\phi^{\alpha}}=\sum_{\beta=1}^{N}\Big(f_{ch}^{\beta}(\boldsymbol{c}^{\alpha})+\frac{1}{2}(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\beta}):\boldsymbol{\mathcal{C}}^{\beta}:(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\beta})-\boldsymbol{\mu}\cdot\boldsymbol{c}^{\beta}\Big)\frac{\partial h^{\beta}}{\partial\phi^{\alpha}}.$$

The Reuss-Sachs Case

In contrast (unless is no dependence on $\boldsymbol{c}^{\alpha}$ in f_{el}^{α}), using a Reuss-Sachs-type approach already entails some notable complications, since the defining requirement $\boldsymbol{\sigma}^{\alpha}=\boldsymbol{\sigma}$ for some $\boldsymbol{\sigma}$ on the phase-specific stresses $\boldsymbol{\sigma}^{\alpha}$ is now fixed in terms of system depending on **two** primal unknowns $(\boldsymbol{\epsilon}^{\alpha},\boldsymbol{c}^{\alpha})$ and one dual one in terms of $\boldsymbol{\sigma}$,

$$\boldsymbol{\sigma}^{\alpha}=\boldsymbol{C}^{\alpha}(\boldsymbol{c}^{\alpha}):\big(\boldsymbol{\epsilon}^{\alpha}-\tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\big)\overset{!}{=}\boldsymbol{\sigma}. \tag{7.194}$$

One can therefore not proceed in a sequential fashion as in the Voigt-Taylor case and instead has to consider the full coupled chemo-mechanical system simultaneously.

Considering for simplicity an a priori reduced formulation and reusing the notation $\mathcal{P}_p$ for the indices of those phases with $h^{\alpha}(\boldsymbol{\phi})>0$, the basic structure of a Newton-step on the full coupled system is of the form

$$\begin{pmatrix}\begin{pmatrix}\mathrm{diag}\big(\frac{\partial\tilde{\boldsymbol{\mu}}^{\alpha}}{\partial\tilde{\boldsymbol{c}}^{\alpha}}\big)_{\alpha\in\mathcal{P}_p} & \begin{pmatrix}-\boldsymbol{I}\\ \vdots\\ -\boldsymbol{I}\end{pmatrix}\\ \begin{pmatrix}h^1\boldsymbol{I} & \dots & h^N\boldsymbol{I}\end{pmatrix} & \boldsymbol{0}\end{pmatrix} & \begin{pmatrix}\mathrm{diag}\big(\frac{\partial\tilde{\boldsymbol{\mu}}^{\alpha}}{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}\big)_{\alpha\in\mathcal{P}_p} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\\ \begin{pmatrix}\mathrm{diag}\big(\frac{\partial\boldsymbol{\sigma}^{\alpha}}{\partial\tilde{\boldsymbol{c}}^{\alpha}}\big)_{\alpha\in\mathcal{P}_p} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix} & \begin{pmatrix}\mathrm{diag}\big(\frac{\partial\boldsymbol{\sigma}^{\alpha}}{\partial\tilde{\boldsymbol{\epsilon}}^{\alpha}}\big)_{\alpha\in\mathcal{P}_p} & \begin{pmatrix}-\boldsymbol{I}\\ \vdots\\ -\boldsymbol{I}\end{pmatrix}\\ \begin{pmatrix}h^1\boldsymbol{I} & \dots & h^N\boldsymbol{I}\end{pmatrix} & \boldsymbol{0}\end{pmatrix}\end{pmatrix}\begin{pmatrix}\begin{pmatrix}(\delta\tilde{\boldsymbol{c}}^{\alpha})_{\alpha\in\mathcal{P}_p}\end{pmatrix}\\ \begin{pmatrix}(\delta\boldsymbol{\epsilon}^{\alpha})_{\alpha\in\mathcal{P}_p}\end{pmatrix}\\ \boldsymbol{\sigma}\end{pmatrix}=\begin{pmatrix}\begin{pmatrix}(-\tilde{\boldsymbol{\mu}}^{\alpha})_{\alpha\in\mathcal{P}_p}\end{pmatrix}\\ \boldsymbol{r}_{\tilde{\boldsymbol{\epsilon}}}\\ \begin{pmatrix}(-\boldsymbol{\sigma}^{\alpha})_{\alpha\in\mathcal{P}_p}\end{pmatrix}\\ \boldsymbol{r}_{\boldsymbol{\epsilon}}\end{pmatrix},$$

i.e. a system consisting of two generalized saddle-point problems[98] coupled by two very sparse off-diagonal blocks.
In addition to the part $\frac{\partial^2 f^\alpha}{\partial (\boldsymbol{c}^\alpha)^2}$ expressed in Equation (7.193), from Equation (7.191), a variation $\mathrm{d}\boldsymbol{\epsilon}^\alpha$ of the phase-specifc strain leads to the variation

$$\frac{\partial \boldsymbol{\mu}^\alpha_{el}}{\partial \boldsymbol{\epsilon}^\alpha} : \mathrm{d}\boldsymbol{\epsilon}^\alpha = -\Big(\boldsymbol{\mathcal{C}}^\alpha(\boldsymbol{c}^\alpha) : \mathrm{d}\boldsymbol{\epsilon}^\alpha\Big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^\alpha}{\partial \boldsymbol{c}^\alpha} + \mathrm{d}\boldsymbol{\epsilon}^\alpha : \Big(\boldsymbol{\epsilon}^\alpha_{el} : \frac{\partial \boldsymbol{\mathcal{C}}^\alpha}{\partial \boldsymbol{c}^\alpha}\Big) = \mathrm{d}\boldsymbol{\epsilon}^\alpha : \frac{\partial \boldsymbol{\sigma}^\alpha}{\partial \boldsymbol{c}^\alpha} \tag{7.195}$$

of the elastic contribution to the chemical potential, whereas, by the phase-specific stress-strain relationship, one trivially has $\frac{\partial \boldsymbol{\sigma}^\alpha}{\partial \boldsymbol{\epsilon}^\alpha} = \boldsymbol{\mathcal{C}}^\alpha$.

As both diagonal subproblems are themselves highly sparse with a known block-diagonal substructure in the "primal" part, it is natural to try to make use of this knowledge by using a block-factorization based approach. There are three relatively obvious choices for doing so.
The first two are to start either analogously to the purely chemical or the purely mechanical case[99], i.e. by performing a block-elimination step either in terms of the inverse of the full chemical or mechanical saddle-point problems. The principle advantage of this approach is that, from an implementation point of view, this semi-sequential treatment allows maintaining at least some of the modularity from the simpler setting whithout an explicit coupling between the chemical and mechanical parts. Its major disadvantage is that since both system containg a coupling between all phases due to the sum-constraint on the $\tilde{\boldsymbol{c}}^\alpha$ resp. the $\boldsymbol{\epsilon}^\alpha$, the first elimination leads to a loss of the remaining block-diagonal structure in the respective other subblock. More precisely, eliminating $\big((\boldsymbol{\epsilon}^\alpha)_{\alpha\mathcal{P}_p}, \boldsymbol{\sigma}\big)$ in terms of $\big((\tilde{\boldsymbol{c}}^\alpha)_{\alpha\mathcal{P}_p}, \tilde{\boldsymbol{\mu}}\big)$ leads in particular to the subblock $\mathrm{diag}\big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\big)_{\alpha\in\mathcal{P}_p}$ being replaced by a fully filled matrix, whereas eliminating $\big((\tilde{\boldsymbol{c}}^\alpha)_{\alpha\mathcal{P}_p}, \tilde{\boldsymbol{\mu}}\big)$ in terms of $\big((\boldsymbol{\epsilon}^\alpha)_{\alpha\mathcal{P}_p}, \boldsymbol{\sigma}\big)$ has the same effect on the previously block-diagonal part $\mathrm{diag}\big(\frac{\partial \boldsymbol{\sigma}^\alpha}{\partial \boldsymbol{\epsilon}^\alpha}\big)_{\alpha\in\mathcal{P}_p}$.
This is avoided by the third approach, which consists in eliminating all primal unknowns in terms of the dual ones, i.e. by performing a block-elimination based on the subblocks

$$\frac{\partial^2 \tilde{f}^\alpha}{\partial (\tilde{\boldsymbol{c}}^\alpha, \boldsymbol{\epsilon}^\alpha)^2} \begin{pmatrix} \delta\tilde{\boldsymbol{c}}^\alpha \\ \delta\boldsymbol{\epsilon}^\alpha \end{pmatrix} = \begin{pmatrix} \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} & \frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \boldsymbol{\epsilon}^\alpha} \\ \frac{\partial \boldsymbol{\sigma}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha} & \frac{\partial \boldsymbol{\sigma}^\alpha}{\partial \boldsymbol{\epsilon}^\alpha} \end{pmatrix} \begin{pmatrix} \delta\tilde{\boldsymbol{c}}^\alpha \\ \delta\boldsymbol{\epsilon}^\alpha \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^\alpha \\ \boldsymbol{\sigma} - \boldsymbol{\sigma}^\alpha \end{pmatrix}$$

for each phase for determining $\big(\tilde{\boldsymbol{c}}^\alpha, \boldsymbol{\epsilon}^\alpha\big)$ as functions of $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\sigma}$, and thus reducing the problem to the one of solving a coupled Schur-complement equation for $(\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma})$. A clear advantage of such a factorization is that it makes the highest use of the underlying sparsity pattern. A potential drawback from an implementation point of view is that it is the "most explicitly coupled" of the three approaches above, and therefore, while likely the most efficient choice, the one which is the trickiest with respect to code-reusability.

In combination with the Reuss-Sachs model, the dependence of both $\boldsymbol{c}^\beta$ and $\boldsymbol{\epsilon}^\beta$ on $\boldsymbol{\phi}$ leads to the same two contributions as before, i.e. $-\boldsymbol{\mu}\cdot\boldsymbol{c}^\beta$ from the chemical part and $-\boldsymbol{\sigma} : \boldsymbol{\epsilon}^\beta$ from the chemical part,

$$\frac{\partial f}{\partial \phi^\alpha} = \sum_{\beta=1}^{N} \Big(f^\beta_{ch}(\boldsymbol{c}^\alpha) + \frac{1}{2}(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\beta) : \boldsymbol{\mathcal{C}}^\beta : (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}^\beta) - \boldsymbol{\mu}\cdot\boldsymbol{c}^\beta - \boldsymbol{\sigma} : \boldsymbol{\epsilon}^\beta \Big) \frac{\partial h^\beta}{\partial \phi^\alpha}.$$

Remark 162. From similar calculations as before, it also follows that the important properties $\boldsymbol{\mu} = \frac{\partial f}{\partial \boldsymbol{c}}$ and $\boldsymbol{\sigma} = \frac{\partial f}{\partial \boldsymbol{\epsilon}}$ still hold with $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ as defined above, despite the dependence of the

[98] The more standard saddle-point structure could again be recovered by "undoing" the division by h^α in the upper lines of each of the subblocks.

[99] The procedure outlined above essentially being a nonlinear (in the $\boldsymbol{c}^\alpha$) variation of the latter approach which does, after performing the linearization in the $\delta\boldsymbol{c}^\alpha$, reduce to the same resulting set of equations.

phase-specific quantities on both $\boldsymbol{c}$ and $\boldsymbol{\epsilon}$. For the chemical potential, this follows from

$$\frac{\partial f}{\partial \boldsymbol{c}} = \sum_\alpha \Big(\frac{\partial f^\alpha}{\partial \boldsymbol{c}^\alpha} \cdot \frac{\partial \boldsymbol{c}^\alpha}{\partial \boldsymbol{c}} + \frac{\partial f^\alpha}{\partial \boldsymbol{\epsilon}^\alpha} : \frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{c}} \Big) h^\alpha(\boldsymbol{\phi}) = \sum_\alpha \Big((\boldsymbol{\mu} + \lambda^\alpha \boldsymbol{e}) \cdot \frac{\partial \boldsymbol{c}^\alpha}{\partial \boldsymbol{c}} + \boldsymbol{\sigma} : \frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{c}} \Big) h^\alpha(\boldsymbol{\phi})$$

combined with $\boldsymbol{e} \cdot \frac{\partial \boldsymbol{c}^\alpha}{\partial \boldsymbol{c}} = \frac{\partial (\boldsymbol{e} \cdot \boldsymbol{c}^\alpha)}{\partial \boldsymbol{c}} = \frac{\partial 1}{\partial \boldsymbol{c}} = 0$ as well as, after extracting the phase-independent terms out of the sums, $\sum_\alpha \frac{\partial \boldsymbol{c}^\alpha}{\partial \boldsymbol{c}} h^\alpha = \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{c}}$ and $\sum_\alpha \frac{\partial \boldsymbol{\epsilon}^\alpha}{\partial \boldsymbol{c}} h^\alpha = \frac{\partial \boldsymbol{\epsilon}}{\partial \boldsymbol{c}} = \mathbf{0}$. The calculation for the stresses is completely analogous. ◇

Remark 163. As the "mechanical" subsystem is linear in both the $\boldsymbol{\epsilon}^\alpha$ and $\boldsymbol{\sigma}$, yet another alternative for solving this system is to proceed in the same manner as for the original description of the Reuss-Sachs model and eliminate the the $\boldsymbol{\epsilon}^\alpha$ as functions of the $\boldsymbol{c}^\alpha$ and $\boldsymbol{\sigma}$ **before** performing the linearization. This first leads to

$$\boldsymbol{\epsilon}^\alpha(\boldsymbol{c}^\alpha, \boldsymbol{\sigma}) = \boldsymbol{S}^\alpha(\boldsymbol{c}^\alpha) : \boldsymbol{\sigma} + \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha),$$

and, combining this with the sum-constraint on the $\boldsymbol{\epsilon}^\alpha$, to the Schur-complement system

$$\boldsymbol{\epsilon} = \sum_{\alpha=1}^{N} \boldsymbol{\epsilon}^\alpha(\boldsymbol{c}^\alpha, \boldsymbol{\sigma}) h^\alpha(\boldsymbol{\phi}) = \boldsymbol{S}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}\big) : \boldsymbol{\sigma} + \tilde{\boldsymbol{\epsilon}}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}\big)$$

for $\boldsymbol{\sigma}$, where $\tilde{\boldsymbol{\epsilon}}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}\big) = \sum_\alpha \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha) h^\alpha(\boldsymbol{\phi})$. It follows that

$$\boldsymbol{\sigma}^\alpha\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big) = \boldsymbol{\sigma}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big) = \boldsymbol{C}_{RS}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}\big) : \Big(\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}\big)\Big)$$

and thus

$$\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha) = \boldsymbol{\epsilon}^\alpha_{el}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big) = \boldsymbol{S}^\alpha(\boldsymbol{c}^\alpha) : \boldsymbol{\sigma}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big),$$

the only difference to the purely mechanical case being that $\boldsymbol{\sigma}$ now in addition depends on all the $\big(\boldsymbol{c}^\alpha\big)_{1 \le \alpha \le N}$. Reinserting these expressions into Equation (7.191) leads to

$$\begin{aligned}
\boldsymbol{\mu}^\alpha =& \boldsymbol{\mu}^\alpha_{ch}(\boldsymbol{c}^\alpha) - \Big(\boldsymbol{C}^\alpha(\boldsymbol{c}^\alpha) : \big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha)\big)\Big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^\alpha}{\partial \boldsymbol{c}^\alpha} + \frac{1}{2}\big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha)\big) : \Big(\big(\boldsymbol{\epsilon}^\alpha - \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha)\big) : \frac{\partial \boldsymbol{C}^\alpha}{\partial \boldsymbol{c}^\alpha}\Big) \\
=& \boldsymbol{\mu}^\alpha_{ch}(\boldsymbol{c}^\alpha) - \boldsymbol{\sigma}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big) : \frac{\partial \tilde{\boldsymbol{\epsilon}}^\alpha}{\partial \boldsymbol{c}^\alpha} \\
&+ \frac{1}{2}\Big(\boldsymbol{S}^\alpha(\boldsymbol{c}^\alpha) : \boldsymbol{\sigma}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big)\Big) : \Big(\Big(\boldsymbol{S}^\alpha(\boldsymbol{c}^\alpha) : \boldsymbol{\sigma}\big(\boldsymbol{\phi}, (\boldsymbol{c}^\alpha)_{1 \le \alpha \le N}, \boldsymbol{\epsilon}\big)\Big) : \frac{\partial \boldsymbol{C}^\alpha}{\partial \boldsymbol{c}^\alpha}\Big),
\end{aligned}$$

and thus a system in the $\boldsymbol{c}^\alpha$ alone.
A subsequent linearization for dealing with the nonlinearity in $\boldsymbol{c}^\alpha$ will then leads to essentially the same system as obtained above by linearization first and then performing a block-elimination of the mechanical quantities in terms of $\tilde{\boldsymbol{c}}^\alpha$ and $\tilde{\boldsymbol{\mu}}$.

◇

The Jump-Condition Based Formulation

Finally, if there are notable differences in the mechanical behavior of the different phases, it may be beneficial to complement the conditions on the chemical potentials with a mechanical model as in Section 7.2.2 or 7.2.4.

The Two-Phase Case In the two-phase case, the simplest choice is to, as in Equation (7.88), impose $\boldsymbol{\epsilon}^1(\boldsymbol{\epsilon}, \boldsymbol{a}) = \boldsymbol{\epsilon} - h^2(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S$ and $\boldsymbol{\epsilon}^2(\boldsymbol{\epsilon}, \boldsymbol{a}) = \boldsymbol{\epsilon} + h^1(\phi)(\boldsymbol{a} \otimes \boldsymbol{n})_S$, and to fix $\boldsymbol{a}$ through the additional condition

$$\boldsymbol{r}_{\boldsymbol{\sigma}} \coloneqq [\![\boldsymbol{\sigma}^\alpha(\boldsymbol{c}^\alpha, \boldsymbol{\epsilon}^\alpha)]\!]^{12} \cdot \boldsymbol{n} = \left(\boldsymbol{\sigma}^2(\boldsymbol{c}^2, \boldsymbol{\epsilon}^2) - \boldsymbol{\sigma}^1(\boldsymbol{c}^1, \boldsymbol{\epsilon}^1)\right) \cdot \boldsymbol{n} \stackrel{!}{=} \boldsymbol{0}. \tag{7.196}$$

Since the $\boldsymbol{\epsilon}^\alpha$ are are both given in terms of a simple explicit expression in terms of $\boldsymbol{\epsilon}$ and $\boldsymbol{a}$, it is clearly convenient to rewrite the system consisting of Equation (7.196) and the equality of the (reduced) chemical potentials as functions of $(\boldsymbol{c}^\alpha, \tilde{\boldsymbol{\mu}})$ and $\boldsymbol{a}$ instead of the $\boldsymbol{\epsilon}^\alpha$ themselves. A Newton-step on this system consists in solving

$$\left(\begin{pmatrix} \frac{\partial \tilde{\boldsymbol{\mu}}^1}{\partial \tilde{\boldsymbol{c}}^1} & \boldsymbol{0} & -\boldsymbol{I} \\ \boldsymbol{0} & \frac{\partial \tilde{\boldsymbol{\mu}}^2}{\partial \tilde{\boldsymbol{c}}^2} & -\boldsymbol{I} \\ h^1\boldsymbol{I} & h^2\boldsymbol{I} & \boldsymbol{0} \\ \left(\frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \tilde{\boldsymbol{c}}^1}\right. & \left.\frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \tilde{\boldsymbol{c}}^2}\right. & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \frac{\partial \tilde{\boldsymbol{\mu}}^1}{\partial \boldsymbol{a}} \\ \frac{\partial \tilde{\boldsymbol{\mu}}^2}{\partial \boldsymbol{a}} \\ \boldsymbol{0} \\ \frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{a}} \end{pmatrix}\right)\left(\begin{pmatrix} \delta\tilde{\boldsymbol{c}}^1 \\ \delta\tilde{\boldsymbol{c}}^2 \\ \tilde{\boldsymbol{\mu}} \\ \delta\boldsymbol{a} \end{pmatrix}\right) = \left(\begin{pmatrix} -\tilde{\boldsymbol{\mu}}^1 \\ -\tilde{\boldsymbol{\mu}}^2 \\ \boldsymbol{r}_{\tilde{\boldsymbol{c}}} \end{pmatrix}\right), \tag{7.197}$$

and thus a relatively simple addition to the purely chemical problem[100]. In addition, from Equation (7.195) together with the expressions for $\boldsymbol{\epsilon}^1$ and $\boldsymbol{\epsilon}^2$, it follows immediately that the $\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \boldsymbol{a}}$ are characterized by

$$\frac{\partial \tilde{\boldsymbol{\mu}}^1}{\partial \boldsymbol{a}} = -h^2 \frac{\partial\left(\boldsymbol{\sigma}^1(\tilde{\boldsymbol{c}}^1, \boldsymbol{a}) \cdot \boldsymbol{n}\right)}{\partial \tilde{\boldsymbol{c}}^1} \quad \text{and} \quad \frac{\partial \tilde{\boldsymbol{\mu}}^2}{\partial \boldsymbol{a}} = h^1 \frac{\partial\left(\boldsymbol{\sigma}^2(\tilde{\boldsymbol{c}}^2, \boldsymbol{a}) \cdot \boldsymbol{n}\right)}{\partial \tilde{\boldsymbol{c}}^2},$$

whereas the remaining new entries are given by

$$\frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \tilde{\boldsymbol{c}}^1} = -\frac{\partial\left(\boldsymbol{\sigma}^1(\boldsymbol{c}^1, \boldsymbol{a}) \cdot \boldsymbol{n}\right)}{\partial \tilde{\boldsymbol{c}}^1} \quad \text{and} \quad \frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \tilde{\boldsymbol{c}}^2} = +\frac{\partial\left(\boldsymbol{\sigma}^2(\boldsymbol{c}^2, \boldsymbol{a}) \cdot \boldsymbol{n}\right)}{\partial \tilde{\boldsymbol{c}}^2}$$

as well as the matrix

$$\frac{\partial \boldsymbol{r}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{a}} = \boldsymbol{n} \cdot \left(h^2(\phi)\boldsymbol{\mathcal{C}}^1(\boldsymbol{c}^1) + h^1(\phi)\boldsymbol{\mathcal{C}}^2(\boldsymbol{c}^2)\right) \cdot \boldsymbol{n}$$

already used in Section 7.2.2 in a non-incremental form in $\boldsymbol{a}$.

Remark 164. It can be noted that this matrix is, after "undoing" the various divisions by the h-functions (i.e. after a multiplication of the first two rows with h^1 and h^2 respectively and the fourth one by h^1h^2 symmetric, as is to be expected from the variational characterization underlying the equality of the chemical potentials as in [25] (see Section 7.1) and the continuity condition on the stresses as proposed in [51] (see Section 7.2.2). ◇

Even though the coupling of the local chemical and mechanical equilibrium conditions now implicitly defines $\boldsymbol{c}^\alpha$ and the $\boldsymbol{\epsilon}^\alpha$ as functions of the four given parameters $(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{\epsilon}, \boldsymbol{n})$, the usual calculations show that this does not really affect the previous contributions in the separate chemical and mechnical case. Firstly, for the driving force, one has

$$\frac{\partial f}{\partial \phi^\alpha} = \sum_\beta f^\beta(\boldsymbol{c}^\beta, \boldsymbol{\epsilon}^\beta) \frac{\partial h^\beta}{\partial \phi^\alpha} + \sum_\beta \frac{\partial f^\beta}{\partial \boldsymbol{c}^\beta} \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \phi^\alpha} + \frac{\partial f}{\partial \boldsymbol{\epsilon}^\beta} : \frac{\partial \boldsymbol{\epsilon}^\beta}{\partial \phi^\alpha} h^\beta.$$

The summation over the first term in the second sum leads as previously to $\sum_\beta -\boldsymbol{\mu} \cdot \boldsymbol{c}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$. The last term consists as in Section 7.2.2 of two contributions, one given by the explicit dependence of the definitions of $\boldsymbol{\epsilon}^1 - h^2(\boldsymbol{\phi})(\boldsymbol{a} \otimes \boldsymbol{n})_S$ and $\boldsymbol{\epsilon}^2 + h^1(\boldsymbol{\phi})(\boldsymbol{a} \otimes \boldsymbol{n})_S$ on $\boldsymbol{\phi}$ and the other one due to

[100] Some points regarding potential solution strategies will be discussed in relation with the multiphase formulation.

the implicit dependence of $\boldsymbol{a}$ on $\boldsymbol{\phi}$. The latter one cancels as the vanishing of the normal stress jump ensures that $\sum_\beta \frac{\partial f^\beta}{\partial \boldsymbol{a}} h^\beta = \mathbf{0}$, whereas the first one results in

$$h^1 \boldsymbol{\sigma}^1 : \big(\boldsymbol{a} \otimes \boldsymbol{n}\big)_S \big(-\frac{\partial h^2}{\partial \phi^\alpha}\big) + h^2 \boldsymbol{\sigma}^2 : \big(\boldsymbol{a} \otimes \boldsymbol{n}\big)_S \frac{\partial h^1}{\partial \phi^\alpha}.$$

Using $(\boldsymbol{a} \otimes \boldsymbol{n})_S = \boldsymbol{\epsilon}^2 - \boldsymbol{\epsilon}^1$, $h^1 + h^2 = 0$ and $\frac{\partial h^1}{\partial \phi^\alpha} = -\frac{\partial h^2}{\partial \phi^\alpha}$, this expression can, as in Section 7.2.2, be rewritten into a more pleasant form as $\sum_\beta -\boldsymbol{\sigma} : \boldsymbol{\epsilon}^\beta \frac{\partial h^\beta}{\partial \phi^\alpha}$, leaving the final expression

$$\frac{\partial f}{\partial \phi^\alpha} = \sum_\beta \Big(f^\beta(\boldsymbol{c}^\beta, \boldsymbol{\epsilon}^\beta) \frac{\partial h^\beta}{\partial \phi^\alpha} - \boldsymbol{\mu} \cdot \boldsymbol{c}^\beta - \boldsymbol{\sigma} : \boldsymbol{\epsilon}^\beta \Big) \frac{\partial h^\beta}{\partial \phi^\alpha}.$$

The remaining contribution to the phasefield equation through $\frac{\partial f}{\partial \boldsymbol{n}}$ can also directly be taken from Equation (7.99), since, even though the $\boldsymbol{c}^\alpha$ now also depend on $\boldsymbol{n}$, one has

$$\sum_\beta \frac{\partial f^\beta}{\partial \boldsymbol{c}^\beta} \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \boldsymbol{n}} h^\beta = \sum_\beta \big(\boldsymbol{\mu} + \lambda^\beta \boldsymbol{e}\big) \cdot \frac{\partial \boldsymbol{c}^\beta}{\partial \boldsymbol{n}} h^\beta = \boldsymbol{\mu} \cdot \sum_\beta \frac{\partial \boldsymbol{c}^\beta}{\partial \boldsymbol{n}} h^\beta = \boldsymbol{\mu} \cdot \frac{\partial \big(\sum_\beta \boldsymbol{c}^\beta h^\beta\big)}{\partial \boldsymbol{n}} = \boldsymbol{\mu} \cdot \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{n}} = \mathbf{0}$$

Finally, making use of the analogous arguments (and in particular $\frac{\partial f}{\partial \boldsymbol{a}} = \mathbf{0}$), one can further verify that $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ continue to satisfy $\boldsymbol{\mu} = \frac{\partial f}{\partial \boldsymbol{c}}$ and $\boldsymbol{\sigma} = \frac{\partial f}{\partial \boldsymbol{\epsilon}}$.

The Multiphase Case Extending the two-phase model above to a multiphase setting is again a highly non-trivial task as it inherits the same difficulties already encountered in the purely mechanical setting in Section 7.2.4 and will only be discussed for the two models in [62] and [63]. The simplest to extend is the one based on a common normal vector as the model was seen to essentially preserve the pleasant mathematical properties from the two-phase case. Using the jump-vector based reformulation with $\boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon} - \big(\boldsymbol{a}^\alpha \otimes \boldsymbol{n}\big)_S$, the system to be solved is obtained by complementing the chemical conditions with the equality of the phase-specific normal stresses

$$\boldsymbol{\sigma}^\alpha(\boldsymbol{c}^\alpha, \boldsymbol{\epsilon}^\alpha) \cdot \boldsymbol{n} = \boldsymbol{n} \cdot \Big(\boldsymbol{\mathcal{C}}^\alpha(\boldsymbol{c}^\alpha) : \big(\boldsymbol{\epsilon} - \big(\boldsymbol{a}^\alpha \otimes \boldsymbol{n}\big)_S - \tilde{\boldsymbol{\epsilon}}^\alpha(\boldsymbol{c}^\alpha)\big)\Big) \overset{!}{=} \boldsymbol{\sigma} \cdot \boldsymbol{n}$$

with the yet to be determined total normal stress $\boldsymbol{\sigma} \cdot \boldsymbol{n}$. For the application of the basic Newton-scheme, it is again convenient to explicitly make use of the expressions for the $\boldsymbol{\epsilon}^\alpha$ in terms of the jump-vectors. In the simpler reduced form, this leads to the system

$$\begin{pmatrix} \begin{pmatrix} \operatorname{diag}\big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \tilde{\boldsymbol{c}}^\alpha}\big)_{\alpha \in \mathcal{P}_p} & \begin{pmatrix} -\boldsymbol{I} \\ \vdots \\ -\boldsymbol{I} \end{pmatrix} \\ \begin{pmatrix} h^1 \boldsymbol{I} & \dots & h^N \boldsymbol{I} \end{pmatrix} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \operatorname{diag}\big(\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \boldsymbol{a}^\alpha}\big)_{\alpha \in \mathcal{P}_p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} \operatorname{diag}\big(\frac{\partial (\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n})}{\partial \tilde{\boldsymbol{c}}^\alpha}\big)_{\alpha \in \mathcal{P}_p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \operatorname{diag}\big(\frac{\partial (\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n})}{\partial \boldsymbol{a}^\alpha}\big)_{\alpha \in \mathcal{P}_p} & \begin{pmatrix} -\boldsymbol{I} \\ \vdots \\ -\boldsymbol{I} \end{pmatrix} \\ \begin{pmatrix} h^1 \boldsymbol{I} & \dots & h^N \boldsymbol{I} \end{pmatrix} & \mathbf{0} \end{pmatrix} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} (\delta \tilde{\boldsymbol{c}}^\alpha)_{\alpha \in \mathcal{P}_p} \\ (\delta \boldsymbol{a}^\alpha)_{\alpha \in \mathcal{P}_p} \end{pmatrix} \\ \tilde{\boldsymbol{\mu}} \\ \boldsymbol{\sigma} \cdot \boldsymbol{n} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} (-\tilde{\boldsymbol{\mu}}^\alpha)_{\alpha \in \mathcal{P}_p} \\ \boldsymbol{r}_{\tilde{\boldsymbol{c}}} \end{pmatrix} \\ \begin{pmatrix} (-\boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n})_{\alpha \in \mathcal{P}_p} \\ \boldsymbol{r}_{\boldsymbol{a}} \end{pmatrix} \end{pmatrix},$$

where, from Equation (7.195) and $\mathrm{d}\boldsymbol{\epsilon}^\alpha = -\big(\mathrm{d}\boldsymbol{a}^\alpha \otimes \boldsymbol{n}\big)_S$, one has $\frac{\partial \tilde{\boldsymbol{\mu}}^\alpha}{\partial \boldsymbol{a}^\alpha} = -\frac{\partial \boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n}}{\partial \tilde{\boldsymbol{c}}^\alpha}$ and $\boldsymbol{r}_{\boldsymbol{\sigma}\cdot\boldsymbol{n}}^\alpha := \boldsymbol{\sigma} \cdot \boldsymbol{n} - \boldsymbol{\sigma}^\alpha \cdot \boldsymbol{n}$, while $\boldsymbol{r}_{\boldsymbol{a}} = \sum_{\alpha=1}^N \boldsymbol{a}^\alpha h^\alpha(\boldsymbol{\phi})$ is the error in the sum-constraint on the average of the jump-vectors.

Remark 165. This system clearly has the same basic structure as the one in the Reuss-Sachs case with the $\boldsymbol{a}^\alpha$ replacing the $\boldsymbol{\epsilon}^\alpha$ and in particular the sum-constraint $\sum_\alpha \boldsymbol{a}^\alpha h^\alpha = \mathbf{0}$ replacing $\sum_\alpha \boldsymbol{\epsilon}^\alpha h^\alpha = \boldsymbol{\epsilon}$. Even though it is in principle somewhat easier to solve due to the smaller number of unknowns in the $\boldsymbol{a}^\alpha$ and $\boldsymbol{\sigma} \cdot \boldsymbol{n}$ as compared to the full values of $\boldsymbol{\epsilon}^\alpha$ and $\boldsymbol{\sigma}$, the considerations underlying its solution are essentially the same as for the Reuss-Sachs based model. ◇

The driving force follows by combining the standard argument for the chemical part with the arguments preceding Equations (7.170) and (7.171)[101] to be given by

$$\frac{\partial}{\partial \phi^\alpha}\Big|_{\boldsymbol{n}} f(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta \left(f_{el}^\beta - \boldsymbol{\mu}\cdot\boldsymbol{c}^\beta + (\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n})\cdot\boldsymbol{a}^\beta \right)\frac{\partial h^\beta}{\partial \phi^\alpha}, \tag{7.198}$$

or, using $\boldsymbol{\epsilon}^\beta = \boldsymbol{\epsilon} - \left(\boldsymbol{a}^\beta \otimes \boldsymbol{n}\right)_S$, by

$$\frac{\partial f_{el}}{\partial \phi^\alpha}\Big|_{\boldsymbol{n}} = \sum_{\beta=1}^{N} \left(f_{el}^\beta(\boldsymbol{\epsilon}^\beta) - \boldsymbol{\mu}\cdot\boldsymbol{c}^\beta - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^\beta \right)\frac{\partial h^\beta}{\partial \phi^\alpha}. \tag{7.199}$$

The differentiation with respect to $\boldsymbol{n}$ leads to the same result as in Equation (7.172), since, despite the additional dependence of the $\boldsymbol{c}^\beta$ on $\boldsymbol{n}$, their total contributions drop out again due to the relation satisfied with respect to the chemical potential. Finally, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ as obtained in the solution of the local quasi-equilibrium conditions above still satisfy, despite the additional implicit dependencies, $\boldsymbol{\mu} = \frac{\partial f}{\partial \boldsymbol{c}}$ and $\boldsymbol{\sigma} = \frac{\partial f}{\partial \boldsymbol{\epsilon}}$[102].

The natural extension of the model by Schneider [61] is obtained by simply admitting an additional dependence on the $\boldsymbol{c}^\alpha$ in the normal continuity conditions

$$[\![\boldsymbol{\sigma}^\alpha]\!]^{\mathrm{R}\alpha}\cdot\boldsymbol{n}^{\mathrm{R}\alpha} = \left(\boldsymbol{\sigma}^\alpha(\boldsymbol{c}^\alpha,\boldsymbol{\epsilon}^\alpha) - \boldsymbol{\sigma}^{\mathrm{R}}(\boldsymbol{c}^{\mathrm{R}},\boldsymbol{\epsilon}^1)\right)\cdot\boldsymbol{n}^{\mathrm{R}\alpha} \stackrel{!}{=} \boldsymbol{0}$$

with respect to the reference phase combined with the previous definition of the phase-specific strains $\boldsymbol{\epsilon}^{\mathrm{R}} = \boldsymbol{\epsilon} - \sum_{\beta\neq\mathrm{R}} \left(\boldsymbol{a}^{\mathrm{R}\beta}\otimes\boldsymbol{n}^{\mathrm{R}\beta}\right)_S h^\beta(\boldsymbol{\phi})$ and $\boldsymbol{\epsilon}^\alpha = \boldsymbol{\epsilon}^{\mathrm{R}} + \left(\boldsymbol{a}^{\mathrm{R}\alpha}\otimes\boldsymbol{n}^{\mathrm{R}\alpha}\right)_S$, $\alpha\neq\mathrm{R}$ in Equation (7.174). In terms of a Newton-step for the $\boldsymbol{c}^\alpha$ and $\boldsymbol{a}^{\mathrm{R}\alpha}$, one has, from Equation (7.195), to consider the additional increments

$$\sum_{\beta\neq\mathrm{R}} \frac{\partial \boldsymbol{\mu}_{el}^1}{\partial \boldsymbol{a}^{\mathrm{R}\beta}}\cdot\delta\boldsymbol{a}^{\mathrm{R}\beta} = -\sum_{\beta\neq\mathrm{R}} \delta\boldsymbol{a}^{\mathrm{R}\beta}\cdot h^\beta \frac{\partial(\boldsymbol{\sigma}^{\mathrm{R}}\cdot\boldsymbol{n}^{\mathrm{R}\beta})}{\partial \boldsymbol{c}^{\mathrm{R}}} \tag{7.200}$$

in the chemical potential for the reference phase, and

$$\begin{aligned}\sum_{\beta\neq\mathrm{R}} \frac{\partial \boldsymbol{\mu}_{el}^\alpha}{\partial \boldsymbol{a}^{\mathrm{R}\beta}}\cdot\delta\boldsymbol{a}^{\mathrm{R}\beta} =& -\sum_{\beta\neq\mathrm{R}} \delta\boldsymbol{a}^{\mathrm{R}\beta}\cdot h^\beta \frac{\partial(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\beta})}{\partial \boldsymbol{c}^\alpha} + \delta\boldsymbol{a}^{\mathrm{R}\alpha}\cdot\frac{\partial(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\alpha})}{\partial \boldsymbol{c}^\alpha}\\ =& \delta\boldsymbol{a}^{\mathrm{R}\alpha}\cdot(1-h^\alpha)\frac{\partial(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\alpha})}{\partial \boldsymbol{c}^\alpha} - \sum_{\beta\neq\mathrm{R},\alpha} \delta\boldsymbol{a}^{\mathrm{R}\beta}\cdot h^\beta\frac{\partial(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\beta})}{\partial \boldsymbol{c}^\alpha}\end{aligned} \tag{7.201}$$

for $\alpha\neq\mathrm{R}$, as well as the increments

$$\frac{\partial(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\alpha})}{\partial \boldsymbol{c}^\alpha}\cdot\delta\boldsymbol{c}^\alpha - \frac{\partial(\boldsymbol{\sigma}^{\mathrm{R}}\cdot\boldsymbol{n}^{\mathrm{R}\alpha})}{\partial \boldsymbol{c}^{\mathrm{R}}}\cdot\delta\boldsymbol{c}^{\mathrm{R}} \tag{7.202}$$

in the normal stress jumps due to the changes in the concentration, whereas the expression for $\sum_{\beta\neq\mathrm{R}} \frac{\partial(\boldsymbol{\sigma}^\alpha-\boldsymbol{\sigma}^{\mathrm{R}})\cdot\boldsymbol{n}^{\mathrm{R}\alpha}}{\partial \boldsymbol{a}^{\mathrm{R}\beta}}\cdot\delta\boldsymbol{a}^{\mathrm{R}\beta}$ is, by linearity in the $\boldsymbol{a}^{\mathrm{R}\beta}$ and Equation (7.176) given by

$$\left(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\left((1-h^\alpha)\boldsymbol{\mathcal{C}}^\alpha + h^\alpha\boldsymbol{\mathcal{C}}^{\mathrm{R}}\right)\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\right)\cdot\delta\boldsymbol{a}^{\mathrm{R}\alpha} - \sum_{\beta\neq\mathrm{R},\alpha} h^\beta\left(\boldsymbol{n}^{\mathrm{R}\alpha}\cdot\left(\boldsymbol{\mathcal{C}}^\alpha - \boldsymbol{\mathcal{C}}^{\mathrm{R}}\right)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\right)\cdot\delta\boldsymbol{a}^{\mathrm{R}\beta}. \tag{7.203}$$

[101] I.e. essentially a combination of the common normal stress for all phases and the constraint equation $\sum_\alpha \boldsymbol{a}^\alpha h^\alpha = 0$.

[102] For example for $\boldsymbol{\mu}$, the potentially problematic new contribution could arise from the dependency of $\boldsymbol{a}^\alpha$ on $\boldsymbol{c}$. Since $\sum_\alpha \boldsymbol{\sigma}^\alpha : \left(\frac{\partial \boldsymbol{a}^\alpha}{\partial c_i}\otimes\boldsymbol{n}\right)h^\alpha = (\boldsymbol{\sigma}\cdot\boldsymbol{n})\cdot\sum_\alpha \frac{\partial \boldsymbol{a}^\alpha}{\partial c_i}h^\alpha = \frac{\partial \sum_\alpha \boldsymbol{a}^\alpha h^\alpha}{\partial c_i} = \boldsymbol{0}$ by the constraint $\sum_\alpha \boldsymbol{a}^\alpha h^\alpha = \boldsymbol{0}$, this dependency drops out.

The total system therefore has a quite different structure from the one based on using a common normal vector. In particular, it consists of a fully filled $(1,2)$ and $(2,2)$-block by Equations (7.201) and (7.203). For this reason, the only sparsity pattern which can be made use of lies in the chemical subblock, i.e. two reasonable alternatives to solving the full system is to either perform a block-elimination on the concentration increments in terms of the jump-vectors and the chemical potential or a full block-elimination on the chemical quantities in terms of the jump-vectors.

The driving force calculation is, except for the standard additional term $-\boldsymbol{\mu}\cdot\boldsymbol{c}^\beta$ essentially the same as the one for Equations (7.170) and (7.171), leading to

$$\frac{\partial}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} f(\boldsymbol{\phi},\boldsymbol{\epsilon}) = \sum_\beta \Big(f^\beta - \boldsymbol{\mu}\cdot\boldsymbol{c}^\beta - \boldsymbol{\sigma}:\boldsymbol{\epsilon}^\beta\Big)\frac{\partial h^\beta}{\partial\phi^\alpha} + \sum_{\beta\neq \mathrm{R}} h^\beta\Big(\big(\boldsymbol{\sigma}^\beta-\boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}\Big)\cdot\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}. \quad (7.204)$$

As before, there is a remaining dependence on the $\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}$. For a full evaluation, one therefore has to obtain the $\frac{\partial \boldsymbol{a}^{\mathrm{R}\beta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}$ from a linearization of the local equilibrium conditions[103]. The non-variational structure of the equilibrium conditions with respect to the $\boldsymbol{a}$ is also the reason why, similarly to the relation $\boldsymbol{\sigma} = \frac{\partial f_{el}}{\partial\boldsymbol{\epsilon}}$ not holding within multiphase regions, the "chemical potential" obtained from the solution of the local conditions above does not actually satisfy $\boldsymbol{\mu} = \frac{\partial f}{\partial \boldsymbol{c}}$[104]. As both points lead to notable complications and there is already a variational consistency in terms of $\boldsymbol{\sigma}$ and $\frac{\partial f}{\partial\boldsymbol{\epsilon}}$, the pragmatic choice is again to ignore this difficulty within the multiphase regions[105].

---

[103] At least assuming that the reference phase does not change as there will usually not be differentiability at these points. Differentiating the equilibrium conditions $\tilde{\boldsymbol{\mu}}^\beta = \tilde{\boldsymbol{\mu}}$ for all β together with the sum-constraint on the $\tilde{\boldsymbol{c}}^\beta$ and $\big(\boldsymbol{\sigma}^\alpha-\boldsymbol{\sigma}^{\mathrm{R}}\big)\cdot\boldsymbol{n}^{\mathrm{R}\beta}=\boldsymbol{0}$ for all $\beta\neq\mathrm{R}$ with respect to ϕ^α with the normal vectors kept fixed one obtains

$$\frac{\partial\tilde{\boldsymbol{\mu}}^\beta}{\partial\tilde{\boldsymbol{c}}^\beta}\frac{\partial\tilde{\boldsymbol{c}}^\beta}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} + \frac{\partial\tilde{\boldsymbol{\mu}}^\beta}{\partial\boldsymbol{a}^{\mathrm{R}\delta}}\frac{\partial\boldsymbol{a}^{\mathrm{R}\delta}}{\partial\phi^\alpha}|_{\boldsymbol{n}^{\alpha\beta}} - \frac{\partial\tilde{\boldsymbol{\mu}}}{\partial\phi^\alpha}|_{\boldsymbol{n}^{\alpha\beta}} = \frac{\partial\tilde{\boldsymbol{\mu}}^\beta}{\partial\boldsymbol{\epsilon}^\beta}:\sum_{\delta\neq\mathrm{R}}\big(\boldsymbol{a}^{\mathrm{R}\delta}\otimes\boldsymbol{n}^{\mathrm{R}\delta}\big)_S\frac{\partial h^\delta}{\partial\phi^\alpha}$$

as well as $\sum_\beta h^\beta\frac{\partial\tilde{\boldsymbol{c}}^\beta}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} = -\sum_\beta \boldsymbol{c}^\beta\frac{\partial h^\beta}{\partial\phi^\alpha}$ from the chemical part which needs to be combined with the conditions

$$\begin{aligned}&-\frac{\partial\boldsymbol{\sigma}^{\mathrm{R}}\cdot\boldsymbol{n}^{\mathrm{R}\beta}}{\partial\tilde{\boldsymbol{c}}^{\mathrm{R}}}\frac{\partial\tilde{\boldsymbol{c}}^{\mathrm{R}}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} + \frac{\partial\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n}^{\mathrm{R}\beta}}{\partial\tilde{\boldsymbol{c}}^\beta}\frac{\partial\tilde{\boldsymbol{c}}^\beta}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}} + \sum_{\delta\neq R}\frac{\partial(\boldsymbol{\sigma}^\beta-\boldsymbol{\sigma}^R)\cdot\boldsymbol{n}^{\mathrm{R}\beta}}{\partial\boldsymbol{a}^{\mathrm{R}\delta}}\frac{\partial\boldsymbol{a}^{\mathrm{R}\delta}}{\partial\phi^\alpha}\Big|_{\boldsymbol{n}^{\alpha\beta}}\\ =&-\Big(\frac{\partial\boldsymbol{\sigma}^\beta\cdot\boldsymbol{n}^{\mathrm{R}\beta}}{\partial\boldsymbol{\epsilon}^\beta}-\frac{\partial\boldsymbol{\sigma}^{\mathrm{R}}\cdot\boldsymbol{n}^{\mathrm{R}\beta}}{\partial\boldsymbol{\epsilon}^{\mathrm{R}}}\Big):\sum_{\delta\neq\mathrm{R}}\big(\boldsymbol{a}^{\mathrm{R}\delta}\otimes\boldsymbol{n}^{\mathrm{R}\delta}\big)_S\frac{\partial h^\delta}{\partial\phi^\alpha}\end{aligned}$$

for all $\beta\neq\mathrm{R}$.

If not based on an interpolation function satisfying $\frac{\partial h^\delta}{\partial\phi^\alpha}=0$ if $h^\delta(\boldsymbol{\phi})=0$, there is a slight additional complication here as compared to the original solution procedure for the local equilibrium conditions, since, even though the coupling in the unknowns is still only through those phases with $h^\beta(\boldsymbol{\phi})>0$, this is not true for the right-hand sides.

[104] Instead, one has

$$\frac{\partial f}{\partial\boldsymbol{c}} = \sum_\alpha(\boldsymbol{\mu}+\lambda^\alpha\boldsymbol{e})\frac{\partial\boldsymbol{c}^\alpha}{\partial\boldsymbol{c}}h^\alpha + \sum_\alpha\boldsymbol{\sigma}^\alpha:\frac{\partial\boldsymbol{\epsilon}^\alpha}{\partial\boldsymbol{c}}h^\alpha.$$

While the first term again simplifies to $\boldsymbol{\mu}$, the second one maintains a dependence of the jump vectors on the total concentration through

$$\begin{aligned}&-\sum_\alpha\sum_{\beta\neq\mathrm{R}}h^\alpha h^\beta\big(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\beta}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{c}} + \sum_{\alpha\neq\mathrm{R}}h^\alpha\big(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{c}} = -\sum_{\beta\neq\mathrm{R}}h^\beta\big(\boldsymbol{\sigma}\cdot\boldsymbol{n}^{\mathrm{R}\beta}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{c}} + \sum_{\alpha\neq\mathrm{R}}h^\alpha\big(\boldsymbol{\sigma}^\alpha\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{c}}\\ =&\sum_{\alpha\neq\mathrm{R}}h^\alpha\Big(\big(\boldsymbol{\sigma}^\alpha-\boldsymbol{\sigma}\big)\cdot\boldsymbol{n}^{\mathrm{R}\alpha}\Big)\cdot\frac{\partial\boldsymbol{a}^{\mathrm{R}\beta}}{\partial\boldsymbol{c}}.\end{aligned}$$

[105] Recall that within two-phase regions, there is no such issue.

Remark 166. One can in principle also construct a chemo-mechanical extension of the model by Tschukin by e.g. simply adding a concentration dependence in his expression

$$\frac{1}{2}\sum_{\alpha=1}^{N}\left(\boldsymbol{\Sigma}^{\alpha}(\boldsymbol{c}^{\alpha}):\left(\boldsymbol{\epsilon}-\tilde{\boldsymbol{\epsilon}}^{\alpha}(\boldsymbol{c}^{\alpha})\right)-\frac{1}{2}\sum_{\beta\neq\alpha}[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta}):\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta}):[\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta})h^{\beta}(\boldsymbol{\phi})\right)h^{\alpha}(\boldsymbol{\phi})$$

for the mechanical free energy density[106] with the $\boldsymbol{c}^{\alpha}$ determined based on the usual chemical equilibrium condition and then defining $\boldsymbol{\sigma}=\frac{\partial f}{\partial\boldsymbol{\epsilon}}$ as in [74].
As one of the motivations of the expressions in [74] seems to have been to obtain a fairly explicit expression of the underlying free energy in terms of the parameters $(\boldsymbol{\phi},\nabla\boldsymbol{\phi},\boldsymbol{\epsilon})$, this is somewhat contrary in spirit as the $\boldsymbol{c}^{\alpha}$ would then be defined implicitly as $\boldsymbol{c}^{\alpha}=\boldsymbol{c}^{\alpha}(\boldsymbol{\phi},\nabla\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{\epsilon})$ and has not been investigated in more detail[107]. ◇

[106] As discussed previously, the term in the double-summation could also be rewritten as

$$\left([\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta})\cdot\boldsymbol{n}^{\alpha\beta}\right)\cdot\left(\boldsymbol{n}^{\alpha\beta}\cdot\bar{\boldsymbol{\mathcal{C}}}^{\alpha\beta}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta})\cdot\boldsymbol{n}^{\alpha\beta}\right)^{-1}\cdot\left([\![\boldsymbol{\Sigma}]\!]^{\alpha\beta}(\boldsymbol{c}^{\alpha},\boldsymbol{c}^{\beta})\cdot\boldsymbol{n}^{\alpha\beta}\right)$$

for avoiding the use of the pseudo-inverse $\bar{\boldsymbol{\mathcal{S}}}^{\alpha\beta}_{\boldsymbol{nn}}$.

[107] It may be interesting to do so though. In this regard, it should be noted that this implicit dependence is, again by the variational characterization of the $\boldsymbol{c}^{\alpha}$, not really problematic here. In particular, the basic form of the expression for the final stress in Equation (7.181) would be maintained, since, despite the a priori additional appearance of the terms $\frac{\partial f}{\partial\boldsymbol{c}^{\alpha}}\cdot\frac{\partial\boldsymbol{c}^{\alpha}}{\partial\boldsymbol{\epsilon}}$, these would again drop out in the sum making use of the equality of the chemical pontials and $\sum_{\alpha}\boldsymbol{c}^{\alpha}h^{\alpha}=1$. For similar reasons, most of the description in Section 7.2.4 is expected to carry over without major modifications.

Chapter 8

Summary

The focus of this thesis has been on some primarily practical implications of different constraints arising in phasefield based models.
Chapter 6 discussed several consequences of the Gibbs-Simplex constraint on the phasefield values. Even though the constraint itself is relatively simple, it has far-reaching numerical consequences. One of these was studied in detail in terms of the spatial discretization for a simple but prototypical two-phase problem, illustrating some interesting relations between the numerical interface width as compared to the parameter ϵ and the associated discrete energetics. The central point of Section 6.3 is, besides a short summary of some by now mostly standard background such as the LROP approach from [40], the description of a simple local projection algorithm in combination with the mobility-matrix based dynamics as in [68]. Even though similar algorithms are likely known in the quadratic programming literature, the author is not aware of any specific source with an explicit description adapted to the particular nature of $\boldsymbol{M}$[1].

In constrast, Chapter 7 considers coupled problems with "internal" phase-specific variables in a chemical, mechanical and chemo-mechanical setting. Subsections 7.1.2 and 7.1.3 are essentially introductory in nature, adding some additional details to the very elegant description of the quantitative free energy model as described in [25]. This is then compared in Subsection 7.1.4 with the now more commonly used description in [56] and [19]. This discussion is in a sense continued in Sections 7.1.5 and 7.1.5, where some practical implications of the choice of "primary" unknown in terms of the concentration or the chemical potential are discussed. In particular, it was argued that, while the latter description can be highly efficient for particularly simple free energy densities, the former approach should be preferred in the general case. Finally, Subsection 7.1.6 considered an extension of the model to a non-isothermal situation, illustrating, among other things, the pleasant interplay of the variational definitions on two classical thermodynamic relations.
A closely related modeling approach in the mechanical case is discussed in Section 7.2. As the conditions fixing the phase-specific quantities is more complex in this case, various different formulations have been suggested in the literature even in the two-phase case. Some advantages and disadvantages of these formulations as well as their links are summarized in Subsections 7.2.2 and 7.2.3 as a basis for the analysis of some of the difficulties encountered by three extensions to the multiphase setting by Schneider and Tschukin in [63], [62] and [74]. As a final topic, Subsection 7.2.5 outlined several practical and theoretical implications obtained by coupling the chemical model from Section 7.1 with the various mechanical models.

[1] There are also surprisingly few explicit descriptions of the Euclidian projection onto the Gibbs-simplex as many algorithms are instead designed to handle weighted projections subject to general equality and inequality constraints. Nevertheless, it is clear that the weight being given by the identity, the equality constraint by $\boldsymbol{e}\cdot\boldsymbol{\phi} = 1$ and the inequality constraints by $\phi^\alpha \geq 0$ allows for some significant simplifications as compared to the general case.

The central aspect of Chapter 7 is not the descriptions of the various models themselves, which are for the most part either well-known or relatively straightforward extensions of already known approaches, but to sensitize to and analyze some of the issues which can arise due to the introduction of more complex models, in particular in the presence of additional "internal" unknowns in terms of phasespecific quantities. Since the advantages of such models for the approximation quality of the related sharp interface modes even in the presence of artificially large interfaces are by now clearly recognized, it is to be expected that their use will become even more widespread in the years to come. As seen on the chemical and mechanical examples considered in Sections 7.1 and 7.2, there are some pitfalls when trying to combine such approaches with the very common two-step formalism for the derivation of new phasefield models in the literature.
On the one hand, the discussion of the multiphase mechanical model in [63] illustrated that independently postulating a phasefield functional and a set of equations to be satisfied by an additional set of unknowns besides the phasefield itself entails the risk of introducing an incompatibility with the standard variational approach. Even though this is not necessarily a problem in itself as the resulting model may still deliver very accurate results, one has to be careful about invoking an intuitively very pleasant but ultimately nonexistent energy minimization principle for its justification.
On the other hand, Chapter 7 also highlighted potential fallacies arising from the derivation of the drivingforces for the phasefield equation through a purely formal partial differentiation of a given functional with respect to ϕ, in particular also in combination with potential changes of primary unknowns. The questions raised by an interdependence between various unknowns and equations are of course well-understood and have mostly been answered a long time ago. Unfortunately, the fact that this allow to justify such a procedure for many of the earlier and simpler models, this seems to have lead to a fairly widespread misconception in the more applied phasefield community that this is simply "what needs to be done" in order to minimize or maximize the underlying functional. One of the purposes of the discussion of the models is therefore also to stress the importance of underlying variational principles or the lack thereof and the additional drivingforce contributions arising due to both local constraints imposed on the phasespecific quantities and global constraints imposed e.g. through a global mass-conservation principle. It is the author's hope that the detailed consideration of some of these issues in the very applied context of Chapter 7 may help to clear up some a priori confusing but ultimately clear-cut questions, in particular for young researches newly discovering the fascinating domain of phasefield modeling.

Appendix A

Calculation of the Discrete One-Dimensional Phasefield Energy and Interface Width

This section will provide some more details on the calculations underlying the one-dimensional discrete profile in subsection 6.2.3, and in particular the expressions (repeated here for convenience)

$$\mathcal{E}(N) = \frac{4\gamma}{\pi^2\epsilon}N + \frac{\epsilon\gamma}{(\Delta x)^2}\frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\sin(\kappa) = \frac{4\gamma}{\pi^2\epsilon}\left(N + \cot\big(\frac{\kappa N}{2}\big)\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}\right) \tag{A.1}$$

for the total energy,

$$\sum_{i=1}^{N-1}\frac{1}{\epsilon}w(\phi_i) = \frac{8}{\pi^2\epsilon}\gamma\left(\frac{1}{4}\Big(1 - \cot^2\big(\frac{\kappa N}{2}\big)\Big)N + \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\cot(\kappa)\right) \tag{A.2}$$

for the total contribution by the bulk potential in Equation (6.42) and

$$\sum_{i=1}^{N-1}\epsilon a_i = \frac{8}{\pi^2\epsilon}\gamma\left(\frac{1}{4}\Big(1 + \cot^2\big(\frac{\kappa N}{2}\big)\Big)N + \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\frac{1}{\sin(\kappa)}\right), \tag{A.3}$$

for the total contribution by the gradient energy in Equation (6.43) as well as the actual number of interface points from Equation (6.47),

$$N = N_{min} = \left\lceil \frac{2}{\kappa}\tan^{-1}\left(\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}\right)\right\rceil. \tag{A.4}$$

A.1 A Quick Recap of the First-Order Analysis

As seen in Subsection 6.2.3, fixing for convenience $\phi_0 = 0$ as the last bulk-point on the left and $\phi_N = 1$ as the first bulk-point on the right, the discrete first-order optimality condition (6.31) **within** the interface (i.e. at all points $1 \le i \le N-1$ for which the strict inequality $0 < \phi_i < 1$ holds) together with imposing the conditions $\phi_0 = 0$ and $\phi_N = 1$ at the transition to the bulk implies

$$\phi_i = \frac{1}{2} - \frac{1}{2}\cos(\kappa i) + \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big)\sin(\kappa i) \tag{A.5}$$

with κ determined by

$$\cos(\kappa) = 1 - \frac{8(\Delta x)^2}{\pi^2\epsilon^2} = 1 - \frac{1}{2}\Big(\frac{4\Delta x}{\pi\epsilon}\Big)^2 \tag{A.6}$$

$$\sin(\kappa) = \sqrt{1 - (1 - \frac{8(\Delta x)^2}{\pi^2\epsilon^2})^2} = \sqrt{2\frac{8(\Delta x)^2}{\pi^2\epsilon^2} - \Big(\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\Big)^2} = \frac{8(\Delta x)^2}{\pi^2\epsilon^2}\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}. \tag{A.7}$$

Remark 167. Note that up to this point, this is simply a consequence of the linear difference equation satisfied wihin the interface and the choice of an (in total) increasing profile going from 0 on the left to 1 on the right and will up to this point work for an arbitrary integer $N > 0$. In order to select the correct N, one still needs take both the underlying demand of a minmimization of the total energy $\Delta x\mathcal{E}(N) = \sum_{i=1}^{N-1}\big(\epsilon a_i + \frac{1}{\epsilon}w_i\big)\Delta x$ and the constraints $0 \le \phi_i \le 1$ on the profile into account. ◇

The positivity constraint $\phi_1 \ge 0$ leads, within the relevant range[1], to the restriction

$$N \le N_{max} = \left\lceil \frac{\pi}{\kappa} \right\rceil \tag{A.8}$$

on the total number of points, whereas the sign-constraint on the multipliers $\mu^{\pm}$ for the box-constraint $0 \le \phi_i \le 1$ lead to the lower bound

$$N \ge N_{min} := \left\lceil \frac{\pi}{\kappa} - 1 \right\rceil \tag{A.9}$$

.

Comparing the bounds it is obvious that this leaves two potential choices for N. Determining the optimal profile is therefore not possible based on the first-order necessary condition alone and requires a more detailed energetic analysis, which is of an inherent interest anyway.

A.2 The Discrete Energetic Analysis

Determining the total energy starting from a given sinusoidal profile is essentially a matter of a simple integration. The only difficulty is that, unlike in the continous case, the correct interpretation of this integration here is actually a summation over a set of discrete sinusoidal values. While this generally leads to - provided one is even able to do this analytically - very cumbersome expressions, it can be achieved in a very elegant fashion in the solenoidal case by the use of **Lagrange's trigonometric identities**

$$\sum_{i=1}^{N-1}\sin(\lambda i) = \frac{1}{2}\cot(\frac{\lambda}{2}) - \frac{\cos\big(\lambda(N-\frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}, \tag{A.10}$$

$$\sum_{i=1}^{N-1}\cos(\lambda i) = -\frac{1}{2} + \frac{\sin\big(\lambda(N-\frac{1}{2})\big)}{2\sin(\frac{\lambda}{2})}. \tag{A.11}$$

In fact, a simple differentiation (based on a finite-dimensional vector of ϕ-values) shows that the first-order optimality condition as in Equation (6.38) corresponds to the minimization of the discrete energy function (the fixed factor Δx essentially being irrelevant for the minimizer, but of course important from a physical point of view)

$$\begin{aligned}\Delta x\mathcal{E}(N) &= \Delta x\epsilon\sum_{i=0}^{N}\frac{1}{2}\Big(\Big(\frac{\phi_i - \phi_{i-1}}{\Delta x}\Big)^2 + \Big(\frac{\phi_{i+1} - \phi_i}{\Delta x}\Big)^2\Big) + \frac{16\Delta x}{\pi^2\epsilon}\sum_{i=0}^{N}\phi_i(1-\phi_i)\\ &= \epsilon\Delta x\sum_{i=1}^{N}\Big(\frac{\phi_i - \phi_{i-1}}{\Delta x}\Big)^2 + \frac{16\Delta x}{\pi^2\epsilon}\sum_{i=1}^{N-1}\phi_i(1-\phi_i)\end{aligned}$$

[1] By periodicity, significantly larger choices of N would in principle also be possible, but would be related to transitions involving several sinusoidals, which is obviously energetically unfavorable.

where the contributions from all $i < 0$ and $i > N$ vanish due to both the zero gradients and the vanishing of the w-term at $\phi \in \{0, 1\}$.
Based on the variational character of the phasefield Equation (6.31) within the interface, it turns out that the total energy is significantly easier to evaluate than the bulk potential or gradient energy contributions themselves. In fact, since

$$\begin{aligned}
&\sum_{i=1}^{N} \big(\frac{\phi_i - \phi_{i-1}}{\Delta x}\big)^2 = \frac{1}{\Delta x}\Big(\sum_{i=1}^{N} \frac{\phi_i - \phi_{i-1}}{\Delta x}\phi_i - \sum_{i=0}^{N-1} \frac{\phi_{i+1} - \phi_i}{\Delta x}\phi_i \Big) \\
=&\frac{1}{\Delta x}\Big(\sum_{i=1}^{N-1} \frac{\phi_i - \phi_{i-1}}{\Delta x}\phi_i - \sum_{i=1}^{N-1} \frac{\phi_{i+1} - \phi_i}{\Delta x}\phi_i \Big) + \frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_1 \big) \\
=&- \sum_{i=1}^{N-1} \frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2}\phi_i + \frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big),
\end{aligned}$$

the total energy can equivalently be rewritten as

$$\begin{aligned}
\mathcal{E}(N) =&\epsilon\Big(- \sum_{i=1}^{N-1} \frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2}\phi_i + \frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big)\Big) + \frac{16}{\pi^2\epsilon} \sum_{i=1}^{N-1} \phi_i(1 - \phi_i) \\
=& \sum_{i=1}^{N-1} \Big(- \epsilon\frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}(1 - \phi_i)\Big)\phi_i + \epsilon\frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big) \\
=&\frac{1}{2} \sum_{i=1}^{N-1} \underbrace{\Big(- 2\epsilon\frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}(1 - 2\phi_i)\Big)}_{=0 \text{ for } 1 \le i \le N-1} \phi_i \\
&+ \frac{1}{2} \sum_{i=1}^{N-1} \frac{16}{\pi^2\epsilon}\phi_i + \epsilon\frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big) \\
=& \sum_{i=1}^{N-1} \frac{8}{\pi^2\epsilon}\phi_i + \epsilon\frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big).
\end{aligned} \tag{A.12}$$

With $\phi_i = \frac{1}{2} - \frac{1}{2}\cos(\kappa i) + c_2 \sin(\kappa i)$ where

$$c_2 \coloneqq \frac{1}{2}\cot\big(\frac{\kappa N}{2}\big) \tag{A.13}$$

and Lagrange's trigonometric identities in Equation (A.10), this can be "integrated" due to

$$\sum_{i=1}^{N-1} \big(\frac{1}{2} - \frac{1}{2}\cos(\kappa i) + c_2 \sin(\kappa i)\big) = \frac{N-1}{2} - \frac{1}{2}\big(- \frac{1}{2} + \frac{\sin\big(\kappa(N - \frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\big) + c_2\big(\frac{1}{2}\cot(\frac{\kappa}{2}) - \frac{\cos\big(\kappa(N - \frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\big)$$

and therefore leads to

$$\begin{aligned}
\mathcal{E}(N) =&\frac{8}{\pi^2\epsilon}\Bigg[\frac{N-1}{2} - \frac{1}{2}\big(- \frac{1}{2} + \frac{\sin\big(\kappa(N - \frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\big) + c_2\big(\frac{1}{2}\cot(\frac{\kappa}{2}) - \frac{\cos\big(\kappa(N - \frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\big)\Bigg] \\
&+ \epsilon\frac{1}{\Delta x}\big(\frac{\phi_N - \phi_{N-1}}{\Delta x}\phi_N - \frac{\phi_1 - \phi_0}{\Delta x}\phi_0 \big).
\end{aligned}$$

Combined with the choice of orientation for the interface here, i.e. $\phi_0 = 0$ and $\phi_N = 1$, one further has

$$\phi_N - \phi_{N-1} = (1 - \cos(\kappa))(\frac{1}{2}) + \sin(\kappa)\big(\frac{1}{2}\sin(\kappa N) + c_2\cos(\kappa N)\big)$$

and thus

$$\mathcal{E} = \frac{8}{\pi^2\epsilon}\Bigg[\frac{N-1}{2} - \frac{1}{2}\Big(-\frac{1}{2} + \frac{\sin\big(\kappa(N-\frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\Big) + c_2\Big(\frac{1}{2}\cot(\frac{\kappa}{2}) - \frac{\cos\big(\kappa(N-\frac{1}{2})\big)}{2\sin(\frac{\kappa}{2})}\Big)\Bigg]$$
$$+ \frac{\epsilon}{(\Delta x)^2}\Big(\frac{1}{2}(1-\cos(\kappa)) + \sin(\kappa)\Big(\frac{1}{2}\sin(\kappa N) + c_2\cos(\kappa N)\Big)\Big).$$

Based on the half-angle formula

$$\cot\Big(\frac{\theta}{2}\Big) = \frac{1+\cos(\theta)}{\sin(\theta)} \tag{A.14}$$

for the cosine, it follows that $c_2 = \frac{1+\cos(\kappa N)}{2\sin(\kappa N)}$, where κ is independent of N. Combining this with the addition formulae

$$\sin\big(\kappa(2N-1)\big) = \sin(2\kappa N)\cos(\kappa) - \cos(2\kappa N)\sin(\kappa), \tag{A.15}$$
$$\cos\big(\kappa(2N-1)\big) = \cos(2\kappa N)\cos(\kappa) + \sin(2\kappa N)\sin(\kappa), \tag{A.16}$$

for the sine and cosine, one obtains

$$\mathcal{E} = \frac{8}{\pi^2\epsilon}\Bigg[\frac{N-1}{2} - \frac{1}{2}\Big(-\frac{1}{2} + \frac{1}{2}\big(\sin(\kappa N)\cot(\frac{\kappa}{2}) - \cos(\kappa N)\big)\Big)$$
$$+ c_2\Big(\frac{1}{2}\cot(\frac{\kappa}{2}) - \frac{1}{2}\big(\cos(\kappa N)\cot(\frac{\kappa}{2}) + \sin(\kappa N)\big)\Big)\Bigg]$$
$$+ \frac{\epsilon}{(\Delta x)^2}\Big(\frac{1}{2}(1-\cos(\kappa)) - \sin(\kappa)\Big(\frac{1}{2}\sin(\kappa N) + c_2\cos(\kappa N)\Big)\Big).$$

A significant simplification can be obtained due to $c_2\sin(\kappa N) = \frac{1}{2}(1+\cos(\kappa N))$ and therefore,

$$c_2\cos(\kappa N) = \frac{\cos(\kappa N) + (1-\sin^2(\kappa N))}{2\sin(\kappa N)} = c_2 - \frac{1}{2}\sin(\kappa N)$$

from which it follows that

$$-\frac{1}{2}\Big(-\frac{1}{2} + \frac{1}{2}\big(\sin(\kappa N)\cot(\frac{\kappa}{2}) - \cos(\kappa N)\big)\Big) + c_2\Big(\frac{1}{2}\cot(\frac{\kappa}{2}) - \frac{1}{2}\big(\cos(\kappa N)\cot(\frac{\kappa}{2}) + \sin(\kappa N)\big)\Big)$$
$$= \frac{1}{4} - \frac{1}{4}\big(\sin(\kappa N) + 2c_2\cos(\kappa N)\big)\cot(\frac{\kappa}{2}) + \frac{1}{4}\cos(\kappa N) + \frac{1}{2}c_2\cot(\frac{\kappa}{2}) - \frac{1}{2}c_2\sin(\kappa N)$$
$$= \frac{1}{4} - \frac{1}{4}\big(\sin(\kappa N) + 2(c_2 - \frac{1}{2}\sin(\kappa N))\big)\cot(\frac{\kappa}{2}) + \frac{1}{4}\cos(\kappa N) + \frac{1}{2}c_2\cot(\frac{\kappa}{2}) - \frac{1}{4}(1+\cos(\kappa N)) = 0.$$

Further making use of

$$\frac{1}{2}\sin(\kappa N) + c_2\cos(\kappa N) = \frac{1}{2}\sin(\kappa N) + c_2 - \frac{1}{2}\sin(\kappa N) = c_2$$

in the last term, the remaining expression

$$\mathcal{E} = \frac{4}{\pi^2\epsilon}(N-1) + \frac{\epsilon}{(\Delta x)^2}\Big(\frac{1}{2}(1-\cos(\kappa)) + c_2\sin(\kappa)\Big)$$

then solely depends on the "width" of the interface and an additional term arising at the transition to the bulk. As $\frac{\epsilon}{(\Delta x)^2}(1-\cos(\kappa)) = \frac{8}{\pi^2\epsilon}$, this reduces to

$$\mathcal{E} = \frac{4}{\pi^2\epsilon}N + \frac{\epsilon}{(\Delta x)^2}c_2\sin(\kappa) = \frac{4}{\pi^2\epsilon}N + \frac{\epsilon}{(\Delta x)^2}\frac{1}{2}\cot\Big(\frac{\kappa N}{2}\Big)\sin(\kappa), \tag{A.17}$$

or, inserting the expression $\sin(\kappa) = \frac{8(\Delta x)^2}{\pi^2\epsilon^2}\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}$ from Equation (A.7), to

$$\mathcal{E} = \frac{4}{\pi^2\epsilon}N + c_2\frac{8}{\pi^2\epsilon}\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1} = \frac{4}{\pi^2\epsilon}N + \frac{4}{\pi^2\epsilon}\cot\Big(\frac{\kappa N}{2}\Big)\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}.$$

With $\cot\left(\frac{\kappa}{2}\right) = \sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}$, one then finally obtains to the expression

$$\mathcal{E} = \frac{4}{\pi^2\epsilon}\Big(N + 2c_2\cot\Big(\frac{\kappa}{2}\Big)\Big) = \frac{4}{\pi^2\epsilon}\Big(N + \cot\Big(\frac{\kappa N}{2}\Big)\cot\Big(\frac{\kappa}{2}\Big)\Big) \tag{A.18}$$

in Equation (6.46).

The evaluation of the individual contributions through the gradient- and bulk energy contributions is unfortunately a little more tedious than that of the total energy as one cannot directly make use of the equation itself to eliminate the quadratic terms in ϕ as in Equation (A.12). Starting with the somewhat simpler bulk-potential term, one has to evaluate

$$\begin{aligned}
\Delta x\sum\frac{1}{\epsilon}w_i =& \frac{16\Delta\bar{x}}{\pi^2}\sum_{i=1}^{N-1}\Big(\frac{1}{2} - \frac{1}{2}\cos(\kappa i) + c_2\sin(\kappa i)\Big)\Big(\frac{1}{2} - \Big(-\frac{1}{2}\cos(\kappa i) + c_2\sin(\kappa i)\Big)\Big)\\
=& \frac{16\Delta\bar{x}}{\pi^2}\sum_{i=1}^{N-1}\frac{1}{4} - \Big(\frac{1}{2}\cos(\kappa i) - c_2\sin(\kappa i)\Big)^2\\
=& \frac{16\Delta\bar{x}}{\pi^2}\sum_{i=1}^{N-1}\frac{1}{4} - \Big(\frac{1}{4}\cos^2(\kappa i) - c_2\sin(\kappa i)\cos(\kappa i) + c_2^2\sin^2(\kappa i)\Big).
\end{aligned}$$

The quadratic terms in the sinusoidal function can be eliminated using the basic identities $\sin(\theta)\cos(\theta) = \frac{1}{2}\sin(2\theta)$, $\sin^2(\theta) = \frac{1-\cos(2\theta)}{2}$ and $\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}$, leading to

$$\begin{aligned}
\Delta x\sum\frac{1}{\epsilon}w_i =& \frac{16\Delta\bar{x}}{\pi^2}\sum_{i=1}^{N-1}\frac{1}{4} - \Big(\frac{1}{4}\frac{1+\cos(2\kappa i)}{2} - \frac{1}{2}c_2\sin(2\kappa i) + c_2^2\frac{1-\cos(2\kappa i)}{2}\Big)\\
=& \frac{8\Delta\bar{x}}{\pi^2\epsilon}\sum_{i=1}^{N-1}\Big(\frac{1}{4} - c_2^2\Big) - \Big(\frac{1}{4} - c_2^2\Big)\cos(2\kappa i) + c_2\sin(2\kappa i).
\end{aligned}$$

This is now in a form where Equation (A.10) can be applied, resulting (with $\lambda = 2\kappa$) in

$$\Delta x\sum\frac{1}{\epsilon}w_i = \frac{8\Delta\bar{x}}{\pi^2\epsilon}\Bigg(\Big(\frac{1}{4} - c_2^2\Big)(N-1) - \Big(\frac{1}{4} - c_2^2\Big)\Big(-\frac{1}{2} + \frac{\sin\big(\kappa(2N-1)\big)}{2\sin(\kappa)}\Big) + c_2\Big(\frac{1}{2}\cot(\kappa) - \frac{\cos\big(\kappa(2N-1)\big)}{2\sin(\kappa)}\Big).$$

Using the addition formulae (A.15) this can further be expanded to

$$\begin{aligned}
\Delta x\sum\frac{1}{\epsilon}w_i =& \frac{8}{\pi^2\epsilon}\Big(\Big(\frac{1}{4} - c_2^2\Big)(N-1) + \Big(c_2^2 - \frac{1}{4}\Big)\Big(-\frac{1}{2} + \frac{1}{2}\sin(2\kappa N)\cot(\kappa) - \frac{1}{2}\cos(2\kappa N)\Big)\\
&+ c_2\Big(\frac{1}{2}\cot(\kappa) - \frac{1}{2}\cos(2\kappa N)\cot(\kappa) - \frac{1}{2}\sin(2\kappa N)\Big)\Big),
\end{aligned}$$

or, with the double angle fomulae

$$\sin(2\kappa N) = 2\sin(\kappa N)\cos(\kappa N) \quad \text{and} \quad \cos(2\kappa N) = \cos^2(\kappa N) - \sin^2(\kappa N), \tag{A.19}$$

to

$$\begin{aligned}\Delta x\sum\frac{1}{\epsilon}w_i =&\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)(N-1)+\frac{1}{2}\big(c_2^2-\frac{1}{4}\big)\big(-1+2\sin(\kappa N)\cos(\kappa N)\cot(\kappa)-\cos^2(\kappa N)+\sin^2(\kappa N)\big)\\&+\frac{1}{2}c_2\big(\cot(\kappa)-(\cos^2(\kappa N)-\sin^2(\kappa N))\cot(\kappa)-2\sin(\kappa N)\cos(\kappa N)\big)\Big)\\=&\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)(N-1)+\big(c_2^2-\frac{1}{4}\big)\big(\sin(\kappa N)\cos(\kappa N)\cot(\kappa)-\cos^2(\kappa N)\big)\\&+c_2\big(\sin^2(\kappa N)\cot(\kappa)-\sin(\kappa N)\cos(\kappa N)\big)\Big).\end{aligned}$$

Making further progress now requires reintroducing the definition $\frac{1}{2}\cot\left(\frac{\kappa N}{2}\right)$ of c_2. First, the half-angle formula (A.14) allows modifying the last term to

$$c_2\big(\sin^2(\kappa N)\cot(\kappa)-\sin(\kappa N)\cos(\kappa N)\big)\Big)=\frac{1}{2}\big(1+\cos(\kappa N)\big)\big(\sin(\kappa N)\cot(\kappa)-\cos(\kappa N)\big)\Big).$$

It further implies on the one hand that

$$c_2\sin^2(\kappa N)=\frac{1}{2}\sin(\kappa N)+\frac{1}{2}\sin(\kappa N)\cos(\kappa N)$$

and on the other hand that

$$\begin{aligned}c_2^2-\frac{1}{4}=&\frac{1}{2}c_2\frac{1+\cos(\kappa N)}{\sin(\kappa N)}-\frac{1}{4}=\frac{2c_2+2c_2\cos(\kappa N)-\sin(\kappa N)}{4\sin(\kappa N)}=\frac{2c_2+2(c_2-\frac{1}{2}\sin(\kappa N))-\sin(\kappa N)}{4\sin(\kappa N)}\\=&\frac{4c_2-2\sin(\kappa N)}{4\sin(\kappa N)}=\frac{c_2}{\sin(\kappa N)}-\frac{1}{2},\end{aligned}$$

and thus the two additional relations

$$\begin{aligned}(c_2^2-\frac{1}{4})\sin(\kappa N)&=c_2-\frac{1}{2}\sin(\kappa N)\\(c_2^2-\frac{1}{4})\sin(\kappa N)\cos(\kappa N)&=c_2\cos(\kappa N)-\frac{1}{2}\sin(\kappa N)\cos(\kappa N).\end{aligned}$$

Combining these observations, one obtains

$$\begin{aligned}\Delta x\sum\frac{1}{\epsilon}w_i =&\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)(N-1)+\big(c_2^2-\frac{1}{4}\big)\big(\sin(\kappa N)\cos(\kappa N)\cot(\kappa)-\cos^2(\kappa N)\big)\\&+c_2\big(\sin^2(\kappa N)\cot(\kappa)-\sin(\kappa N)\cos(\kappa N)\big)\Big)\\=&\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)(N-1)-\big(c_2^2-\frac{1}{4}\big)\cos^2(\kappa N)\\&+\big(c_2\cos(\kappa N)-\frac{1}{2}\sin(\kappa N)\cos(\kappa N)\big)\cot(\kappa)\\&-c_2\sin(\kappa N)\cos(\kappa N)+\big(\frac{1}{2}\sin(\kappa N)+\frac{1}{2}\sin(\kappa N)\cos(\kappa N)\big)\cot(\kappa)\Big)\\=&\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)(N-1)-\big(c_2^2-\frac{1}{4}\big)\cos^2(\kappa N)+c_2\cos(\kappa N)\cot(\kappa)\\&-c_2\sin(\kappa N)\cos(\kappa N)+\frac{1}{2}\sin(\kappa N)\cot(\kappa)\Big).\end{aligned}$$

Again based on the half-angle formula (A.14), it further follows that $c_2\cos(\kappa N)=c_2-\frac{1}{2}\sin(\kappa N)$ and replacing $\cos^2(\kappa N)$ with $1-\sin^2(\kappa N)$ leads to the major simplification,

$$\Delta x\sum\frac{1}{\epsilon}w_i=\frac{8}{\pi^2\epsilon}\Big(\big(\frac{1}{4}-c_2^2\big)N+\big(c_2^2-\frac{1}{4}\big)\sin^2(\kappa N)+c_2\cot(\kappa)-c_2\sin(\kappa N)\cos(\kappa N)\Big),$$

and then finally with

$$\left(c_2^2-\frac{1}{4}\right)\sin^2(\kappa N)-c_2\sin(\kappa N)\cos(\kappa N)=\left(c_2-\frac{1}{2}\sin(\kappa N)\right)\sin(\kappa N)-\left(c_2-\frac{1}{2}\sin(\kappa N)\right)\sin(\kappa N)=0$$

to the expression

$$\Delta x\sum\frac{1}{\epsilon}w_i=\frac{8}{\pi^2\epsilon}\left(\left(\frac{1}{4}-c_2^2\right)N+c_2\cot(\kappa)\right) \tag{A.20}$$

for the contributions by the bulk-potential already given in Equation (6.42) resp. (A.2).
The energetic contribution of the gradient energy density term in Equation (A.3) resp. (6.43) can in principle be recovered by a similarly tedious calculation using a slight variation of the arguments above, but, as one already disposes of an epression for the total energy and the contribution by the bulk-potential, simply by taking the difference between the expression in Equation (A.1) and Equation (A.2).

Based on the expression for the total energy in Equation (A.1), it is now an easy matter to determine the value of N with the lowest interface energy. Taking the difference between the energies for two successive values of N shows that

$$\mathcal{E}(N+1)-\mathcal{E}(N)=\frac{4}{\pi^2\epsilon}+\frac{\epsilon}{(\Delta x)^2}(c_2(N+1)-c_2(N))\sin(\kappa),$$

where $c_2(N)$ is defined in Equation (A.13). Based on

$$c_2(N+1)-c_2(N)=\frac{1}{2}\left(\cot\left(\frac{\kappa(N+1)}{2}\right)-\cot\left(\frac{\kappa N}{2}\right)\right)$$

and

$$\cot\left(\frac{\kappa(N+1)}{2}\right)=\frac{\cot\left(\frac{\kappa N}{2}\right)\cot\left(\frac{\kappa}{2}\right)-1}{\cot\left(\frac{\kappa N}{2}\right)+\cot\left(\frac{\kappa}{2}\right)},$$

one has

$$\begin{aligned}2(c_2(N+1)-c_2(N))=&\frac{\cot\left(\frac{\kappa N}{2}\right)\cot\left(\frac{\kappa}{2}\right)-1}{\cot\left(\frac{\kappa N}{2}\right)+\cot\left(\frac{\kappa}{2}\right)}-\cot\left(\frac{\kappa N}{2}\right)=-\frac{1+\cot^2\left(\frac{\kappa N}{2}\right)}{\cot\left(\frac{\kappa N}{2}\right)+\cot\left(\frac{\kappa}{2}\right)}\\=&-\frac{1}{\sin^2\left(\frac{\kappa N}{2}\right)\left(\cot\left(\frac{\kappa N}{2}\right)+\cot\left(\frac{\kappa}{2}\right)\right)}\end{aligned}$$

and therefore

$$\mathcal{E}(N+1)-\mathcal{E}(N)=\frac{4}{\pi^2\epsilon}-\frac{\epsilon}{2(\Delta x)^2}\frac{1}{\sin^2\left(\frac{\kappa N}{2}\right)\left(\cot\left(\frac{\kappa N}{2}\right)+\cot\left(\frac{\kappa}{2}\right)\right)}\sin(\kappa). \tag{A.21}$$

This expression is initially decreasing for small N and will eventually start increasing as N passes the optimal value of N whose zero, treating N as continuous, would be achieved [2] for (see the discussion preceding Equation (6.40) for details on the elimination of the $\tan^{-1}$)

$$N_{cont}=\frac{2}{\kappa}\tan^{-1}\left(\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2}-1}\right)=\frac{\pi}{\kappa}-1. \tag{A.22}$$

[2] While this can be shown more rigorously treating N as continuous and considering $\frac{\partial^2\mathcal{E}}{\partial N^2}$, the monotonicity is intuitively obvious as very thin interfaces will spread due to the very high gradient energy density whereas very broad interfaces will contract due to the high bulk energy density. Treating N as continuous, both the actual zero of the difference and the sign of the second derivative can be found by using tangent half-angle substitution $t=\tan\left(\frac{\kappa N}{2}\right)$ and basic trigonometric identities to be given by $t=\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2}-1}$, and therefore the expression in Equation (A.22). Since the correct "guess" is already available, this calculation is skipped here for shortness.

In fact, using Equations (A.6) and (A.7) as well as $\kappa N_{cont} = \pi - \kappa$, one has

$$\sin^2\left(\frac{\kappa N_{cont}}{2}\right) = \frac{1-\cos(\kappa N_{cont})}{2} = \frac{1-\cos(\pi-\kappa)}{2} = \frac{1+\cos(\kappa)}{2} = 1 - \frac{4(\Delta x)^2}{\pi^2\epsilon^2} = \frac{4(\Delta x)^2}{\pi^2\epsilon^2}\left(\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1\right)$$

and $\cot\left(\frac{\kappa}{2}\right) = \frac{1+\cos(\kappa)}{\sin(\kappa)} = \sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}$ and by Equation (A.22) $\cot\left(\frac{\kappa N_{cont}}{2}\right) = \sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}^{-1}$. It follows that the denominator in the second term in Equation (A.21) reduces to

$$\frac{4(\Delta x)^2}{\pi^2\epsilon^2}\left(\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1\right)\left(\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}^{-1} + \sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}\right)$$

Combining this with Equation (A.7), one obtains

$$\begin{aligned}\mathcal{E}(N_{cont}+1) - \mathcal{E}(N_{cont}) =& \frac{4}{\pi^2\epsilon} - \frac{\epsilon}{2(\Delta x)^2}\frac{\frac{8(\Delta x)^2}{\pi^2\epsilon^2}\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}}{\frac{4(\Delta x)^2}{\pi^2\epsilon^2}\left(\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1\right)\left(\sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}^{-1} + \sqrt{\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1}\right)} \\ =& \frac{4}{\pi^2\epsilon} - \frac{\frac{4}{\pi^2\epsilon}}{\frac{4(\Delta x)^2}{\pi^2\epsilon^2}\left(1 + \left(\frac{\pi^2\epsilon^2}{4(\Delta x)^2} - 1\right)\right)} = 0.\end{aligned}$$

From the global point of view, it follows that the optimal N for enforcing a transition region compatible with the first-order necessary condition is given by $\lceil N_{cont}\rceil = \lceil \frac{\pi}{\kappa} - 1\rceil$ if $\frac{\pi}{\kappa}$ is not integer since the lower choice $\lfloor N_{cont}\rfloor$ is still in the region where the energy is smaller when adding an additional point. In contrast, if $\frac{\pi}{\kappa}$ is integer, the energies for both $N = \lfloor N_{cont}\rfloor$ and $N = \lceil N_{cont}\rceil$ coincide, meaning that from a discrete perspective (in N), the choice is indeterminate since both are energetically optimal.

A.3 The Local Analysis and the Second-Order Conditions

While the analysis in the previous section was primarily concerned with the analysis of the energetics in terms of the discrete parameter N, this does not exclude the possibility of several local minima for two successive values of N. This section is therefore more focused on the local stability in terms of continuous variations of the ϕ_i.

As recalled in Section A.1, the two values $N_{min} := \lceil \frac{\pi}{\kappa} - 1\rceil$ and $N_{max} = \lceil \frac{\pi}{\kappa}\rceil$ together with the profile in Equation (A.5) define two admissible solutions to the first-order necessary condition for a local minimizer of the phasefield equation. While it was seen in the previous section that the lower choice $N = N_{min}$ usually is the one with the lower total energy (unless $\frac{\pi}{\kappa}$ is integer, in which case both energies are equal), this does neither imply that the second choice $N = N_{max}$ is not a local minimizer nor, without further argument, that the profile for $N = N_{min}$ is in fact also a local minimizer and not just a critical point of the energy $\mathcal{E}$. Investigating this question requires including second-order information. Even though the question of second-order necessary and sufficient conditions for equality- and inequality-constrained problems is a well-studied one (see e.g. [13] and [46]), due to the fact that the energy $\mathcal{E}$ is a quadratic form in ϕ and that the bound-constraint $0 \le \phi_i \le 1$ is the only relevant restriction on ϕ it seems preferable to argue directly on the equation itself instead of using theorems primarily aimed at more difficult settings. The following argument is essentially a slight variation and extension of the proof of the closely related Theorem 16.4 in [54] for quadratic programming problems with linear side constraints in order to avoid the convexity assumption on the energy in that theorem[3].

[3] As there is a large body of (primarily algorithmic) literature on such indefinite quadratic programming problems, one could certainly also find theorems specifically adapted to this situation. Since the argument below is quite instructive for the particular problem considered here, in particular for $\frac{\pi}{\kappa}$ integer, no particular effort has been made though for finding a specific reference where the required conclusions are stated in a simple and explicit form.

As the phasefield energy $\mathcal{E}$ is quadratic in ϕ, the quadratic expansion

$$\mathcal{E}(\boldsymbol{\phi}+\delta\boldsymbol{\phi}) = \mathcal{E}(\boldsymbol{\phi}) + \mathcal{E}'(\boldsymbol{\phi})\cdot\delta\boldsymbol{\phi} + \frac{1}{2}\delta\boldsymbol{\phi}\cdot\mathcal{E}''(\boldsymbol{\phi})\cdot\delta\boldsymbol{\phi} \tag{A.23}$$

is exact around any given value of $\boldsymbol{\phi}$ and for any direction $\delta\boldsymbol{\phi}$. $\mathcal{E}'(\boldsymbol{\phi})$ is precisely the left-hand side of by Equation (6.38), such that, if the ϕ-profile solves the first-order necessary condition, one has

$$\mathcal{E}'(\boldsymbol{\phi})\cdot\delta\boldsymbol{\phi} = \sum_i \Big(-2\gamma\epsilon\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon}\gamma(1-2\phi_i)\Big)\delta\phi_i = \sum_i \big(\mu_i^- - \mu_i^+\big)\delta\phi_i,$$

together with the complementarity conditions $\mu_i^- \geq 0$, $\mu_i^- = 0$ if $\phi_i > 0$ resp. $\mu_i^+ \geq 0$, $\mu_i^+ = 0$ if $\phi_i = 1$. From this, $\mathcal{E}(\boldsymbol{\phi}+\delta\boldsymbol{\phi})$ in any such point can be rewritten as

$$\mathcal{E}(\boldsymbol{\phi}+\delta\boldsymbol{\phi}) = \mathcal{E}(\boldsymbol{\phi}) + \sum_i \big(\mu_i^- - \mu_i^+\big)\delta\phi_i + \frac{1}{2}\delta\boldsymbol{\phi}\cdot\big(\mathcal{E}''(\boldsymbol{\phi})\delta\boldsymbol{\phi}\big), \tag{A.24}$$

where $E''(\boldsymbol{\phi}) =: \boldsymbol{A}$ is the (constant) matrix characterized by the homogeneous form

$$(\boldsymbol{A}\boldsymbol{\psi})_i = -2\gamma\epsilon\frac{\psi_{i+1}-2\psi_i+\psi_{i-1}}{(\Delta x)^2} - \frac{32}{\pi^2\epsilon}\gamma\psi_i. \tag{A.25}$$

In order to show that a given phasefield profile satisfying the FONC is indeed a local minimizer (not necessarily strict), what needs to be shown is that

$$\mathcal{E}(\boldsymbol{\phi}+t\boldsymbol{d}) - \mathcal{E}(\boldsymbol{\phi}) = t\sum_i \big(\mu_i^- - \mu_i^+\big)d_i + \frac{1}{2}t^2\boldsymbol{d}\cdot\big(\mathcal{E}''(\boldsymbol{\phi})\boldsymbol{d}\big) \geq 0 \tag{A.26}$$

for all admissible directions $\boldsymbol{d}$ and $t > 0$ sufficiently small. A first important observation is that, if any multiplier $\mu_i^\pm$ is strictly positive (i.e. the constraint is strongly active), it suffices to consider directions $\boldsymbol{d}$ such that $d_i = 0$ for any such i. In fact, assuming e.g. $\mu_i^- > 0$, the complementarity condition enforces that $\phi_i = 0$ and thus the only admissible directions are such that $d_i \geq 0$. Similarly, d_i is necessarily non-positive if $\mu_i^+ > 0$. From this, it follows that $\big(\mu_i^- - \mu_i^+\big)d_i \geq 0$, and, if an entry d_i for any strongly active constraint is non-zero, will actually be strictly positive. Combined with the quadratic term being $\mathcal{O}(t^2)$, any such direction leads to a strict increase in the energy for sufficiently small $t > 0$, regardless of the values of the remaining d_j, $j \neq i$.
The only admissible directions in which the energy could potentially locally decrease are thus the ones which can vary freely if $0 < \phi_i < 1$ and, if there are weakly active constraint with $\phi_i = 0$ and $\mu_i^- = 0$ or $\phi_i = 1$ and $\mu_i^+ = 0$, the respective sign-restriction at these points. In addition, since the linear term $\sum_i \big(\mu_i^- - \mu_i^+\big)d_i$ vanishes for all such directions, one has

$$\mathcal{E}(\boldsymbol{\phi}+t\boldsymbol{d}) - \mathcal{E}(\boldsymbol{\phi}) = t^2\boldsymbol{d}\cdot\big(\mathcal{E}''(\boldsymbol{\phi})\boldsymbol{d}\big), \tag{A.27}$$

i.e. the sign of the energy difference depends only upon that of the quadratic term in $\boldsymbol{d}$.
All constraints in the bulk regions $i < 0$ and $i > N$ are strictly active (with the multiplier taking the value $\frac{16\gamma}{\pi^2\epsilon}$). Furthermore, from the derivation of the lower bound on N in Equation (6.40), the multipliers μ_0^- resp. μ_N^+ at the outermost points are strictly positive provided $N > \frac{\pi}{\kappa} - 1$ and only become zero if $N = \frac{\pi}{\kappa} - 1$. One therefore has to distinguish two cases:

1. If $\frac{\pi}{\kappa}$ is not an integer, both by N_{min} and N_{max} are strictly larger than $\frac{\pi}{\kappa} - 1$ and all active constraints are strongly active. From this, it follows that it suffices to focus on variations within the inner interface region, i.e. the interval $1 \leq i \leq N-1$, with all relevant search directions $\boldsymbol{d}$ satisfying $d_i = 0$, for $i \leq 0$ and $i \geq N$ and d_i arbitrary for $1 \leq i \leq N-1$. In

combination with Equation (A.27), this reduces the question of being a local minimizer to the subblock of the matrix $\boldsymbol{A} = \mathcal{E}''(\boldsymbol{\phi})$ corresponding to values $1 \leq i \leq N-1$ being positive (semi-)definite, resp. by Equation (A.25), to all eigenvalues of the linear difference equation

$$-2\gamma\epsilon\frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{(\Delta x)^2} - \frac{32}{\pi^2\epsilon}\gamma\psi_i \quad , 1 \leq i \leq N-1,$$

subject to the homogeneous Dirichlet boundary condition $\psi_0 = \psi_N = 0$ being non-negative. It is well-known that the eigenfunctions for such an operator are given by $\psi_i = c\sin\left(\frac{k\pi}{N}i\right), 1 \leq i \leq N$ with $1 \leq k \leq N-1$. Using $\sin\left(\frac{k\pi}{N}(i \pm 1)\right) = \sin\left(\frac{k\pi}{N}i\right)\cos\left(\frac{k\pi}{N}\right) \pm \cos\left(\frac{k\pi}{N}i\right)\sin\left(\frac{k\pi}{N}\right)$, it is easy to see that, for each k, the corresponding eigenvalue λ_k is given by

$$\lambda_k = 4\gamma\epsilon\frac{1 - \cos\left(\frac{k\pi}{N}\right)}{(\Delta x)^2} - \frac{32}{\pi^2\epsilon}\gamma.$$

Since $0 < \frac{k\pi}{N} < \pi$ for $1 \leq k \leq N-1$ and the cosine is decreasing on this interval, the lowest eigenvalue is given by $4\gamma\epsilon\frac{1-\cos\left(\frac{\pi}{N}\right)}{(\Delta x)^2} - \frac{32}{\pi^2\epsilon}\gamma = \frac{4\gamma\epsilon}{(\Delta x)^2}\left(1 - \cos\left(\frac{\pi}{N}\right) - \frac{8(\Delta x)^2}{\pi^2\epsilon^2}\right)$, which is non-negative if

$$\cos\left(\frac{\pi}{N}\right) \leq 1 - \frac{8(\Delta x)^2}{\pi^2\epsilon^2} = \cos(\kappa).$$

On the relevant domain, $\cos\left(\frac{\pi}{N}\right)$ is monotonically decreasing with N, from which it follows that $\lambda_1 \geq 0$ if $\frac{\pi}{N} \geq \kappa$ i.e. if $N \leq \frac{\pi}{\kappa}$. Since $\frac{\pi}{\kappa}$ was assumed non-integer, this inequality is strictly satisfied for $N = N_{min} = \lceil\frac{\pi}{\kappa} - 1\rceil$, whereas it is strictly violated for $N = N_{max} = \lceil\frac{\pi}{\kappa}\rceil$. It follows that the profile for $N = N_{min}$ is a strict local (and actually also global under the "constraint" of having an actual interface by the analysis in the last section) minimizer of $\mathcal{E}$. In contrast, the profile obtained for $N = N_{max}$ is not a local minimizer since adding an aribtrarily small multiple of the first eigenfunction to the profile - and thus in particular breaking its symmetry - will lead to a strict decrease of the energy.

2. If $\frac{\pi}{\kappa}$ is integer, the situation is somewhat more complex but also considerable more interesting. In this case, the two possible choices for N are given by $N_{min} = \lceil\frac{\pi}{\kappa} - 1\rceil = \frac{\pi}{\kappa} - 1$ and $N_{max} = \lceil\frac{\pi}{\kappa}\rceil = \frac{\pi}{\kappa}$. For the latter choice, both relevant multipliers μ_0^- and μ_N^+ are strictly positive with a value given, by Equation (6.39) combined with $\cot\left(\frac{\pi}{2}\right) = 0$ and equation (A.6) by

$$\mu_0^- = \mu_N^+ = \gamma\epsilon\frac{\cos(\kappa) - 1}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon^2}\gamma = \gamma\epsilon\frac{-\frac{8(\Delta x)^2}{\pi^2\epsilon}}{(\Delta x)^2} + \frac{16}{\pi^2\epsilon^2}\gamma = \frac{8}{\pi^2\epsilon^2}\gamma,$$

i.e. a value which corresponds precisely to half the for the multipliers in the bulk. As the constraints are therefore strictly active, one can apply the same analysis as for the previous case to conclude that the profile for $N = N_{max}$ is in fact a local minimum, but with the smallest eigenvalue of the corresponding eigenvalue problem being zero due to $\cos\left(\frac{\pi}{N}\right) = \cos(\kappa)$.
In contrast, for the choice $N = \frac{\pi}{2} - 1$, $\mu_0^- = \mu_N^+ = 0$. This implies that the values ϕ_0 and ϕ_N can also be varied without a first-order increase in the energy, but the entries d_0 and d_N of any admissible direction do have to satisfy a sign-restriction in order to remain compatible with the box-constraints. Here it is convenient to separately consider two types of search directions.

- The first one is formed by those directions for which at most one of the outermost interface points ϕ_0 or ϕ_N would move away from the respective constraint, i.e. directions with either $d_0 \geq 0$ and $d_N = 0$ or $d_0 = 0$ and $d_N \leq 0$. These directions are a subset of all the directions $\boldsymbol{d}$ for which at most $N_{min} + 1 = N_{max} = \frac{\pi}{\kappa}$ values are free

to vary, i.e. of those directions where either $d_0 = d_{N+1} = 0$ but d_i, $1 \le i \le N$ can be chosen freely or where d_i, $0 \le i \le N-1$ can be chosen freely but $d_{-1} = d_N = 0$. This "relaxed" problem is easily seen to reduce precisely to the eigenvalue analysis performed for $N = N_{max}$ and it follows that, except for those directions generated by negative multiples of $\sin\left(\frac{\pi}{N_{max}} i\right)$ (compatible with $d_0 = 0$ and $d_{N_{min}} = d_{N_{max}-1} < 0$) or positive multiples of $\sin\left(\frac{\pi}{N_{max}}(i-1)\right)$ (compatible with $d_0 > 0$ and $d_{N_{min}+1} = d_{N_{max}} = 0$) of the eigenfunction $\sin\left(\frac{\pi}{N_{max}} i\right)$ resp. its translate lead to a strict increase in the energy. As these particular eigenfunctions leave the energy invariant, it follows that $\boldsymbol{d} \cdot \left(\mathcal{E}''(\boldsymbol{\phi})\boldsymbol{d}\right) \ge 0$ for all such directions.

- It remains to verify that it is not possible to decrease the energy by choosing a direction with both $d_0 > 0$ and $d_{N_{min}} < 0$ corresponding to a simultaneous expansion of the inner interface in two directions. This cannot be verified using the same arguement, as fixing only $d_i = 0$ for $i \le -1$ and $i \ge N_{min} + 1 = N_{max}$ would lead to an eigenvalue problem on a domain which is "broader" than the maximal stable number of points given by N_{max} and would actually allow for an eigenfunction $\sin\left(\frac{\pi}{N_{max}+1}(i+1)\right)$, $-1 \le i \le N_{max}$ with a negative eigenvalue. Here the constaints on d_0 and $d_{N_{min}}$ are decisive, as these essentially exclude this type of direction since it either has to increase or decrease both ϕ_0 and $\phi_{N_{min}}$ and is therefore always incompatible with the one of the sign restriction.
  More precisely, from the analysis for $N = N_{max}$, the function $\hat{d}_i = \sin\left(\frac{\pi}{N_{max}} i\right)$, $0 \le i \le N_{max}$ is an eigenfunction with eigenvalue 0 for the restriction of the operator $\mathcal{E}''(\boldsymbol{\phi})$ to the range of indices $1 \le i \le N_{max} - 1$. In contrast, when broadening the region of interest by increasing the inner interval under consideration to $0 \le i \le N_{max}$ and extending this function (still denoted by $\hat{\boldsymbol{d}}$) with $\hat{d}_{-1} = 0$, this function is not an eigenfunction anymore, but satisfies (similarly to the homogeneous form of Equation (6.39) defining μ_0^-)

$$\begin{aligned} \left(\mathcal{E}''(\boldsymbol{\phi})\hat{\boldsymbol{d}}\right)_{i,1\le i\le N_{max}-1} &= 0 \quad \text{and} \\ \left(\mathcal{E}''(\boldsymbol{\phi})\hat{\boldsymbol{d}}\right)_{i,1\le i\le N_{max}-1} &= 2\gamma\epsilon\frac{d_1}{(\Delta x)^2} = 2\gamma\epsilon\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}, \end{aligned} \tag{A.28}$$

  i.e. the homogeneous difference operator applied to this extension of the eigenfunction still leaves the previous (for $N = N_{min}$) interior of the interface unaffected, but leads a positive contribution to the cell $i = 0$. An arbitrary search direction $\boldsymbol{d} = (d_i)_{0\le i\le N_{min}}$ can be decomposed as

$$d_i = \underbrace{d_i - \frac{d_{N_{min}}}{\sin\left(\frac{\pi}{N_{max}} N_{min}\right)}\hat{d}_i}_{=:d'_i} + \frac{d_{N_{min}}}{\sin\left(\frac{\pi}{N_{max}} N_{min}\right)} d_i = d'_i - \frac{d_{N_{min}}}{\sin\left(\frac{\pi}{N_{max}}\right)}\hat{d}_i,$$

  where use was made of $\sin\left(\frac{\pi}{N_{max}} N_{min}\right) = \sin\left(\frac{\pi}{N_{max}}(N_{max}-1)\right) = \sin\left(\pi - \frac{\pi}{N_{max}}\right) = \sin\left(\frac{\pi}{N_{max}}\right)$ and where $d'_{N_{min}} = 0$ by the choice of prefactor for $\hat{\boldsymbol{d}}$.
  Since $d_{N_{min}}$ is negative by assumption and $\sin\left(\frac{\pi}{N_{max}}\right)$ is positive, this decomposes any such direction $\boldsymbol{d}$ as the sum of one vector $\boldsymbol{d'}$ with $d'_{N_{min}} = 0$, $d'_0 = d_0 + c\sin\left(\frac{\pi}{N_{max}}\right) > d_0 > 0$ and a positive multiple $c = -\frac{d_{N_{min}}}{\sin\left(\frac{\pi}{N_{max}}\right)}$ of the extension of the eigenvector $\hat{\boldsymbol{d}}$.

From Equation (A.28), it further follows that

$$\begin{aligned}\left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_{1\le i\le N_{min}} &= \left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)_{1\le i\le N_{min}} \quad \text{and}\\ \left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_0 &= \left(\mathcal{E}''(\phi)\boldsymbol{d}'\right) + 2\gamma\epsilon c\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}\end{aligned} \tag{A.29}$$

since $\hat{\boldsymbol{d}}$ does not affect the results of the operator for any index $i > 0$. Taking the scalar product with $\boldsymbol{d}$, one has

$$\begin{aligned}\boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}\right) =& d_0\left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_0 + \sum_{i=1}^{N_{min}} d_i\left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_i\\ =& d_0\left(\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)_0 + 2\gamma\epsilon c\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}\right) + \sum_{i=1}^{N_{min}} d_i\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)_i\\ =& \underbrace{2\gamma\epsilon c\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}d_0}_{>0} + \boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right) > \boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right).\end{aligned}$$

Again with Equation (A.29), $\boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)$ can be expanded in a similar fashion using the symmetry of $\mathcal{E}''(\phi)$ as

$$\begin{aligned}\boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right) =& \left(\mathcal{E}''(\phi)\boldsymbol{d}\right)\cdot\boldsymbol{d}' = \left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_0 d_0' + \sum_{i=1}^{N_{min}}\left(\mathcal{E}''(\phi)\boldsymbol{d}\right)_i d_i'\\ =& \left(\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)_0 + 2\gamma\epsilon c\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}\right)d_0' + \sum_{i=1}^{N_{min}}\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)_i d_i'.\end{aligned}$$

The term $2\gamma\epsilon c\frac{\sin\left(\frac{\pi}{N_{max}}\right)}{(\Delta x)^2}d_0'$ is again strictly positive, such that $\boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}\right)$ can finally be esimated from below as $\boldsymbol{d}\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}\right) > \boldsymbol{d}'\cdot\left(\mathcal{E}''(\phi)\boldsymbol{d}'\right)$. Since the last entry of $\boldsymbol{d}'$ is in addition zero by construction, the sign of this term in fact only depdends on the sign of this quadratic form restricted to the subvector $\left(\boldsymbol{d}'\right)_{0\le i\le N_{min}-1}$, i.e. on vectors which have at most one additional non-zero entry as compared to the original innner transition region. This is precisely the question already investigated in the previous point, finally showing that the "standard" profile with $N = N_{min}$ is also a local (non-strict) minimum.

Bibliography

[1] Patrick Altschuh. Skalenübergreifende analyse makroporöser membranen im kontext digitaler zwillinge, 2020. Dissertation, Karlsruher Institut für Technologie (KIT).

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Clarendon Press, Oxford, 2000.

[3] Kais Ammar, Benoît Appolaire, Georges Cailletaud, and Samuel Forest. Combining phase field approach and homogenization methods for modelling phase transformation in elastoplastic media. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 18(5-6):485–523, 2009.

[4] PG Kubendran Amos, Ephraim Schoof, Daniel Schneider, and Britta Nestler. Chemo-elastic phase-field simulation of the cooperative growth of mutually-accommodating widmanstätten plates. *Journal of Alloys and Compounds*, 767:1141–1154, 2018.

[5] PG Kubendran Amos, Ephraim Schoof, Nick Streichan, Daniel Schneider, and Britta Nestler. Phase-field analysis of quenching and partitioning in a polycrystalline fe-c system under constrained-carbon equilibrium condition. *Computational Materials Science*, 159:281–296, 2019.

[6] Hedy Attouch, Giuseppe Buttazzo, and Gérard Michaille. *Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*. SIAM, Philadelphia, Pa., 2. ed. edition, 2014.

[7] Jean-Pierre Aubin and Hélène Frankowska. *Set-valued analysis*. Birkhäuser, Boston, Mass., 2009.

[8] H. Bauschke and P. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer, Cham, 2 edition, 2017.

[9] Marouen Ben Said, Michael Selzer, Britta Nestler, Daniel Braun, Christian Greiner, and Harald Garcke. A phase-field approach for wetting phenomena of multiphase droplets on solid surfaces. *Langmuir*, 30(14):4033–4039, 2014.

[10] Benzi, Michele and Golub, Gene H and Liesen, Jörg. Numerical solution of saddle point problems. *Acta numerica*, 14:1–137, 2005.

[11] Bertsekas, Dimitri P. *Nonlinear programming*. Athena scientific, Belmont, Mass., 1999.

[12] L. Blank et al. Allen-Cahn and Cahn-Hilliard variational inequalities solved with optimization techniques. *International Series of Numerical Mathematics*, 160:21–35, 2012.

[13] F. B. Bonnans et al. *Numerical Optimization - Theoretical and Practical Aspects*. Universitext. Springer, Berlin, 2 edition, 2006.

[14] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer, New York, 2000.

[15] A. Braides. *Gamma-convergence for beginners*, volume 22 of *Oxford lecture series in mathematics and its applications*. Oxford University Press, Oxford, 2002.

[16] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, New York, NY, 2010.

[17] Céa, Jean and Glowinski, Roland. Sur des méthodes d'optimisation par relaxation. *Revue française d'automatique informatique recherche opérationnelle. Mathématique*, 7(R3):5–31, 1973.

[18] Armen G. Chačaturijan. *Theory of structural transformations in solids*. Wiley, New York [u.a.], 1983.

[19] Abhik Choudhury and Britta Nestler. Grand-potential formulation for multicomponent phase transformations combined with thin-interface asymptotics of the double-obstacle potential. *Physical Review E*, 85(2):021602, 2012.

[20] Clarke, F. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, London, 2013.

[21] Klaus Deckelnick, Gerhard Dziuk, and Charles M Elliott. Computation of geometric partial differential equations and mean curvature flow. *Acta numerica*, 14:139, 2005.

[22] N Dinh and V Jeyakumar. Farkas' lemma: three decades of generalizations for mathematical optimization. *Top*, 22(1):1–22, 2014.

[23] A Durga, Patrick Wollants, and Nele Moelans. Evaluation of interfacial excess contributions in different phase-field models for elastically inhomogeneous systems. *Modelling and simulation in materials science and engineering*, 21(5):055018, 2013.

[24] A Durga, Patrick Wollants, and Nele Moelans. A quantitative phase-field model for two-phase elastically inhomogeneous systems. *Computational Materials Science*, 99:81–95, 2015.

[25] J Eiken, B Böttger, and I Steinbach. Multiphase-field approach for multicomponent alloys with extrapolation scheme for numerical application. *Physical review E*, 73(6):066122, 2006.

[26] I. Ekeland and R. Témam. *Convex analysis and variational problems*. Classics in Applied Mathematics. SIAM, Philadelphia, Pa., 1999.

[27] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementary Problems*, volume 1 of *Springer Series in Operations Research*. Springer, New York, 2007.

[28] Avner Friedman. *Variatonal Principles and Free Boundary Problems*. Dover Publications, Mineola, NY, dover ed. edition, 2010.

[29] H. Garcke et al. Allen-Cahn systems with volume constraints. *Mathematical Models and Methods in Applied Sciences*, 18(8):1347–1381, 2008.

[30] C. Gräser. Convex minimization and phase field models, 2011. Dissertation Freie Universität Berlin.

[31] C. Gräser, R. Kornhuber, and U. Sack. Time discretizations of anistropic Allen-Cahn equations. *IMA Journal of Numerical Analysis*, 33(48):1226–1244, 2013.

[32] Carsten Gräser and Ralf Kornhuber. On preconditioned Uzawa-type iterations for a saddle point problem with inequality constraints. In *Domain decomposition methods in science and engineering XVI*, pages 91–102. Springer, Berlin, Heidelberg, 2007.

[33] Carsten Gräser and Oliver Sander. Truncated nonsmooth Newton multigrid methods for simplex-constrained minimization problems. *Preprint 384, IGPM Aachen*, 2014.

[34] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 69 of *Classics in Applied Mathematics*. SIAM, Philadelphia, Pa., 2011.

[35] Christoph Herrmann, Ephraim Schoof, Daniel Schneider, Felix Schwab, Andreas Reiter, Michael Selzer, and Britta Nestler. Multiphase-field model of small strain elasto-plasticity according to the mechanical jump conditions. *Computational Mechanics*, 62(6):1399–1412, 2018.

[36] Johannes Hötzer. *Massiv-parallele und großskalige Phasenfeldsimulationen zur Untersuchung der Mikrostrukturentwicklung*. Schriftenreihe des Instituts für Angewandte Materialien ; Band 70. KIT Scientific Publishing, Karlsruhe, 2017. Dissertation, Karlsruher Institut für Technologie (KIT).

[37] K. Ito and K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control. SIAM, Philadelphia, Pa., 2008.

[38] Boško S Jovanović and Endre Süli. *Analysis of finite difference schemes: for linear partial differential equations with generalized solutions*, volume 46. Springer Science & Business Media, London, 2013.

[39] Michael Kellner. Modellierung mehrkomponentiger Materialsysteme für die Phasenfeldmethode und Analyse der simulierten Mikrostrukturen, 2020. Dissertation, Karlsruher Institut für Technologie (KIT).

[40] Seong Gyoon Kim, Dong Ik Kim, Won Tae Kim, and Yong Bum Park. Computer simulations of two-dimensional and three-dimensional ideal grain growth. *Physical Review E*, 74(6):061605, 2006.

[41] Kim, Seong Gyoon and Kim, Won Tae and Suzuki, Toshio. Interfacial compositions of solid and liquid in a phase-field model with finite interface thickness for isothermal solidification in binary alloys. *Physical Review E*, 58(3):3316–3323, 1998.

[42] Kim, Seong Gyoon and Kim, Won Tae and Suzuki, Toshio. Phase-field model for binary alloys. *Physical Review E*, 60(6):7186–7197, 1999.

[43] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications*. Classics in Applied Mathematics. SIAM, Philadelphia, Pa., 2000.

[44] Cornelius Lanczos. *The variational principles of mechanics*. Dover, New York, 2012.

[45] Stephan Luckhaus and Luciano Modica. The gibbs-thompson relation within the gradient theory of phase transitions. *Archive for Rational Mechanics and Analysis*, 107(1):71–83, 1989.

[46] Luenberger, David G and Ye, Yinyu and others. *Linear and nonlinear programming*, volume 2. Springer, Cham, 1984.

[47] G. Dal Maso. *An Introduction to Gamma-Convergence*. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser, Boston, 1993.

[48] W. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, Cambridge, 2000.

[49] L. Modica. Gradient theory of phase transitions with boundary contact energy. *Annales de l'I.H.P.*, 4(5):487–512, 1987.

[50] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, 1987.

[51] J Mosler, O Shchyglo, and H Montazer Hojjat. A novel homogenization method for phase field approaches based on partial rank-one relaxation. *Journal of the Mechanics and Physics of Solids*, 68:251–266, 2014.

[52] Britta Nestler, Harald Garcke, and Björn Stinner. Multicomponent alloy solidification: phase-field modeling and simulations. *Physical Review E*, 71(4):041609, 2005.

[53] Britta Nestler, Frank Wendler, Michael Selzer, Björn Stinner, and Harald Garcke. Phase-field model for multiphase systems with preserved volume fractions. *Physical Review E*, 78(1):011604, 2008.

[54] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, 2 edition, 2006.

[55] James M Ortega and Werner C Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30. Siam, Philadelphia, Pa., 1970.

[56] M. Plapp. Unified derivation of phase-field models for alloy solidification from a grand-potential functional. *Physical Review E*, 84(031601):1–15, 2011.

[57] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, Berlin, 2009.

[58] Ruszczyński, Andrzej P and Ruszczynski, Andrzej. *Nonlinear optimization*, volume 13. Princeton university press, Princeton, 2006.

[59] Alexander A Samarskii. *The theory of difference schemes*, volume 240. CRC Press, New York, 2001.

[60] Lavinia Sarbu. *Primal-dual active set methods for Allen-Cahn variational inequalities*. PhD thesis, University of Sussex, 2010.

[61] Daniel Schneider. Phasenfeldmodellierung mechanisch getriebener grenzflächenbewegungen in mehrphasigen systemen, 2016. Dissertation, Karlsruher Institut für Technologie (KIT).

[62] Daniel Schneider, Ephraim Schoof, Oleg Tschukin, Andreas Reiter, Christoph Herrmann, Felix Schwab, Michael Selzer, and Britta Nestler. Small strain multiphase-field model accounting for configurational forces and mechanical jump conditions. *Computational Mechanics*, 61(3):277–295, 2018.

[63] Daniel Schneider, Felix Schwab, Ephraim Schoof, Andreas Reiter, Christoph Herrmann, Michael Selzer, Thomas Böhlke, and Britta Nestler. On the stress calculation within phase-field approaches: a model for finite deformations. *Computational Mechanics*, 60(2):203–217, 2017.

[64] Daniel Schneider, Oleg Tschukin, Abhik Choudhury, Michael Selzer, Thomas Böhlke, and Britta Nestler. Phase-field elasticity model based on mechanical jump conditions. *Computational Mechanics*, 55(5):887–901, 2015.

[65] Ephraim Schoof. Chemomechanische modellierung der wärmebehandlung von stählen mit der phasenfeldmethode, 2020. Dissertation, Karlsruher Institut für Technologie (KIT).

[66] Ephraim Schoof, Christoph Herrmann, Nick Streichhan, Michael Selzer, Daniel Schneider, and Britta Nestler. On the multiphase-field modeling of martensitic phase transformation in dual-phase steel using j2-viscoplasticity. *Modelling and Simulation in Materials Science and Engineering*, 27(2):025010, 2019.

[67] Michael Selzer. Mechanische und strömungsmechanische topologieoptimierung mit der phasenfeldmethode, 2014. Dissertation, Karlsruher Institut für Technologie (KIT).

[68] I Steinbach and F Pezzolla. A generalized field method for multiphase transformations using interface fields. *Physica D: Nonlinear Phenomena*, 134(4):385–393, 1999.

[69] Ingo Steinbach and Markus Apel. Multi phase field model for solid state transformation with elastic strain. *Physica D: Nonlinear Phenomena*, 217(2):153–160, 2006.

[70] Philipp Steinmetz. Simulation der bei der gerichteten erstarrung ternärer eutektika entstehenden mikrostruktur mit der phasenfeldmethode, 2017. Dissertation, Karlsruher Institut für Technologie (KIT).

[71] Bjorn Stinner, Britta Nestler, and Harald Garcke. A diffuse interface model for alloys with multiple components and phases. *SIAM Journal on Applied Mathematics*, 64(3):775–799, 2004.

[72] Ying Sun and Christoph Beckermann. Sharp interface tracking using the phase-field equation. *Journal of Computational Physics*, 220(2):626–653, 2007.

[73] R. Temam. *Navier-Stokes Equations - Theory and Numerical Analysis*. AMS Chelsea Publishing, Providence, 2001.

[74] Oleg Tschukin. Phase-field modelling of welding and of elasticity-dependent phase transformations, 2017. Dissertation, Karlsruher Institut für Technologie (KIT).

[75] Oleg Tschukin, Alexander Silberzahn, Michael Selzer, Prince GK Amos, Daniel Schneider, and Britta Nestler. Concepts of modeling surface energy anisotropy in phase-field approaches. *Geothermal Energy*, 5(1):19, 2017.

[76] R. S. Varga. *Matrix Iterative Analysis*, volume 27 of *Springer Series in Computational Mathematics*. Springer, Berlin, 2000.

[77] P. S. Vassilevski. *Multilevel Block Factorization Preconditioners*. Springer, New York, 2008.

[78] Alexander Vondrous. Grain growth behavior and efficient large scale simulations of recrystallization with the phase-field method, 2014. Dissertation, Karlsruher Institut für Technologie (KIT).

[79] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer, Berlin, 6 edition, 1980.

[80] Zeidler, E. *Nonlinear functional analysis and its applications: III: variational methods and optimization*. Springer Science & Business Media, New York, 2013.