Numerical methods of localization of Wannier functions in modeling of Photonic Crystals

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik der
Karlsruher Institut für Technologie (KIT)
genehmigte

DISSERTATION

von

Dipl. - Math. Tatiana Bulovyatova

aus Novosibisrk, Russland

Tag der mündlichen Prüfung: 26. Januar 2011
Referent: Prof. Dr. Willy Dörfler
Korreferent: Prof. Dr. Christian Wieners

Contents

Chapter 1

Introduction

1.1 Abstract

Photonic Crystals, or PCs, are artificially developed optical materials with periodic dielectricity. They represent an innovative type of structures with tailor designed properties and could be considered as semiconductors of the next generation. They are hoped to become the first step to a technological revolution in computing, communication, optics and other spheres.

The flow of light is much faster then the electric current, and the idea to use laser light instead of electricity in information processing and transmission is very tempting. Investigation of new optical materials permitting to capture, to direct and to control the propagation of the electromagnetic waves is a challenge for physicists and mathematicians. Theoretically, an appropriate modification of material composition, symmetry, dielectric periodicity with embedding of defect structures can provide a creation of a semiconductor applicable to any specific purpose. But because of their microscopic scale comparable with the wavelength of light, Photonic Crystals are very complicated to fabricate. For this reason a mathematical simulation plays such a significant role.

For description of light propagation, different techniques have been applied by researchers. For example, the Finite Difference Time Domain (FDTD) approach was considered by A.

Taflove in [Taf88]. Another route is to implement Galerkin-type methods with an appropriate set of basis functions, like, for instance, plane waves. However, these ways do not seem to be very useful in modeling of complicated structures with piecewise-continuous dielectric permittivity because of huge computational resources that they require.
It would be preferable if a simulation method could take into account the Photonic Crystal structure that it deals with. Fortunately, it is possible. There exist basis functions for a Galerkin approximation which already contain information about dielectric and symmetric properties of a particular PC — these are the eigenfunctions of this PC called the Bloch waves. Although these functions are wide-spread and non-localized, one can use the advantages that they give to build other, well-localized bases.
The Wannier functions have been introduced by G. Wannier in [Wan37] for description of insulating crystals and afterwards have been successfully employed in modeling of electronic orbitals. The idea to draw an analogy between Photonic Crystals and real crystals and to apply the same functions for PCs first appeared in 2003 in a work of the group of K. Busch [BMGM$^+$03] and at the same time in a paper of Whittaker and Croucher [WC03]. The Wannier functions have several advantages which make them very attractive for PC simulation: first, they form a complete orthonormal basis applicable for Galerkin-type methods; second, being calculated by definition via the Bloch waves, they already contain the information about the layout of a given PC; and third, they are not unique and thus can be localized — the existence of exponentially decaying Wannier functions is proved. The last property is significant: due to the localization of a basis set, the computations would require less resources which is highly desirable.

The aim of this Ph.D. thesis is to summarize and unify knowledge about the Wannier functions and to explore several algorithms for construction of maximally localized Wannier functions in application to Photonic Crystals. In Chapter 2 we state the problem to consider within the thesis — the Transverse Magnetic for of the Maxwell's equations — and analyse the properties of its eigenfuctions, the Bloch waves. In Chapter 3 we

define the Wannier functions as the result of application of the inverse Floquet transform to the Bloch waves. We study the existence of real, (anti-) symmetric, exponentially decaying Wannier functions for different types of crystals and give a method of computing such functions explicitly as the eigenfunctions of the projected position operator in one-dimensional crystals. This method is implemented numerically. In Chapter 4 we introduce the spread functional as a criterion of localization of the Wannier functions and explore its properties. Chapter 5 is dedicated to an algorithm of minimization of the spread functional given by Marzari and Vanderbilt and based on unitary transform of the Bloch waves. This algorithm we implement numerically in 1D and 2D. In Chapter 6 we consider unitary transform of the Wannier functions and the method of simultaneous diagonalization of a set of matrices developed by Gygi, Fatterbert and Schwegler. The numerical results in 1D and 2D are also given. Chapter 7 describes a more general approach then in Chapter 6 — minimization of the spread under sum-unitary transform of the Wannier functions which we first introduce in the framework of this thesis. Again, we compute it numerically for 1D and 2D Photonic Crystals. Chapter 8 refers to the application of the Wannier functions in physics — particularly, in simulation of Photonic Crystals with point defects. Finally, the conclusions of the work are given in Chapter 9. Some numerical aspects of the presented methods one can find in Appendix.

1.2 Acknowledgment

I appreciate this opportunity to devote the present thesis to my husband Dr. Alexander Bulovyatov who has inspired me.

I want to thank foremost my scientific advisor Prof. Dr. Willy Dörfler who supported and guided my work. He could always find time to discuss the difficulties and the achievements, and his help was really priceless.

I express my gratitude to the German Research Foundation (DFG) for the Ph.D.-scholarship which was an important financial assistance during my work.

The Research Training Group where I spent so much time — GRK1294 — is also worth to be mentioned. It was a pleasure to work with such smart and nice people, and I thank them for our everyday office life, fruitful discussions and especially the coffee-breaks.

Last, but not least, I would like to say a huge Thank You to my parents Liubov and Sergey Abramov who have always loved and encouraged me and did their best for my education.

1.3 List of notations

Symbol	*Meaning*
$\delta_{m,n}$	the Kronecker symbol: equals to 1, if $m = n$, and 0 otherwise
$\delta(f)$	the Dirac distribution: generalized function with a property $\delta(f) = f(0)$ for continuous f and $\int_{-\infty}^{\infty} f(\mathrm{r})\delta(\mathrm{r} - a)\,\mathrm{d}\,\mathrm{r} = f(a)$
$\mathcal{U}f$	Floquet transform acting on a function f
d	the dimensionality of the problem (here 1 or 2)
$\mathrm{r} \in \mathbb{R}^d$	a space vector, d-dimensional
WSC	Wigner-Seitz cell — a fundamental domain in $\mathbb{R}^d$: here $[0, 1]^d$
Γ	periodicity lattice in real space; here $\Gamma = \mathbb{Z}^d$
$\mathrm{R} \in \Gamma$	a node of the lattice Γ
∇	a d-dimensional gradient operator w.r.t. variable r
Δ	a Laplacian operator $\nabla \cdot \nabla$ w.r.t. vector r
$\mathrm{k} \in \mathbb{R}^d$	a dual space vector, also called a crystal momentum, d-dimensional
BZ	the (first) Brillouin zone — a fundamental domain in the reciprocal space: here $[-\pi, \pi]^d$
Γ^*	reciprocal lattice; here $\Gamma^* = 2\pi\mathbb{Z}^d$
$\mathrm{K} \in \Gamma^*$	a node of the reciprocal lattice Γ^*
∇_{k}	a d-dimensional gradient operator w.r.t. variable k
MP	Monkhorst-Pack mesh in BZ: a mesh shifted by half-meshsize from the boundary
S_{k}	a stencil of the closest neighbour mesh nodes for a node k
$\varepsilon(\cdot)$	crystal permittivity, periodic w.r.t. Γ: $\varepsilon(\mathrm{r}) = \varepsilon(\mathrm{r} + \mathrm{R})$
$\langle \cdot, \cdot \rangle_{D,\varepsilon}$	an inner product $\langle f, g \rangle_{D,\varepsilon} = \int_D f^*(\mathrm{r})\varepsilon(\mathrm{r})g(\mathrm{r})\,\mathrm{d}\,\mathrm{r}$; D is either WSC or $\mathbb{R}^d$

Symbol	*Meaning*
$\langle\cdot,\cdot\rangle_D$	a shortend notation for $\langle\cdot,\cdot\rangle_{D,\varepsilon}$ in cases when it is obvious
$\mathrm{L}^2_\varepsilon(D)$	function space $\mathrm{L}^2(D)$ with a weighted inner product $\langle\cdot,\cdot\rangle_\varepsilon$; D is WSC or $\mathbb{R}^d$
$\lambda_{n,\mathrm{k}}$	the n-th eigenvalue of **TM** problem corresponding to the n-th Bloch wave
$\psi_{n,\mathrm{k}}(\cdot)$	the n-th Bloch wave with crystal momentum k
$u_{n,\mathrm{k}}(\cdot)$	the periodic part of the n-th Bloch wave: $u_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}}\psi_{n,\mathrm{k}}(\mathrm{r})$
$w_{n,0}(\cdot)$	the n-th Wannier function centered at 0
$w_{n,\mathrm{R}}(\cdot)$	the n-th Wannier function centered at $\mathrm{R}\in\Gamma$; $w_{n,\mathrm{R}}(\mathrm{r}) = w_{n,0}(\mathrm{r}-\mathrm{R})$
β	a composite band
N_β	the dimensionality of the composite band β
β_ψ	a set of Bloch waves corresponding to the eigenvalues of a composite band β
P^{k}	projection onto the span of Bloch waves of a band β, $P^{\mathrm{k}} = \sum_{n=n_1}^{n_1+N_\beta-1} \psi_{n,\mathrm{k}} \otimes \psi_{n,\mathrm{k}}$
P_u^{k}	projection onto the periodic parts of Bloch waves of a band β, $P_u^{\mathrm{k}} = \sum_{n=n_1}^{n_1+N_\beta-1} u_{n,\mathrm{k}} \otimes u_{n,\mathrm{k}}$
P	projection onto the span of Wannier functions of a band β, $P = \sum_{n=n_1}^{n_1+N_\beta-1} w_{n,0} \otimes w_{n,0}$
Q	projection onto the span of other bands except β: $Q = \mathrm{Id} - P$
$\widehat{R}$	projected position operator, $\widehat{R} = P\mathrm{r}P$
Ω	spread functional — used as a measure of localization of a set of (Wannier) functions
Ω_{I}	gauge invariant part of the spread functional
Ω_{D}	band-diagonal part of the spread functional
Ω_{OD}	band-off-diagonal part of the spread functional
$\Omega_{\mathrm{D,OD}}$	$\Omega_{\mathrm{D}} + \Omega_{\mathrm{OD}}$
$\Omega_{\mathrm{I,OD}}$	$\Omega_{\mathrm{I}} + \Omega_{\mathrm{OD}}$

Symbol	*Meaning*
$\bar{\mathrm{r}}_n$	a special notation for $\langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \in \mathbb{R}^d$
$\langle r^2 \rangle_n$	a special notation for $\langle r^2 w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \in \mathbb{R}_+$
$\bar{\mathrm{r}}_n^{\triangle}$	a discrete analogue for $\bar{\mathrm{r}}_n$ (approximation)
$\langle r^2 \rangle_n^{\triangle}$	a discrete analogue for $\langle r^2 \rangle_n$
U^{k}	a unitary transform of the Bloch waves of a band β, a matrix $N_\beta \times N_\beta$
W^{R}	a sum-unitary transform of the Wannier functions of a band β, a matrix $N_\beta \times N_\beta$

Chapter 2

Bloch waves

2.1 Maxwell's equations. Transverse Magnetic and Transverse Electric problems

At a macroscopic level a *Photonic Crystal* is a low-loss dielectric medium (one-, two- or three-dimensional) with periodic electric permittivity, without free charges and non-conducting. In some sense it is an optical analogue of an electrical wire where the information is transported by photons instead of electrons.

From mathematical point of view, a Photonic Crystal is an infinite linear periodic medium where the propagation of light is described by the *time-harmonic Maxwell's equations*:

$$-\nabla \times \mathbf{E}(\mathrm{r}) = \frac{\mathrm{i}\omega}{c}\mathbf{H}(\mathrm{r}), \tag{2.1.1}$$

$$\nabla \times \mathbf{H}(\mathrm{r}) = \frac{\mathrm{i}\omega}{c}\varepsilon(\mathrm{r})\mathbf{E}(\mathrm{r}), \tag{2.1.2}$$

$$\nabla \cdot (\varepsilon(\mathrm{r})\mathbf{E}(\mathrm{r})) = 0. \tag{2.1.3}$$

$$\nabla \cdot \mathbf{H}(\mathrm{r}) = 0, \tag{2.1.4}$$

Here $\mathbf{E}$ is the *electric field*, $\mathbf{H}$ is the *magnetic field*, ω is frequency, c is the speed of light and ε is the *electric permittivity* of the medium. Here we restrict electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ to be *time-independent*, which means that the medium properties depend

on position in space r only. Most of the described Photonic Crystals have no magnetic properties, for this reason in our considerations the *magnetic permeability* is assumed to be 1 everywhere (and thus excluded from the equations above). We also assume that the electric permittivity ε is scalar and independent of the third space coordinate r_3. Therefore, **E** and **H** do not depend on it either.

Since the Maxwell's equations (2.1.1)-(2.1.4) are formulated for a vector

$$(\mathbf{E}, \mathbf{H}) = (E_1, E_2, E_3, H_1, H_2, H_3),$$

an exclusion of the third space coordinate r_3 means that the waves propagate in the plain $(\mathrm{r}_1, \mathrm{r}_2)$ and therefore it is sufficient to study a vector of the form $(E_1, E_2, 0, 0, 0, H_3)$ or $(0, 0, E_3, H_1, H_2, 0)$ [Kuc01]. Thus the Maxwell's equations are getting decomposed into a direct sum of two well-known forms:

$$\begin{aligned} -\nabla \cdot \left(\frac{1}{\varepsilon(\mathrm{r})} \nabla \psi(\mathrm{r}) \right) &= \frac{\omega^2}{c^2} \psi(\mathrm{r}) \quad \textit{(Transverse Electric)}, \\ -\Delta \psi(\mathrm{r}) &= \frac{\omega^2}{c^2} \varepsilon(\mathrm{r}) \psi(\mathrm{r}) \quad \textit{(Transverse Magnetic)}. \end{aligned}$$

The first equation (**TE**) describes polarized waves which have the electric field orthogonal to the the r_3 axis while the magnetic field propagates along it, in this case $\psi = H_3$. Analogously, the second equation (**TM**) describes the waves with magnetic field orthogonal to r_3 and the electric field parallel to it, and $\psi = E_3$. Here the space vector r and the operators ∇ and Δ are two-dimensional in terms of two remaining space coordinates r_1 and r_2:

$$\mathrm{r} = \begin{pmatrix} \mathrm{r}_1 \\ \mathrm{r}_2 \end{pmatrix}, \quad \nabla = \begin{pmatrix} \frac{\partial}{\partial \mathrm{r}_1} \\ \frac{\partial}{\partial \mathrm{r}_2} \end{pmatrix}, \quad \Delta = \frac{\partial^2}{\partial \mathrm{r}_1^2} + \frac{\partial^2}{\partial \mathrm{r}_2^2}.$$

In the framework of this thesis we will follow the second form (*Transverse Magnetic* problem). We will take it up as an *eigenvalue problem* in the following sense: for a fixed piecewise periodic function ε consider a partial differential operator with periodic coefficient $L_{\mathbf{TM}} = -\Delta/\varepsilon$ and find its eigenvalues $\lambda = \omega^2/c^2$ and the corresponding eigenfunctions ψ:

$$-\Delta \psi(\mathrm{r}) = \lambda \varepsilon(\mathrm{r}) \psi(\mathrm{r}). \tag{2.1.5}$$

Considerations analogous to those given here may also be implemented for the *Transverse Electric* case.

2.2 Basic notations

In this section we introduce some notations which will be used in all parts of our work. Let d be a dimensionality of a crystal, it can be equal to 1 or 2.

A Photonic Crystal that we consider is a medium with a certain periodicity. In practice it means that cylinders of a particular dielectric material are embedded into another dielectric material so that they form a periodic structure. Let us consider a uniform square periodic lattice with the centers of the cylinders in the midpoints between its nodes. Without loss of generality, let it be an integer-valued lattice $\mathbb{Z}^d$ in real coordinate space $\mathbb{R}^d$. This lattice is often called *Bravais lattice* (henceforth denoted by "Γ"). The smallest cell, or *a fundamental domain*, of this periodic structure is a unit square. To be precise, let us consider the square $[0,1]^d$. In physics this is called *Wigner-Seitz cell* (in our further notations "WSC"). This domain is called fundamental, because the whole space $\mathbb{R}^2$ is formed of an infinite number of its repetitions so that

$$\mathbb{R}^d = \bigcup_{\mathrm{R}\in\Gamma} (\mathrm{WSC} + \mathrm{R})$$

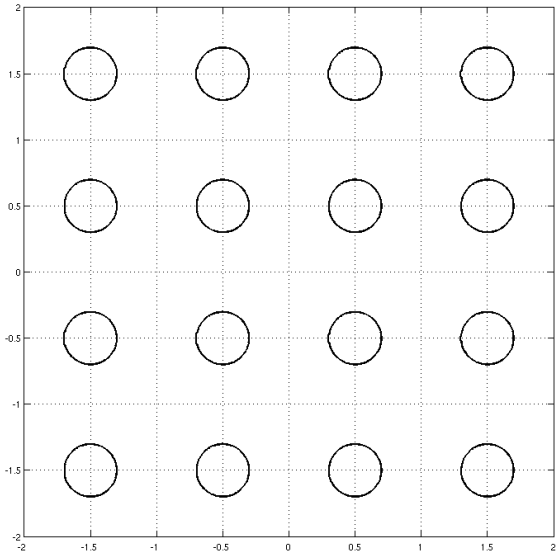
and every vector $\mathrm{r} \in \mathbb{R}^d$ either belongs to the Wigner-Seitz cell $\mathrm{r} \in \mathrm{WSC}$ or has a "representative" there: $\mathrm{r} = \mathrm{r}' + \mathrm{R}$ with $\mathrm{r}' \in \mathrm{WSC}$ and $\mathrm{R} \in \Gamma$. Therefore, a Photonic Crystal periodic with respect to Γ is completely described by its WSC and can be easily reconstructed from it.

Now we define the spaces which we are going to work with. Let c_0 be the center of the Wigner-Seitz cell ($\frac{1}{2}$ in 1D or $\left[\frac{1}{2}, \frac{1}{2}\right]$ in 2D). Let ε be a positive piecewise constant periodic function, in particular, for some $\rho_0 < \frac{1}{2}$, let

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & \mathrm{r} \in \mathrm{WSC}, \quad |\mathrm{r} - c_0| > \rho_0 \\ \varepsilon_0 \neq 1, \varepsilon_0 > 0, & |\mathrm{r} - c_0| \leq \rho_0 \end{cases} \tag{2.2.1}$$

and $\varepsilon(\mathrm{r}+\mathrm{R}) = \varepsilon(\mathrm{r})$ for $\mathrm{R} \in \Gamma$ (see Figure 2.1).

Figure 2.1: A Photonic Crystal structure



Define ε-weighted inner products by

$$\langle f, g\rangle_{\mathrm{WSC},\varepsilon} = \int_{\mathrm{WSC}} f^*(\mathrm{r})\varepsilon(\mathrm{r})g(\mathrm{r})\,\mathrm{d}\,\mathrm{r}, \tag{2.2.2}$$

$$\langle f, g\rangle_{\mathbb{R}^d,\varepsilon} = \int_{\mathbb{R}^d} f^*(\mathrm{r})\varepsilon(\mathrm{r})g(\mathrm{r})\,\mathrm{d}\,\mathrm{r}. \tag{2.2.3}$$

Let $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ be the space $\mathrm{L}^2(\mathbb{R}^d)$ with the inner product $\langle\cdot,\cdot\rangle_{\mathbb{R}^d,\varepsilon}$ and analogously $\mathrm{L}^2_\varepsilon(\mathrm{WSC})$ be the space $\mathrm{L}^2(\mathrm{WSC})$ with the inner product $\langle\cdot,\cdot\rangle_{\mathrm{WSC},\varepsilon}$. Obviously, these are Hilbert spaces with corresponding norms $\|f\|_{\mathbb{R}^d,\varepsilon} = \sqrt{\langle f, f\rangle_{\mathbb{R}^d,\varepsilon}}$ and $\|f\|_{\mathrm{WSC},\varepsilon} = \sqrt{\langle f, f\rangle_{\mathrm{WSC},\varepsilon}}$. The mapping $f \mapsto f/\sqrt{\varepsilon}$ is an isometry between $\mathrm{L}^2(\Omega)$ and $\mathrm{L}^2_\varepsilon(\Omega)$ where Ω is either WSC or $\mathbb{R}^d$.

For the cases where the applicable domain of the functions f and g is evident, we will use simplified notations $\langle f, g\rangle_\varepsilon$ and $\|f\|_\varepsilon$.

2.3 Floquet theory and the Bloch waves

Now let us introduce a transform $\mathcal{U}$ on the Bravais lattice which is called *Floquet transform*. Let $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$, then

$$[\mathcal{U}f](\mathrm{r},\mathrm{k}) := \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f(\mathrm{r}-\mathrm{R}). \tag{2.3.1}$$

This is an analogue of the Fourier transform in the periodic case. Note that in contrast to the Fourier transform, the resulting function $\mathcal{U}f$ depends not only on the dual variable k, but on r as well. k is called *quasi-momentum*.

We review shortly the main properties of this transform; for a more detailed consideration one can refer to the literature, e.g. [Kuc01].

First, the *Floquet transform is quasi-periodic by* r *with respect to* Γ, this property is called *Floquet condition*:

$$[\mathcal{U}f](\mathrm{r}+\mathrm{R},\mathrm{k}) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,[\mathcal{U}f](\mathrm{r},\mathrm{k}), \quad \mathrm{R}\in\Gamma. \tag{2.3.2}$$

Second, the *Floquet transform is periodic by the quasi-momentum* k:

$$[\mathcal{U}f](\mathrm{r},\mathrm{k}+\mathrm{K}) = [\mathcal{U}f](\mathrm{r},\mathrm{k}), \quad \mathrm{K}\in\Gamma^* = 2\pi\Gamma. \tag{2.3.3}$$

Here we introduce a reciprocal lattice $\Gamma^* = 2\pi\Gamma$, and a corresponding fundamental domain $[-\pi,\pi]^d$, called *Brillouin zone* (or *first Brillouin zone* in some literature sources). Henceforth we will use notation "BZ" for it.

From these two properties we can conclude that a resulting function $\mathcal{U}f$ can be defined on a finite domain WSC $\times$ BZ $= [0,1]^d \times [-\pi,\pi]^d$ and reconstructed on its exterior by periodicity with use of (2.3.2) and (2.3.2).

Third, the *Floquet transform commutes with a periodic partial differential operator*. Let L_r be such an operator, the index "r" shows that it differentiates in space variable r. Then we easily get

$$[\mathcal{U}(L_\mathrm{r}f)](\mathrm{r},\mathrm{k}) = [L_\mathrm{r}(\mathcal{U}f)](\mathrm{r},\mathrm{k}). \tag{2.3.4}$$

On the right hand side the function $\mathcal{U}f$ depends on both variables r and k and can be considered as a set of r-depending functions $[\mathcal{U}f]_{\mathrm{k}}$, so the differential operator L_{r} is applied to each of them. Using the same formalism on the left hand side, we have a *decomposition* of the differential operator into a *direct integral* of k-dependent operators:

$$L_{\mathrm{r}} = \int_{\mathrm{BZ}} L_{\mathrm{r}}^{(\mathrm{k})} \,\mathrm{d}\,\mathrm{k},$$

with $L_{\mathrm{r}}^{(\mathrm{k})}$ having the same periodicity as L_{r}.
Finally, the fourth property of the Floquet transform will be formulated in the following theorem.

Theorem 2.1 (Invertibility of the Floquet transform, [Kuc01])**.** *The Floquet transform*

$$\begin{aligned} \mathcal{U}: \qquad & \mathrm{L}^2_\varepsilon\left(\mathbb{R}^d\right) \mapsto \mathrm{L}^2\left(\mathrm{BZ}, \mathrm{L}^2_\varepsilon(\mathrm{WSC})\right), \\ [\mathcal{U}f]_{\mathrm{k}}(\mathrm{r}) \quad &= \quad \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f(\mathrm{r}-\mathrm{R}) \end{aligned}$$

is an isometric isomorphism. Its inverse is given by

$$\left[\mathcal{U}^{-1}g\right](\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} g_{\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k}, \tag{2.3.5}$$

where V_{BZ} is the volume of the Brillouin zone and $g_{\mathrm{k}}(\mathrm{r})$ is extended to $\mathrm{r} \in \mathbb{R}^d$ according to the Floquet condition (2.3.2)*.*

Here the space $\mathrm{L}^2\left(\mathrm{BZ}, \mathrm{L}^2_\varepsilon(\mathrm{WSC})\right)$ has an inner product

$$\langle f, g\rangle_{\mathrm{L}^2(\mathrm{BZ},\mathrm{L}^2_\varepsilon(\mathrm{WSC}))} = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \langle f_{\mathrm{k}}, g_{\mathrm{k}}\rangle_{\mathrm{WSC},\varepsilon} \,\mathrm{d}\,\mathrm{k}.$$

The proof of this theorem can be found in [Kuc01, Theorem 7.2, p. 220].

The Transverse Magnetic problem (2.1.5) represents a system of partial differential equations with periodic coefficients. The operator of the problem $L_{\mathbf{TM}} = -\Delta/\varepsilon$ is an example

of a differential operator L_{r} considered above. Therefore, it can be decomposed under the Floquet transform into an infinite set of k-dependent operators:

$$L_{\mathbf{TM}} = \mathcal{U}^{-1}\left(\mathcal{U}\left(L_{\mathbf{TM}}\right)\right) = \mathcal{U}^{-1}\left(\mathcal{U}\left(\frac{-\Delta}{\varepsilon}\right)\right) = \int_{\mathrm{BZ}} \frac{-\Delta^{(\mathrm{k})}}{\varepsilon}\,\mathrm{d\,k} = \int_{\mathrm{BZ}} L^{(\mathrm{k})}_{\mathbf{TM}}\,\mathrm{d\,k}. \quad (2.3.6)$$

For each of them the eigenvalue problem reads:

$$-\Delta^{(\mathrm{k})}\psi_{\mathrm{k}}(\mathrm{r}) \;\; = \;\; \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})\psi_{\mathrm{k}}(\mathrm{r}).$$

Here the Laplacian $\Delta^{(\mathrm{k})}$ is the same operator as an ordinary Laplacian Δ (in sense of differentiation by spacial variable r). The index (k) just shows that its eigenvalues depend on k either and this way represent a decomposition of the solution. Having this in mind, we omit the index in further discussion.

Since the operators $L^{(\mathrm{k})}_{\mathbf{TM}}$ in (2.3.6) are elliptic, each of them has a compact resolvent and therefore a discrete spectrum:

$$0 < \lambda_{1,\mathrm{k}} \le \lambda_{2,\mathrm{k}} \le \ldots \lambda_{n,\mathrm{k}} \le \ldots.$$

The sequence is unbounded and the multiplicity of each eigenspace is finite. As it is well known [Kuc01], $\lambda_{n,\mathrm{k}}$ as continuous functions of k exhibit *band gap structure*. It means that the spectrum of $L_{\mathbf{TM}}$ consists of closed intervals $[\lambda_{N_1}, \lambda_{N_2}]$ called *bands*: for n between N_1 and $N_2 - 1$

$$\max_{\mathrm{k}\in\mathrm{BZ}} \lambda_{n,\mathrm{k}} \ge \min_{\mathrm{k}\in\mathrm{BZ}} \lambda_{n+1,\mathrm{k}}$$

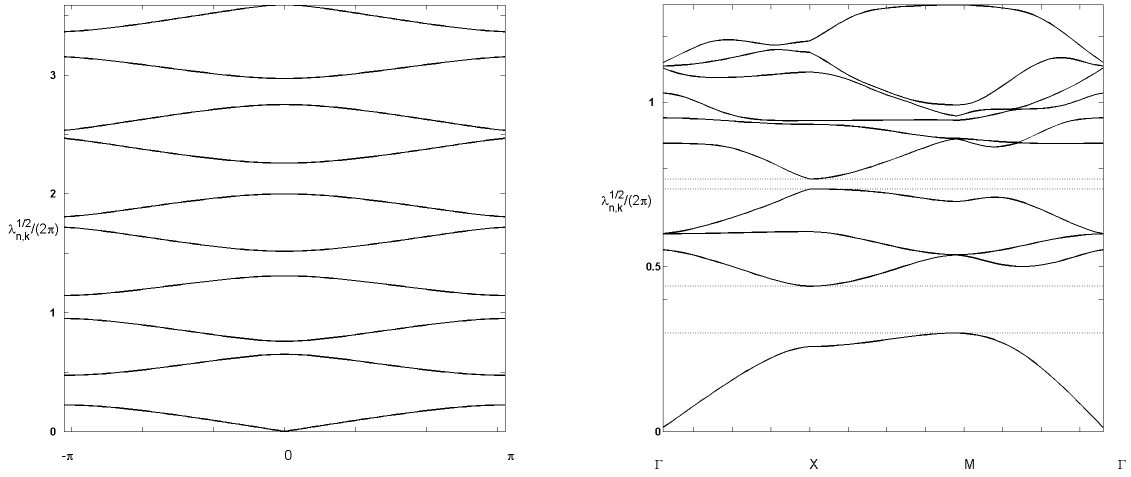
and the λ_n as functions of k cannot be separated from each other. Their graphs on a plain (k, λ) can intersect or coincide. A band is called *simple*, if it contains only one such a function ($N_2 = N_1$), or *complex*, if there are two or more functions ($N_2 > N_1$).
For $\varepsilon \neq const$ there may exist *gaps* between the bands, namely, intervals on λ-axis which consist of such values λ^* that for $\mathrm{k} \in \mathrm{BZ}$ they do not belong to a spectrum of any $L^{(\mathrm{k})}_{\mathbf{TM}}$:

$$\max_{\mathrm{k}\in\mathrm{BZ}} \lambda_{n,\mathrm{k}} < \lambda^* < \min_{\mathrm{k}\in\mathrm{BZ}} \lambda_{n+1,\mathrm{k}}.$$

From the physical point of view it means that electromagnetic waves of a frequency $\omega = \sqrt{c^2\lambda^*}$ cannot propagate into the Photonic Crystal.

An analytical proof of existence of band gaps for high-contrast media was given by Figotin and Kuchment in [FK94]. For an illustration of the photonic band gap structure see Figure 2.2 (for numerical details see Appendix, Section 10.2).

Figure 2.2: Band Gaps in a Photonic Crystal: 1D (left) and 2D (right)



We say that two functions $\psi_{n,\mathrm{k}}$, $\psi_{m,\mathrm{k}}$ which correspond to the eigenvalues $\lambda_{n,\mathrm{k}}$ and $\lambda_{m,\mathrm{k}}$ "belong to one band", if the eigenvalues $\lambda_{n,\mathrm{k}}, \lambda_{m,\mathrm{k}}$ which they comply with belong to one band, and this sentence actually makes sense only in this context, because the eigen*functions* do not form any bandstructure.

According to *Floquet-Bloch theory*, the solutions of (2.1.5) permit the *Floquet-Bloch ansatz*:

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}} u_{n,\mathrm{k}}(\mathrm{r}), \tag{2.3.7}$$

where the functions $u_{n,\mathrm{k}}(\mathrm{r})$ are periodic in r with respect to Γ: $u_{n,\mathrm{k}}(\mathrm{r}+\mathrm{R}) = u_{n,\mathrm{k}}(\mathrm{r})$ for all $\mathrm{R} \in \Gamma$. For a fixed k, functions $\{\psi_{n,\mathrm{k}}\}_{n\in\mathbb{N}}$ are the eigenfunctions of the corresponding operator $L_{\mathbf{TM}}^{(\mathrm{k})} = -\Delta^{(\mathrm{k})}/\varepsilon$. These functions are called *Bloch waves*.

2.4 Properties of the Bloch waves

We start with a well-known orthogonality property:

Lemma 2.1.

$$\int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} = V_{\mathrm{BZ}}\delta_{\mathrm{R},0}.$$

Proof. Indeed, if $\mathrm{R} = 0$,

$$\int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot 0}\,\mathrm{d}\,\mathrm{k} \;=\; \int_{\mathrm{BZ}} 1\,\mathrm{d}\,\mathrm{k} = V_{\mathrm{BZ}}.$$

If $\mathrm{R} \neq 0$, the integral $\int_{\mathrm{BZ}}$ consists of a sum of two integrals taken by half of BZ with opposite signs of k:

$$\begin{aligned}
\int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} &= \int_{(\mathrm{BZ}/2)^-} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} + \int_{(\mathrm{BZ}/2)^+} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} \\
&= \int_{(\mathrm{BZ}/2)^+} \mathrm{e}^{\mathrm{i}(-\mathrm{k})\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} + \int_{(\mathrm{BZ}/2)^+} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\,\mathrm{d}\,\mathrm{k} \\
&= \int_{(\mathrm{BZ}/2)^+} \left(\mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} + \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\right)\mathrm{d}\,\mathrm{k} \\
&= 2\int_{(\mathrm{BZ}/2)^+} \cos(\mathrm{k}\cdot\mathrm{R})\,\mathrm{d}\,\mathrm{k} = 0.
\end{aligned}$$

Together this proves the assistance. □

Let us make a close acceptation which will be useful later:

$$\sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} = V_{\mathrm{BZ}}\delta(\mathrm{k}). \tag{2.4.1}$$

This equality should be understood in the following sense. On the right hand side we have a Dirac distribution δ, therefore, we do not speak here about *values*. On the left hand side there is a series which actually does not converge absolutely. But in sense of *tempered distributions* [Rud91] it can be considered as a Floquet transform of identity, which is δ: $\hat{1} \mapsto \delta$. Of course, we should not forget that $1 \notin \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ and a Floquet transform can be applied to it only in the distributional sense. For intuitive understanding we write the

following sequence of formulas which can be made mathematically rigorous with some effort:

$$\begin{aligned} 1 = \mathcal{U}^{-1}\left(\mathcal{U}[1]\right) &= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\, \mathrm{d}\,\mathrm{k}, \\ \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} &= 0 \quad \text{for } \mathrm{k} \neq 0. \end{aligned}$$

The same equality (2.4.1) is mentioned in [Koh59] as a "fact" which the author uses in his computations.

Now we summarize the most important properties of the Bloch waves in the following theorem.

Theorem 2.2 (Properties of the Bloch waves)**.** *For* $\mathrm{k} \in \mathrm{BZ}$ *the operator* $L^{(\mathrm{k})}_{\boldsymbol{TM}} = -\Delta^{(\mathrm{k})}/\varepsilon$ *has a complete set of eigenfunctions* $\psi_{n,\mathrm{k}}$ *with eigenvalues* $\lambda_{n,\mathrm{k}}$*. This functions have the following properties:*

1. *They are periodic by* k *with respect to the lattice* Γ^**:*

$$\psi_{n,\mathrm{k}+\mathrm{K}}(\mathrm{r}) = \psi_{n,\mathrm{k}}(\mathrm{r}), \quad \mathrm{K} \in \Gamma^*.$$

2. *They are quasi-periodic by* r *with respect to the lattice* Γ*:*

$$\psi_{n,\mathrm{k}}(\mathrm{r}+\mathrm{R}) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\psi_{n,\mathrm{k}}(\mathrm{r}), \quad \mathrm{R} \in \Gamma.$$

3. *They belong to* $\mathrm{L}^2_\varepsilon(\mathrm{WSC})$ *and are orthonormal by means of its inner product:*

$$\left\langle \psi_{m,\mathrm{k}}, \psi_{n,\mathrm{k}'} \right\rangle_{\mathrm{WSC},\varepsilon} = \delta_{m,n}\delta_{\mathrm{k},\mathrm{k}'}.$$

4. *Extended by quasi-periodicity to the whole space* $\mathbb{R}^d$*, the* $\psi_{n,\mathrm{k}}$ *are orthogonal in the sense of distributions:*

$$\left\langle \psi_{m,\mathrm{k}}, \psi_{n,\mathrm{k}'} \right\rangle_{\mathbb{R}^d,\varepsilon} = V_{\mathrm{BZ}}\delta_{m,n}\delta(\mathrm{k}-\mathrm{k}'),$$

where V_{BZ} *is the volume of the Brillouin zone.*

5. Any function $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ admits an expansion

$$f(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} f_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k},$$

with $\psi_{n,\mathrm{k}}$ extended to $\mathbb{R}^d$ by Property 2 and $f_{n,\mathrm{k}} = \langle \psi_{n,\mathrm{k}}, [\mathcal{U}f]_{\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon}$.

Proof. (After [RS78, p. 304] and [DLP+11])
Orthonormality of Bloch waves in WSC follows from the fact that they do form a complete set of eigenfunctions of an elliptic operator $L^{(\mathrm{k})}_{\mathbf{TM}}$ (see [Kuc01], [RS78]). The periodic properties 1-2 follow from the definition of the Floquet transform since $L^{(\mathrm{k})}_{\mathbf{TM}} = \mathcal{U}(L_{\mathbf{TM}})$. Let us show k- and n-orthogonality of extended Bloch waves.

$$\begin{aligned}
\left\langle \psi_{m,\mathrm{k}}, \psi_{n,\mathrm{k}'} \right\rangle_{\mathbb{R}^d,\varepsilon} &= \int_{\mathbb{R}^d} \psi^*_{m,\mathrm{k}}(\mathrm{r}) \varepsilon(\mathrm{r}) \psi_{n,\mathrm{k}'}(\mathrm{r}) \,\mathrm{d}\,\mathrm{r} \\
&= \sum_{\mathrm{R}\in\Gamma} \int_{\mathrm{WSC}} \psi^*_{m,\mathrm{k}}(\mathrm{r}-\mathrm{R}) \varepsilon(\mathrm{r}-\mathrm{R}) \psi_{n,\mathrm{k}'}(\mathrm{r}-\mathrm{R}) \,\mathrm{d}(\mathrm{r}-\mathrm{R}) \\
&= \int_{\mathrm{WSC}} \left(\sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}(\mathrm{k}-\mathrm{k}')\cdot\mathrm{R}} \right) \psi^*_{m,\mathrm{k}}(\mathrm{r}) \varepsilon(\mathrm{r}) \psi_{n,\mathrm{k}'}(\mathrm{r}) \,\mathrm{d}\,\mathrm{r} \\
&\quad \text{(according to (2.4.1))} \\
&= V_{\mathrm{BZ}} \delta(\mathrm{k}-\mathrm{k}') \left\langle \psi_{m,\mathrm{k}}, \psi_{n,\mathrm{k}'} \right\rangle_{\mathrm{WSC},\varepsilon} \\
&= V_{\mathrm{BZ}} \delta(\mathrm{k}-\mathrm{k}') \delta_{m,n}.
\end{aligned}$$

Let us consider the Property 5. From the definition of the Floquet transform (2.3.1), $\mathcal{U}f \in \mathrm{L}^2_\varepsilon(\mathrm{WSC}\times\mathrm{BZ})$ for any $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$. As far as the Bloch waves $\{\psi_{n,\mathrm{k}}\}_{n\in\mathbb{N}}$ form an orthonormal basis in this space, $\mathcal{U}f$ can be expanded as

$$[\mathcal{U}f]_{\mathrm{k}}(\mathrm{r}) = \sum_n \langle \psi_{n,\mathrm{k}}, [\mathcal{U}f]_{\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \, \psi_{n,\mathrm{k}}(\mathrm{r}) = \sum_n f_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r})$$

with $f_{n,\mathrm{k}} = \langle \psi_{n,\mathrm{k}}, [\mathcal{U}f]_{\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon}$ for $\mathrm{r} \in \mathrm{WSC}$.
To obtain f, we apply the inverse Floquet transform (2.3.5):

$$f(\mathrm{r}) = \mathcal{U}^{-1}[\mathcal{U}f](\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} [\mathcal{U}f]_{\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k} = \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} f_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k}.$$

Here $\mathrm{r} \in \mathbb{R}^d$ and $\psi_{n,\mathrm{k}}$ are extended by the Property 2. A rigorous proof of the Property 5 (in particular, of the convergence of the sum) one can find in [DLP$^+$11]. Theorem 2.2 is proven. □

Remark 2.1. *The coefficients $f_{n,\mathrm{k}}$ from Property 5 of Theorem 2.2 are defined for* k *in the Brilloiun zone only. By Property 1 they can be considered as periodic in* k *with respect to Γ^*.*

Remark 2.2. *The extended Bloch waves $\psi_{n,\mathrm{k}}$ do not belong to $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$. The inner product in Property 4 can be computed only formally and understood in the sense of distributions.*

2.5 Periodic functions $u_{n,\mathrm{k}}$

In our analysis it is sometimes preferable to deal with periodic functions. That is why we turn from the Bloch waves $\psi_{n,\mathrm{k}}$ to their periodic parts $u_{n,\mathrm{k}}$. Recall (2.3.7):

$$u_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{r}}\psi_{n,\mathrm{k}}(\mathrm{r}).$$

For functions $u_{n,\mathrm{k}}$ we update the eigenvalue problem (2.1.5):

$$\begin{aligned} -\Delta\left(\mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}u_{\mathrm{k}}(\mathrm{r})\right) &= \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})\mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}u_{\mathrm{k}}(\mathrm{r}), \\ \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}\left(-\Delta u_{\mathrm{k}}(\mathrm{r}) - 2\mathrm{i}\left(\mathrm{k}\cdot\nabla\right)u_{\mathrm{k}}(\mathrm{r}) + k^2 u_{\mathrm{k}}(\mathrm{r})\right) &= \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})\mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}u_{\mathrm{k}}(\mathrm{r}) \end{aligned}$$

and thus

$$-\left(\nabla + \mathrm{i}\mathrm{k}\right)\cdot\left(\nabla + \mathrm{i}\mathrm{k}\right)u_{\mathrm{k}}(\mathrm{r}) = \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})u_{\mathrm{k}}(\mathrm{r}). \tag{2.5.1}$$

Note that functions $u_{n,\mathrm{k}}$ and $\psi_{n,\mathrm{k}}$ correspond to *the same* eigenvalue $\lambda_{n,\mathrm{k}}$.
The operator to examine now is $L^{(\mathrm{k})}_{\mathbf{TM},u} = -\left(\nabla + \mathrm{i}\mathrm{k}\right)\cdot\left(\nabla + \mathrm{i}\mathrm{k}\right)$. As we see, the replacement of $\psi_{n,\mathrm{k}}$ by $u_{n,\mathrm{k}}$ gives us a k-dependent operator, but a k-independent domain of definition. For the functions $u_{n,\mathrm{k}}$ we formulate a theorem equivalent to Theorem 2.2:

Theorem 2.3 (Properties of the functions $u_{n,\mathrm{k}}$)**.** *For every* $\mathrm{k} \in \mathrm{BZ}$ *the operator* $L^{(\mathrm{k})}_{\boldsymbol{TM},u} = -(\nabla + \mathrm{ik}) \cdot (\nabla + \mathrm{ik})$ *has a complete set of eigenfunctions* $u_{n,\mathrm{k}}(\mathrm{r})$ *with eigenvalues* $\lambda_{n,\mathrm{k}}$*. This functions have the following properties:*

1. *They are quasi-periodic by* k *with respect to the lattice* Γ^**:*

$$u_{n,\mathrm{k}+\mathrm{K}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{iK}\cdot\mathrm{r}} u_{n,\mathrm{k}}(\mathrm{r}), \quad \mathrm{K} \in \Gamma^*.$$

2. *They are periodic by* r *with respect to the lattice* Γ*:*

$$u_{n,\mathrm{k}}(\mathrm{r}+\mathrm{R}) = u_{n,\mathrm{k}}(\mathrm{r}), \quad \mathrm{R} \in \Gamma.$$

3. *They belong to* $\mathrm{L}^2_\varepsilon(\mathrm{WSC})$ *and are orthonormal by means of its inner product:*

$$\left\langle u_{m,\mathrm{k}}, u_{n,\mathrm{k}'} \right\rangle_{\mathrm{WSC},\varepsilon} = \delta_{m,n}\delta_{\mathrm{k},\mathrm{k}'}.$$

4. *Being extended by periodicity to the whole space* $\mathbb{R}^d$*, they form an orthogonal set:*

$$\left\langle u_{m,\mathrm{k}}, u_{n,\mathrm{k}'} \right\rangle_{\mathbb{R}^d,\varepsilon} = V_{\mathrm{BZ}}\delta_{m,n}\delta(\mathrm{k}-\mathrm{k}'),$$

   where V_{BZ} *is the volume of the Brillouin zone.*

5. *Any function* $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ *admits an expansion*

$$f(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} f^u_{n,\mathrm{k}} u_{n,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k},$$

   with $u_{n,\mathrm{k}}$ *extended to* $\mathbb{R}^d$ *by Property 2 and* $f^u_{n,\mathrm{k}} = \langle u_{n,\mathrm{k}}, [\mathcal{U}f]_{\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon}$*.*

Proof. A mapping T such that $Tf_{\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}} f_{\mathrm{k}}(\mathrm{r})$ is an isometry $T : \mathrm{L}^2_\varepsilon(\Omega) \mapsto \mathrm{L}^2_\varepsilon(\Omega)$ for Ω being WSC or $\mathbb{R}^d$. Indeed, for any $f_{\mathrm{k}}, g_{\mathrm{k}} \in \mathrm{L}^2_\varepsilon(\Omega)$

$$\langle Tf_{\mathrm{k}}, Tg_{\mathrm{k}} \rangle_{\Omega,\varepsilon} = \left\langle \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}} f_{\mathrm{k}}, \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}} g_{\mathrm{k}} \right\rangle_{\Omega,\varepsilon} = \langle f_{\mathrm{k}}, g_{\mathrm{k}} \rangle_{\Omega,\varepsilon}.$$

Consequently, the functions $u_{n,\mathrm{k}} = T\psi_{n,\mathrm{k}}$ inherit orthogonality from $\psi_{n,\mathrm{k}}$.

Periodic properties of $u_{n,\mathrm{k}}$ can be easily obtained from those of Theorem 2.2 and its definition. Obviously, since $u_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}}\psi_{n,\mathrm{k}}(\mathrm{r})$, for $\mathrm{K} \in \Gamma^*$ and $\mathrm{R} \in \Gamma$ we have:

$$\begin{aligned} u_{n,\mathrm{k+K}}(\mathrm{r}) &= \mathrm{e}^{-\mathrm{i(k+K)}\cdot\mathrm{r}}\psi_{n,\mathrm{k+K}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{iK}\cdot\mathrm{r}}\mathrm{e}^{-\mathrm{ik}\cdot\mathrm{r}}\psi_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{-\mathrm{iK}\cdot\mathrm{r}}u_{n,\mathrm{k}}(\mathrm{r}), \\ u_{n,\mathrm{k}}(\mathrm{r+R}) &= \mathrm{e}^{-\mathrm{ik}\cdot(\mathrm{r+R})}\psi_{n,\mathrm{k}}(\mathrm{r+R}) = \mathrm{e}^{-\mathrm{ik}\cdot(\mathrm{r+R})}\mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}}\psi_{n,\mathrm{k}}(\mathrm{r}) = u_{n,\mathrm{k}}(\mathrm{r}). \end{aligned}$$

Theorem 2.3 is proven. □

Chapter 3

Wannier functions

3.1 Definition and properties

In the previous chapter we have made all necessary preparations to define the main subject of this thesis — the *Wannier functions*. Formally, these are functions obtained from the Bloch waves by application of the inverse Floquet transform:

$$w_{n,0}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \psi_{n,\mathrm{k}}(\mathrm{r})\, \mathrm{d}\,\mathrm{k}. \tag{3.1.1}$$

Here the index "0" indicates that the functions defined this way are centered at $\mathrm{r} = 0$. More generally, Wannier function centered at $\mathrm{R} \in \Gamma$ are determined as

$$w_{n,\mathrm{R}}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi_{n,\mathrm{k}}(\mathrm{r})\, \mathrm{d}\,\mathrm{k} = w_{n,0}(\mathrm{r} - \mathrm{R}). \tag{3.1.2}$$

These functions originally appeared in the chemical literature for modeling of electron orbitals in crystal solids. They were described by Gregory Wannier [Wan37] first in 1937 and were named after him "Wannier orbitals", or just "Wannier functions". The idea to apply the same tool to Photonic Crystals by analogy to normal crystals was stated in 2003 by two independent groups of scientists in their papers [BMGM$^+$03] and [WC03].
By definition of the (inverse) Floquet transform, Wannier functions belong to $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$. They can be understood as some "average" electromagnetic field at a certain node R of

the Bravais lattice.

Theorem 3.1 (Orthonormality of the Wannier functions)**.** *Wannier functions form a complete basis in* $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ *which is* R− *and* $n-$*orthonormal:*

$$\left\langle w_{m,\mathrm{R}}, w_{n,\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} = \delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}.$$

Proof. First let us recall Lemma 2.1:

$$\frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \,\mathrm{d}\,\mathrm{k} = \delta_{\mathrm{R},0}.$$

From the definition of Wannier functions and the orthonormality of Bloch waves we have:

$$\begin{aligned}
\left\langle w_{m,\mathrm{R}}, w_{n,\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} &= \left\langle \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi_{m,\mathrm{k}} \,\mathrm{d}\,\mathrm{k}, \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'} \psi_{n,\mathrm{k}'} \,\mathrm{d}\,\mathrm{k}' \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}(\mathrm{k}\cdot\mathrm{R}-\mathrm{k}'\cdot\mathrm{R}')} \frac{1}{V_{\mathrm{BZ}}} \left\langle \psi_{m,\mathrm{k}}, \psi_{n,\mathrm{k}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \,\mathrm{d}\,\mathrm{k}\,\mathrm{d}\,\mathrm{k}' \\
&= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}(\mathrm{k}\cdot\mathrm{R}-\mathrm{k}'\cdot\mathrm{R}')} \delta_{m,n}\delta(\mathrm{k}-\mathrm{k}') \,\mathrm{d}\,\mathrm{k}\,\mathrm{d}\,\mathrm{k}' \\
&= \delta_{m,n} \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot(\mathrm{R}-\mathrm{R}')} \,\mathrm{d}\,\mathrm{k} \\
&= \delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}.
\end{aligned}$$

Now we want to show completeness. Let $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$. From Theorem 2.2,

$$f(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} f_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k}$$

with coefficients $f_{n,\mathrm{k}}$ periodic in k with respect to Γ^*. Applying Fourier expansion yields:

$$f_{n,\mathrm{k}} = \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f_{n,\mathrm{R}}$$

with

$$f_{n,\mathrm{R}} = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f_{n,\mathrm{k}} \,\mathrm{d}\,\mathrm{k}.$$

Therefore,

$$\begin{aligned} f(\mathrm{r}) &= \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} f_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r}) \, \mathrm{d}\,\mathrm{k} \\ &= \frac{1}{V_{\mathrm{BZ}}} \sum_n \sum_{\mathrm{R}\in\Gamma} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f_{n,\mathrm{R}} \psi_{n,\mathrm{k}}(\mathrm{r}) \, \mathrm{d}\,\mathrm{k} \\ &= \sum_n \sum_{\mathrm{R}\in\Gamma} f_{n,\mathrm{R}} w_{n,\mathrm{R}}(\mathrm{r}). \end{aligned}$$

To finish the proof it remains to show that $f_{n,\mathrm{R}} = \langle w_{n,\mathrm{R}}, f \rangle_{\mathbb{R}^d,\varepsilon}$.

$$\begin{aligned} \langle w_{n,\mathrm{R}}, f \rangle_{\mathbb{R}^d,\varepsilon} &= \left\langle \frac{1}{V_{\mathrm{BZ}}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi_{n,\mathrm{k}} \, \mathrm{d}\,\mathrm{k}, \frac{1}{V_{\mathrm{BZ}}} \sum_m f_{m,\mathrm{k}'} \psi_{m,\mathrm{k}'} \, \mathrm{d}\,\mathrm{k}' \right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \frac{1}{V_{\mathrm{BZ}}} \sum_m \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} f_{m,\mathrm{k}'} \left[\frac{1}{V_{\mathrm{BZ}}} \left\langle \psi_{n,\mathrm{k}}, \psi_{m,\mathrm{k}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \right] \mathrm{d}\,\mathrm{k} \, \mathrm{d}\,\mathrm{k}' \\ &= \frac{1}{V_{\mathrm{BZ}}} \sum_m \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \left(\sum_{\mathrm{R}'\in\Gamma} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'} f_{m,\mathrm{R}'} \right) \delta_{n,m} \delta(\mathrm{k}-\mathrm{k}') \, \mathrm{d}\,\mathrm{k} \, \mathrm{d}\,\mathrm{k}' \\ &= \sum_{\mathrm{R}'\in\Gamma} f_{n,\mathrm{R}'} \left[\frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot(\mathrm{R}-\mathrm{R}')} \, \mathrm{d}\,\mathrm{k} \right] \\ &= \sum_{\mathrm{R}'\in\Gamma} f_{n,\mathrm{R}'} \delta_{\mathrm{R},\mathrm{R}'} = f_{n,\mathrm{R}}. \end{aligned}$$

So, we see that any function $f \in \mathrm{L}^2_\varepsilon(\mathbb{R}^d)$ allows a representation of the form $f(\mathrm{r}) = \sum_n \sum_{\mathrm{R}\in\Gamma} f_{n,\mathrm{R}} w_{n,\mathrm{R}}(\mathrm{r})$ with the coefficients $f_{n,\mathrm{R}} = \langle w_{n,\mathrm{R}}, f \rangle_{\mathbb{R}^d,\varepsilon}$, which means the completeness of the orthonormal basis $w_{n,\mathrm{R}}$ in the space $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$. Theorem 3.2 is proven. □

3.2 Phase indeterminacy

Wannier functions represent a perfect basis for description of various effects in Photonic Crystals, e.g. defect modes. They can be used for Galerkin approximation which will be discussed later in Chapter 8. Since they are constructed from the Bloch waves of a particular Photonic Crystal, they *already contain* implicitly the information about the crystal structure. In general the Wannier functions appear to decay slowly and show

erratic behaviour at a distance from the center R. But we can repair it using *nonuniqueness* of Bloch waves which originates from freedom of phase choice. For a simple band n_0 (and an "isolated" Bloch wave) a phase factor is $\mathrm{e}^{\mathrm{i}\theta_{n_0}(\mathrm{k})}$. In other words, if $\psi_{n_0,\mathrm{k}}$ is a Bloch wave of the band, then $\mathrm{e}^{\mathrm{i}\theta_{n_0}(\mathrm{k})}\psi_{n_0,\mathrm{k}}$ is evidently also a Bloch wave – namely, also an n_0-eigenfunction of the operator $L^{(\mathrm{k})}_{\mathbf{TM}}$ (2.3.6):

$$L^{(\mathrm{k})}_{\mathbf{TM}}\mathrm{e}^{\mathrm{i}\theta_{n_0}(\mathrm{k})}\psi_{n_0,\mathrm{k}} \;=\; \lambda_{n_0,\mathrm{k}}\mathrm{e}^{\mathrm{i}\theta_{n_0}(\mathrm{k})}\psi_{n_0,\mathrm{k}}. \tag{3.2.1}$$

For a complex band where several eigenvalue states $\lambda_{N_1,\mathrm{k}},\ldots,\lambda_{N_2,\mathrm{k}}$ are mixed up such a nonuniqueness is given by a unitary transform U^{k} of the set of the Bloch waves of this band:

$$\psi_{n,\mathrm{k}} \;\mapsto\; \sum_{m=N_1}^{N_2} \mathrm{U}^{\mathrm{k}}_{mn}\psi_{n,\mathrm{k}}. \tag{3.2.2}$$

In this sense we can talk about *generalized Bloch waves* which represent a class of functions defined uniquely up to a unitary transform of their set. The corresponding Wannier functions are determined as

$$w_{n,\mathrm{R}}(\mathrm{r}) \;=\; \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \sum_{m=N_1}^{N_2} \mathrm{U}^{\mathrm{k}}_{mn}\psi_{m,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}, \tag{3.2.3}$$

which means that they are also *non-unique*.

This important feature gives a possibility to influence the localization properties of the Wannier function basis by choosing proper phases for the Bloch waves. Slowly and oscillating decaying functions can be substituted by the ones which decay faster and have monotone "tails" away from the center. In particular, one can obtain *exponentially localized* Wannier functions — namely, decaying like an exponential function at infinity.

The next section is dedicated to the question of existence of exponentially localized Wannier functions. First, we will prove the existence in a case of one-dimensional crystals with symmetries following the classical paper of W. Kohn [Koh59]. Second, we generalize this result to arbitrary-dimensional crystals with symmetries using the investigation of J.

des Cloizeaux [dC63], [dC64a], [dC64b]. Finally, we will adduce the proof of A. Nenciu [Nen83] of the existence of exponentially decaying functions in a case of crystals without symmetries in arbitrary dimensions.
Once the proofs will be given, our aim in the framework of the present thesis will be to construct localized Wannier functions using various methods.

3.3 Existence of exponentially localized Wannier functions

The existence of exponentially decaying Wannier functions is a very attractive property which is, however, not so evident. From 1959 mathematicians proved it for several cases starting from a very restricted one and making the result more and more general.
The Wannier functions were first introduced as electron orbitals in periodic crystalline solids [Wan37]. The motion of an electron in such crystals is described by a Schrödinger equation

$$[-\Delta + V(\mathrm{r})]\,\psi(\mathrm{r}) = \lambda\psi(\mathrm{r})$$

with a periodic potential V. Thus, the Bloch waves appeared as the eigenfunctions of the Schrödinger operator $-\Delta + V$, and the theory we are going to review in this section was developed for this case. However, the problem we study in the framework of this thesis is the Transverse Magnetic form (2.1.5) of the Maxwell's equations with an elliptic operator $-\Delta/\varepsilon$ where ε is periodic. The analogy between these two operators was considered in detail by Kuchment [Kuc01], [Kuc08]. In particular, he calls the **TM** problem a "Schrödinger-type" spectral problem with ε playing a role of a *metric*. One can show that there exists a Hilbert space isometry between $L^2(D)$ and $L^2_\varepsilon(D)$ where D is a Wigner-Seitz cell $[0,1]^d$ or a space $\mathbb{R}^d$, [Klo04, Section 2.3]. This isometry is given by the mapping

$$\begin{aligned} \Phi(D) : L^2(D) &\rightarrow L^2_\varepsilon(D), \\ \psi &\mapsto \Phi\psi = \psi/\sqrt{\varepsilon}. \end{aligned}$$

Although in this section we review the theory for usual crystals — as it was originally developed by the authors we refer to — we will mention that it can be adopted to Photonic Crystals as well.

3.3.1 Kohn's proof for 1D crystals with inversion symmetry

The first rigorous proof was introduced by Walter Kohn in [Koh59] for an isolated band in a *one-dimensional crystal with inverse symmetry*. Consider a 1D Schrödinger equation:

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}\,x^2} + V(x)\right]\psi(x) = \lambda\psi(x)$$

with a periodic, symmetric and non-constant potential V:

$$V(x+L) = V(x) = V(-x) \neq \text{const.}$$

Let $\psi_{n,\mathrm{k}}$ be its *Bloch waves* (see Section 2.3, 2.4), i.e. solutions with the quasi-periodicity condition

$$\psi_{n,\mathrm{k}}(L) = \mathrm{e}^{\mathrm{i}\mathrm{k}L}\psi_{n,\mathrm{k}}(0), \quad \psi'_{n,\mathrm{k}}(L) = \mathrm{e}^{\mathrm{i}\mathrm{k}L}\psi'_{n,\mathrm{k}}(0)$$

for real k and normalized:

$$\frac{2\pi}{L}\int_0^L |\psi_{n,\mathrm{k}}(x)|^2\,\mathrm{d}\,x = 1.$$

The corresponding eigenvalue curves $\mathrm{k} \mapsto \lambda_{n,\mathrm{k}}$ are called *energy bands*.

The Wannier functions $w_{n,0}$ are defined as localized linear combinations of all Bloch waves of a given band n, namely:

$$w_{n,0}(x) = \frac{L}{2\pi}\int_{-\pi/L}^{\pi/L} \psi_{n,\mathrm{k}}(x)\,\mathrm{d}\,\mathrm{k},$$

or, more general,

$$w_{n,\mathrm{R}}(x) = w_{n,0}(x - \mathrm{R}L) = \frac{L}{2\pi}\int_{-\pi/L}^{\pi/L} \mathrm{e}^{-\mathrm{i}\mathrm{k}\mathrm{R}L}\psi_{n,\mathrm{k}}(x)\,\mathrm{d}\,\mathrm{k}.$$

To let these integrals exist, we consider a priori that the phases of $\psi_{n,\mathrm{k}}$ are chosen so that they are piecewise continuous in k in the interval $[-\pi/L, \pi/L]$ (the Brillouin zone).

In the framework of this chapter we observe only the key points of Kohn's paper; for further details see the original source [Koh59].
Let us denote with the symbol $\widetilde{f}$ a complex extension of a real function f. The energy bands $\lambda_{n,\mathrm{k}}$ can be studied as functions of a *complex* variable $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$. Kohn shows that in a vicinity of the real axis the function $\mathrm{k} \mapsto \widetilde{\lambda}(\mathrm{k})$ is analytic, symmetric and periodic in k:

$$\begin{aligned} \widetilde{\lambda}(\mathrm{k}) &= \widetilde{\lambda}(-\mathrm{k}), \\ \widetilde{\lambda}(\mathrm{k}^*) &= \widetilde{\lambda}(\mathrm{k})^*, \\ \widetilde{\lambda}\left(\mathrm{k} + \frac{2\pi}{L}\right) &= \widetilde{\lambda}(\mathrm{k}). \end{aligned}$$

The image of $\widetilde{\lambda}$ lies on a Riemannian surface S with an infinite series of sheets S_n connected with each other in branch points apart from the real axis. The bands $\lambda_{n,\mathrm{k}'}$ are the traces of $\widetilde{\lambda}(\mathrm{k})$ on the sheet S_n for k real. One can pass continuously from one band to another in the complex plane.
The most important result of Kohn's observations is the following theorem.

Theorem 3.2 (Existence and uniqueness of real, (anti-) symmetric and exponentially localized Wannier function, W. Kohn)**.** *Consider a one-dimensional Schrödinger equation*

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}\,x^2} + V(x)\right]\psi(x) = \lambda\psi(x)$$

with a periodic, symmetric and non-constant potential V: $V(x+L) = V(x) = V(-x)$. For every band number n there exists one and only one Wannier function $w_{n,0}$ which has all three of the following properties:

- *it is real;*

- *it is either symmetric or antisymmetric about either* $x = 0$ *or* $x = L/2$*;*
- *it falls off exponentially as*

$$w_{n,0}(x) \sim \mathrm{e}^{-A_n|x|}$$

for some $A_n \in (0, \infty)$*.*

Proof. [Koh59]. To prove the theorem, let us introduce a more general class of functions than the Bloch waves. Let $\widetilde{\psi}$

- satisfy

$$\left[-\frac{\mathrm{d}^2}{\mathrm{d}\,x^2} + V(x)\right]\widetilde{\psi}(x) = \widetilde{\lambda}\widetilde{\psi}(x) \tag{3.3.1}$$

with complex $\widetilde{\lambda}$,

- be quasi-periodic in the sense

$$\widetilde{\psi}(L) = \gamma\widetilde{\psi}(0), \quad \widetilde{\psi}'(L) = \gamma\widetilde{\psi}'(0), \qquad \gamma \in \mathbb{C}, \tag{3.3.2}$$

- be normalized in a manner which for $|\gamma| = 1$ reduces to

$$\frac{2\pi}{L}\int_0^L |\widetilde{\psi}(x)|^2\,\mathrm{d}\,x = 1. \tag{3.3.3}$$

For $\gamma = \mathrm{e}^{\mathrm{i}kL}$ these are just ordinary Bloch waves.
For a given $\widetilde{\lambda}$, we now search for such γ that (3.3.1) and (3.3.2) are fulfilled. $\widetilde{\psi}$ which satisfy (3.3.1) can be written as

$$\widetilde{\psi}(x, \widetilde{\lambda}) = a_1\widetilde{\psi}^{(1)}(x, \widetilde{\lambda}) + a_2\widetilde{\psi}^{(2)}(x, \widetilde{\lambda}) \tag{3.3.4}$$

with

$$\widetilde{\psi}^{(1)}(0, \widetilde{\lambda}) = 1, \qquad \left(\widetilde{\psi}^{(1)}\right)'(0, \widetilde{\lambda}) = 0,$$
$$\widetilde{\psi}^{(2)}(0, \widetilde{\lambda}) = 0, \qquad \left(\widetilde{\psi}^{(2)}\right)'(0, \widetilde{\lambda}) = 1.$$

Substituting (3.3.4) into (3.3.2) gives the ratio:

$$\frac{a_1}{a_2} = \frac{\widetilde{\psi}^{(2)}(L,\widetilde{\lambda})}{\gamma - \widetilde{\psi}^{(1)}(L,\widetilde{\lambda})} = \frac{\widetilde{\psi}^{(2)}(L,\widetilde{\lambda})}{\frac{1}{2}\left(\gamma - \frac{1}{\gamma}\right)}. \tag{3.3.5}$$

By eliminating a_1 and a_2 we obtain the following relation for γ:

$$\gamma^2 - 2\mu(\widetilde{\lambda})\gamma + 1 = 0, \tag{3.3.6}$$

where

$$\mu(\widetilde{\lambda}) = \frac{1}{2}\left(\widetilde{\psi}^{(1)}(L,\widetilde{\lambda}) + \left(\widetilde{\psi}^{(2)}\right)'(L,\widetilde{\lambda})\right) \tag{3.3.7}$$

is a single-valued, entire function of $\widetilde{\lambda}$, [Kra35].

Now we want to replace the variable $\widetilde{\lambda}$ with γ in order to study the properties of $\widetilde{\psi}$ depending on γ. By (3.3.6), μ relates to γ as

$$\mu = \frac{1}{2}\left(\gamma + \frac{1}{\gamma}\right). \tag{3.3.8}$$

Thus,

$$\widetilde{\lambda} = \widetilde{\lambda}(\mu) = \widetilde{\lambda}\left(\frac{\gamma + 1/\gamma}{2}\right)$$

can be considered as a function of γ. For this reason, we may henceforth omit the notation $\widetilde{\lambda}$ and write $\bar{\psi}(x,\gamma)$, $\bar{\psi}^{(1)}(L,\gamma)$, $\bar{\psi}^{(2)}(L,\gamma)$, having in mind that it assists at these expressions implicitly:

$$\widetilde{\psi}(\cdot,\lambda) = \widetilde{\psi}(\cdot,\lambda(\gamma)) \mapsto \bar{\psi}(\cdot,\gamma). \tag{3.3.9}$$

So, $\bar{\psi}$ is the same function of x as $\widetilde{\psi}$, the difference between them is only in the second variable.

With the equation (3.3.5), $\bar{\psi}$ as well as $\bar{\psi}^{(1)}$ and $\bar{\psi}^{(2)}$ can be considered as functions over two variables x and γ:

$$\bar{\psi}(x,\gamma) = \frac{1}{C^{1/2}}\left[\bar{\psi}^{(2)}(L,\gamma)\bar{\psi}^{(1)}(x,\gamma) + \frac{1}{2}\left(\gamma - \frac{1}{\gamma}\right)\bar{\psi}^{(2)}(x,\gamma)\right],$$

where C is chosen to satisfy the normalization requirement (3.3.3).
Let us denote

$$\chi(x,\gamma) = \bar{\psi}^{(2)}(L,\gamma)\bar{\psi}^{(1)}(x,\gamma) + \frac{1}{2}\left(\gamma - \frac{1}{\gamma}\right)\bar{\psi}^{(2)}(x,\gamma), \tag{3.3.10}$$

so that

$$\bar{\psi}(x,\gamma) = \frac{\chi(x,\gamma)}{C^{1/2}}. \tag{3.3.11}$$

After simple computations one can conclude that C must have a form

$$C = -\frac{4\pi}{L}\bar{\psi}^{(2)}(L,\gamma)\frac{\mathrm{d}\,\mu}{\mathrm{d}\,\widetilde{\lambda}},$$

and therefore (3.3.11) turns to

$$\bar{\psi}(x,\gamma) = \frac{\chi(x,\gamma)}{\left(-\frac{4\pi}{L}\bar{\psi}^{(2)}(L,\gamma)\frac{\mathrm{d}\,\mu}{\mathrm{d}\,\widetilde{\lambda}}\right)^{1/2}}. \tag{3.3.12}$$

We investigate the properties of the analytic function $\gamma \mapsto \bar{\psi}(\cdot,\gamma)$. From (3.3.12), it can have singularities at the points where $\mathrm{d}\,\mu/\,\mathrm{d}\,\widetilde{\lambda} = 0$ or where $\bar{\psi}^{(2)}(L,\gamma) = 0$. Kohn shows that such points are $\gamma = \pm 1$. He proves that one can find Laurent series for $\bar{\psi}$ depending on properties of $\bar{\psi}^{(2)}(L,\pm 1)$:

(A) if both $\bar{\psi}^{(2)}(L,1)$ and $\bar{\psi}^{(2)}(L,-1)$ are zero or both nonzero, then $\bar{\psi}(\cdot,\gamma)$ permits a Laurent expansion of the form

$$\bar{\psi}(x,\gamma) = \sum_{\mathrm{R}=-\infty}^{\infty} \alpha^{\mathrm{R}}(x)\gamma^{\mathrm{R}},$$

(B) while if one of $\bar{\psi}^{(2)}(L,1)$ or $\bar{\psi}^{(2)}(L,-1)$ is zero and the other is nonzero, the Laurent expansion is

$$\bar{\psi}(x,\gamma) = \sum_{\mathrm{R}=-\infty}^{\infty} \beta^{\mathrm{R}}(x)\gamma^{\mathrm{R}+1/2}.$$

In both cases the series converge in a ring

$$\mathrm{e}^{-LA} < |\gamma| < \mathrm{e}^{LA}$$

for some $A \in (0, \infty)$. Let us substitute $\gamma = \mathrm{e}^{\mathrm{i}\mathrm{k}L}$:

$$\bar{\psi}(x, \gamma) = \bar{\psi}\left(x, \mathrm{e}^{\mathrm{i}\mathrm{k}L}\right).$$

Then $\gamma = 1$ corresponds to $\mathrm{k} = 0$ and $\gamma = -1$ to $\mathrm{k} = \pi/L$.

From [Koh59] we know that $\bar{\psi}$ is a multivalued function of k whose branch points coincide with those of $\widetilde{\lambda}$. At each branch point four Riemannian sheets are connected; two of them correspond to a band n and the other two to a band $n+1$; two functions related to the same band differ only by sign.

Let $\widehat{\psi}_{n,\mathrm{k}}$ be branches of $\bar{\psi}$. We need to observe some properties of these functions which will be used further. Consider the same cases as before (**A** and **B**).

Case A: Both $\widehat{\psi}_{n,0}(0)$ and $\widehat{\psi}_{n,\pi/L}(0)$ are zero or both nonzero. We have the Fourier expansion

$$\widehat{\psi}_{n,\mathrm{k}}(x) = \sum_{\mathrm{R}=-\infty}^{\infty} \alpha_n^{\mathrm{R}}(x)\mathrm{e}^{\mathrm{i}\mathrm{k}\mathrm{R}L}$$

which represents an analytic function of $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$ in a strip $|\mathrm{k}''| < A_n$. Obviously, $\widehat{\psi}_{n,\mathrm{k}'}$ as a function of a real variable k' is periodic with period $\dfrac{2\pi}{L}$:

$$\widehat{\psi}_{n,\mathrm{k}'+\frac{2\pi}{L}}(x) = \widehat{\psi}_{n,\mathrm{k}'}(x).$$

Case B: One of $\widehat{\psi}_{n,0}(0)$ or $\widehat{\psi}_{n,\pi/L}(0)$ is zero and the other is nonzero. The Fourier expansion

$$\widehat{\psi}_{n,\mathrm{k}}(x) = \sum_{\mathrm{R}=-\infty}^{\infty} \beta_n^{\mathrm{R}}(x)\mathrm{e}^{\mathrm{i}\mathrm{k}\frac{2\mathrm{R}+1}{2}L}$$

represents an analytic function of $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$ in a strip $|\mathrm{k}''| < A_n$. In this case $\widehat{\psi}_{n,\mathrm{k}'}$ is antiperiodic by k' with period $\dfrac{2\pi}{L}$:

$$\widehat{\psi}_{n,\mathrm{k}'+\frac{2\pi}{L}}(x) = -\widehat{\psi}_{n,\mathrm{k}'}(x).$$

In addition to the described **Case A** and **Case B**, another classification of possible cases will also be useful.

Case 1: $\widehat{\psi}_{n,0}(0) \neq 0$. One can check that

$$\begin{aligned} \widehat{\psi}^*_{n,\mathrm{k}'}(x) &= \widehat{\psi}_{n,\mathrm{k}'}(-x), \\ \widehat{\psi}_{n,-\mathrm{k}'}(x) &= \widehat{\psi}_{n,\mathrm{k}'}(-x). \end{aligned}$$

Case 2: $\widehat{\psi}_{n,0}(0) = 0$. Then

$$\begin{aligned} \widehat{\psi}^*_{n,\mathrm{k}'}(x) &= \widehat{\psi}_{n,\mathrm{k}'}(-x), \\ \widehat{\psi}_{n,-\mathrm{k}'}(x) &= -\widehat{\psi}_{n,\mathrm{k}'}(-x). \end{aligned}$$

Let us now switch to the Wannier functions which are defined via *arbitrary* Bloch waves $\psi_{n,\mathrm{k}}$:

$$w_{n,\mathrm{R}}(x) = w_{n,0}(x - \mathrm{R}L) = \frac{L}{2\pi} \int_{-\pi/2}^{\pi/2} \mathrm{e}^{-\mathrm{ikR}L} \psi_{n,\mathrm{k}}(x)\, \mathrm{d\,k}$$

We will fix the phases of the Bloch waves according to their initial values in the point $x = 0$.

(1) If $\psi_{n,0}(0) \neq 0$, choose the phase so that

- $\psi_{n,\mathrm{k}}(0)$ is *real* for real k;
- $\psi_{n,\mathrm{k}}$ is *analytic* for complex k in a strip $|\,\mathrm{Im(k)}| < A_n$ with some finite and positive A_n.

Then it is equal (apart from a possible factor -1) to $\widehat{\psi}_{n,\mathrm{k}}(x)$ described in **Case 1**. Therefore,

$$\begin{aligned} \psi^*_{n,\mathrm{k}'}(x) &= \psi_{n,\mathrm{k}'}(-x), \\ \psi_{n,-\mathrm{k}'}(x) &= \psi_{n,\mathrm{k}'}(-x), \end{aligned}$$

and this yields

$$\begin{aligned} w^*_{n,0}(x) &= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi^*_{n,\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,-\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = w_{n,0}(x), \\ w_{n,0}(-x) &= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,\mathrm{k}'}(-x)\,\mathrm{d}\,\mathrm{k}' = \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,-\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = w_{n,0}(x). \end{aligned}$$

Consequently, $w_{n,0}$ is real and symmetric about $x = 0$.

(2) If $\psi_{n,0}(0) = 0$, let the phase be such that

- $\psi_{n,\mathrm{k}}(0)$ is *purely imaginary* for real k;
- $\psi_{n,\mathrm{k}}(x)$ is *analytic* for complex k in a strip $|\,\mathrm{Im(k)}| < A_n$ with some finite and positive A_n.

Then it coincides with $\pm\mathrm{i}\widehat{\psi}_{n,\mathrm{k}}$ described in **Case 2**. Consequently,

$$\begin{aligned} \psi^*_{n,\mathrm{k}'}(x) &= -\psi_{n,\mathrm{k}'}(-x), \\ \psi_{n,-\mathrm{k}'}(x) &= -\psi_{n,\mathrm{k}'}(-x), \end{aligned}$$

and thus

$$\begin{aligned} w^*_{n,0}(x) &= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi^*_{n,\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = -\frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,-\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = w_{n,0}(x), \\ w_{n,0}(-x) &= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,\mathrm{k}'}(-x)\,\mathrm{d}\,\mathrm{k}' = \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2}\psi_{n,-\mathrm{k}'}(x)\,\mathrm{d}\,\mathrm{k}' = -w_{n,0}(x). \end{aligned}$$

Hence, $w_{n,0}$ is real and antisymmetric about $x = 0$.

What about the asymptotic behaviour of the Wannier functions?
Again, take the phases of the Bloch waves as described in **(1)** or **(2)** depending on whether $\psi_{n,0}(0)$ equals to 0 or not. Then our Bloch waves $\psi_{n,\mathrm{k}}$ are connected with $\widehat{\psi}_{n,\mathrm{k}}(x)$ as

$$\begin{aligned} \psi_{n,\mathrm{k}}(x) &= \epsilon\widehat{\psi}_{n,\mathrm{k}}(x), \\ \epsilon &= \pm 1 \text{ or } \pm\,\mathrm{i}. \end{aligned}$$

Let us examine two combinations.

(a) Both $\psi_{n,0}(0)$ and $\psi_{n,\pi/L}(0)$ are zero or both nonzero. This is related to **Case A** above. We have the Fourier expansion

$$\psi_{n,\mathrm{k}}(x) = \epsilon \sum_{\mathrm{R}=-\infty}^{\infty} \alpha_n^{\mathrm{R}}(x) \mathrm{e}^{\mathrm{ikR}L}$$

which represents an analytic function of $\mathrm{k} = \mathrm{k}' + \mathrm{ik}''$ in a strip $|\mathrm{k}''| < A_n$. But from the definition of the Wannier functions we obtain:

$$\psi_{n,\mathrm{k}}(x) = \sum_{\mathrm{R}=-\infty}^{\infty} \mathrm{e}^{\mathrm{ikR}L} w_{n,\mathrm{R}}(x).$$

Thus we immediately see that $w_{n,\mathrm{R}}(x) \equiv \epsilon \alpha_n^{\mathrm{R}}(x)$. But $\alpha_n^{\mathrm{R}}(x)$ are the coefficients of the Laurent expansion of $\psi_{n,\mathrm{k}}$ and this series was convergent for $|\mathrm{k}''| < A_n$. This implies $\lim_{\mathrm{R}\to\infty} \alpha_n^{\mathrm{R}}(x)\mathrm{e}^{\mathrm{ikR}L} = 0$ for this range of k''. This is equivalent to

$$\lim_{x\to\infty} w_{n,0}(x)\mathrm{e}^{qx} \;=\; \begin{cases} 0, & q < A_n \\ \infty, & q > A_n \end{cases} .$$

Indeed, in the convergence domain it holds:

$$\begin{aligned} 0 &= \lim_{\mathrm{R}\to\infty} \alpha_n^{\mathrm{R}}(x)\mathrm{e}^{\mathrm{ikR}L} = \lim_{\mathrm{R}\to\infty} \frac{1}{\epsilon} w_{n,\mathrm{R}}(x)\mathrm{e}^{\mathrm{ikR}L} \\ &= \lim_{\mathrm{R}\to\infty} \frac{1}{\epsilon} w_{n,0}(x - \mathrm{R}L)\mathrm{e}^{\mathrm{ikR}L} = \lim_{y\to\infty} \frac{1}{\epsilon} w_{n,0}(-y)\mathrm{e}^{qy} \\ &= \lim_{y\to\infty} w_{n,0}(y)\mathrm{e}^{qy}, \end{aligned}$$

where $q < A_n$.

(b) One of $\psi_{n,0}(0)$ or $\psi_{n,\pi/L}(0)$ is zero and the other is nonzero. Shifting the coordinates (in x axis) by a half-period $L/2$, one can prove that in this case $\psi_{n,0}(L/2)$ and $\psi_{n,\pi/L}(L/2)$ are both zero or both nonzero and therefore we have case (a) for the coordinate origin $x = L/2$. Note that if we shifted the coordinate origin from $x = 0$ to $x = L/2$, then the

(anti-)symmetry center is also shifted into this point.

No other choice of phase except a possible trivial factor ± 1 leads to a Wannier function which is simultaneously real and (anti-)symmetric. Indeed, let us suppose that another set of Bloch waves leads to real and (anti-)symmetric exponentially decaying Wannier functions. This set is connected with the just observed by a phase shift:

$$\bar{\psi}_{n,\mathrm{k}}(x) = \mathrm{e}^{\mathrm{i}\theta_n(k)}\psi_{n,\mathrm{k}}(x).$$

Representing the phase as

$$\mathrm{e}^{\mathrm{i}\theta_n(k)} = \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,\mathrm{R}}\mathrm{e}^{-\mathrm{i}\mathrm{R}L\mathrm{k}},$$

we can rewrite the corresponding Wannier function as

$$\begin{aligned}
\bar{w}_{n,0}(x) &= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2} \bar{\psi}_{n,\mathrm{k}}(x)\,\mathrm{d}\,\mathrm{k} \\
&= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2} \mathrm{e}^{\mathrm{i}\theta_n(\mathrm{k})}\psi_{n,\mathrm{k}}(x)\,\mathrm{d}\,\mathrm{k} \\
&= \frac{L}{2\pi}\int_{-\pi/2}^{\pi/2} \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,\mathrm{R}}\mathrm{e}^{-\mathrm{i}\mathrm{R}L\mathrm{k}}\psi_{n,\mathrm{k}}(x)\,\mathrm{d}\,\mathrm{k} \\
&= \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,\mathrm{R}} w_{n,\mathrm{R}}(x).
\end{aligned}$$

As far as the $w_{n,\mathrm{R}}$ are real and orthonormal, and $\bar{w}_{n,0}$ is also real by assumption,

$$\bar{w}_{n,0}(x) = \bar{w}^*_{n,0}(x) = \sum_{\mathrm{R}=-\infty}^{\infty} a^*_{n,\mathrm{R}} w_{n,\mathrm{R}}(x),$$

we have: $a_{n,\mathrm{R}} = a^*_{n,\mathrm{R}}$, hence $a_{n,\mathrm{R}}$ is real. Analogously, as soon as $\{w_{n,\mathrm{R}}\}_{\mathrm{R}\in\mathbb{Z}}$ and $\bar{w}_{n,0}$ are

(anti-)symmetric,

$$\begin{aligned}
\pm\bar{w}_{n,0}(x) &= \bar{w}_{n,0}(-x) \\
&= \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,\mathrm{R}} w_{n,\mathrm{R}}(-x) \\
&= \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,\mathrm{R}} w_{n,0}(-x-\mathrm{R}L) \\
&= \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,-\mathrm{R}} w_{n,0}(-x+\mathrm{R}L) \\
&= \pm \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,-\mathrm{R}} w_{n,0}(x-\mathrm{R}L) \\
&= \pm \sum_{\mathrm{R}=-\infty}^{\infty} a_{n,-\mathrm{R}} w_{n,\mathrm{R}L}(x),
\end{aligned}$$

hence, $a_{n,-\mathrm{R}} = \pm a_{n,\mathrm{R}}$.

Substituting this into the representation of the phase yields:

$$\mathrm{e}^{\mathrm{i}\theta_n(k)} = \begin{cases} a_{n,0} + 2\sum_{\mathrm{R}=1}^{\infty} a_{n,\mathrm{R}} \cos(\mathrm{R}L\mathrm{k}), & a_{n,-\mathrm{R}} = a_{n,\mathrm{R}} \\ -2\mathrm{i}\sum_{\mathrm{R}=1}^{\infty} a_{n,\mathrm{R}} \sin(\mathrm{R}L\mathrm{k}), & a_{n,-\mathrm{R}} = -a_{n,\mathrm{R}} \end{cases}.$$

In the first case the phase is real and consequently, the factor $\mathrm{e}^{\mathrm{i}\theta_n(k)}$ equals ± 1. The second case can be ruled out by setting $\mathrm{k} = 0$. Theorem (3.2) is proven. □

Remark 3.1. *Let us turn to the Transverse Magnetic problem*

$$-\frac{\mathrm{d}^2}{\mathrm{d}\,x^2}\widetilde{\psi}(x) = \widetilde{\lambda}\varepsilon(x)\widetilde{\psi}(x)$$

instead of (3.3.1)*. In this case* (3.3.3) *holds and* (3.3.2) *should be understood in sense*

$$\frac{2\pi}{L}\int_0^L \widetilde{\psi}^*(x)\varepsilon(x)\widetilde{\psi}(x)\,\mathrm{d}\,x = 1.$$

Then we can represent $\widetilde{\psi}$ *in form* (3.3.4) *and again obtain the relation* (3.3.6)*. It means that we can adopt the theorem for the* ***TM*** *problem and get the same result — namely, the existence and uniqueness of real, (anti-) symmetric, exponentially localized Wannier functions.*

Remark 3.2. *In a multidimensional case, if a band is simple (namely, it contains only one energy state λ), there exists a Bloch wave which is analytic in* $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$ *in a strip* $|\mathrm{k}''| < A$. *But if the band is composite (it consists of several intersecting eigenvalue branches $\lambda_{N_1}, \ldots, \lambda_{N_2}$), the energy $\widetilde{\lambda}$ is a multivalued function of a complex variable* $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$*; the points where several states touch or intersect each other are branch points of this function and also branch points for the Bloch function* $\mathrm{k} \mapsto \widetilde{\psi}(\mathrm{k})$ *which therefore cannot be analytic in* k. *This remark indicates that Kohn's results cannot be trivially generalized to a multidimensional crystal and that another approach is required.*

Remark 3.3. *For a composite band, a simultaneous choice of a phase of each Bloch wave of the band is nothing but a unitary transform (for every* k*) of the whole set of waves corresponding to the band with regard to some fixed initial Bloch waves:*

$$\psi_{n,\mathrm{k}} \mapsto \sum_{m=n_1}^{n_2} \mathrm{U}^{\mathrm{k}}_{mn} \psi^0_{m,\mathrm{k}},$$

where $n_1, \ldots, n_2$ is a set of indices which determine the composite band.

3.3.2 Des Cloizeaux's proof for multidimensional crystals with inversion symmetry

The next step to the proof of existence of exponentially localized Wannier functions was given by des Cloizeaux in [dC64a] and [dC64b] for *crystals with a center of inversion in arbitrary dimensions*. Let us take up this approach in details.

Consider the Schrödinger equation in a crystal of dimensionality $d \geq 1$ and periodic with respect to some lattice Γ (for instance, $\Gamma = \mathbb{Z}^d$):

$$[-\Delta + V(\mathrm{r})]\psi(\mathrm{r}) = \lambda\psi(\mathrm{r}). \tag{3.3.13}$$

Here the potential V is assumed to be periodic: $V(\mathrm{r} + \mathrm{R}) = V(\mathrm{r})$ for $\mathrm{R} \in \Gamma$.

As we know from the fundamental works [RS78] (Chapter XIII.16) and [Kuc01], the

eigenfunctions $\psi_{n,\mathrm{k}}$ of the operator $-\Delta + V$ lie in $\mathrm{L}^2(\mathrm{WSC} \times \mathrm{BZ})$ with WSC being the Wigner-Seitz cell of the crystal, BZ being the first Brillouin zone.
Following [dC64a], consider an isolated band β — by which we mean the following.

Definition 3.1. *Let β be a set of N_β energy branches $\lambda_{n,\mathrm{k}} = \lambda_n(\mathrm{k})$, $n = n_1 \dots n_{N_\beta}$, which are the eigenvalues of the Schrödinger equation noticed above. If the ranges of the branches (i.e. projections of their graphs onto the λ-axis) create a continuous segment in the spectrum and are separated from above and from below by spectral gaps, then we call the set β an isolated band.*
If $N_\beta = 1$, the band is called simple. If $1 < N_\beta < \infty$, the band is complex or composite.

For an illustration of a band gap structure see Figure 2.2. On this figure we can derive two isolated bands: a simple band $n = 1$ and a complex band $n \in [2, 4]$. As we see, they are separated by gaps from the other parts of the spectrum presented on the λ-axis.

Notation 3.1. *Let β_ψ be a set of N_β independent Bloch waves $\psi_{n,\mathrm{k}}$ which correspond to the eigenvalues $\lambda_{n,\mathrm{k}}$ of band β. Here "independent" means that the phases of the Bloch waves are somehow fixed, and there are no Bloch waves in the set obtained from other waves of the set by a phase shift.*

Remark 3.4. *The requirement for N_β to be finite excludes the case of a constant potential $V =$const because then all the eigenvalue branches touch and thus it is not possible to single out an isolated band.*

Without loss of generality, assume that the Bloch waves of the set β_ψ are enumerated from $n = 1$ to N_β.

Let us introduce a projection operator P^{k}, periodic in k, which characterizes the subspace of the Bloch waves $\psi_{n,\mathrm{k}}$ belonging to β_ψ. For real k this operator can be defined as

$$P^{\mathrm{k}} = \sum_{n=1}^{N_\beta} \psi_{n,\mathrm{k}} \otimes \psi_{n,\mathrm{k}} \tag{3.3.14}$$

or represented by its matrix elements

$$[P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} = \sum_{n=1}^{N_\beta} \psi^*_{n,\mathrm{k}}(\mathrm{r})\psi_{n,\mathrm{k}}(\mathrm{r}')$$

which are invariant with respect to a phase transform of the wave functions $\psi_{n,\mathrm{k}} \mapsto \mathrm{e}^{\mathrm{i}\theta_n(\mathrm{k})}\psi_{n,\mathrm{k}}$:

$$\left[P^{\mathrm{k}}\right]_{\mathrm{r},\mathrm{r}'} \;\mapsto\; \sum_{n=1}^{N_\beta} \mathrm{e}^{-\mathrm{i}\theta_n(\mathrm{k})}\psi^*_{n,\mathrm{k}}(\mathrm{r})\mathrm{e}^{\mathrm{i}\theta_n(\mathrm{k})}\psi_{n,\mathrm{k}}(\mathrm{r}') = [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}. \tag{3.3.15}$$

This operator acts on a function $f \in \mathrm{L}^2(\mathbb{R}^d)$ as

$$P^{\mathrm{k}}\left[f(\mathrm{r})\right] = \sum_{n=1}^{N_\beta} \left\langle \psi_{n,\mathrm{k}}, f \right\rangle_{\mathbb{R}^d} \psi_{n,\mathrm{k}}(\mathrm{r})$$

with inner product

$$\langle f, g \rangle_{\mathbb{R}^d} = \int_{\mathbb{R}^d} f^*(\mathrm{r}) g(\mathrm{r}) \,\mathrm{d}\,\mathrm{r}.$$

For complex $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$, the operator P^{k} can be extended to a strip $|\mathrm{k}''| < A$ for some constant $A > 0$ as

$$P^{\mathrm{k}} = \sum_{n=1}^{N_\beta} \psi_{n,\mathrm{k}^*} \otimes \psi_{n,\mathrm{k}}, \qquad \text{or}$$

$$\left[P^{\mathrm{k}}\right]_{\mathrm{r},\mathrm{r}'} = \sum_{n=1}^{N_\beta} \psi^*_{n,\mathrm{k}^*}(\mathrm{r})\psi_{n,\mathrm{k}}(\mathrm{r}').$$

Let us define another operator P^{k}_u, proportional to P^{k} and projecting onto the span of periodic parts of the Bloch waves:

$$P^{\mathrm{k}}_u = \frac{1}{V_{\mathrm{WSC}}} \sum_{n=1}^{N_\beta} u_{n,\mathrm{k}^*} \otimes u_{n,\mathrm{k}}. \tag{3.3.16}$$

Here V_{WSC} is the volume of the Wigner-Seitz cell and

$$\begin{aligned} u_{n,\mathrm{k}}(\mathrm{r}) &= \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{r}}\psi_{n,\mathrm{k}}(\mathrm{r}), \\ u_{n,\mathrm{k}}(\mathrm{r}+\mathrm{R}) &= u_{n,\mathrm{k}}(\mathrm{r}) \text{ for } \mathrm{R}\in\Gamma. \end{aligned}$$

P^{k} and P_u^{k} are closely related by the following equality:

$$[P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} = V_{\mathrm{WSC}}\cdot\mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot(\mathrm{r}-\mathrm{r}')}[P_u^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}.$$

In [dC64a] P_u^{k} is proved to be analytic in a narrow strip $|\mathrm{k}''| < A$. A straightforward consequence of this result is analyticity of the matrix elements $[P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}$ in the same domain. Let us define a projection operator P as

$$P = \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}} P^{\mathrm{k}}\,\mathrm{d}\,\mathrm{k}. \tag{3.3.17}$$

Its matrix elements

$$P_{\mathrm{r},\mathrm{r}'} = \int_{\mathrm{BZ}}[P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}\,\mathrm{d}\,\mathrm{k} \tag{3.3.18}$$

are continuous in r and r′. The operator is periodic with respect to lattice translations of the crystal:

$$P_{\mathrm{r}+\mathrm{R},\mathrm{r}'+\mathrm{R}'} = P_{\mathrm{r},\mathrm{r}'}, \quad \mathrm{R},\mathrm{R}'\in\Gamma.$$

By its definition, P can be expressed in the form

$$P_{\mathrm{r},\mathrm{r}'} = \sum_{n,\mathrm{R}} w^*_{n,\mathrm{R}}(\mathrm{r})w_{n,\mathrm{R}}(\mathrm{r}'), \tag{3.3.19}$$

where $w_{n,\mathrm{R}}$ are the Wannier functions corresponding to the Bloch modes of the set β_ψ:

$$w_{n,\mathrm{R}}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}}\mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\psi_{n,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}.$$

The relation (3.3.19) is easy to show:

$$\begin{aligned}
\sum_{n,\mathrm{R}} w^*_{n,\mathrm{R}}(\mathrm{r}) w_{n,\mathrm{R}}(\mathrm{r}') &= \sum_{n,\mathrm{R}} \frac{1}{V_{\mathrm{BZ}}^2} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi^*_{n,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}} \psi_{n,\mathrm{k}'}(\mathrm{r}')\,\mathrm{d}\,\mathrm{k}' \\
&= \frac{1}{V_{\mathrm{BZ}}} \sum_n \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \left(\frac{1}{V_{\mathrm{BZ}}} \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}(\mathrm{k}-\mathrm{k}')\cdot\mathrm{R}} \right) \psi^*_{n,\mathrm{k}}(\mathrm{r}) \psi_{n,\mathrm{k}'}(\mathrm{r}')\,\mathrm{d}\,\mathrm{k}\,\mathrm{d}\,\mathrm{k}' \\
&\quad \text{(recall the property (2.4.1))} \\
&= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_n \psi^*_{n,\mathrm{k}}(\mathrm{r}) \psi_{n,\mathrm{k}}(\mathrm{r}')\,\mathrm{d}\,\mathrm{k} \\
&= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}\,\mathrm{d}\,\mathrm{k} = P_{\mathrm{r},\mathrm{r}'}.
\end{aligned}$$

In our further considerations we will need the following result.

Theorem 3.3 (Analyticity of periodic functions and asymptotic properties of their Fourier coefficients)**.** *Let f be a periodic function of a d-dimensional complex vector $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$, admitting real vectors K_j $(j = 1...d)$ as periods: $f(\mathrm{k} + \mathrm{K}_j) = f(\mathrm{k})$. Let R be translational vectors of the reciprocal lattice defined by vectors R_l $(l = 1, ..., d)$ so that*

$$\begin{aligned}
\mathrm{R} &= \sum_l \alpha_l \mathrm{R}_l \quad (\alpha_l \in \mathbb{Z}), \\
\mathrm{K}_j \cdot \mathrm{R}_l &= 2\pi\delta_{jl}.
\end{aligned}$$

Then, if f is analytic in a domain $|k''| < A$ with some constant $A > 0$, it can be expanded in a convergent Fourier series:

$$f(\mathrm{k}) = \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} g_{\mathrm{R}}$$

and the Fourier coefficients satisfy the condition:

$$\lim_{\mathrm{R}\to\infty} \mathrm{e}^{\epsilon A\cdot\mathrm{R}} g_{\mathrm{R}} = 0$$

for some $0 < \epsilon < 1$.

Conversely, if the coefficients g_{R} of a Fourier series have this asymptotic behaviour, the

series converges in the strip $|k''| < A$ *and its sum is an analytic function of* k.

Different variations of this assertion appear in literature, in particular, the result is known as *Paley-Wiener theorem*. Its proof, which is very simple, can be found, for instance, in [dC64a, p. 692].

Theorem 3.3 implies that, due to analyticity of $[P^{\mathrm{k}}]_{\mathrm{r,r'}}$ in a vicinity of the real values of k ($|\mathrm{k}''| < A$), the matrix elements $P_{\mathrm{r,r'}}$ decrease exponentially with $|\mathrm{r} - \mathrm{r}'| \to \infty$.
But unfortunately, as it is pointed out in [MV97, p. 12856], this result does not immediately imply exponential localization of the Wannier functions. However, it is possible to prove the possibility of choosing such Wannier functions, as des Cloizeaux does in his paper [dC64b]. Let us review it shortly.

The Bloch waves can be represented as Fourier sums of the Wannier functions:

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} w_{n,\mathrm{R}}(\mathrm{r}).$$

With the help of the Theorem 3.3, if $\psi_{n,\mathrm{k}}$ is analytic in $\mathrm{k} = \mathrm{k}' + \mathrm{ik}''$ in the strip $|\mathrm{k}''| < A$ with a constant $A > 0$, then its Fourier coefficients $w_{n,\mathrm{R}}(\mathrm{r})$ satisfy conditions of the form:

$$\lim_{\mathrm{R}'\to\infty} \mathrm{e}^{\epsilon A \mathrm{R}'} w_{n,\mathrm{R}}(\mathrm{r} + \mathrm{R}') = 0$$

with $0 < \epsilon < 1$. The converse is also true. Thus the problem is reduced to the construction of Bloch waves which are analytic in k in a strip centered on the real axis.

For a *simple band* ($N_\beta = 1$), these functions can be constructed explicitly. Introduce trial Wannier functions $w^{\mathrm{trial}}_{n,0}$ which are arbitrarily normalizable with $\mathrm{e}^{A\mathrm{r}} \left|w^{\mathrm{trial}}_{n,0}(\mathrm{r})\right| < \infty$ and have a bounded support. These can be, for example, true Wannier functions with a

cutoff:

$$w_{n,0}^{\mathrm{trial}}(\mathrm{r}) = \begin{cases} w_{n,0}(\mathrm{r}), & |\mathrm{r}| < C, \\ \theta(\mathrm{r}), & C \le |\mathrm{r}| \le C+\delta, \\ 0, & |\mathrm{r}| > C+\delta, \end{cases}$$

with some $C > 0$, some small $\delta > 0$ and a smooth function θ such that $\theta(\mathrm{r})_{|\mathrm{r}|=C} = w_{n,0}(\mathrm{r})_{|\mathrm{r}|=C}$ and $\theta(\mathrm{r})_{|\mathrm{r}|=C+\delta} = 0$.

Apply the following construction:

1) $\phi_{n,\mathrm{k}}(\mathrm{r}) = \int_{\mathbb{R}^d} [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} w_{n,0}^{\mathrm{trial}}(\mathrm{r}')\,\mathrm{d}\,\mathrm{r}'$, or equivalently $\phi_{n,\mathrm{k}}(\mathrm{r}) = \left\langle \psi_{n,\mathrm{k}}, w_{n,0}^{\mathrm{trial}} \right\rangle_{\mathbb{R}^d} \psi_{n,\mathrm{k}}(\mathrm{r})$
(projection of trial functions onto the span of Bloch modes)

2) $G^{\mathrm{k}} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} \left(w_{n,\mathrm{R}_0}^{\mathrm{trial}}(\mathrm{r})\right)^* w_{n,\mathrm{R}_0}^{\mathrm{trial}}(\mathrm{r}')\,\mathrm{d}\,\mathrm{r}'\,\mathrm{d}\,\mathrm{r}$, which is the same as
$[G^{\mathrm{k}}]_{mn} = \langle \phi_{m,\mathrm{k}}, \phi_{n,\mathrm{k}} \rangle_{\mathbb{R}^d}$
(orthonormalization)

3) $\psi_{n,\mathrm{k}}^{\mathrm{new}}(\mathrm{r}) = \left[G^{\mathrm{k}}\right]^{-1/2} \phi_{n,\mathrm{k}}(\mathrm{r})$
(new Bloch modes)

The integrals converge uniformly, therefore, the functions $\mathrm{k} \mapsto \phi_{n,\mathrm{k}}$ and $\mathrm{k} \mapsto G^{\mathrm{k}}$ are analytic in k for $|\mathrm{k}''| < A$ where A is the width of the strip of analyticity of $[P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'}$. Moreover, by its definition, G^{k} is real and non-negative for real k, and, if G^{k} is strictly positive, this remains true for some strip $|\mathrm{k}''| \le A_0 < A$ with $A_0 > 0$ and constant. Consequently, $\left[G^{\mathrm{k}}\right]^{-1/2}$ is also analytic in this strip and, finally, $\psi_{n,\mathrm{k}}^{\mathrm{new}}$ is analytic — and this is just what we want. The corresponding Wannier functions

$$w_{n,\mathrm{R}}^{\mathrm{new}}(\mathrm{r}) = \frac{1}{\sqrt{V_{\mathrm{BZ}}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi_{n,\mathrm{k}}^{\mathrm{new}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}$$

fall off exponentially at infinity.

Positivity of G^{k} will be discussed further.

Remark 3.5. *The construction described above can be interpreted as choosing a phase for the Bloch waves:*

$$\psi_{n,\mathrm{k}} \mapsto \left[G^{\mathrm{k}}\right]^{-1/2} \left\langle \psi_{n,\mathrm{k}}, w_{n,0}^{trial} \right\rangle_{\mathbb{R}^d} \psi_{n,\mathrm{k}} = \mathrm{e}^{\mathrm{i}\theta_n(\mathrm{k})} \psi_{n,\mathrm{k}},$$

since this transform keeps the orthonormality.

Let us now consider the case of a *complex band*, or a *composite group* of eigenvalue branches ($N_\beta > 1$). In such a group the branches of $\lambda_{n,\mathrm{k}}$ are connected with each other by degeneracies but separated from all the lower and higher branches by gaps. We introduce the *generalized Bloch waves* (in [dC64b] they are named *quasi-Bloch waves*):

$$\widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r}) = \sum_{m=1}^{N_\beta} \mathrm{U}_{mn}^{\mathrm{k}} \psi_{m,\mathrm{k}}(\mathrm{r}),$$

where U^{k} is a unitary matrix of size $N_\beta \times N_\beta$ for every k. The set of $\widetilde{\psi}_{n,\mathrm{k}}$ inherits the following properties of $\psi_{n,\mathrm{k}}$:

1. $\widetilde{\psi}_{n,\mathrm{k}}$ also form a complete orthonormal set in $\mathrm{L}^2(\mathrm{WSC} \times \mathrm{BZ})$.

2. For every $\mathrm{k} \in \mathrm{BZ}$ they span the same space $S_\beta(\mathrm{k})$ of functions formed by the Bloch waves $\psi_{n,\mathrm{k}}$ of the set β_ψ (Notation 3.1), namely:

$$\begin{aligned} S_\beta(\mathrm{k}) &= \left\{ f_{\mathrm{k}}(\mathrm{r}) = \sum_{n=1}^{N_\beta} \alpha_{n,\mathrm{k}} \psi_{n,\mathrm{k}}(\mathrm{r}) \mid \psi_{n,\mathrm{k}} \in \beta_\psi,\ \alpha_{n,\mathrm{k}} \in \mathbb{C} \right\} \\ &= \left\{ g_{\mathrm{k}}(\mathrm{r}) = \sum_{n=1}^{N_\beta} \gamma_{n,\mathrm{k}} \widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r}) \mid \gamma_{n,\mathrm{k}} \in \mathbb{C} \right\}. \end{aligned}$$

3. They are periodic in k with the same period: $\widetilde{\psi}_{n,\mathrm{k}+\mathrm{K}}(\mathrm{r}) = \widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r})$, $\mathrm{K} \in \Gamma^*$.

These three properties allow us to call the new set "quasi-*Bloch* waves", or "generalized *Bloch* waves".

Accordingly, we introduce *generalized Wannier functions* as being obtained by

$$\widetilde{w}_{n,\mathrm{R}}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{R})} \widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r})\, \mathrm{d}\,\mathrm{k}.$$

They possess the following properties:

1. $\{\widetilde{w}_{n,\mathrm{R}}\}_{n,\mathrm{R}}$ form a complete orthonormal set in $\mathrm{L}^2(\mathbb{R}^d)$.

2. $\{\widetilde{w}_{n,\mathrm{R}}\}_{n,\mathrm{R}}$ span the space S_β formed by the Wannier functions $w_{n,\mathrm{R}}$ which correspond to the set β_ψ:

$$\begin{aligned} S_\beta &= \left\{ f(\mathrm{r}) = \sum_{n=1}^{N_\beta} \sum_{\mathrm{R}\in\Gamma} \alpha_{n,\mathrm{R}} w_{n,\mathrm{R}}(\mathrm{r}) \mid w_{n,\mathrm{R}} = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{ik}\cdot\mathrm{R}} \psi_{n,\mathrm{k}}\, \mathrm{d}\,\mathrm{k},\ \psi_{n,\mathrm{k}} \in \beta_\psi \right\} \\ &= \left\{ g(\mathrm{r}) = \sum_{n=1}^{N_\beta} \sum_{\mathrm{R}\in\Gamma} \gamma_{n,\mathrm{R}} \widetilde{w}_{n,\mathrm{R}}(\mathrm{r}) \right\}. \end{aligned}$$

3. The functions $\widetilde{w}_{n,\mathrm{R}}$ corresponding to a site R can be obtained from those attached to the lattice origin $\mathrm{R} = 0$ by a rotation or a reflection of the lattice. In particular, $\widetilde{w}_{n,\mathrm{R}}(\mathrm{r}) = \widetilde{w}_{n,0}(\mathrm{r} - \mathrm{R})$.

In terms of the introduced notations, the goal is to prove the existence of generalized Wannier functions which decay exponentially at infinity.

As in a simple band case, by applying Theorem 3.3 to Fourier series, we conclude that existence of a set of exponentially decreasing Wannier functions is equivalent to the possibility to build quasi-Bloch waves which are analytic in a domain $|\mathrm{k}''| < A$ with some positive constant A. Again, our plan is to prove such a possibility by construction, and to do so we need a generalization of the method described for a simple band. Let us review it briefly.

First, choose trial Wannier functions $w_{n,0}^{\mathrm{trial}}$ which satisfy the properties of generalized Wannier functions and, moreover, have exponential tails. For example, these may be

polynomial functions of r for $|\mathrm{r}| \leq C$ and zero for $|\mathrm{r}| > C$, $C > 0$. As before, unnormalized trial waves are determined by

$$\phi_{n,\mathrm{k}}(\mathrm{r}) \;=\; \int_{\mathbb{R}^d} [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} w_{n,0}^{\mathrm{trial}}(\mathrm{r}')\,\mathrm{d}\,\mathrm{r}'.$$

Their orthonormalization matrix turns out to be

$$G^{\mathrm{k}} \;=\; \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} [P^{\mathrm{k}}]_{\mathrm{r},\mathrm{r}'} \left(w_{n,0}^{\mathrm{trial}}(\mathrm{r})\right)^* w_{n,0}^{\mathrm{trial}}(\mathrm{r}')\,\mathrm{d}\,\mathrm{r}'\,\mathrm{d}\,\mathrm{r}.$$

And finally, the quasi-Bloch waves are introduced by putting

$$\widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r}) \;=\; \left[G^{\mathrm{k}}\right]^{-1/2} \phi_{n,\mathrm{k}}(\mathrm{r}).$$

$\phi_{n,\mathrm{k}}$ and G^{k} are analytic functions of k in a strip $|\mathrm{k}''| < A$. The question is, however, whether $[G^{\mathrm{k}}]^{-1/2}$ is analytic or not.

By its definition, G^{k} is non-negative for real values of k, and thus we can define the non-negative function $[G^{\mathrm{k}}]^{-1/2}$. Furthermore, if G^{k} is strictly positive for k real, then $[G^{\mathrm{k}}]^{-1/2}$ can be continued analytically into a complex-space domain $|\mathrm{k}''| < A_0 \leq A$ with some constant $A_0 > 0$. In this case the quasi-Bloch waves $\widetilde{\psi}_{n,\mathrm{k}}$ are also analytic in the same domain. By Theorem 3.3, the corresponding Wannier functions

$$\widetilde{w}_{n,\mathrm{R}} = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \widetilde{\psi}_{n,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}$$

have exponential decay at infinity.

In both simple and complex bands, the open question which may (or may not) provide us with the analyticity of (quasi-) Bloch waves is *positivity* of G^{k} for real k. Henceforth, to make the description shorter, we write just "Bloch waves", but for a complex band have in mind "quasi-Bloch waves".

We already have a result of Kohn [Koh59] for centrosymmetric one-dimensional crystals. Des Cloizeaux [dC64b] removes the restriction of inverse symmetry in 1D and proves the possibility to construct analytic Bloch waves.

Let $\mathrm{k} = \mathrm{k}_0$ be a point in which G^{k} vanishes. As an analytic function, G^{k} can be expanded in a convergent series in the vicinity of k_0:

$$G^{\mathrm{k}} = a^2(\mathrm{k} - \mathrm{k}_0)^{2p}[1 + \epsilon(\mathrm{k} - \mathrm{k}_0)].$$

Here $a > 0$, $p \in \mathbb{N}$ (namely, the exponent $2p$ is even), $\epsilon(\mathrm{k})$ is analytic and $\epsilon(\mathrm{k}_0) = 0$. Thus,

$$[G^{\mathrm{k}}]^{1/2} = \pm a(\mathrm{k} - \mathrm{k}_0)^p[1 + \epsilon(\mathrm{k} - \mathrm{k}_0)]^{1/2}.$$

$[G^{\mathrm{k}}]^{1/2}$ can be chosen to be analytic in $\mathrm{k} = \mathrm{k}_0$. Therefore, if a crystal is linear, there exist exponentially localized Wannier functions.

In d-dimensional crystals, if $G^{\mathrm{k}_0} = 0$, the function $[G^{\mathrm{k}}]^{1/2}$ has a singularity in this point, and its effect often cannot be eliminated. However, it is possible to overcome this difficulty if the crystal has a center of inversion in R_0, in particular, $V(\mathrm{R}_0 - \mathrm{r}) = V(\mathrm{R}_0 + \mathrm{r})$. Without loss of generality, R_0 may be zero.

In this case trial Wannier functions $w^{\mathrm{trial}}_{n,\mathrm{R}_0}$ are symmetric or antisymmetric with respect to this point [dC64b]. The Bloch functions can obey the form:

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = s_{n,\mathrm{k}}(\mathrm{r}) + \mathrm{i}a_{n,\mathrm{k}}(\mathrm{r}),$$

where the functions $s_{n,\mathrm{k}}$ and $a_{n,\mathrm{k}}$ are real and respectively symmetric or antisymmetric. This representation implies that

$$\begin{aligned} G^{\mathrm{k}} &= \left\langle s_{n,\mathrm{k}}, w^{\mathrm{trial}}_{n,\mathrm{R}_0} \right\rangle^2_{\mathbb{R}^d}, \quad \text{if } w^{\mathrm{trial}}_{n,\mathrm{R}_0} \text{ is symmetric;} \\ G^{\mathrm{k}} &= \left\langle a_{n,\mathrm{k}}, w^{\mathrm{trial}}_{n,\mathrm{R}_0} \right\rangle^2_{\mathbb{R}^d}, \quad \text{if } w^{\mathrm{trial}}_{n,\mathrm{R}_0} \text{ is antisymmetric.} \end{aligned}$$

In such a way, $[G^{\mathrm{k}}]^{1/2}$ is analytic and real.

As a consequence, we have shown the existence of exponentially localized Wannier functions (i) for 1D crystals in general and (ii) for multidimensional crystals under a restriction of inverse symmetry.

Remark 3.6. *The approach of des Cloizeaux can be applied in the same manner to the Transverse Magnetic problem* (2.1.5) *instead of* (3.3.13)*, with the corresponding substitution of the inner product:* $\langle \cdot, \cdot \rangle_{\mathbb{R}^d} \mapsto \langle \cdot, \cdot \rangle_{\mathbb{R}^d, \varepsilon}$.

3.3.3 Nenciu's proof for crystals without inversion symmetry

The projection method of des Cloizeaux was later used by Gheorghe Nenciu in [Nen83] to generalize the proof for *crystals without an inversion center*.
Consider a nondegenerating band. Let us split a space vector r into two components:

$$\mathrm{r} = \rho + \mathrm{R},$$

where $\rho \in \mathrm{WSC}$, $\mathrm{R} \in \Gamma$. Represent the Bloch waves as

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}} u_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\rho} u_{n,\mathrm{k}}(\rho) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} v_{n,\mathrm{k}}(\rho).$$

The definition of Bloch waves is not unique because of a k-dependent phase factor; let it be included into notation $\widehat{u}_{n,\mathrm{k}}(\rho)$ and the corresponding $\widehat{v}_{n,\mathrm{k}}(\rho)$ and $\widehat{\psi}_{n,\mathrm{k}}(\mathrm{r})$.
The Wannier functions turn out to be

$$\widehat{w}_{n,0}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \widehat{v}_{n,\mathrm{k}}(\rho)\, \mathrm{d}\,\mathrm{k}.$$

By Theorem 3.3, the functions $\widehat{w}_{n,0}$ are exponentially decaying, if and only if their Fourier sums $\widehat{v}_{n,\mathrm{k}}(\rho)$ are periodic and analytic in k. To construct functions which satisfy these two conditions, we need the following result.

Theorem 3.4 (G. Nenciu)**.** *Let X be a separable Hilbert space and $Q(\mathrm{k}) : X \to X$ be a projection operator, dependent on the d-dimensional complex vector $\mathrm{k} = \mathrm{k}' + \mathrm{i}\mathrm{k}''$ and analytic in a strip $|\mathrm{k}''| < A$ $(A > 0)$. Let $Q(\mathrm{k})$ satisfy the conditions:*

$$\begin{aligned} Q(\mathrm{k}) &= Q^*(\mathrm{k}) \qquad \textit{for } \mathrm{k} \in \mathbb{R}^d, \\ Q(\mathrm{k}) &= Q(\mathrm{k} + 2\pi\mathrm{K}) \qquad \textit{for } |\mathrm{k}''| < A,\ \mathrm{K} \in \mathbb{Z}^d, \\ \dim(Q(\mathrm{k})) &= 1. \end{aligned}$$

If there exists an antilinear involution $\theta : X \to X$ such that

$$\theta Q(\mathrm{k})\theta = Q(-\mathrm{k}) \qquad \textit{for } \mathrm{k} \in \mathbb{R}^d,$$

then there exists a bounded operator-valued function $B(\mathrm{k}) : X \to X$*, with bounded inverse, analytic in the strip* $|\mathrm{k}''| < A$ *and satisfying*

$$\begin{aligned} Q(\mathrm{k}) &= B(\mathrm{k})Q(0)B^{-1}(\mathrm{k}), \qquad B(0) = 1 \quad \textit{for } |k''| < A, \\ B^{-1}(\mathrm{k}) &= B^*(\mathrm{k}) \qquad \textit{for } \mathrm{k} \in \mathbb{R}^d, \\ B(\mathrm{k})Q(0) &= B(\mathrm{k} + 2\pi\mathrm{K})Q(0) \qquad \textit{for } |\mathrm{k}''| < A,\ \mathrm{K} \in \mathbb{Z}^d. \end{aligned}$$

A rigorous proof of this theorem can be found in [Nen83].

Applying Theorem 3.4, we now build now analytic and periodic in k functions $\widehat{v}_{n,\mathrm{k}}(\rho)$. Let $Q_n(\mathrm{k})$ be the orthogonal projection in $\mathrm{L}^2_\varepsilon(\mathrm{WSC})$ (and therefore a restriction to BZ of $Q(\mathrm{k})$ from Theorem 3.4). Evidently, $\dim(Q(\mathrm{k})) = 1$ — this results from the nondegeneracy of the band β. Let θ be the involution representing complex conjugation such that $\theta\psi_{n,\mathrm{k}} = \psi_{n,-\mathrm{k}}$ up to a phase factor. $Q(\mathrm{k})$ and θ satisfy the conditions of Theorem 3.4; therefore, a function $B(\mathrm{k})$ can be determined.
Consider the phase factor of $v_{n,0}$ fixed. Then functions $\widehat{v}_{n,\mathrm{k}} = B(\mathrm{k})v_{n,0}$ are analytic and periodic in k.
The existence of exponentially decaying Wannier functions of nondegenerated bands in arbitrary multidimensional crystals is proven.

3.4 Projected position operator in 1D

In this section we will consider a special case when the exponentially decaying Wannier functions can be found explicitly.
Consider a 1D crystal with inversion symmetry. As it is noticed in [dC64a], one-dimensional crystals have in general *only simple bands*. We fix an n-th energy band.
As before, let P_n be a projection operator onto the subspace S_n of this band:

$$P_n = \sum_{\mathrm{R}\in\Gamma} w_{n,\mathrm{R}} \otimes w_{n,\mathrm{R}},$$

where $\{w_{n,\mathrm{R}}\}_{\mathrm{R}}$ are the Wannier functions of the n-th band:

$$w_{n,\mathrm{R}}(x) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{ikR}} \psi_{n,\mathrm{k}}(x)\, \mathrm{d\,k}.$$

It acts on a function $f \in \mathrm{L}^2(\mathbb{R})$ as

$$P_n f = \sum_{\mathrm{R}\in\Gamma} \langle w_{n,\mathrm{R}}, f\rangle_{\mathbb{R}}\, w_{n,\mathrm{R}}.$$

Define $\widehat{R}_n$ as a projected position operator:

$$\widehat{R}_n = P_n x P_n, \tag{3.4.1}$$

$$\widehat{R}_n f = \sum_{\mathrm{R},\mathrm{R}'} \left\langle x w_{n,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}} \left\langle w_{n,\mathrm{R}'}, f\right\rangle_{\mathbb{R}} w_{n,\mathrm{R}}. \tag{3.4.2}$$

S. Kivelson in his article [Kiv82] has shown that exponentially localized Wannier functions of the band n are eigenfunctions of the operator $\widehat{R}_n$:

$$\widehat{R}_n w^{\mathrm{loc}}_{n,\mathrm{R}}(x) = \mathrm{R} w^{\mathrm{loc}}_{n,\mathrm{R}}(x). \tag{3.4.3}$$

We are going to examine this assertion.

Recall the translational invariance of the crystal and the orthogonality of the Wannier functions:

$$w_{n,\mathrm{R}}(x) = w_{n,\mathrm{R}-\mathrm{R}'}(x-\mathrm{R}'), \quad \text{or} \quad w_{n,\mathrm{R}'}(x) = w_{n,0}(x-\mathrm{R}'),$$

$$\left\langle w_{n,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}} = \delta_{\mathrm{R},\mathrm{R}'}.$$

One can write the following sequence of equations:

$$\begin{aligned}
\left\langle \widehat{R}_n w_{n,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}} &= \left\langle P_n x P_n w_{n,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}} \\
&= \left\langle P_n(\mathrm{R}' + (x-\mathrm{R}'))P_n w_{n,\mathrm{R}-\mathrm{R}'}(\cdot-\mathrm{R}'), w_{n,0}(\cdot-\mathrm{R}')\right\rangle_{\mathbb{R}} \\
&= \mathrm{R}' \left\langle w_{n,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}} + \left\langle \widehat{R}_n w_{n,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}} \\
&= \delta_{\mathrm{R},\mathrm{R}'}\mathrm{R} + \left\langle \widehat{R}_n w_{n,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}}.
\end{aligned}$$

We want to show that the second term can vanish. By definition of the projected position operator,

$$\widehat{R}_n w_{n,\mathrm{R}} = x w_{n,\mathrm{R}}.$$

According to Kohn's theorem 3.2, the Wannier functions can be (uniquely) chosen to be real and (anti-)symmetric $w^{\rm loc}_{n,{\rm R}}(x) = \pm w^{\rm loc}_{n,-{\rm R}}(-x)$. (In addition, such functions are exponentially localized). Thus,

$$\begin{aligned}
\left\langle \widehat{R}_n w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}} &= \left\langle x w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}} \\
&= (\pm 1)^2 \left\langle -x w^{\rm loc}_{n,-{\rm R}}(-\cdot), w^{\rm loc}_{n,0}(-\cdot) \right\rangle_{\mathbb{R}} \\
&= -\left\langle x w^{\rm loc}_{n,0}, w^{\rm loc}_{n,-{\rm R}} \right\rangle_{\mathbb{R}} \\
&= -\left\langle (x-{\rm R}) w^{\rm loc}_{n,{\rm R}}(\cdot - {\rm R}), w^{\rm loc}_{n,0}(\cdot - {\rm R}) \right\rangle_{\mathbb{R}} + {\rm R}\delta_{{\rm R},0} \\
&= -\left\langle x w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}},
\end{aligned}$$

and, comparing the first and the last lines, we conclude:

$$\left\langle \widehat{R}_n w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}} = 0.$$

Consequently, for the exponentially decaying Wannier functions it holds

$$\left\langle \widehat{R}_n w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{n,{\rm R}'} \right\rangle_{\mathbb{R}} = \delta_{{\rm R},{\rm R}'}{\rm R}.$$

Consider another isolated band m with its own set of Wannier functions $w_{m,\rho}$, $m \neq n$. From the definition of $\widehat{R}_n$,

$$\begin{aligned}
\widehat{R}_n w_{m,\rho} &= \sum_{{\rm R},{\rm R}'} \left\langle x w_{n,{\rm R}}, w_{n,{\rm R}'} \right\rangle \left\langle w_{n,{\rm R}'}, w_{m,\rho} \right\rangle w_{n,{\rm R}}(x) \\
&= \sum_{{\rm R},{\rm R}'} \left\langle x w_{n,{\rm R}}, w_{n,{\rm R}'} \right\rangle \delta_{m,n}\delta_{\rho,{\rm R}'} w_{n,{\rm R}}(x) \\
&= 0.
\end{aligned}$$

Therefore, for an arbitrary l, R, R$'$ it holds:

$$\left\langle \widehat{R}_n w^{\rm loc}_{n,{\rm R}}, w^{\rm loc}_{l,{\rm R}'} \right\rangle = \delta_{n,l}\delta_{{\rm R},{\rm R}'}{\rm R}.$$

The last equation is the matrix equivalent of the eigenvalue problem:

$$\widehat{R}_n w^{\rm loc}_{n,{\rm R}}(x) = {\rm R} w^{\rm loc}_{n,{\rm R}}(x)$$

Finally, let us formulate a theorem to sum up the result of this section.

Theorem 3.5 (The eigenfunctions of the projected position operator, S. Kivelson)**.** *Let n be an isolated band for a 1D periodic crystal with inversion symmetry. Let P_n be a projection operator onto the subspace of this band:*

$$P_n = \sum_{\mathrm{R}\in\Gamma} w_{n,\mathrm{R}} \otimes w_{n,\mathrm{R}}$$

and let $\widehat{R}_n$ be a projected position operator:

$$\widehat{R}_n = P_n x P_n.$$

Then the exponentially localized Wannier functions of the band are the eigenfunctions of the operator $\widehat{R}_n$:

$$\widehat{R}_n w^{loc}_{n,\mathrm{R}}(x) = \mathrm{R} w^{loc}_{n,\mathrm{R}}(x).$$

A. Nenciu and G. Nenciu [NN82] have shown that the operator $\widehat{R}_n$ has discrete spectrum and the functions from Theorem 3.5 are its only eigenfunctions. In [CNN08] the authors H. D. Cornean, A. Nenciu and G. Nenciu generalize this result for arbitrary one-dimensional crystals, not necessarily periodic — with a potential which tends to zero as the variable tends to infinity, — and prove that the eigenfunctions of $\widehat{R}_n$ are maximally localized Wannier functions in this case also.

3.5 Numerical computation of maximally localized Wannier functions in a 1D Photonic Crystal

Our aim is to compute the eigenfunctions of the projected position operator — which are at the same time the Kohn's localized Wannier functions — numerically. The formulas below should be applied to each band separately, for this reason the band index n is further omitted.

The projected position operator $\widehat{R} = PxP$ is independent on the choice of phases of the

Bloch waves because the operator P does not depend on them (3.3.15). Thus, let us fix some initial set of Wannier functions w_{R}^0; it is not yet important how we do it. Let the functions we are looking for be $w_{\mathrm{R}}^{\mathrm{loc}}$. Expand them into the basis $\{w_{\mathrm{R}}^0\}_{\mathrm{R}\in\mathbb{Z}}$:

$$w_{\mathrm{R}}^{\mathrm{loc}}(x) = \sum_{\mathrm{R}'\in\Gamma} \alpha_{\mathrm{R}'}^{(\mathrm{R})} w_{\mathrm{R}'}^0(x). \tag{3.5.1}$$

The eigenvalue problem to solve can be then reformulated:

$$\begin{aligned} \widehat{R} w_{\mathrm{R}}^{\mathrm{loc}} &= \mathrm{R} w_{\mathrm{R}}^{\mathrm{loc}}, \\ \sum_{\rho,\rho'} \left\langle x w_\rho^0, w_{\rho'}^0 \right\rangle_{\mathbb{R},\varepsilon} \left\langle w_{\rho'}^0, \sum_{\mathrm{R}'\in\Gamma} \alpha_{\mathrm{R}'}^{(\mathrm{R})} w_{\mathrm{R}'}^0 \right\rangle_{\mathbb{R},\varepsilon} w_\rho^0 &= \mathrm{R} \sum_\rho \alpha_\rho^{(\mathrm{R})} w_\rho^0. \end{aligned}$$

Let $X^{\rho,\rho'} = \left\langle x w_\rho^0, w_{\rho'}^0 \right\rangle_{\mathbb{R},\varepsilon}$, then we have:

$$\begin{aligned} \sum_{\rho,\rho'} X^{\rho,\rho'} \sum_{\mathrm{R}'\in\Gamma} \alpha_{\mathrm{R}'}^{(\mathrm{R})} \delta_{\rho',\mathrm{R}'} w_\rho^0 &= \mathrm{R} \sum_\rho \alpha_\rho^{(\mathrm{R})} w_\rho^0, \\ \sum_{\rho,\rho'} X^{\rho,\rho'} \alpha_{\rho'}^{(\mathrm{R})} w_\rho^0 &= \mathrm{R} \sum_\rho \alpha_\rho^{(\mathrm{R})} w_\rho^0, \\ \sum_{\rho'} X^{\rho,\rho'} \alpha_{\rho'}^{(\mathrm{R})} &= \mathrm{R} \alpha_\rho^{(\mathrm{R})}. \end{aligned}$$

Therefore, the eigenvalue problem for the projected position operator is reduced to the following matrix equation:

$$X\alpha^{(\mathrm{R})} = \mathrm{R}\alpha^{(\mathrm{R})}. \tag{3.5.2}$$

We know a priori the eigenvalues of X to be $\mathrm{R}\in\mathbb{Z}$, so when we find its eigenvectors $\alpha^{(\mathrm{R})}$, the corresponding eigenfunctions of the operator $\widehat{R}$ can be easily found:

$$w_{\mathrm{R}}^{\mathrm{loc}}(x) = \sum_{\mathrm{R}'\in\Gamma} \alpha_{\mathrm{R}'}^{(\mathrm{R})} w_{\mathrm{R}'}^0(x).$$

How to calculate $X^{\rho,\rho'}$? Recall:

$$\begin{aligned} X^{\mathrm{R}\mathrm{R}'} &= \left\langle x w_{\mathrm{R}}^0, w_{\mathrm{R}'}^0 \right\rangle_{\mathbb{R},\varepsilon} \\ &= \left\langle (x - \mathrm{R}' + \mathrm{R}') w_{\mathrm{R}-\mathrm{R}'}^0(\cdot - \mathrm{R}'), w_0^0(\cdot - \mathrm{R}') \right\rangle_{\mathbb{R},\varepsilon} \\ &= \left\langle x w_{\mathrm{R}-\mathrm{R}'}^0, w_0^0 \right\rangle_{\mathbb{R},\varepsilon} + \mathrm{R}\delta_{\mathrm{R},\mathrm{R}'}. \end{aligned}$$

Now we want to use some facts which will be described in detail later. Since we need them only for the numerical implementation, let us omit the explanations in this section and take only the results. In Section 5.1 it will be pointed out that

$$\left\langle x w^0_{\mathrm{R}-\mathrm{R}'}, w^0_0 \right\rangle_{\mathbb{R},\varepsilon} = \frac{\mathrm{i}}{2\pi} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}(\mathrm{R}-\mathrm{R}')} \left\langle u^0_{\mathrm{k}}, \partial_{\mathrm{k}} u^0_{\mathrm{k}} \right\rangle_{\mathrm{WSC},\varepsilon} \mathrm{d}\,\mathrm{k},$$

where $u^0_{\mathrm{k}}(x) = \mathrm{e}^{-\mathrm{i}\mathrm{k}x}\psi^0_{\mathrm{k}}(x)$ are periodic in x with respect to the nodes R: $u^0_{\mathrm{k}}(x+\mathrm{R}) = u^0_{\mathrm{k}}(x)$. So far we have considered *continuous* problem, but now we turn to numerics. To do so, we have to approximate X. After [MV97] (which will be considered in Section 5.2), the last expression can be approximated by

$$\frac{\mathrm{i}}{2\pi} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}(\mathrm{R}-\mathrm{R}')} \left\langle u^0_{\mathrm{k}}, \partial_{\mathrm{k}} u^0_{\mathrm{k}} \right\rangle_{\mathrm{WSC},\varepsilon} \mathrm{d}\,\mathrm{k} \approx -\frac{1}{4\pi} \sum_{\mathrm{k}=1}^{N_{\mathrm{k}}} \mathrm{e}^{\mathrm{i}\mathrm{k}(\mathrm{R}-\mathrm{R}')} \operatorname{Im} \ln \frac{M^{\mathrm{k},1}}{M^{\mathrm{k},-1}}$$

with

$$M^{\mathrm{k},\mathrm{s}} = \left\langle u^0_{\mathrm{k}}, u^0_{\mathrm{k}+\mathrm{s}} \right\rangle_{\mathrm{WSC},\varepsilon}.$$

Introduce another notation:

$$\widetilde{M}^{\mathrm{k}} = \operatorname{Im} \ln \frac{M^{\mathrm{k},1}}{M^{\mathrm{k},-1}} = \operatorname{Im} \ln \frac{\left\langle u^0_{\mathrm{k}}, u^0_{\mathrm{k}+1} \right\rangle_{\mathrm{WSC}}}{\left\langle u^0_{\mathrm{k}}, u^0_{\mathrm{k}-1} \right\rangle_{\mathrm{WSC},\varepsilon}}.$$

Then

$$\left\langle x w^0_{\mathrm{R}-\mathrm{R}'}, w^0_0 \right\rangle_{\mathbb{R}^d} \approx -\frac{1}{4\pi} \sum_{\mathrm{k}=1}^{N_{\mathrm{k}}} \mathrm{e}^{\mathrm{i}\mathrm{k}(\mathrm{R}-\mathrm{R}')} \widetilde{M}^{\mathrm{k}}$$

and thus

$$X^{\mathrm{R}\mathrm{R}'} \approx -\frac{1}{4\pi} \sum_{\mathrm{k}=1}^{N_{\mathrm{k}}} \mathrm{e}^{\mathrm{i}\mathrm{k}(\mathrm{R}-\mathrm{R}')} \widetilde{M}^{\mathrm{k}} + \mathrm{R}\delta_{\mathrm{R},\mathrm{R}'}. \tag{3.5.3}$$

Now we have the formula to solve the eigenvalue problem (3.5.2) numerically. First, we need the initial functions u^0_{k}. These can be obtained as the numerical solution of (2.5.1). What we actually do here is implementing of the Finite Element method; for a description see Appendix, Section 2. Then we fix the amount $2L+1$ of considered R-nodes (this

means, L to the left and L to the right from R = 0) and compute matrix X by (3.5.3). The eigenvector $\alpha^{(\mathrm{R})}$ of (3.5.2) can be now found in Matlab with use of the subroutine "eig". The eigenfunctions $w_{\mathrm{R}}^{\mathrm{loc}}$ of the projected position operator $\widehat{R}$ are then determined by (3.5.1).

We implement the algorithm with the following input data:

1. The Photonic Crystal:
   - nodes R are integer from -50 to 50 ($L = 50$);
   - in the Wigner-Seitz cell $[0, 1]$

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, 0.41) \cup (0.59, 1], \\ 11.56, & x \in [0.41, 0.59]; \end{cases}$$

2. Mesh in x-space: $N_x = 51$, $h_x = 0.02$;
3. Monkhorst-Pack mesh in k-space as in Section 5.2: $N_{\mathrm{k}} = 100$ $h_{\mathrm{k}} = 2\pi/100$;
4. Linear Finite Elements to compute u_{k}^0;
5. Implemented for the first 4 bands ($n = 1, 2, 3, 4$).

As we see on Figures 3.1-3.4, the resulting functions look as expected: real, (anti-) symmetric about 0 or 1/2 and exponentially localized.

Figure 3.1: Eigenfunction of $\widehat{R}_1$

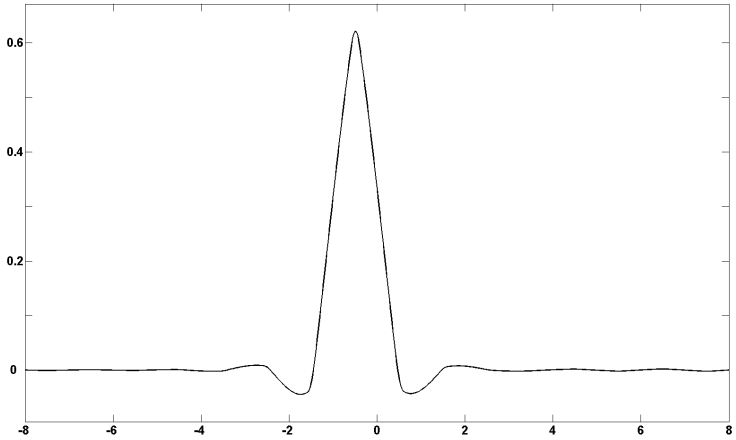


Figure 3.2: Eigenfunction of $\widehat{R}_2$

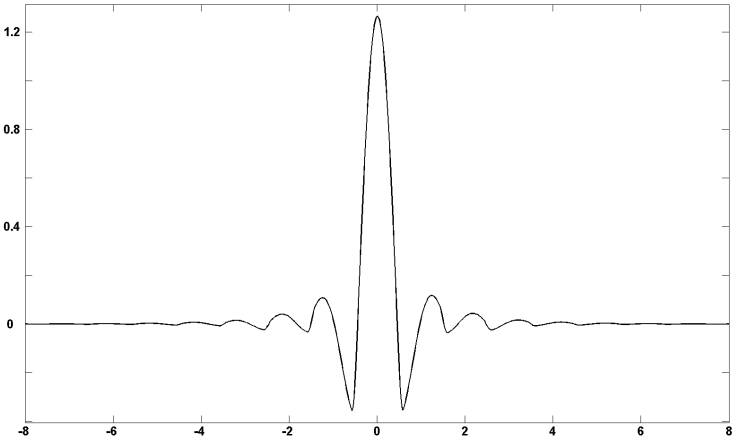


Figure 3.3: Eigenfunction of $\widehat{R}_3$

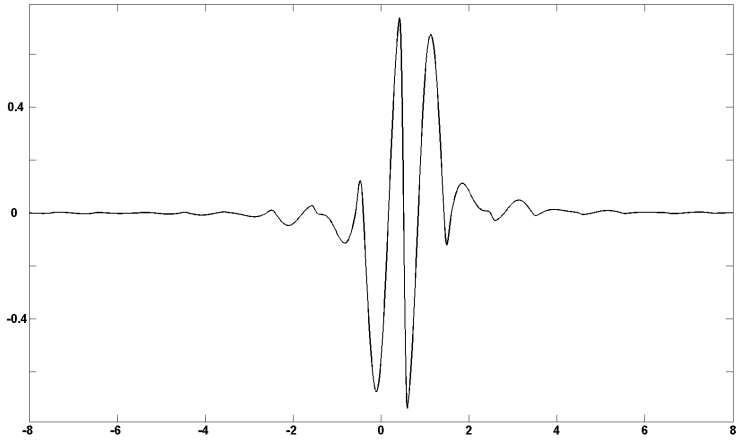
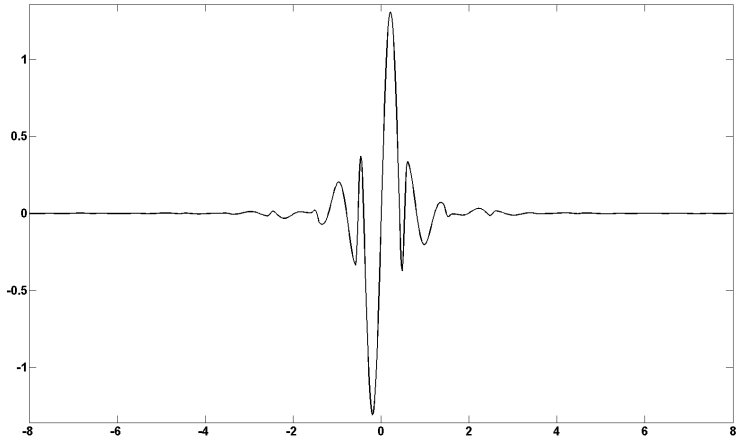


Figure 3.4: Eigenfunction of $\widehat{R}_4$



Chapter 4

Spread functional

4.1 Localization criterion

In the previous chapter we have observed the proofs of existence of exponentially decaying Wannier functions for periodic crystal structures. However, we have no general recipe to obtain them, so, we have to talk about some "optimally" or "maximally" localized functions. But how can one estimate, how "good" they are?

To discuss localization, we need to define a measure of the total spread of the Wannier function basis such that its minimum could serve as a criterion of localized functions. For a given basis $\{w_{n,0}\}_n$ define:

$$\Omega(w) = \sum_n \left[\left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right]. \tag{4.1.1}$$

This *spread functional* was suggested by Nicola Marzari and David Vanderbilt in [MV97]. Following them, we shorten the notations:

$$\left\langle r^2 \right\rangle_n = \left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon}, \tag{4.1.2}$$

$$\bar{\mathrm{r}}_n = \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \quad (\text{a } d\text{-dimensional vector}), \tag{4.1.3}$$

so that

$$\Omega = \sum_n \left[\left\langle r^2 \right\rangle_n - |\bar{\mathrm{r}}_n|^2 \right].$$

A functional of this type comes from chemistry literature as a criteria of "localized molecule orbitals" — since the whole theory of the Wannier functions actually origins from the modeling of electron orbitals.

To realize why such a functional was chosen as a criterion of localization of a Wannier function set, let us turn to some analogies from probability theory.

The term

$$\langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon}$$

has a meaning of *mathematical expectation* of r for given probability density $f(\mathrm{r}) = \varepsilon(\mathrm{r}) w_{n,0}^2(\mathrm{r})$:

$$\mathcal{E}(\mathrm{r}) = \int_{\mathbb{R}^d} \mathrm{r} f(\mathrm{r}) \, \mathrm{d}\, \mathrm{r} = \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} .$$

This term shows where the function $w_{n,0}^2$ (and herewith $w_{n,0}$) "is most probably located". Actually, we are interested in $w_{n,0}$, but $w_{n,0}^2$ (which has the same localization region) is positive and thus can be considered as probability density.

So, $\langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon}$ can be interpreted as the *center of n-th Wannier function.*

The expression

$$\left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2$$

is a *dispersion* of r with the same density f:

$$\begin{aligned} \mathcal{D}(\mathrm{r}) &= \mathcal{E}(|\mathrm{r}|^2) - [\mathcal{E}(\mathrm{r})]^2 \\ &= \int_{\mathbb{R}^d} |\mathrm{r}|^2 f(\mathrm{r}) \, \mathrm{d}\, \mathrm{r} - \left| \int_{\mathbb{R}^d} \mathrm{r} f(\mathrm{r}) \, \mathrm{d}\, \mathrm{r} \right|^2 \\ &= \left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 . \end{aligned}$$

This shows the deviation of r from the expected center $\mathcal{E}$, namely, how wide is the interval where the $w_{n,0}^2$ "most probably lie".

So, the functional Ω can be understood as a sum of the *second moments*, or *variances* of the band functions and therefore can serve as a measure of their deviations from the center — namely, the measure of *localization.*

4.2 Lattice shift

Let us examine the behaviour of the components of the spread by a lattice shift $\mathrm{r} \mapsto \mathrm{r}+\mathrm{R}$:

$$\begin{aligned}
\bar{\mathrm{r}}_n = \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} &\mapsto \langle (\mathrm{r}-\mathrm{R}) w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \\
&= \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} + \mathrm{R} \\
&= \bar{\mathrm{r}}_n + \mathrm{R}; \\
\left\langle r^2 \right\rangle_n = \left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} &\mapsto \left\langle (\mathrm{r}+\mathrm{R})^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \left\langle r^2 w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} + 2 \langle \mathrm{r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \cdot \mathrm{R} + R^2 \\
&= \left\langle r^2 \right\rangle_n + 2\bar{\mathrm{r}}_n \cdot \mathrm{R} + |\mathrm{R}|^2.
\end{aligned}$$

The shift of the lattice (or, in other words, of the zero point) is equivalent to "movement" of the Wigner-Seitz cell — namely, to the choice, which Wannier functions belong to the fundamental domain. Notice that Ω itself keeps unchanged:

$$\begin{aligned}
\Omega = \sum_n \left(\left\langle r^2 \right\rangle_n - |\bar{\mathrm{r}}_n|^2 \right) &\mapsto \sum_n \left(\left\langle r^2 \right\rangle_n + 2\bar{\mathrm{r}}_n \cdot \mathrm{R} + R^2 - |\bar{\mathrm{r}}_n + \mathrm{R}|^2 \right) \\
&= \sum_n \left(\left\langle r^2 \right\rangle_n + 2\bar{\mathrm{r}}_n \cdot \mathrm{R} + R^2 - |\bar{\mathrm{r}}_n|^2 - 2\bar{\mathrm{r}}_n \cdot \mathrm{R} - R^2 \right) \\
&= \sum_n \left(\left\langle r^2 \right\rangle_n - |\bar{\mathrm{r}}_n|^2 \right).
\end{aligned}$$

4.3 Decomposition of the spread functional

For deeper understanding of the mathematical nature of the spread functional, it is useful to decompose it into the following parts:

$$\Omega = \Omega_{\mathrm{D}} + \Omega_{\mathrm{OD}} + \Omega_{\mathrm{I}}$$

such that

$$\Omega_{\mathrm{D}} = \sum_{n}\sum_{\mathrm{R}\neq 0}\left|\langle \mathrm{r}w_{n,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2, \tag{4.3.1}$$

$$\Omega_{\mathrm{OD}} = \sum_{n,m\neq n}\sum_{\mathrm{R}\in\Gamma}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2, \tag{4.3.2}$$

$$\Omega_{\mathrm{I}} = \sum_{n}\left[\left\langle r^2\right\rangle_n - \sum_{m,\mathrm{R}}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2\right]. \tag{4.3.3}$$

The terms Ω_{D} and Ω_{OD} are called *band-diagonal* and *band-off-diagonal* correspondingly. It is evident that their sum

$$\Omega_{\mathrm{D,OD}} = \sum_{n}\sum_{\mathrm{R}\neq 0, m\neq n}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2 \tag{4.3.4}$$

is positive.

For the third term Ω_{I} we formulate the following result.

Theorem 4.1 (Invariant part of the spread functional)**.** *The functional Ω_{I} is positive and gauge invariant, namely, invariant with respect to a unitary transform of the Bloch waves* (3.2.2)*:*

$$\psi_{n,\mathrm{k}}(\mathrm{r}) \mapsto \sum_{m=1}^{N_\beta} \mathrm{U}^{\mathrm{k}}_{mn}\psi_{m,\mathrm{k}}(\mathrm{r}).$$

Proof. Consider a composite energy band β in a crystal. Let P be a projection operator (3.3.17) as in Section 3.3.2 onto a subspace of Wannier functions corresponding to this band:

$$P = \sum_{n=n_1}^{n_{N_\beta}}\sum_{\mathrm{R}\in\Gamma} w_{n,\mathrm{R}}\otimes w_{n,\mathrm{R}}.$$

Let $Q = \mathrm{Id} - P$ be a projection onto all other bands. Then Ω_{I} can be rewritten as

$$\begin{aligned}\Omega_{\mathrm{I}} &= \sum_{n}\sum_{\alpha=1}^{d}\langle r_\alpha Q r_\alpha w_{n,0}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\\ &= \sum_{\alpha=1}^{d}\mathrm{tr}\left[P r_\alpha Q r_\alpha\right],\end{aligned}$$

where d is the dimensionality of the crystal [MV97]. We immediately see that $\Omega_{\rm I}$ is positive.

Having shown in Chapter 2 that the operator P is invariant under the considered transform of the Bloch waves (3.3.15), we conclude that Q is also invariant. The required property follows from the representation of $\Omega_{\rm I}$. Theorem 4.1 is proven.

□

Remark 4.1. *Since all three components $\Omega_{\rm I}$, $\Omega_{\rm D}$ and $\Omega_{\rm OD}$ of the spread functional are positive, Ω is also always positive. Its value is bounded from below by the constant $\Omega_{\rm I}$.*

The following combination can be useful:

$$\Omega_{\rm I,OD} \;=\; \Omega_{\rm I} + \Omega_{\rm OD} = \sum_n \left[\left\langle r^2 \right\rangle_n - \sum_{\rm R\in\Gamma} \left| \langle {\rm r} w_{n,\rm R}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right]. \tag{4.3.5}$$

Clearly, it is also positive.

4.4 Minimum of the spread in 1D

In a 1D crystal, let us choose a set of Wannier functions $w^{\rm loc}_{n,\rm R}$ to be eigenfunctions of projected position operator (as in Section 2.3). For them it holds

$$\left\langle \widehat{R}_m w^{\rm loc}_{m,\rm R}, w^{\rm loc}_{n,\rm R'} \right\rangle_{\mathbb{R}^d,\varepsilon} = \delta_{m,n}\delta_{\rm R,R'}{\rm R}.$$

Therefore,

$$\begin{aligned} \Omega_{\rm D,OD} \;&=\; \sum_n \sum_{\rm R\neq 0, m\neq n} \left| \left\langle {\rm r} w^{\rm loc}_{m,\rm R}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \\ &=\; \sum_n \sum_{\rm R\neq 0, m\neq n} \left| \left\langle \widehat{R}_m w^{\rm loc}_{m,\rm R}, w^{\rm loc}_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \\ &=\; 0. \end{aligned}$$

In other words, the non-invariant part of Ω vanishes on this set. From here we make a remarkable conclusion:

In a 1D case, the eigenfunctions of the projected position operator are maximally localized Wannier functions in sense of their spread.

From Theorem 3.2 we know their properties described by Kohn: real, (anti-)symmetric about 0 or 1/2 and exponentially decaying (see Figures 3.1-3.4).

4.5 Minimization methods

Unfortunately, we have no exact solutions for crystals of higher dimensions. Our aim is to employ and implement different minimization methods to obtain localized Wannier functions numerically.

An efficient algorithm based on a conjugate gradient method was presented by N. Marzari and D. Vanderbilt in [MV97]. They suggested to switch to the level of Bloch waves since the computational domain WSC $\times$ BZ $= [0,1]^d \times [-\pi,\pi]^d$ is bounded in both variables r and k. The spread functional Ω is reformulated in terms of functions $u_{n,\mathrm{k}}$ and their unitary transforms U^{k}. The authors could find an explicit formula for the gradient G_{U} of Ω from

$$\langle Y, G_{\mathrm{U}} \rangle_M = \mathrm{d}\,\Omega(\mathrm{U})[Y] = \lim_{t\to 0} \frac{\Omega(\mathrm{U}\mathrm{e}^{tY}) - \Omega(\mathrm{U})}{t}$$

and applied conjugate gradient minimization to find

$$\min_{\alpha} \Omega(\mathrm{U}\mathrm{e}^{-\alpha G_{\mathrm{U}}})$$

with positive (small) α. This algorithm is known to be one of the best so far, and we will consider it in detail in Chapter 5.

Another, intuitively simple approach was suggested by F. Gygi, J.-L. Fatterbert and E. Schwegler in their paper [GFS03]. This one deals with unitary transformation of the Wannier functions themselves without referring to the Bloch waves. Instead of Ω_{I} introduced

in (4.3.3), another part of the spread is shown to be invariant under this transform. The minimization of the non-invariant part is reduced to a problem of simultaneous diagonalization of several matrices. This method will be discussed in Chapter 6.

Finally, in Chapter 7 we present a new method based on sum-unitary transform of the Wannier functions which has been developed in a framework of this thesis.

Chapter 5

Marzari-Vanderbilt method

The method we plan to review in this chapter was first presented by N. Marzari and D. Vanderbilt in [MV97] and further developed in [SMV01]. Being relatively new, it became, however, very popular very soon and is considered today as the most efficient way to localize Wannier functions.

5.1 Ω as a functional over U

When one talks about numerical integration, the computational domain is naturally considered to be bounded. However, in the definition of the spread functional we have two integrals over an unbounded domain $\mathbb{R}^d$: $\langle r^2 \rangle_n$ and $\bar{\mathrm{r}}_n$.
Let us turn to the Bloch waves level where we have the following theorem.

Theorem 5.1 (Representation of $\bar{\mathrm{r}}_n$ and $\langle r^2 \rangle_n$, [Blo62])**.** *For the periodic parts $u_{n,\mathrm{k}}$ of the Bloch waves it holds:*

$$\begin{aligned} \langle \mathrm{r} w_{n,\mathrm{R}}, w_{m,0} \rangle_{\mathbb{R}^d,\varepsilon} &= \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \langle u_{n,\mathrm{k}}, \nabla_{\mathrm{k}} u_{m,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \, \mathrm{d}\,\mathrm{k}, \\ \langle r^2 w_{n,\mathrm{R}}, w_{m,0} \rangle_{\mathbb{R}^d,\varepsilon} &= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \langle \nabla_{\mathrm{k}} u_{n,\mathrm{k}}, \nabla_{\mathrm{k}} u_{m,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \, \mathrm{d}\,\mathrm{k}. \end{aligned}$$

The proof can be found in [Blo62] or [Klo04].

From Theorem 5.1 follows that the spread functional can be rewritten as

$$\Omega = \sum_n \left[\frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \langle \nabla_{\mathrm{k}} u_{n,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \, \mathrm{d}\,\mathrm{k} - \left| \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \langle u_{n,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \, \mathrm{d}\,\mathrm{k} \right|^2 \right].$$

In Chapter 2 we established that Bloch waves as eigenfunctions of the operator $L_{\mathbf{TM}}$ are unique only up to a unitary transform of a band:

$$\psi_{n,\mathrm{k}} \mapsto \sum_{m=n_1}^{n_{N_\beta}} \mathrm{U}^{\mathrm{k}}_{m,n} \psi_{m,\mathrm{k}}, \qquad u_{n,\mathrm{k}} \mapsto \sum_{m=n_1}^{n_{N_\beta}} \mathrm{U}^{\mathrm{k}}_{m,n} u_{m,\mathrm{k}}.$$

With this freedom in mind, we fix some initial set $u^0_{n,\mathrm{k}}$ and rewrite the spread as depending only on U:

$$\begin{aligned} \Omega(\mathrm{U}) = \sum_n \Bigg[& \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_{m,m'} \left\langle \nabla_{\mathrm{k}} \left(\mathrm{U}^{\mathrm{k}}_{m,n} u^0_{m,\mathrm{k}} \right), \nabla_{\mathrm{k}} \left(\mathrm{U}^{\mathrm{k}}_{m',n} u^0_{m',\mathrm{k}} \right) \right\rangle_{\mathrm{WSC},\varepsilon} \mathrm{d}\,\mathrm{k} - \\ & - \left| \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_{m,m'} \left\langle \mathrm{U}^{\mathrm{k}}_{m,n} u^0_{m,\mathrm{k}}, \nabla_{\mathrm{k}} \left(\mathrm{U}^{\mathrm{k}}_{m',n} u^0_{m',\mathrm{k}} \right) \right\rangle_{\mathrm{WSC},\varepsilon} \mathrm{d}\,\mathrm{k} \right|^2 \Bigg]. \end{aligned}$$

5.2 Discretized Ω in the Brillouin zone

As far as we plan to compute Ω numerically, we have to discretize it.

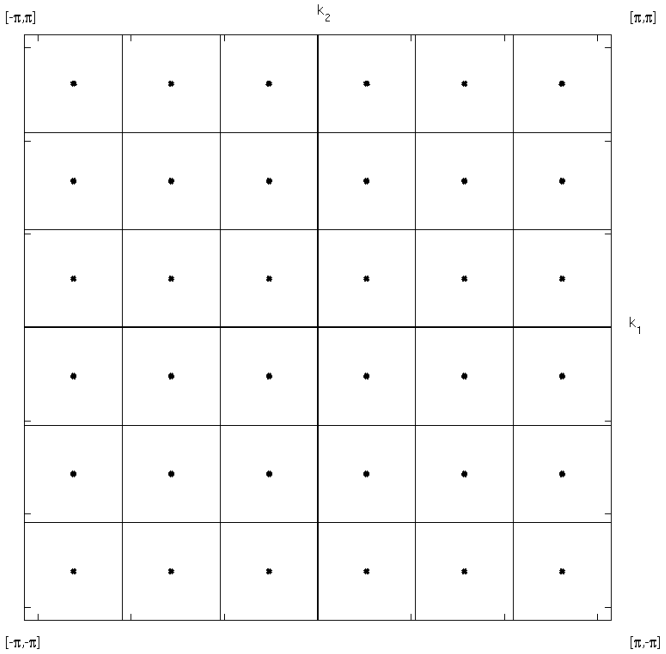
Following [MP76], we introduce a *Monkhorst-Pack mesh* MP in BZ which can be described as follows. Let N_{k} be a number of nodes in every axis direction, and let $h_{\mathrm{k}} = 2\pi/N_{\mathrm{k}}$ be a meshsize. The Monkhorst-Pack mesh has nodes in points with coordinates of the form $(-2\pi + h_{\mathrm{k}}/2) + jh_{\mathrm{k}}$ with integer j. Thus, the nodes are shifted from the boundary by half-meshsize (see Figure 5.1). The mesh consists of N^d_{k} nodes.

What advantages does this mesh give? First, it simplifies the periodic boundary condition: recall, $u_{n,-\pi} = u_{n,\pi} \mathrm{e}^{2\mathrm{i}\pi\cdot\mathrm{r}}$, therefore, the mesh contains only "unique" points. Second, it is symmetric about the axes which gives some numerical advantages that we will show later.

Let

$$\frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{d}\,\mathrm{k} \approx \frac{1}{(N_{\mathrm{k}})^d} \sum_{\mathrm{k}\in\mathrm{MP}}.$$

Figure 5.1: Monkhorst-Pack mesh in the Brillouin zone



Let S_{k} be a "stencil" of closest neighbouring nodes for a node k:

$$S_{\mathrm{k}} = \{-h_{\mathrm{k}}, h_{\mathrm{k}}\} \quad \text{in 1D}, \tag{5.2.1}$$

$$S_{\mathrm{k}} = \left\{ \begin{pmatrix} -h_{\mathrm{k}} \\ 0 \end{pmatrix}, \begin{pmatrix} h_{\mathrm{k}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -h_{\mathrm{k}} \end{pmatrix}, \begin{pmatrix} 0 \\ h_{\mathrm{k}} \end{pmatrix} \right\} \quad \text{in 2D}. \tag{5.2.2}$$

The stencil is the same for all nodes and independent on k. After [MV97], we approximate by finite differences

$$\nabla_{\mathrm{k}} f(\mathrm{k}) \approx \sum_{\mathrm{s}\in S_{\mathrm{k}}} a_s \mathrm{s} \left(f(\mathrm{k}+\mathrm{s}) - f(\mathrm{k}) \right), \tag{5.2.3}$$

$$|\nabla_{\mathrm{k}} f(\mathrm{k})|^2 \approx \sum_{\mathrm{s}\in S_{\mathrm{k}}} a_s \left(f(\mathrm{k}+\mathrm{s}) - f(\mathrm{k}) \right)^2 \tag{5.2.4}$$

with some weights a_s which satisfy the consistency condition:

$$\sum_{\mathrm{s}\in S_{\mathrm{k}}} a_s \mathrm{s} \otimes \mathrm{s} = \mathrm{Id}.$$

From here we obtain the following

Lemma 5.1. *For any vector* $\mathrm{s}' \in \mathbb{R}^d$,

$$\sum_{\mathrm{s}\in S_\mathrm{k}} a_s |\mathrm{s}\cdot\mathrm{s}'|^2 = |\mathrm{s}'|^2.$$

Proof. This can be shown easily:

$$\sum_{\mathrm{s}\in S_\mathrm{k}} a_s |\mathrm{s}\cdot\mathrm{s}'|^2 = \sum_{\mathrm{s}\in S_\mathrm{k}} a_s (\mathrm{s}\cdot\mathrm{s}')(\mathrm{s}\cdot\mathrm{s}')^* = \mathrm{s}' \left(\sum_{\mathrm{s}\in S_\mathrm{k}} a_s \mathrm{s}\otimes\mathrm{s} \right) (\mathrm{s}')^* = \mathrm{s}'\mathrm{Id}(\mathrm{s}')^* = |\mathrm{s}'|^2.$$

□

The terms of the spread functional can be now approximated as

$$\bar{\mathrm{r}}_n \approx \frac{\mathrm{i}}{N_\mathrm{k}^d} \sum_{\mathrm{k}\in\mathrm{MP}} \sum_{\mathrm{s}\in S_\mathrm{k}} a_s \mathrm{s} \left[\langle u_{n,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle - 1 \right] = \bar{\mathrm{r}}_n^{\triangle}, \tag{5.2.5}$$

$$\left\langle r^2 \right\rangle_n \approx \frac{1}{N_\mathrm{k}^d} \sum_{\mathrm{k}\in\mathrm{MP}} \sum_{\mathrm{s}\in S_\mathrm{k}} a_s \left[2 - 2\,\mathrm{Re}\left(\langle u_{n,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle \right) \right] = \left\langle r^2 \right\rangle_n^{\triangle}. \tag{5.2.6}$$

Here we denote the approximations with a symbol "$\cdot^{\triangle}$. These expressions tend to $\bar{\mathrm{r}}_n$ and $\langle r^2 \rangle_n$ as $N_\mathrm{k} \to \infty$, which is equivalent to $h_\mathrm{k} \to 0$.
Let

$$M_{mn}^{\mathrm{k,s}} = \langle u_{m,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle_{\mathrm{WSC},\varepsilon}. \tag{5.2.7}$$

We note some properties of M:

1. $\left[M^{\mathrm{k+s},-\mathrm{s}} \right]^{\mathrm{H}} = M^{\mathrm{k,s}}$. Indeed,

$$\begin{aligned} \left[M^{\mathrm{k+s},-\mathrm{s}} \right]_{mn}^{\mathrm{H}} &= \left[M_{nm}^{\mathrm{k+s},-\mathrm{s}} \right]^* \\ &= \langle u_{n,\mathrm{k+s}}, u_{m,\mathrm{k+s-s}} \rangle^* \\ &= \langle u_{m,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle = M_{mn}^{\mathrm{k,s}}. \end{aligned}$$

2. $M^{\mathrm{k+K,s}} = M^{\mathrm{k,s}}$ for $\mathrm{K} \in \Gamma^*$. We can easily show:

$$\begin{aligned} M_{mn}^{\mathrm{k+K,s}} &= \langle u_{m,\mathrm{k+K}}, u_{n,\mathrm{k+K+s}} \rangle \\ &= \left\langle \mathrm{e}^{-\mathrm{iK\cdot r}} u_{m,\mathrm{k}}, \mathrm{e}^{-\mathrm{iK\cdot r}} u_{n,\mathrm{k+s}} \right\rangle \\ &= \langle u_{m,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle = M_{mn}^{\mathrm{k,s}}. \end{aligned}$$

The equations for $\bar{\mathrm{r}}_n$ and $\langle r^2 \rangle_n$ can be rewritten as

$$\begin{aligned} \bar{\mathrm{r}}_n^{\triangle} &= \frac{\mathrm{i}}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \mathrm{s} \left[M_{nn}^{\mathrm{k,s}} - 1 \right], \\ \left\langle r^2 \right\rangle_n^{\triangle} &= \frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \left[2 - 2 \operatorname{Re} \left(M_{nn}^{\mathrm{k,s}} \right) \right]. \end{aligned}$$

With these approximations we have, however, a small problem. They do not transform by a lattice shift $\mathrm{r} \mapsto \mathrm{r} + \mathrm{R}$ as their continuous analogues do. From Section 4.2, we require for the approximations:

$$\begin{aligned} \bar{\mathrm{r}}_n^{\triangle} &\mapsto \bar{\mathrm{r}}_n^{\triangle} + \mathrm{R}, \\ \left\langle r^2 \right\rangle_n^{\triangle} &\mapsto \left\langle r^2 \right\rangle_n^{\triangle} + 2\bar{\mathrm{r}}_n^{\triangle} \cdot \mathrm{R} + R^2. \end{aligned}$$

The approximations suggested above do not fulfill these conditions, however, they can be modified a little to become satisfactory. Marzari and Vanderbilt suggested the following update of (5.2.5) and (5.2.6):

$$\bar{\mathrm{r}}_n^{\triangle} = -\frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \mathrm{s} \arg M_{nn}^{\mathrm{k,s}}, \tag{5.2.8}$$

$$\left\langle r^2 \right\rangle_n^{\triangle} = \frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \left[1 - \left| M_{nn}^{\mathrm{k,s}} \right|^2 + \left(\arg M_{nn}^{\mathrm{k,s}} \right)^2 \right]. \tag{5.2.9}$$

A detailed derivation of these formulas can be found in [Klo04, Section 4.5]; we also refer to Appendix (Section 10.4) of this thesis.

These new expressions still approximate $\bar{\mathrm{r}}_n$ and $\langle r^2 \rangle_n$ as $N_{\mathrm{k}} \to \infty$, but also satisfy the lattice shift conditions. Together they give an approximation of the spread functional:

$$\Omega \approx \Omega^{\triangle} = \sum_n \left(\frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \left[1 - \left| M_{nn}^{\mathrm{k,s}} \right|^2 + \left(\arg M_{nn}^{\mathrm{k,s}} \right)^2 \right] - \left| \frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} a_s \mathrm{s} \arg M_{nn}^{\mathrm{k,s}} \right|^2 \right).$$

Now Ω formally depends on the variable M, but the information about unitary transforms of Bloch waves is implicitly contained:

$$\begin{aligned} M^{\mathrm{k,s}}_{mn} &= \langle u_{m,\mathrm{k}}, u_{n,\mathrm{k+s}} \rangle_{\mathrm{WSC},\varepsilon} \\ &= \left\langle \sum_{m'} \mathrm{U}^{\mathrm{k}}_{m'm} u^0_{m',\mathrm{k}}, \sum_{n'} \mathrm{U}^{\mathrm{k+s}}_{n'n} u^0_{n',\mathrm{k+s}} \right\rangle_{\mathrm{WSC},\varepsilon} \\ &= \sum_{m'} \sum_{n'} \left(\mathrm{U}^{\mathrm{k}}_{mm'}\right)^* \left\langle u^0_{m',\mathrm{k}}, u^0_{n',\mathrm{k+s}} \right\rangle_{\mathrm{WSC},\varepsilon} \mathrm{U}^{\mathrm{k+s}}_{n'n} \\ &= \sum_{m'} \sum_{n'} \left(\mathrm{U}^{\mathrm{k}}_{mm'}\right)^* M^{0;\mathrm{k,s}}_{mn} \mathrm{U}^{\mathrm{k+s}}_{n'n} \\ &= \left[\left(\mathrm{U}^{\mathrm{k}}\right)^{\mathrm{H}} M^{0;\mathrm{k,s}} \mathrm{U}^{\mathrm{k+s}} \right]_{mn}, \end{aligned}$$

where the set u^0 is fixed and

$$M^{0;\mathrm{k,s}}_{mn} = \left\langle u^0_{m,\mathrm{k}}, u^0_{n,\mathrm{k+s}} \right\rangle_{\mathrm{WSC},\varepsilon}.$$

Therefore, M transforms as follows:

$$\begin{aligned} u_{n,\mathrm{k}} &\mapsto \sum_{m} \mathrm{U}^{\mathrm{k}}_{mn} u_{m,\mathrm{k}}, \\ M^{\mathrm{k,s}} &\mapsto \left(\mathrm{U}^{\mathrm{k}}\right)^{\mathrm{H}} M^{\mathrm{k,s}} \mathrm{U}^{\mathrm{k+s}}. \end{aligned}$$

Let us update the formulas for Ω_{I}, Ω_{D} and Ω_{OD}. First, we need to simplify

$$\begin{aligned} \sum_{\mathrm{R}\in\Gamma} \left| \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 &= \sum_{\mathrm{R}\in\Gamma} \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle^*_{\mathbb{R}^d,\varepsilon} \\ &= \sum_{\mathrm{R}\in\Gamma} \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} \langle u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \,\mathrm{d}\,\mathrm{k} \cdot \\ &\quad \cdot \frac{(-\mathrm{i})}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{ik}'\cdot\mathrm{R}} \left\langle u_{m,\mathrm{k}'}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}'} \right\rangle^*_{\mathrm{WSC},\varepsilon} \mathrm{d}\,\mathrm{k}' \\ &= \frac{1}{V^2_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \int_{\mathrm{BZ}} \langle u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \left\langle u_{m,\mathrm{k}'}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}'} \right\rangle^*_{\mathrm{WSC},\varepsilon} \\ &\quad \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{i}(\mathrm{k}-\mathrm{k}')\cdot\mathrm{R}} \,\mathrm{d}\,\mathrm{k}\,\mathrm{d}\,\mathrm{k}'. \end{aligned}$$

As we stated in (2.4.1),

$$\frac{1}{V_{\mathrm{BZ}}} \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} = \delta(\mathrm{k}).$$

With this in place, we continue:

$$\sum_{\mathrm{R}\in\Gamma}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2 \;=\; \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}}\left|\langle u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}}\rangle_{\mathrm{WSC},\varepsilon}\right|^2 \mathrm{d}\,\mathrm{k}.$$

Now we discretize it using the same scheme as before:

$$\sum_{\mathrm{R}\in\Gamma}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2 \;\approx\; \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}}\sum_{\mathrm{s}\in S_{\mathrm{k}}} a_s \left|M_{mn}^{\mathrm{k,s}} - \delta_{m,n}\right|^2.$$

Assuming the approximation (5.2.8) for $\bar{\mathrm{r}}_n$, we finally obtain:

$$\sum_{\mathrm{R}\in\Gamma}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2 \approx \begin{cases} \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s \left(\arg M_{nn}^{\mathrm{k,s}}\right)^2, & m = n, \\ \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s \left|M_{mn}^{\mathrm{k,s}}\right|^2, & m \neq n. \end{cases}$$

Now we can write down discretized formulas for the decomposition parts of the spread functional.

$$\begin{aligned} \Omega_{\mathrm{I}} &= \sum_n \left[\left\langle r^2\right\rangle_n - \sum_{m,\mathrm{R}}\left|\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}\right|^2\right] \\ &\approx \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s\left[1 - |M_{nn}^{\mathrm{k,s}}|^2 + (\arg M_{nn}^{\mathrm{k,s}})^2 - (\arg M_{nn}^{\mathrm{k,s}})^2 - \sum_{m\neq n}|M_{mn}^{\mathrm{k,s}}|^2\right] \\ &= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s\left[1 - \sum_m |M_{mn}^{\mathrm{k,s}}|^2\right] = \Omega_{\mathrm{I}}^{\triangle}. \end{aligned}$$

Remark 5.1. *Note that the approximation $\Omega_{\mathrm{I}}^{\triangle}$ is also gauge invariant as its continuous analogue Ω_{I} (Theorem (4.1)) since it can be written in form*

$$\Omega_{\mathrm{I}}^{\triangle} = \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s \,\mathrm{tr}\left(P^{\mathrm{k}} Q^{\mathrm{k+s}}\right)$$

with P^{k} defined as (3.3.14) and $Q^{\mathrm{k}} = \mathrm{Id} - P^{\mathrm{k}}$.

For $\Omega_{\rm D}$ we have:

$$\begin{aligned}
\Omega_{\rm D} &= \sum_n \sum_{\rm R\neq 0} \left| \langle {\rm r} w_{n,{\rm R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \\
&= \sum_n \left[\sum_{\rm R\in\Gamma} \left| \langle {\rm r} w_{n,{\rm R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 - \left| \langle {\rm r} w_{n,0}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right] \\
&\approx \sum_n \left[\frac{1}{N_{\rm k}^d} \sum_{\rm k,s} a_s (\arg M_{nn}^{\rm k,s})^2 - |\bar{\rm r}_n|^2 \right] \\
&= \frac{1}{N_{\rm k}^d} \sum_n \sum_{\rm k,s} a_s \left[(\arg M_{nn}^{\rm k,s})^2 + \arg M_{nn}^{\rm k,s} \cdot \bar{\rm r}_n \right].
\end{aligned}$$

To continue, we will use the following trick:

$$\begin{aligned}
\sum_{\rm k} \left(\sum_{\rm s} a_s |{\rm s}\cdot\bar{\rm r}_n|^2 \right) &= \sum_{\rm k} |\bar{\rm r}_n|^2 = N_{\rm k}^d |\bar{\rm r}_n|^2, \\
\sum_{\rm k,s} a_s {\rm s}\cdot\bar{\rm r}_n \arg M_{nn}^{\rm k,s} &= -N_{\rm k}^d |\bar{\rm r}_n|^2, \\
\Longrightarrow \quad 0 &= \sum_{\rm k,s} a_s \left(|{\rm s}\cdot\bar{\rm r}_n|^2 + {\rm s}\cdot\bar{\rm r}_n \arg M_{nn}^{\rm k,s} \right).
\end{aligned}$$

Adding this to the last expression, we obtain:

$$\begin{aligned}
\Omega_{\rm D} &\approx \frac{1}{N_{\rm k}^d} \sum_n \sum_{\rm k,s} a_s \left[(\arg M_{nn}^{\rm k,s})^2 + 2 \arg M_{nn}^{\rm k,s} {\rm s}\cdot\bar{\rm r}_n + |{\rm s}\cdot\bar{\rm r}_n|^2 \right] \\
&= \frac{1}{N_{\rm k}^d} \sum_n \sum_{\rm k,s} a_s \left[\arg M_{nn}^{\rm k,s} + {\rm s}\cdot\bar{\rm r}_n \right]^2 = \Omega_{\rm D}^{\triangle}.
\end{aligned}$$

Finally, for the last term we write:

$$\begin{aligned}
\Omega_{\rm OD} &= \sum_{n,m\neq n} \sum_{\rm R\in\Gamma} \left| \langle {\rm r} w_{m,{\rm R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \\
&\approx \frac{1}{N_{\rm k}^d} \sum_{n,m\neq n} \sum_{\rm k,s} a_s |M_{nn}^{\rm k,s}|^2 = \Omega_{\rm OD}^{\triangle}.
\end{aligned}$$

For a construction $\Omega_{\rm I,OD}$ we write:

$$\begin{aligned}
\Omega_{\rm I,OD} &= \Omega_{\rm I} + \Omega_{\rm OD} \approx \frac{1}{N_{\rm k}^d} \sum_n \sum_{\rm k,s} a_s \left[1 - \sum_m |M_{mn}^{\rm k,s}|^2 + \sum_{m\neq n} |M_{mn}^{\rm k,s}|^2 \right] \\
&= \frac{1}{N_{\rm k}^d} \sum_n \sum_{\rm k,s} a_s \left[1 - |M_{nn}^{\rm k,s}|^2 \right] = \Omega_{\rm I,OD}^{\triangle}.
\end{aligned}$$

Alltogether,

$$\Omega \approx \sum_n \sum_{\mathrm{k,s}} a_s \left[1 + \left(\arg M_{nn}^{\mathrm{k,s}} + \mathrm{s} \cdot \bar{\mathrm{r}}_n\right)^2 - |M_{nn}^{\mathrm{k,s}}|^2\right] = \Omega^{\triangle}. \tag{5.2.10}$$

5.3 Discrete gradient of Ω

It would be too costly to search for the minimum of the spread without the information about the descent direction. Several minimization algorithms are considered in [BMR$^+$00] and the most effective ones are gradient methods. For this reason, in this section we want to formulate explicitly a gradient of the spread functional for later use.
Consider Taylor's expansion of Ω:

$$\begin{aligned} \Omega(\mathrm{U}_0 + t\mathrm{U}) &= \Omega(\mathrm{U}_0) + t \sum_{\mathrm{k}} \frac{\partial \Omega(\mathrm{U}_0)}{\partial \mathrm{U}^{\mathrm{k}}} \mathrm{U}^{\mathrm{k}} + O(t^2) \\ &= \Omega(\mathrm{U}_0) + t \left\langle \mathrm{U}, \nabla\Omega(\mathrm{U}_0) \right\rangle_{\mathcal{F}} + O(t^2). \end{aligned}$$

From this we define a gradient of Ω to be

$$\left\langle \mathrm{U}, G(\mathrm{U}_0) \right\rangle_{\mathcal{F}} = \left\langle \mathrm{U}, \nabla\Omega(\mathrm{U}_0) \right\rangle_{\mathcal{F}} = \mathrm{d}\,\Omega(\mathrm{U}_0)[\mathrm{U}] = \lim_{t \to 0} \frac{\Omega(\mathrm{U}_0 + t\mathrm{U}) - \Omega(\mathrm{U}_0)}{t}.$$

Here $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ is some "inner product" between objects like U. An explicit form of this product is so far not important. Let

$$\triangle f(\mathrm{U}) = f(\mathrm{U} + \triangle \mathrm{U}) - f(\mathrm{U}).$$

Compute a discretized expression for

$$\triangle\Omega = \Omega(\mathrm{U}_0 + t\mathrm{U}) - \Omega(\mathrm{U}_0).$$

At this point we must mention that $t\mathrm{U}$ is not allowed to be arbitrary. The matrices $\mathrm{U}_0^{\mathrm{k}} + t\mathrm{U}^{\mathrm{k}}$ must be unitary, as the initial $\mathrm{U}_0^{\mathrm{k}}$ are. From the properties of matrices we know that one unitary matrix can be obtained from another by multiplication by an exponential of a skew-hermitian matrix: namely, if $\mathrm{U}_0^{\mathrm{k}}$ is unitary $\left(\mathrm{U}_0^{\mathrm{k}}\right)^{\mathrm{H}} \mathrm{U}_0^{\mathrm{k}} = \mathrm{Id}$, and Y^{k} is skew-hermitian $\left(Y^{\mathrm{k}}\right)^{\mathrm{H}} = -Y^{\mathrm{k}}$, then $\mathrm{U}_0^{\mathrm{k}} \exp\left(Y^{\mathrm{k}}\right)$ is unitary.

As far as we search for a first-order approximation of $\triangle\Omega$, it will be enough to take only the linear part of matrix exponent: $\exp\left(Y^{\mathrm{k}}\right) \approx \mathrm{Id} + tY^{\mathrm{k}}$.
Alltogether, consider a perturbation of U_0 as $\mathrm{U}_0(\mathrm{Id} + tY)$ with a skew-hermitian Y.

Since we have rewritten Ω in terms of M (5.2.10), let us start with $\triangle M$.

$$\begin{aligned}
M^{\mathrm{k,s}} = \left(\mathrm{U}_0^{\mathrm{k}}\right)^{\mathrm{H}} M^{0;\mathrm{k,s}}\mathrm{U}_0^{\mathrm{k+s}} \quad &\mapsto \quad \left(\mathrm{U}_0^{\mathrm{k}}(\mathrm{Id} + tY^{\mathrm{k}})\right)^{\mathrm{H}} M^{0;\mathrm{k,s}}\mathrm{U}_0^{\mathrm{k+s}}(\mathrm{Id} + tY^{\mathrm{k+s}}) \\
&= \left(\mathrm{Id} + tY^{\mathrm{k}}\right)^{\mathrm{H}} M^{\mathrm{k,s}}(\mathrm{Id} + tY^{\mathrm{k+s}}) + O(t^2) \\
&\approx M^{\mathrm{k,s}} - tY^{\mathrm{k}}M^{\mathrm{k,s}} + tM^{\mathrm{k,s}}Y^{\mathrm{k+s}}; \\
\triangle M^{\mathrm{k,s}} &= -tY^{\mathrm{k}}M^{\mathrm{k,s}} + tM^{\mathrm{k,s}}Y^{\mathrm{k+s}} \\
&= -tY^{\mathrm{k}}M^{\mathrm{k,s}} + t\left(Y^{\mathrm{k+s}}M^{\mathrm{k+s,-s}}\right)^{\mathrm{H}}.
\end{aligned}$$

For the hemitian conjugate,

$$\left(M^{\mathrm{k,s}}\right)^{\mathrm{H}} = \left(\mathrm{U}_0^{\mathrm{k+s}}\right)^{\mathrm{H}} M^{0;\mathrm{k,s}}\mathrm{U}_0^{\mathrm{k}} \quad \mapsto \quad \left(M^{\mathrm{k,s}}\right)^{\mathrm{H}} - tY^{\mathrm{k+s}}\left(M^{\mathrm{k,s}}\right)^{\mathrm{H}} + t\left(M^{\mathrm{k,s}}\right)^{\mathrm{H}}Y^{\mathrm{k}} + O(t^2);$$

from here

$$\begin{aligned}
\triangle\left(M^{\mathrm{k,s}}\right)^{\mathrm{H}} &= -tY^{\mathrm{k+s}}M^{\mathrm{k+s,-s}} + t\left(Y^{\mathrm{k}}M^{\mathrm{k,s}}\right)^{\mathrm{H}} \\
&= \left(\triangle M^{\mathrm{k,s}}\right)^{\mathrm{H}}.
\end{aligned}$$

Let us introduce some new notations:

$$\begin{aligned}
R_{mn}^{\mathrm{k,s}} &= M_{mn}^{\mathrm{k,s}}\left(M_{nn}^{\mathrm{k,s}}\right)^{*}; \\
\widetilde{R}_{mn}^{\mathrm{k,s}} &= \frac{M_{mn}^{\mathrm{k,s}}}{M_{nn}^{\mathrm{k,s}}}; \\
q_n^{\mathrm{k,s}} &= \arg M_{nn}^{\mathrm{k,s}} + \mathrm{s}\cdot\bar{\mathrm{r}}_n; \\
T_{mn}^{\mathrm{k,s}} &= \widetilde{R}_{mn}^{\mathrm{k,s}}q_n^{\mathrm{k,s}}.
\end{aligned}$$

Recall the decomposition of the spread functional: $\Omega = \Omega_{\mathrm{I,OD}} + \Omega_{\mathrm{D}}$. Perturbations give:

$$\begin{aligned}
\triangle\Omega_{\mathrm{I,OD}} &= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \triangle\left(1-|M_{nn}^{\mathrm{k,s}}|^2\right)\\
&= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \left[-\triangle\left(|M_{nn}^{\mathrm{k,s}}|^2\right)\right]\\
&= -\frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \left[\left|M_{nn}^{\mathrm{k,s}}+\triangle M_{nn}^{\mathrm{k,s}}\right|^2 - |M_{nn}^{\mathrm{k,s}}|^2\right]\\
&= -\frac{2}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \,\mathrm{Re}\left(\left[M_{nn}^{\mathrm{k,s}}\right]^*\triangle M_{nn}^{\mathrm{k,s}}\right)+O(t^2)\\
&\approx \frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \,\mathrm{Re}\left(\left[M_{nn}^{\mathrm{k,s}}\right]^*\left(\left[Y^{\mathrm{k}}M^{\mathrm{k,s}}\right]_{nn}+\left[Y^{\mathrm{k+s}}M^{\mathrm{k+s,-s}}\right]_{nn}^*\right)\right)\\
&= \frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \left(\mathrm{Re}\left(\left[Y^{\mathrm{k}}R^{\mathrm{k,s}}\right]_{nn}\right)+\mathrm{Re}\left(\left[Y^{\mathrm{k+s}}R^{\mathrm{k+s,-s}}\right]_{nn}\right)\right)\\
&= \frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \,\mathrm{Re}\left(\mathrm{tr}\left[Y^{\mathrm{k}}R^{\mathrm{k,s}}\right]\right)+\frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \,\mathrm{Re}\left(\mathrm{tr}\left[Y^{\mathrm{k+s}}R^{\mathrm{k+s,-s}}\right]\right)\\
&= \frac{4t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \,\mathrm{Re}\left(\mathrm{tr}\left[Y^{\mathrm{k}}R^{\mathrm{k,s}}\right]\right).
\end{aligned}$$

Accordingly,

$$\begin{aligned}
\triangle\Omega_{\mathrm{D}} &= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \triangle\left(\left[\arg M_{nn}^{\mathrm{k,s}}+\mathrm{s}\cdot\bar{\mathrm{r}}_n\right]^2\right)\\
&= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \triangle\left(\left[q_n^{\mathrm{k,s}}\right]^2\right)\\
&= \frac{1}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s \left(\left[q_n^{\mathrm{k,s}}+\triangle q_n^{\mathrm{k,s}}\right]^2-\left[q_n^{\mathrm{k,s}}\right]^2\right)\\
&= \frac{2}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\triangle q_n^{\mathrm{k,s}}\\
&= \frac{2}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\left[\triangle\left(\arg M_{nn}^{\mathrm{k,s}}\right)+\triangle\left(\mathrm{s}\cdot\bar{\mathrm{r}}_n\right)\right]\\
&= \frac{2}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\triangle\left(\arg M_{nn}^{\mathrm{k,s}}\right)+\frac{2}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\triangle\left(\mathrm{s}\cdot\bar{\mathrm{r}}_n\right).
\end{aligned}$$

Let us examine the terms separately.

$$\begin{aligned}
\triangle\left(\arg M_{nn}^{\mathrm{k,s}}\right) &= \triangle\left(\operatorname{Im}\ln M_{nn}^{\mathrm{k,s}}\right)\\
&= \operatorname{Im}\left(\triangle\ln M_{nn}^{\mathrm{k,s}}\right)\\
&= \operatorname{Im}\left(\ln\left(M_{nn}^{\mathrm{k,s}}+\triangle M_{nn}^{\mathrm{k,s}}\right)-\ln\left(M_{nn}^{\mathrm{k,s}}\right)\right)\\
&= \operatorname{Im}\left(\ln\left(1+\frac{\triangle M_{nn}^{\mathrm{k,s}}}{M_{nn}^{\mathrm{k,s}}}\right)\right)\\
&= \operatorname{Im}\frac{\triangle M_{nn}^{\mathrm{k,s}}}{M_{nn}^{\mathrm{k,s}}}+O(t^3)\\
&\approx \operatorname{Im}\frac{-t\left[Y^{\mathrm{k}}M^{\mathrm{k,s}}\right]_{nn}-t\left[Y^{\mathrm{k+s}}M^{\mathrm{k+s,-s}}\right]_{nn}^{*}}{M_{nn}^{\mathrm{k,s}}}\\
&= -t\left[Y^{\mathrm{k}}\widetilde{R}^{\mathrm{k,s}}\right]_{nn}+t\left[Y^{\mathrm{k+s}}\widetilde{R}^{\mathrm{k+s,-s}}\right]_{nn};\\
\triangle\left(\mathrm{s}\cdot\bar{\mathrm{r}}_n\right) &= \triangle\left(-\frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}',\mathrm{s}'}a_{s'}\mathrm{s}\cdot\mathrm{s}'\arg M_{nn}^{\mathrm{k}',\mathrm{s}'}\right)\\
&= -\frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}',\mathrm{s}'}a_{s'}\mathrm{s}\cdot\mathrm{s}'\triangle\left(\arg M_{nn}^{\mathrm{k}',\mathrm{s}'}\right)\\
&= \frac{t}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}',\mathrm{s}'}a_{s'}\mathrm{s}\cdot\mathrm{s}'\left(\operatorname{Im}\left(\left[Y^{\mathrm{k}'}\widetilde{R}^{\mathrm{k}',\mathrm{s}'}\right]_{nn}\right)-\operatorname{Im}\left(\left[Y^{\mathrm{k}'+\mathrm{s}'}\widetilde{R}^{\mathrm{k}'+\mathrm{s}',-\mathrm{s}'}\right]_{nn}\right)\right).
\end{aligned}$$

Note that

$$\begin{aligned}
q_n^{\mathrm{k+s,-s}} &= \arg\left(M_{nn}^{\mathrm{k,s}}\right)^{*}-\mathrm{s}\cdot\bar{\mathrm{r}}_n\\
&= -\arg M_{nn}^{\mathrm{k,s}}\mathrm{s}\cdot\bar{\mathrm{r}}_n\\
&= -q_n^{\mathrm{k,s}},\\
\sum_{\mathrm{k,s}}\mathrm{s}q_n^{\mathrm{k,s}} &= \sum_{\mathrm{k,s}}a_s\mathrm{s}\left(\arg M_{nn}^{\mathrm{k,s}}+\mathrm{s}\cdot\bar{\mathrm{r}}_n\right)\\
&= \sum_{\mathrm{k,s}}a_s\mathrm{s}\arg M_{nn}^{\mathrm{k,s}}+\sum_{\mathrm{k,s}}a_s\mathrm{s}\left(\mathrm{s}\cdot\bar{\mathrm{r}}_n\right)\\
&= -N_{\mathrm{k}}^d\bar{\mathrm{r}}_n+N_{\mathrm{k}}^d\left(\sum_{\mathrm{s}}a_s\mathrm{s}\otimes\mathrm{s}\right)\cdot\bar{\mathrm{r}}_n\\
&= -N_{\mathrm{k}}^d\bar{\mathrm{r}}_n+N_{\mathrm{k}}^d\bar{\mathrm{r}}_n\\
&= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\triangle\Omega_{\mathrm{D}} &= -\frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\,\mathrm{Im}\left[Y^{\mathrm{k}}\widetilde{R}^{\mathrm{k,s}}\right]_{nn} \\
&\quad -\frac{2t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k+s,-s}}\,\mathrm{Im}\left[Y^{\mathrm{k+s}}\widetilde{R}^{\mathrm{k+s,-s}}\right]_{nn} \\
&\quad +\frac{2t}{N_{\mathrm{k}}^{2d}}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\mathrm{s}\cdot\sum_{\mathrm{k}',\mathrm{s}'} a_{s'}\mathrm{s}'\,\mathrm{Im}\left[Y^{\mathrm{k}}\widetilde{R}^{\mathrm{k,s}}\right]_{nn} \\
&\quad +\frac{2t}{N_{\mathrm{k}}^{2d}}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\mathrm{s}\cdot\sum_{\mathrm{k}',\mathrm{s}'} a_{s'}(-\mathrm{s}')\,\mathrm{Im}\left[Y^{\mathrm{k+s}}\widetilde{R}^{\mathrm{k+s,-s}}\right]_{nn} \\
&= -\frac{4t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s q_n^{\mathrm{k,s}}\,\mathrm{Im}\left[Y^{\mathrm{k}}\widetilde{R}^{\mathrm{k,s}}\right]_{nn} \\
&= -\frac{4t}{N_{\mathrm{k}}^d}\sum_n\sum_{\mathrm{k,s}} a_s\,\mathrm{Im}\left[Y^{\mathrm{k}}T^{\mathrm{k,s}}\right]_{nn} \\
&= -\frac{4t}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s\,\mathrm{Im}\left(\mathrm{tr}\left[Y^{\mathrm{k}}T^{\mathrm{k,s}}\right]\right).
\end{aligned}$$

Finally, we obtain the following formula for the discretized spread functional:

$$\begin{aligned}
\triangle\Omega &= \triangle\Omega_{\mathrm{I,OD}} + \triangle\Omega_{\mathrm{D}} \\
&= \frac{4t}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s\,\mathrm{Re}\left(\mathrm{tr}\left[Y^{\mathrm{k}}R^{\mathrm{k,s}}\right]\right) - \frac{4t}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s\,\mathrm{Im}\left(\mathrm{tr}\left[Y^{\mathrm{k}}T^{\mathrm{k,s}}\right]\right).
\end{aligned}$$

Now it is time to define the inner product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ mentioned earlier. Let it be a Frobenius-type product:

$$\langle A, B\rangle_{\mathcal{F}} = \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}}\mathrm{tr}\left(\left[A^{\mathrm{k}}\right]^{\mathrm{H}} B^{\mathrm{k}}\right).$$

Then we define a gradient of Ω as

$$\langle Y, G\rangle_{\mathcal{F}} = \lim_{t\to 0}\frac{\triangle\Omega}{t}$$

for any skew-hermitian W.

From previous calculations we have then:

$$\begin{aligned}
\langle Y, G\rangle_{\mathcal{F}} &= \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} G^{\mathrm{k}}\right) \\
&\quad (\textit{as far as } Y^{\mathrm{k}} \textit{ is skew-hermitian, } \left[Y^{\mathrm{k}}\right]^{\mathrm{H}} = -Y^{\mathrm{k}}) \\
&= -\frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{tr}\left(Y^{\mathrm{k}} G^{\mathrm{k}}\right) \\
&= \frac{4}{N_{\mathrm{k}}^d}\sum_{\mathrm{k,s}} a_s \left(\mathrm{Re}\left(\mathrm{tr}\left[Y^{\mathrm{k}} R^{\mathrm{k,s}}\right]\right) - \mathrm{Im}\left(\mathrm{tr}\left[Y^{\mathrm{k}} T^{\mathrm{k,s}}\right]\right)\right).
\end{aligned}$$

Let us introduce "asymmetric" and "symmetric" operators $\mathcal{A}[B]$ and $\mathcal{S}[B]$:

$$\begin{aligned}
\mathcal{A}[B] &= -\frac{1}{2}\left(B - B^{\mathrm{H}}\right), \\
\mathcal{S}[B] &= \frac{\mathrm{i}}{2}\left(B + B^{\mathrm{H}}\right).
\end{aligned}$$

Then

$$\begin{aligned}
\langle Y, \mathcal{A}[B]\rangle_{\mathcal{F}} &= -\frac{1}{2}\left(\langle Y, B\rangle_{\mathcal{F}} - \left\langle Y, B^{\mathrm{H}}\right\rangle_{\mathcal{F}}\right) \\
&= -\frac{1}{2N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \left(\mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} B^{\mathrm{k}}\right) - \mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} \left[B^{\mathrm{k}}\right]^{\mathrm{H}}\right)\right) \\
&= -\frac{1}{2N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \left(-\mathrm{tr}\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right) - \mathrm{tr}^*\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right)\right) \\
&= \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{Re}\left(\mathrm{tr}\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right)\right);
\end{aligned}$$

$$\begin{aligned}
\langle Y, \mathcal{S}[B]\rangle_{\mathcal{F}} &= \frac{\mathrm{i}}{2}\left(\langle Y, B\rangle_{\mathcal{F}} + \left\langle Y, B^{\mathrm{H}}\right\rangle_{\mathcal{F}}\right) \\
&= \frac{\mathrm{i}}{2N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \left(\mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} B^{\mathrm{k}}\right) + \mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} \left[B^{\mathrm{k}}\right]^{\mathrm{H}}\right)\right) \\
&= \frac{\mathrm{i}}{2N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \left(-\mathrm{tr}\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right) + \mathrm{tr}^*\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right)\right) \\
&= -\frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{Im}\left(\mathrm{tr}\left(Y^{\mathrm{k}} B^{\mathrm{k}}\right)\right).
\end{aligned}$$

Gathering together all these pieces of information, we conclude:

$$\begin{aligned}
\langle Y, G\rangle_{\mathcal{F}} &= 4\sum_{\mathrm{s}} a_s \left(\langle Y, \mathcal{A}[R^{\cdot,\mathrm{s}}]\rangle_{\mathcal{F}} + \langle Y, \mathcal{S}[T^{\cdot,\mathrm{s}}]\rangle_{\mathcal{F}}\right) \\
&= 4\sum_{\mathrm{s}} a_s \left\langle Y, \mathcal{A}[R^{\cdot,\mathrm{s}}] + \mathcal{S}[T^{\cdot,\mathrm{s}}]\right\rangle_{\mathcal{F}}.
\end{aligned}$$

From here we find the gradient:

$$\begin{aligned} G^{\mathrm{k}} &= 4\sum_{\mathrm{s}} a_s \left(\mathcal{A}[R^{\mathrm{k,s}}] + \mathcal{S}[T^{\mathrm{k,s}}]\right) \\ &= 2\sum_{\mathrm{s}} a_s \left(-R^{\mathrm{k,s}} + \left[R^{\mathrm{k,s}}\right]^{\mathrm{H}} + \mathrm{i}T^{\mathrm{k,s}} + \mathrm{i}\left[T^{\mathrm{k,s}}\right]^{\mathrm{H}}\right). \end{aligned}$$

Obviously, it is also skew-hermitian.

Let us take $Y^{\mathrm{k}} = -\alpha G^{\mathrm{k}}$ with some infinitesimal $\alpha \geq 0$. Then

$$\begin{aligned} \mathrm{d}\,\Omega(\mathrm{U})[Y] &= \langle Y, G\rangle_{\mathcal{F}} = \frac{1}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{tr}\left(\left[Y^{\mathrm{k}}\right]^{\mathrm{H}} G^{\mathrm{k}}\right) = -\frac{\alpha}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \mathrm{tr}\left(\left[G^{\mathrm{k}}\right]^{\mathrm{H}} G^{\mathrm{k}}\right) \\ &= -\frac{\alpha}{N_{\mathrm{k}}^d}\sum_{\mathrm{k}} \|G^{\mathrm{k}}\|_F^2 \leq 0, \end{aligned}$$

where $\|\cdot\|_F$ is Frobenius matrix norm.

As we see, such a choice of W makes Ω decay or at least not grow. It means that if we transform U as

$$\mathrm{U}^{\mathrm{k}} \mapsto \mathrm{U}^{\mathrm{k}} \exp\left(Y^{\mathrm{k}}\right) = \mathrm{U}^{\mathrm{k}} \exp\left(-\alpha G^{\mathrm{k}}\right),$$

we can minimize the spread.

5.4 A choice of the initial set U

In the algorithm described in this chapter, we have skipped an important point — the choice of initial values for U. Meanwhile, it has a big influence on efficiency of the method, namely, on how fast a localized solution can be found and how good it will be.

Ω as a functional over U has erratic behaviour, and therefore conjugate gradient minimization depends crucially on how far from the (global) minimum it starts. With a bad initial guess, the algorithm can be trapped in a local minimum. Wannier functions obtained by this minimum can have unphysical oscillations. Marzari and Vanderbilt notify that the proposed method is not secured of such a possibility, but they also point out, that at a global minimum $\left|q_n^{\mathrm{k,s}}\right| \ll \pi$, while at a false (local) minimum some of its values approach

π.

We also remember that maximally localized Wannier functions are expected to be real, apart from a trivial overall phase, while a false local minimum of Ω often is complex-valued.

So, we conclude that initial matrices U^{k} should be "good". But what does it mean?

First, they must be smooth in k. Indeed, as far as we use finite differences in k-mesh, we expect to obtain a proper approximation of the functions $u_{n,\mathrm{k}}$. Second, they have to satisfy periodicity conditions on the Bloch waves $\psi_{n,\mathrm{k}}$:

$$\begin{aligned} \psi_{n,\mathrm{k}} &= \psi_{n,\mathrm{k+K}}, \\ \sum_m \mathrm{U}^{\mathrm{k}}_{mn}\psi^0_{m,\mathrm{k}} = \psi_{n,\mathrm{k}} &= \psi_{n,\mathrm{k+K}} = \sum_m \mathrm{U}^{\mathrm{k+K}}_{mn}\psi^0_{m,\mathrm{k+K}} = \sum_m \mathrm{U}^{\mathrm{k+K}}_{mn}\psi^0_{m,\mathrm{k}}, \\ \mathrm{U}^{\mathrm{k}} &= \mathrm{U}^{\mathrm{k+K}}. \end{aligned}$$

An intuitively natural way is to take identity matrices $\mathrm{U}^{\mathrm{k}} = \mathrm{Id}$. As we will show later, this choice is rather good, but can be also improved. With this goal in mind, we put the third condition: initial values for U are desirable to look roughly like localized Wannier functions.

A nice example of such localized functions can be Gaussian bell functions centered in the Wigner-Seitz cell:

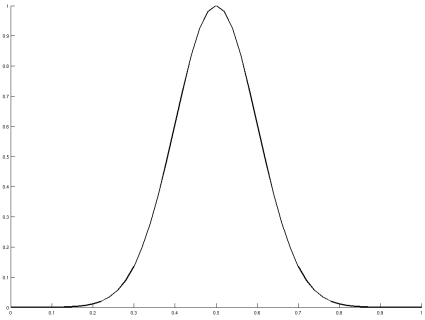
$$g_n(\mathrm{r}) = a_n \exp\left(-\frac{(\mathrm{r}-\mathrm{b}_n)^2}{2c_n^2}\right),$$

where a_n is a peak height of n-th Gaussian, b_n is the position of its center and c_n is a value related to its full width at half maximum (FWHM) as FWHM$= 2\sqrt{2\ln 2}\mathrm{c}_n$ (see Figure 5.2).

Let us consider such functions as an initial guess for Wannier functions, with $a_n = 1$, $\mathrm{b}_n \in \mathrm{WSC} = [0,1]^d$ and closer to its center, and c_n^2 taken such that FWHM is comparable to the width of WSC.

We project the Gaussians onto the span of Bloch waves, orthonormalize the projections, and take the result as starting Bloch waves — or, equivalently, construct such matrices

Figure 5.2: A Gaussian bell function



U^{k} which rotate fixed initial Bloch waves $\psi^0_{n,\mathrm{k}}$ to this starting guess. Algorithmically it can be written down as follows.

Construction of matrices U^{k} based on Gaussian bells

1. Construct a random $d \times N_\beta$-dimensional set b such that $\mathrm{b}_n \in [0.1, 0, 9]^d$ for every n;

2. Construct a random $d \times N_\beta$-dimensional set c with $\mathrm{c}_n \in [0.1, \min(\mathrm{b}_n, 1 - \mathrm{b}_n)]^d$;

3. Define d-dimensional Gaussian bells:

$$g_n(\mathrm{r}) = \exp\left(-\frac{(\mathrm{r} - \mathrm{b}_n)^2}{2c_n^2}\right);$$

4. Project them onto the span of Bloch waves:

$$\phi_{l,\mathrm{k}} = \sum_{m=1}^{N_\beta} \left\langle \psi^0_{m,\mathrm{k}}, g_l \right\rangle \psi^0_{m,\mathrm{k}};$$

5. Orthonormalize these functions to obtain a starting guess for Bloch waves:

$$\varphi_{n,\mathrm{k}} = \sum_{l=1}^{N_\beta} \left[S_{\mathrm{k}}^{-1/2}\right]_{ln} \phi_{l,\mathrm{k}},$$

where $S_{\mathrm{k};ln} = \langle \phi_{l,\mathrm{k}}, \phi_{n,\mathrm{k}} \rangle$ (this must be strictly positive according to [dC64b]);

6. Finally, construct the matrices U^{k} using two equalities:

$$\begin{aligned}
\varphi_{n,\mathrm{k}} &= \sum_{m=1}^{N_\beta} \mathrm{U}^{\mathrm{k}}_{mn} \psi^0_{m,\mathrm{k}}, \\
\varphi_{n,\mathrm{k}} &= \sum_{l=1}^{N_\beta} \left[S_{\mathrm{k}}^{-1/2} \right]_{ln} \phi_{l,\mathrm{k}} \\
&= \sum_{l=1}^{N_\beta} \left[S_{\mathrm{k}}^{-1/2} \right]_{ln} \sum_{m=1}^{N_\beta} \left\langle \psi^0_{m,\mathrm{k}}, g_l \right\rangle \psi^0_{m,\mathrm{k}} \\
&= \sum_{m=1}^{N_\beta} \left(\sum_{l=1}^{N_\beta} \left[S_{\mathrm{k}}^{-1/2} \right]_{ln} \left\langle \psi^0_{m,\mathrm{k}}, g_l \right\rangle \right) \psi^0_{m,\mathrm{k}}.
\end{aligned}$$

Therefore, assign

$$\mathrm{U}^{\mathrm{k}}_{mn} = \sum_{l=1}^{N_\beta} \left[S_{\mathrm{k}}^{-1/2} \right]_{ln} \left\langle \psi^0_{m,\mathrm{k}}, g_l \right\rangle .$$

Such a construction suggested in [MV97] is not new — it comes from chemistry literature on modeling electron orbitals. As we will show in the next section, this starting guess leads to quite good results.

5.5 Minimization algorithm

In this section we describe explicitly the algorithm of spread minimization based on the conjugate gradient method [She94].

The subroutines

(p) Preparational computations

At the very beginning compute:

p1) periodic parts of Bloch waves $u^0_{n,\mathrm{k}}$ in Wigner-Seitz cell

p2) $M^{0;\mathrm{k},\mathrm{s}}_{m,n} = \left\langle u^0_{m,\mathrm{k}}, u^0_{n,\mathrm{k+s}} \right\rangle_{\varepsilon,\mathrm{WSC}}$

p3) some initial values for U^k

(t) Transformations of matrices

Subroutines to transform U as $\mathrm{U}(Y)$ and M as $M(\mathrm{U})$:

t1) for given Y, $\mathrm{U}^{\mathrm{k}} \mapsto \mathrm{U}^{\mathrm{k}} \exp\left(Y^{\mathrm{k}}\right)$

t2) for given U, $M^{\mathrm{k,s}} \mapsto \left(\mathrm{U}^{\mathrm{k}}\right)^{\mathrm{H}} M^{\mathrm{k,s}} \mathrm{U}^{\mathrm{k+s}}$

(s) Spread functional Ω

Input: M.

s1) $\bar{\mathrm{r}}_n = -\dfrac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k}} \sum_{\mathrm{s}\in S_{\mathrm{k}}} a_s \mathrm{s} \arg M_{nn}^{\mathrm{k},s}$

s2) $\Omega = \dfrac{1}{N_{\mathrm{k}}^d} \sum_n \sum_{\mathrm{k,s}} a_s \left(\left[\arg M_{nn}^{\mathrm{k,s}} + \mathrm{s} \cdot \bar{\mathrm{r}}_n\right]^2 + 1 - |M_{nn}^{\mathrm{k,s}}|^2 \right)$

(g) Gradient G

Input: M.

g1) $R_{mn}^{\mathrm{k,s}} = M_{mn}^{\mathrm{k,s}} \left(M_{nn}^{\mathrm{k,s}}\right)^*$

g2) $\widetilde{R}_{mn}^{\mathrm{k,s}} = \dfrac{M_{mn}^{\mathrm{k,s}}}{M_{nn}^{\mathrm{k,s}}}$

g3) calculate $\bar{\mathrm{r}}_n$ as in **(s1)**, then $q_n^{\mathrm{k,s}} = \arg M_{nn}^{\mathrm{k,s}} + \mathrm{s} \cdot \bar{\mathrm{r}}_n$

g4) $T_{mn}^{\mathrm{k,s}} = \widetilde{R}_{mn}^{\mathrm{k,s}} q_n^{\mathrm{k,s}}$

g5) $G^{\mathrm{k}} = 2 \sum_{\mathrm{s}} a_s \left(-R^{\mathrm{k,s}} + \left[R^{\mathrm{k,s}}\right]^{\mathrm{H}} + \mathrm{i} T^{\mathrm{k,s}} + \mathrm{i} \left[T^{\mathrm{k,s}}\right]^{\mathrm{H}} \right)$

The minimization procedure

Input: U, tolerance δ, maximal number of iterations m_{iter}.
Conjugate gradient method [She94]:

m01) for initial set U, compute M as in **(t2)**

m02) for M, compute G as described in **(g)**

m03) let $F_{\text{last}} = G$

m1) for fixed G, find $\alpha_0 \geq 0$ which minimizes $\Omega(M(\mathrm{U}(-\alpha G)))$; here first compute $\mathrm{U}(\alpha G)$ as in **(t1)**, then $M(\mathrm{U})$ as in **(t2)** and finally $\Omega(M)$ as in **(s)**

m2) update $\mathrm{U} \mapsto \mathrm{U}(-\alpha_0 G)$ as in **(t1)**

m3) let $F = G(M(\mathrm{U}))$ using subroutines **(t2)** and **(g)**

m4) Polac-Ribière formula: $\beta = \max\left(0, \dfrac{\langle F, F - F_{\text{last}}\rangle_{\mathcal{F}}}{\langle F_{\text{last}}, F_{\text{last}}\rangle_{\mathcal{F}}}\right)$

m5) refine $G \mapsto F + \beta G$ and set $F_{\text{last}} = F$

Perform loop **(m1)**-**(m5)** until Ω cannot reduce the tolerance δ in 3 consequitive steps, or until $\alpha_0 = 0$, or until the number of iterations achieves the prescribed maximum m_{iter}.

Localized Wannier functions

Input: $u^0_{n,\mathrm{k}}$, U.

w1) for the last values of U, compute the functions $u_{n,\mathrm{k}}$ as $u_{n,\mathrm{k}} = \sum_m \mathrm{U}^{\mathrm{k}}_{mn} u^0_{m,\mathrm{k}}$

w2) $w_{n,0}(\mathrm{r}) = \dfrac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}} u_{n,\mathrm{k}}(\mathrm{r})$

5.6 Computational results

1D case

In 1D all bands are isolated and thus can be considered separately. Let us take the first four bands.

We take the following input data:

1. The Photonic Crystal: in the Wigner-Seitz cell $[0, 1]$

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, 0.41) \cup (0.59, 1], \\ 11.56, & x \in [0.41, 0.59]; \end{cases}$$

2. Mesh in x-space: $N_x = 51$, $h_x = 0.02$;

3. Monkhorst-Pack mesh in k-space as in Section 5.2: $N_{\mathrm{k}} = 100$, $h_{\mathrm{k}} = 2\pi/100$;

4. Stencil (5.2.1), weights $a_1 = a_2 = 1/(2h_{\mathrm{k}}^2)$;

5. Linear Finite Elements to compute $u^0_{n,\mathrm{k}}$ (for details see Appendix, Section 10.2.1);

6. Implemented for the first 4 bands ($n = 1, 2, 3, 4$).

If we start the algorithm of Section 5.4 with the initial guess $\mathrm{U}^{\mathrm{k}} = \mathrm{Id} = 1$, the results look as follows:

n	Ω_0	$\Omega_{\min}$	iterations	time
1	7.527	0.036	52	2 sec
2	0.148	0.147	6	1 sec
3	6.684	0.174	52	3 sec
4	2.711	0.139	52	2 sec

For the starting U^{k} constructed from Gaussian bells we obtain:

n	Ω_0	$\Omega_{\min}$	iterations	time
1	24.098	0.036	52	3 sec
2	7.773	0.147	52	3 sec
3	3.109	0.174	52	2 sec
4	24.003	0.139	52	3 sec

The corresponding Wannier functions are illustrated in Figures 5.3-5.14.

2D case

In 2D we consider the following Photonic Crystal: In the Wigner-Seitz cell $[0, 1]$

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 > 0.18^2, \\ 11.56, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 \leq 0.18^2. \end{cases}$$

The band structure corresponding to this crystal is illustrated in Figure 2.2. The first band is isolated (simple), the following tree bands a gathered together in a composite. Therefore, we consider them in groups: 1) $n = 1$; 2) $n = 2, 3, 4$.
We take the following input data:

1. mesh in r-space: $N_x = 31$, $h_x = 0.033$;
2. Monkhorst-Pack mesh in k-space as in Section 5.2 (see Figure 5.1): $N_{\mathrm{k}} = 40$ $h_{\mathrm{k}} = 2\pi/40$;
3. stencil (5.2.2), weights $a_1 = a_2 = a_3 = a_4 = 1/(2h_{\mathrm{k}}^2)$;
4. linear Finite Elements to compute $u^0_{n,\mathrm{k}}$ (for details see Appendix, Section 10.2.2);
5. implemented for the first 4 bands ($n = 1, 2, 3, 4$).

We start the algorithm with the initial guess $\mathrm{U}^{\mathrm{k}} = \mathrm{Id}$. For the isolated band ($n = 1$) it is just 1, for the composite group ($n = 2, 3, 4$) it is a 3×3 identity matrix.
The localization process looks as follows:

n	Ω_0	$\Omega_{\min}$	iterations	time
1	8.667	0.084	30	4 min 7 sec
2-4	26.552	0.510	73	20 min 21 sec

Again, as in 1D also, we remark the localization and the symmetries which corroborate the theoretical expectations of Section 3.2.

Another initial guess is U^k constructed from gaussian bells.
The localization process looks as follows:

n	Ω_0	$\Omega_{\min}$	iterations	time
1	0.996	0.951	8	1 min 48 sec
2-4	8.896	1.637	30	8 min 45 sec

The corresponding Wannier functions are illustrated in Figures 5.15-5.20.

Figure 5.3: 1D, band 1: Initial Wannier function $w^0_{1,0}$

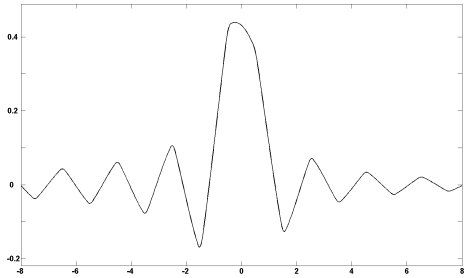


Figure 5.4: Localized $w_{1,0}$ and the minimization of Ω, $\mathrm{U^k} = \mathrm{Id}$

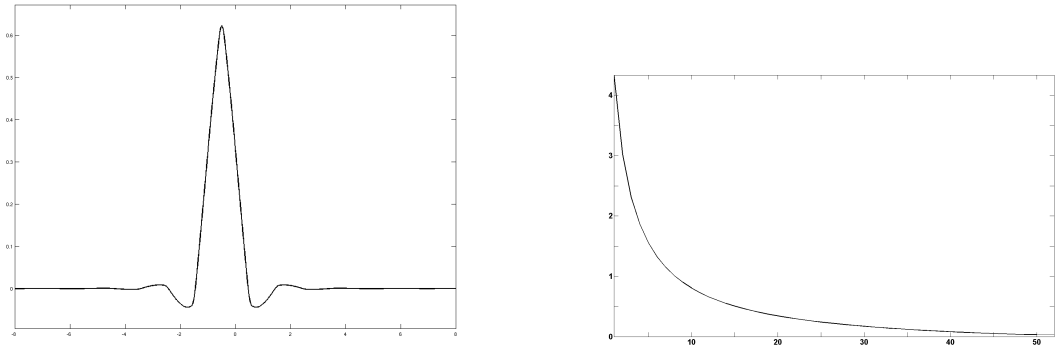


Figure 5.5: Localized $w_{1,0}$ and the minimization of Ω, $\mathrm{U^k}$ gaussian

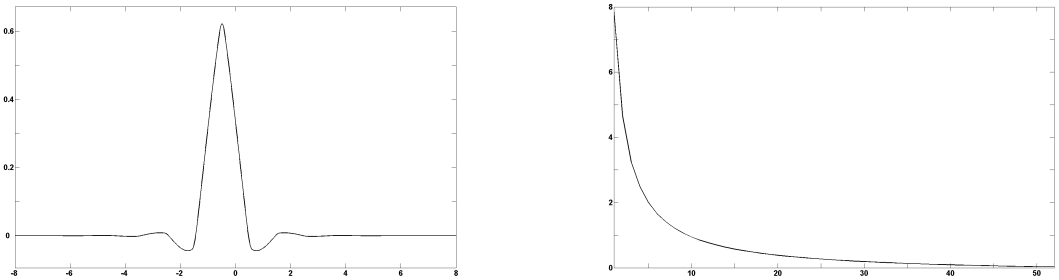


Figure 5.6: 1D, band 2: Initial Wannier function $w^0_{2,0}$

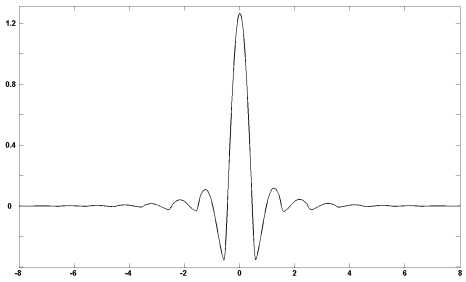


Figure 5.7: Localized $w_{2,0}$ and the minimization of Ω, $\mathrm{U^k} = \mathrm{Id}$

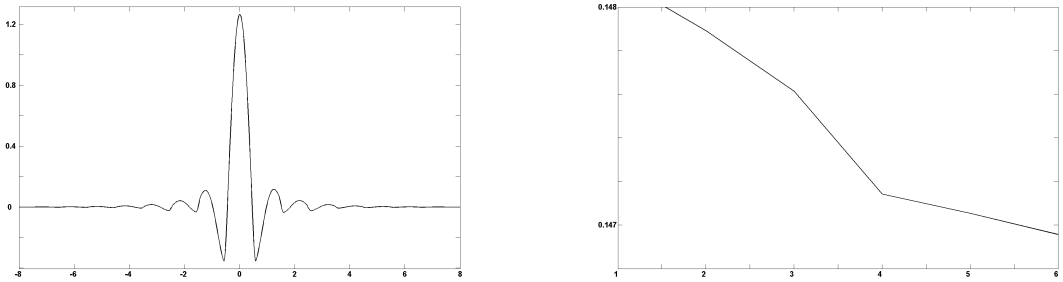


Figure 5.8: Localized $w_{2,0}$ and the minimization of Ω, $\mathrm{U^k}$ gaussian

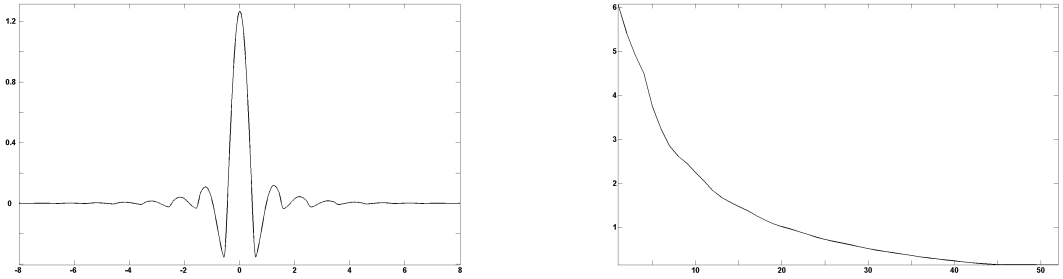


Figure 5.9: 1D, band 3: Initial Wannier function $w^0_{3,0}$

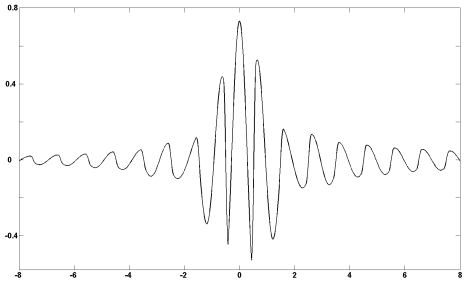


Figure 5.10: Localized $w_{3,0}$ and the minimization of Ω, $\mathrm{U^k} = \mathrm{Id}$

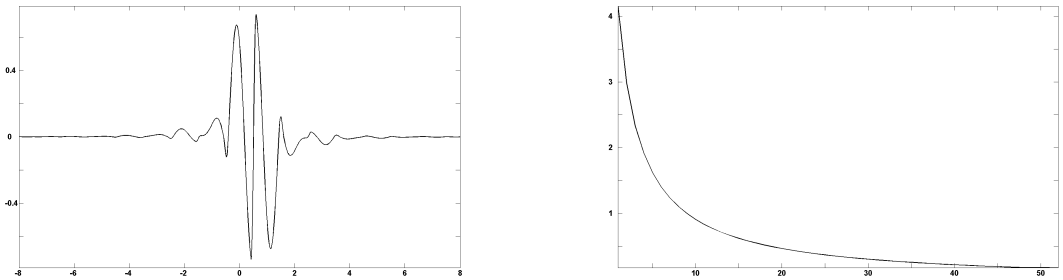


Figure 5.11: Localized $w_{3,0}$ and the minimization of Ω, $\mathrm{U^k}$ gaussian

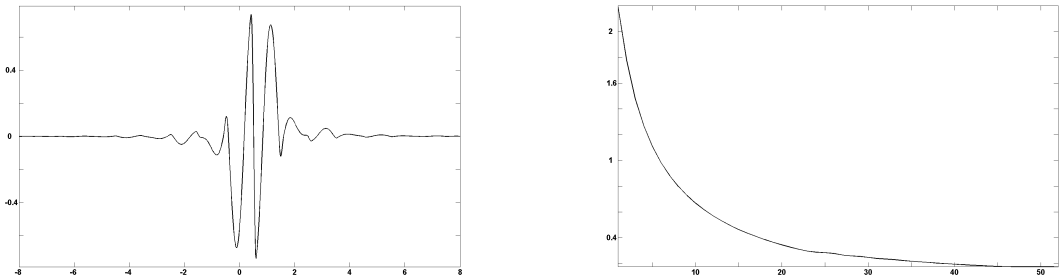


Figure 5.12: 1D, band 4: Initial Wannier function $w^0_{4,0}$

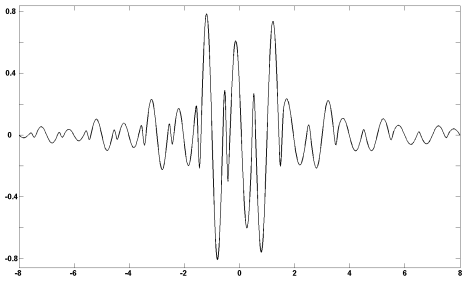


Figure 5.13: Localized $w_{4,0}$ and the minimization of Ω, $\mathrm{U^k} = \mathrm{Id}$

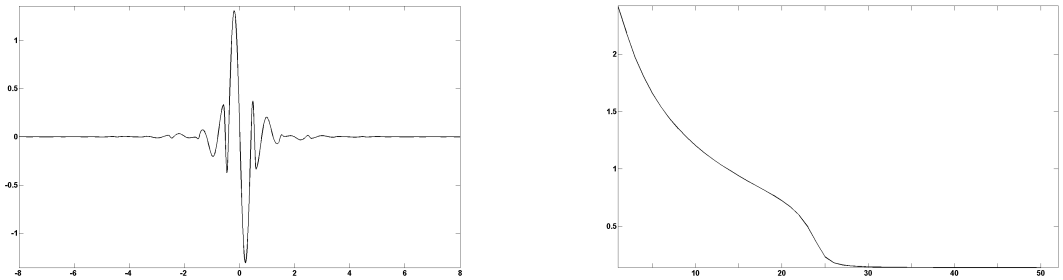


Figure 5.14: Localized $w_{4,0}$ and the minimization of Ω, $\mathrm{U^k}$ gaussian

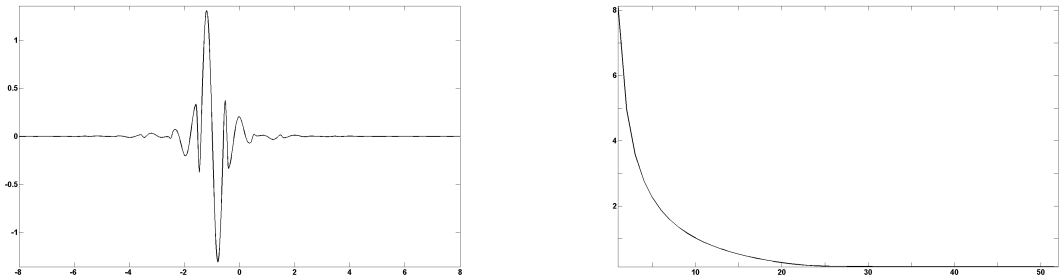


Figure 5.15: 2D, band 1 (composite 1): Initial Wannier function $w^0_{1,0}$

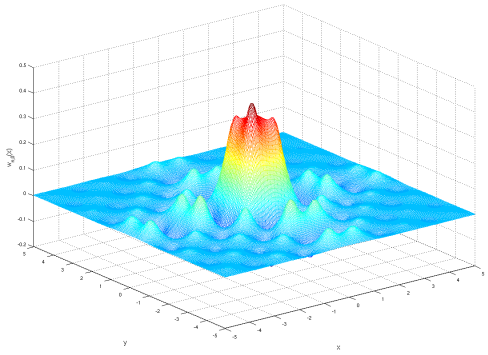


Figure 5.16: Localized $w_{1,0}$ and the minimization of Ω, $\mathrm{U}^{\mathrm{k}} = \mathrm{Id}$

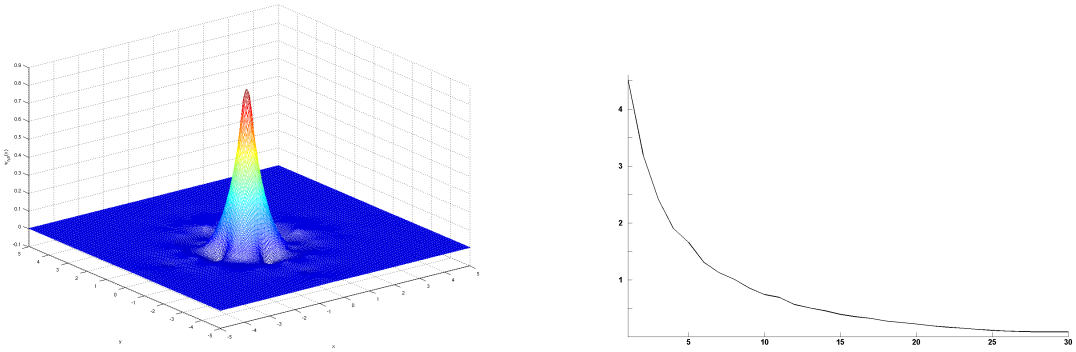


Figure 5.17: Localized $w_{1,0}$ and the minimization of Ω, U^{k} gaussian

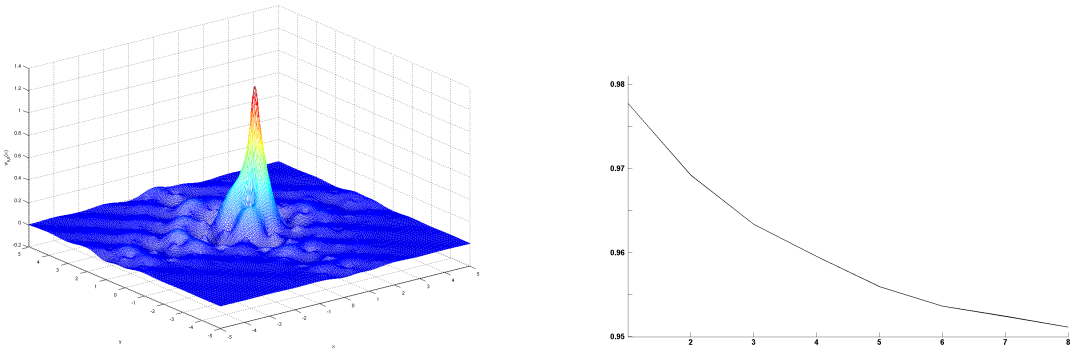
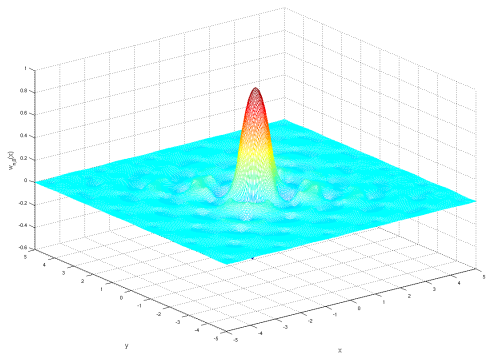
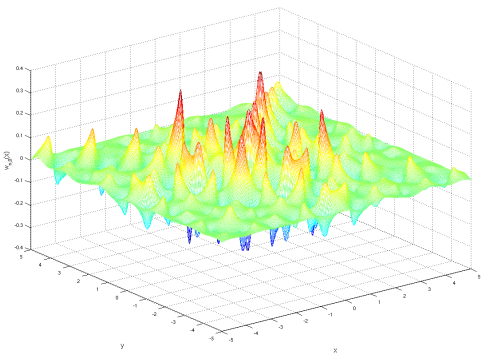


Figure 5.18: 2D, composite 2-4: Initial Wannier functions $w^0_{2,0}$, $w^0_{3,0}$, and $w^0_{4,0}$





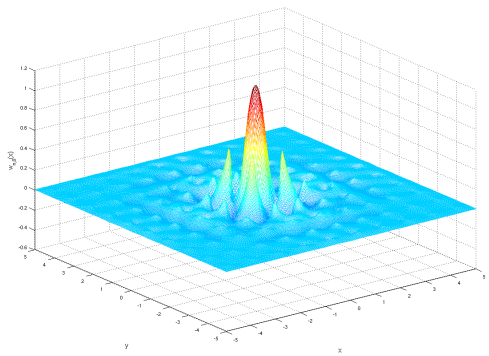


Figure 5.19: 2D, composite 2-4: localized Wannier functions $w_{2,0}$, $w_{3,0}$ and $w_{4,0}$ and minimization of Ω; $\mathrm{U}^{\mathrm{k}} = \mathrm{Id}$

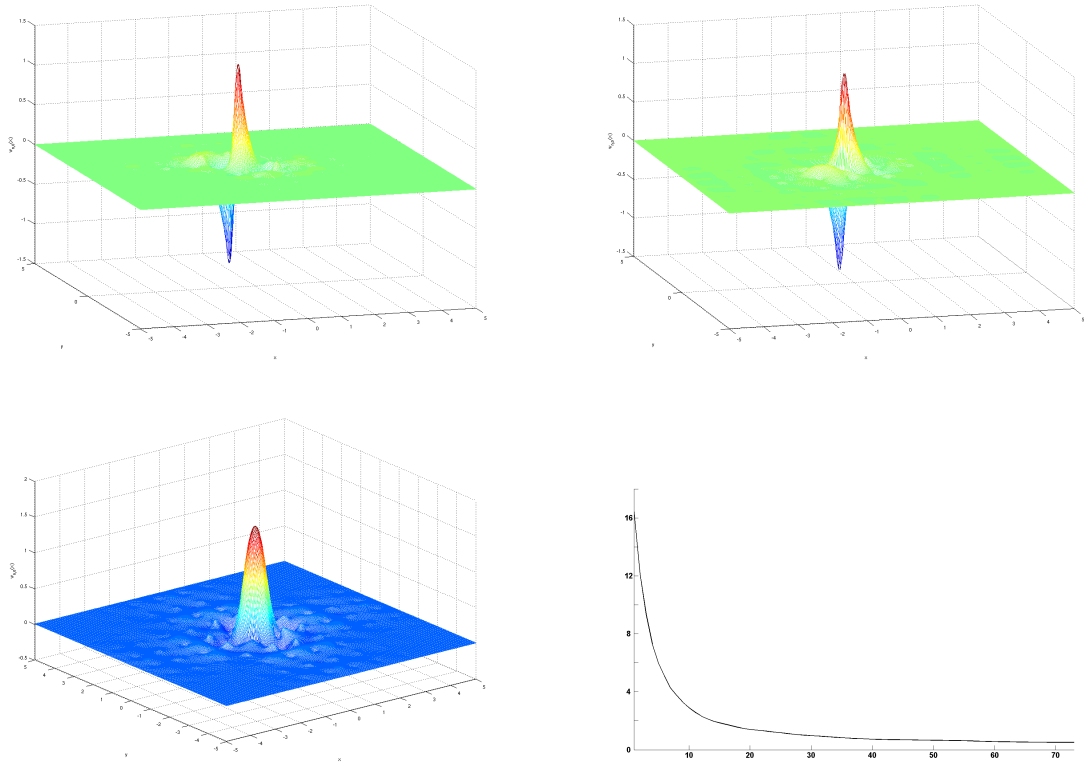
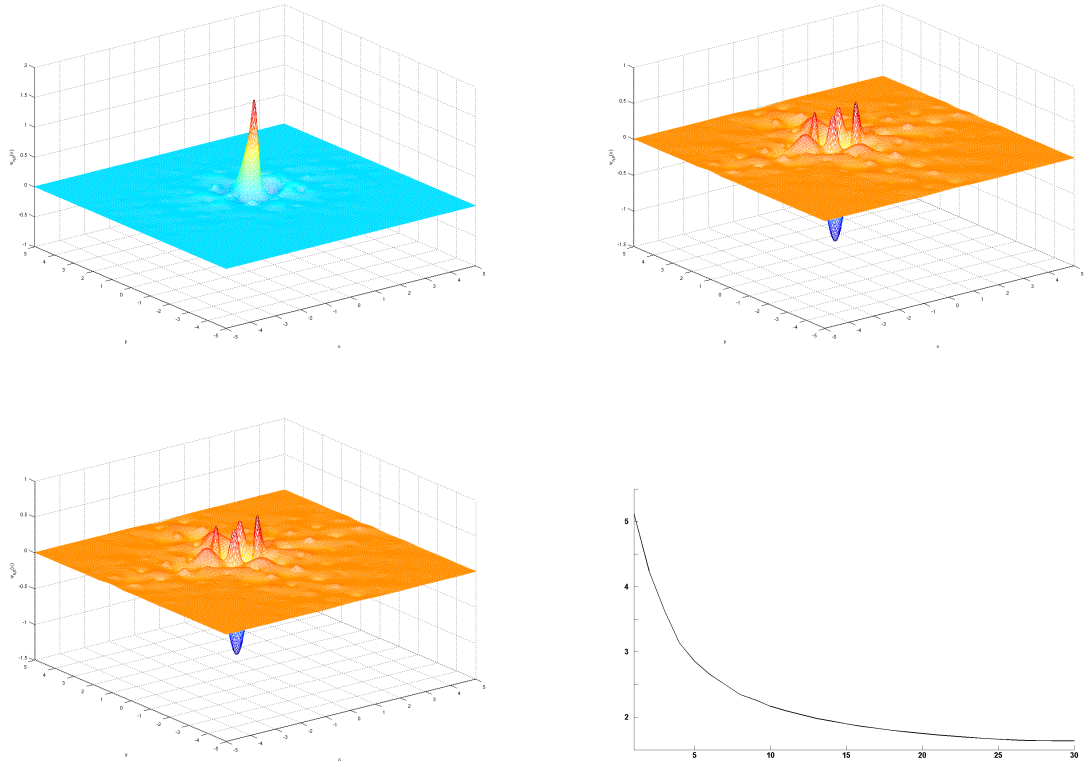


Figure 5.20: 2D, composite 2-4: localized Wannier function $w_{2,0}$, $w_{3,0}$ and $w_{4,0}$ and minimization of Ω; U^k gaussian



Chapter 6

Unitary transform of Wannier functions

6.1 Description of the method

Besides the algorithms which minimize the spread functional at Bloch level, we also review the methods dealing with Wannier functions explicitly. The transformations here are intuitively more transparent, but we must satisfy strong requirements to be sure that the computations are actually allowed.

An elegant method of minimizing Ω was presented by F. Gygi, J.-L. Fatterbert and E. Schwegler in their paper [GFS03]. Having in mind (3.1.2), we can speak here only about the functions $w_{n,0}$ centered at 0 and thus omit index "0". Recall, the spread functional is introduced as follows:

$$\Omega(w) = \sum_n \left[\left\langle r^2 w_n, w_n \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \langle \mathrm{r} w_n, w_n \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right].$$

For a fixed composite of N_β bands taken into account the summation is over $n \in \{1, \ldots, N_\beta\}$. In this chapter we will mention any sum over n or m to be from 1 to N_β.

Defining hermitian $N_\beta \times N_\beta$ matrices A, B by

$$\begin{aligned} A_{mn} &= \langle \mathrm{r} w_m, w_n \rangle_{\mathbb{R}^d,\varepsilon}\,, \\ B_{mn} &= \left\langle r^2 w_m, w_n \right\rangle_{\mathbb{R}^d,\varepsilon}\,, \end{aligned}$$

we rewrite the spread as

$$\Omega(w) = \mathrm{tr}(B) - \sum_n |A_{nn}|^2\,.$$

Note that the elements A_{mn} are d-dimensional vectors.

In contrast to the Marzari-Vanderbilt method observed in Chapter 5, the authors of this method suggest to apply a unitary transform *to the set of Wannier functions* instead of the set of corresponding Bloch waves:

$$\widetilde{w}_n = \sum_m X_{mn} w_m, \quad X \text{ unitary.} \tag{6.1.1}$$

According to the definition (3.1.1), $\widetilde{w}_n$ are also Wannier functions since

$$\widetilde{w}_n(\mathrm{r}) = \sum_m X_{mn} w_m(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_m X_{mn} \psi_{m,\mathrm{k}}(\mathrm{r})\, \mathrm{d}\,\mathrm{k}$$

and in Section 3.2 we illustrated that $\sum_m X_{mn}\psi_{m,\mathrm{k}}$ with unitary X are also Bloch waves. We point out that the transform of Wannier functions suggested in this chapter is very restricted; a more general type of transformation (7.1.4) will be considered in Chapter 7.

The spread functional is not invariant under (6.1.1):

$$\begin{aligned} \Omega(\widetilde{w}) &= \sum_n \left[\left\langle r^2 \widetilde{w}_n, \widetilde{w}_n \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \langle \mathrm{r} \widetilde{w}_n, \widetilde{w}_n \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right] \\ &= \sum_n \left[\left\langle r^2 \sum_m X_{mn} w_m, \sum_{m'} X_{m'n} w_{m'} \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \left\langle \mathrm{r} \sum_m X_{mn} w_m, \sum_{m'} X_{m'n} w_{m'} \right\rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right] \\ &= \sum_n \left[\sum_m X^*_{mn} \sum_{m'} X_{m'n} \left\langle r^2 w_m, w_{m'} \right\rangle_{\mathbb{R}^d,\varepsilon} - \left| \sum_m X^*_{mn} \sum_{m'} X_{m'n} \langle \mathrm{r} w_m, w_{m'} \rangle_{\mathbb{R}^d,\varepsilon} \right|^2 \right] \\ &= \mathrm{tr}(X^{\mathrm{H}} B X) - \sum_n [X^{\mathrm{H}} A X]^2_{nn} \end{aligned}$$

and in general $\Omega(\widetilde{w}) \neq \Omega(w)$.

The problem that we want to solve can be formulated in the following way:

Problem 6.1. *For given hermitian matrices A, B find a unitary transformation X that minimizes the functional $\Omega = \mathrm{tr}(X^{\mathrm{H}}BX) - \sum_n [X^{\mathrm{H}}AX]^2_{nn}$.*

But the trace stays invariant under a unitary transformation, therefore it suffices to maximize the term $\sum_n [X^{\mathrm{H}}AX]^2_{nn}$. As far as the Frobenius norm

$$\|A\|_F^2 = \mathrm{tr}(A^{\mathrm{H}}A)$$

is also invariant under a unitary transformation, we can go further:

$$\sum_n [X^{\mathrm{H}}AX]^2_{nn} = \sum_{m,n} [X^{\mathrm{H}}AX]^2_{mn} - \sum_{m,n\neq m} [X^{\mathrm{H}}AX]^2_{mn} = \|A\|_F^2 - \sum_{m,n\neq m} [X^{\mathrm{H}}AX]^2_{mn}$$

and thus maximization of $\sum_n [X^{\mathrm{H}}AX]^2_{nn}$ is equivalent to minimization of $\sum_{m,n\neq m} [X^{\mathrm{H}}AX]^2_{mn}$. Both together lead us to the problem of diagonalization of a matrix A. Consequently, the new formulation of Problem 6.1 is:

Problem 6.2. *For a given hermitian matrix A find unitary X such that the matrix $X^{\mathrm{H}}AX$ is diagonal.*

6.2 Algorithm in 1D: eigenvalue problem

In a 1D case we have nothing else but the *eigenvalue problem* for A. The algorithm of minimizing the spread of the Wannier functions under a unitary transformation of the basis is now reduced to the following steps.

The minimization procedure

1) Compute the matrix A defined by its elements $A_{mn} = \langle \mathrm{r}w_m^0, w_n^0 \rangle_{\mathbb{R}^d,\varepsilon}$;

2) Solve the eigenvalue problem $AX = XD$;

3) Update the Wannier functions basis: $w_n = \sum_m X_{mn} w_m^0$.

To obtain a numerical result, one has to deal with a *finite* set of Wannier functions. The method is described in [GFS03] just for this requirement. Therefore it can be recommended for application to composite bands which generically occur in higher dimensions. It can also give a result for several isolated bands considered together as a group. For a simple isolated band the problem of diagonalization does not make sense since in that case the matrix A is 1×1 and thus already diagonal.

Note that a unitary transform of a finite number of Wannier functions keeps the orthogonality of the whole basis. Indeed, let the functions $w_{N_1}, \ldots, w_{N_2}$ be transformed and the others not. Then for $m, n \in [N_1, N_2]$ we have

$$\begin{aligned} \langle w_m, w_n \rangle_{\mathbb{R}^d,\varepsilon} &= \sum_{m'=N_1}^{N_2} X^*_{mm'} \sum_{n'=N_1}^{N_2} X_{nn'} \left\langle w^0_{m'}, w^0_{n'} \right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \sum_{m'=N_1}^{N_2} X^{\mathrm{H}}_{m'm} \sum_{n'=N_1}^{N_2} X_{nn'} \delta_{m',n'} \\ &= \left[X^{\mathrm{H}} X\right]_{mn} = \delta_{mn}. \end{aligned}$$

For $m \in [N_1, N_2]$ and $n \notin [N_1, N_2]$,

$$\left\langle w_m, w^0_n \right\rangle_{\mathbb{R}^d,\varepsilon} = \sum_{m'=N_1}^{N_2} X^{\mathrm{H}}_{m'm} \delta_{m',n} = 0.$$

Remark 6.1. *The results of this method do not guarantee to be the same as in the previous chapter. Here we apply another approach — the unitary transformation of a Wannier function set, while a transform corresponding to that of Bloch waves is completely different. Thus, here we are looking for a solution in another manifold, and the result is not expected to be the same.*

Remark 6.2. *Note that in the algorithm suggested above only the second component of the spread functional gets minimized:* $-\sum_n |\langle \mathrm{r} w_n, w_n \rangle|^2$. *The first summand is invariant under a unitary transform.*

Remark 6.3. *Formally the algorithm given above looks similar to that of Section 3.5: first we compute a matrix with entries of a form* $\langle xw_{\cdot}^0, w_{\cdot}^0\rangle_{\mathbb{R}^d,\varepsilon}$*, then find its eigenvectors and construct localized functions. However, there is an essential difference: in one case (Section 3.5) the band* n *is fixed and the eigenvalues are enumerated with respect to nodes* R*; while in the other case (this section) the node* $\mathrm{R} = 0$ *is fixed and we consider several bands. So, being resembling in form, these are two totally different methods.*

6.3 Algorithm in 2D: simultaneous diagonalization

In a two-dimensional case the minimization of the spread is not so obvious. Instead of one matrix A as in 1D, we now have two:

$$A_{mn} = \langle \mathrm{r} w_m, w_n\rangle_{\mathbb{R}^d,\varepsilon} = \left\langle \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} w_m, w_n \right\rangle_{\mathbb{R}^d,\varepsilon} = \begin{pmatrix} A^{(1)}_{mn} \\ A^{(2)}_{mn} \end{pmatrix}.$$

To write it explicitly,

$$\begin{aligned} A^{(1)}_{mn} &= \int_{\mathbb{R}}\int_{\mathbb{R}} r_1 w^*_m(r_1,r_2)\varepsilon(r_1,r_2)w_n(r_1,r_2)\,\mathrm{d}\,r_1\,\mathrm{d}\,r_2, \\ A^{(2)}_{mn} &= \int_{\mathbb{R}} r_2 \int_{\mathbb{R}} w^*_m(r_1,r_2)\varepsilon(r_1,r_2)w_n(r_1,r_2)\,\mathrm{d}\,r_1\,\mathrm{d}\,r_2. \end{aligned}$$

Matrices $A^{(1)}$ and $A^{(2)}$ have to be diagonalized — or, to be precise, maximally diagonalized — *simultaneously*. If the matrices were commuting ones, there would be a unitary transform to diagonalize both of them, but in general they do not commute. Therefore, they cannot be brought to diagonal form with one unitary transform. For this reason we discuss *maximally diagonal* matrices.

J.-F. Cardoso and A. Souloumiac in [CS96] introduce a Jacobi angle technique for simultaneous diagonalization of a set of any matrices, including non-commuting ones. An object to minimize is a sum of the squares of offdiagonal elements:

$$\Phi(X) = \sum_{m,n\neq m} \left(\left| \left[X^{\mathrm{H}} A^{(1)} X\right]_{mn} \right|^2 + \left| \left[X^{\mathrm{H}} A^{(2)} X\right]_{mn} \right|^2 \right)$$

which comes from the previous section.

A unitary transform X is represented as a product of plane rotations. Define a rotation matrix $R^{(i,j;c,s)}$ which can be obtained from an identity matrix by a replacement of the following elements:

$$\begin{pmatrix} R_{ii}^{(i,j;c,s)} & R_{ij}^{(i,j;c,s)} \\ R_{ji}^{(i,j;c,s)} & R_{jj}^{(i,j;c,s)} \end{pmatrix} = \begin{pmatrix} c^* & -s^* \\ s & c \end{pmatrix} \qquad c, s \in \mathbb{C}, \quad |c|^2 + |s|^2 = 1.$$

Introduce a functional $O(i, j; c, s)$ as

$$\begin{aligned} O(i,j;c,s) &= \Phi\left(R^{(i,j;c,s)}\right) \\ &= \sum_{m,n\neq m} \left(\left| \left[\left(R^{(i,j;c,s)}\right)^{\mathrm{H}} A^{(1)} R^{(i,j;c,s)} \right]_{mn} \right|^2 + \left| \left[\left(R^{(i,j;c,s)}\right)^{\mathrm{H}} A^{(2)} R^{(i,j;c,s)} \right]_{mn} \right|^2 \right). \end{aligned}$$

The goal is, for every choice of $i \neq j$, to find Jacobi angles c, s which minimize $O(i, j; c, s)$. The resulting matrices $R^{(i,j;c,s)}$ are then multiplied to construct X.

In [CS96] the values of c, s were given explicitly. For an arbitrary matrix A define a vector

$$h^{(i,j)}(A) = \left(A_{ii} - A_{jj}, \quad A_{ij} + A_{ji}, \quad \mathrm{i}(A_{ji} - A_{ij})\right).$$

Then a real symmetric 3×3 matrix $G(i, j)$ is given by

$$G(i,j) = \mathrm{Re}\left(\left[h^{(i,j)}(A^{(1)})\right]^{\mathrm{H}} \otimes h^{(i,j)}(A^{(1)}) + \left[h^{(i,j)}(A^{(2)})\right]^{\mathrm{H}} \otimes h^{(i,j)}(A^{(2)}) \right).$$

Let $g^{(i,j)} = \left(g_1^{(i,j)}, g_2^{(i,j)}, g_3^{(i,j)}\right)^{\mathrm{T}}$ be an eigenvector of $G(i, j)$ associated to its largest eigenvalue. Without loss of generality, let $\|g^{(i,j)}\| = 1$ and $g_1^{(i,j)} \geq 0$. (Here $\|a\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$ is just an Euclidean norm of a 3-dimensional vector). Cardoso and Souloumiac proved in [CS96], that for any matrices $A^{(1)}$, $A^{(2)}$ the Jacobi angles which minimize $O(i, j; c, s)$ are equal to

$$c = \sqrt{\frac{g_1^{(i,j)} + 1}{2}}, \quad s = \frac{g_2^{(i,j)} - \mathrm{i} g_3^{(i,j)}}{\sqrt{2\left(g_1^{(i,j)} + 1\right)}}.$$

Now let us sum it all up.

The algorithm [CS96], [GFS03]

1. For an initial set of N_β Wannier functions w_n^0, compute matrices $A^{(1)}$ and $A^{(2)}$.
2. Initialize $X = \mathrm{Id}$ of size $N_\beta \times N_\beta$.
3. *(loop)* For $i = 1 \dots N_\beta - 1$ and $j = i + 1 \dots N_\beta$ apply the following steps:
   - compute the vectors $h^{(i,j)}(A^{(1)})$ and $h^{(i,j)}(A^{(2)})$;
   - construct 3×3 matrix $G(i,j)$;
   - find the largest eigenvalue of $G(i,j)$ and take the corresponding eigenvector $g^{(i,j)}$;
   - apply $g^{(i,j)} = g^{(i,j)} \cdot \operatorname{sign} g_1^{(i,j)} / \|g^{(i,j)}\|$;
   - compute Jacobi angles c and s;
   - construct an $N_\beta \times N_\beta$ rotation matrix $R^{(i,j;c,s)}$;
   - apply $R^{(i,j;c,s)}$ to $A^{(1)}$ and $A^{(2)}$: $A^{(k)} = \left[R^{(i,j;c,s)}\right]^{\mathrm{H}} A^{(k)} R^{(i,j;c,s)}$, $k = 1, 2$;
   - update the global unitary transform: $X = X R^{(i,j;c,s)}$.
4. Compute $\Phi(X) = \sum_{m,n \neq m} \left(\left| \left[A^{(1)}\right]_{mn} \right|^2 + \left| \left[A^{(2)}\right]_{mn} \right|^2 \right)$.
5. Repeat the steps 3-4 until $\Phi(X)$ does not decrease by more than some given tolerance at one loop.
6. Update the Wannier functions $w_n = \sum_{m=1}^{N_\beta} X_{mn} w_m^0$.

Remark 6.4. *According to [Blo62], initial matrices $A^{(j)}$ $(j = 1, 2)$ can be determined as:*

$$A_{mn}^{(j)} = \frac{\mathrm{i}}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \int_0^1 u^*_{m,(k_1,k_2)}(r_1, r_2) \varepsilon(r_1, r_2) \frac{\partial}{\partial k_j} u_{n,(k_1,k_2)}(r_1, r_2) \, \mathrm{d}\, r_1 \, \mathrm{d}\, r_2 \, \mathrm{d}\, k_1 \, \mathrm{d}\, k_2.$$

It means that in computation for the initial data we can use periodic functions $u^0_{n,\mathrm{k}}$ instead of w_n^0 and integration over bounded domains WSC *and* BZ.

6.4 Computational results

Although it looks nice and well in theory, the numerical results of the described algorithm are not so good as we would hope. We will demonstrate it below.

1D case

Let us consider a 1D Photonic Crystal with periodicity R = 1, permittivity $\varepsilon_1 = 1$ with a strip of width 0.18 of permittivity $\varepsilon_2 = 11.56$ in the center of the unit interval:

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, 0.41) \cup (0.59, 1], \\ 11.56, & x \in [0.41, 0.59]; \end{cases}$$

The figures were obtained with te following input data:

1. Mesh in x-space: $N_x = 51$, $h_x = 0.02$;
2. Initial Wannier functions w_n^0 computed as in Appendix, Section 10.3, from functions $u_{n,\mathrm{k}}^0$ obtained by FE method as in Appendix, Section 10.2.1;
3. Implemented for the first 10 bands ($n = 1, \ldots, 10$).

For such a crystal the term $\mathrm{tr}(B)$ is much larger than $\sum_n |A_{nn}|^2$: 46.799 against 1.175. It means that the larger term of the spread cannot be minimized. After diagonalization of A the second part grows by 16 percent, but is of minor importance. Totally, Ω reduces only to 99.492 percent:

$$\begin{aligned} \Omega(w^0) &= 46.799 - 1.175 = 45.625, \\ \Omega(w^{\mathrm{loc}}) &= 46.799 - 1.406 = 45.393. \end{aligned}$$

The resulting Wannier functions are not well-localized, see Figure 6.1.
Neither for another number of bands, nor for another crystal, the situation becomes better.

2D case

Consider a 2D Photonic Crystal with a periodic cell $[0,1]^2$ and the following permittivity distribution therein:

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 > 0.18^2, \\ 11.56, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 \leq 0.18^2 \end{cases}$$

(see Figure 2.1). Take first two isolated bands: for $n = 1$ it is a simple band for $n = 2, 3, 4$ there is a composite band. Alltogether, we have four starting Wannier functions. Input data:

1. Mesh in x-space: $N_x = 31$, $h_x = 0.033$;
2. Initial Wannier functions w_n^0 computed as in Appendix, Section 10.3, from functions $u_{n,\mathrm{k}}^0$ obtained by FE method as in Appendix, Section 10.2.2;
3. Implemented for the first 2 composite bands grouped together: $n = 1$ and $n = 2, 3, 4$.

Again, like in 1D, the invariant part of the spread functional is much more significant than the one we maximize:

$$\begin{aligned} \Omega(w^0) &= 36.512 - 0.470 = 36.042, \\ \Omega(w^{\mathrm{loc}}) &= 36.512 - 0.532 = 35.979. \end{aligned}$$

The matrices $A^{(1)}$ and $A^{(2)}$ get maximally diagonalized very fast — in 4 iterations. Nevertheless, it does not help a lot to localize the Wannier functions. The results one can see on Figure 6.2.

6.5 Conclusions

The technique presented in this chapter deals with unitary transformations of a set of Wannier functions. Thanks to elegant simplifications we obtain an algorithm which can be easily implemented. It does not require any information about a descent direction of

the spread functional and gives quick results.

But unfortunately, due to the serious restriction (6.1.1) on the transform, these results are not satisfactory. The main problem of this approach is the *invariance* of $\sum_n \langle r^2 w_n, w_n \rangle_{\mathbb{R}^d,\varepsilon}$ under such a unitary transform. This part of the spread functional is much larger that the rest, and therefore any unitary transform of w_n does not make the spread decrease significantly. The authors of the method do not report about this problem in [GFS03].

Figure 6.1: Wannier functions in 1D, bands 1-4: Non-localized (left) and localized under unitary transform minimization (right)

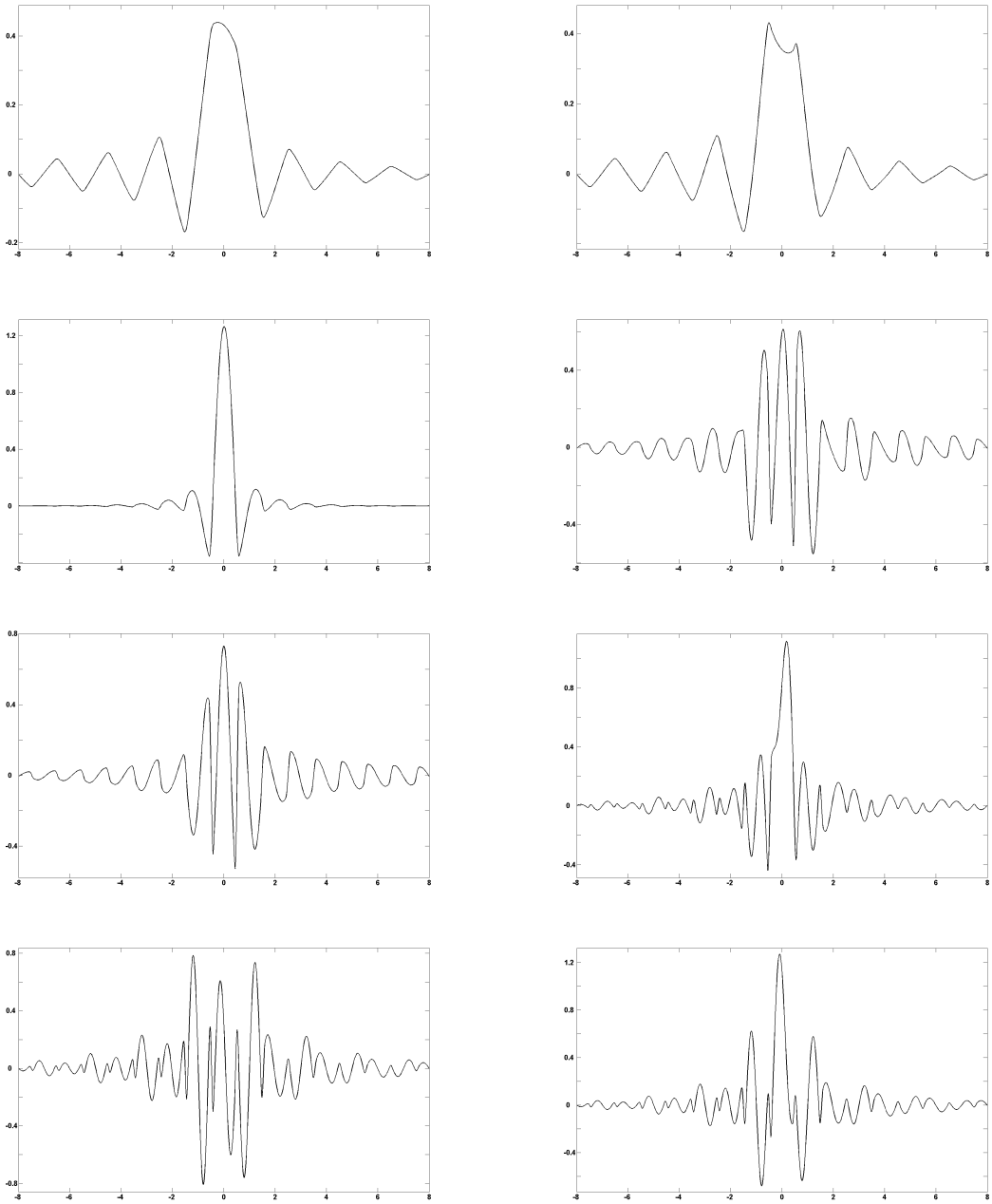
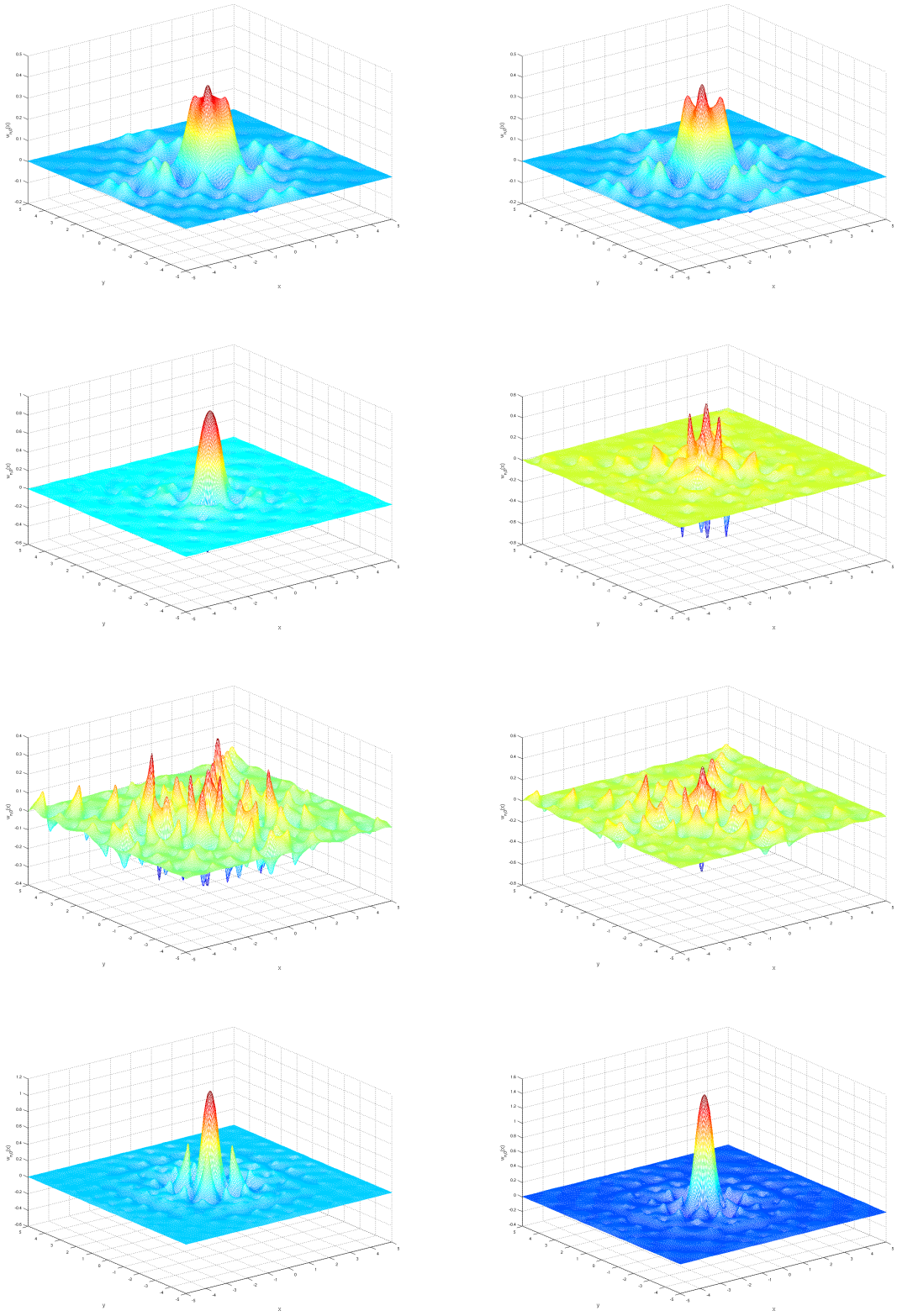


Figure 6.2: Wannier functions in 2D, bands 1-4: Non-localized (left) and localized under unitary transform minimization (right)



Chapter 7

Sum-unitary transform of the Wannier functions

7.1 Fourier transform of U

As we have mentioned in Section 3.2, nonuniqueness of Wannier functions originates from phase indeterminacy of the corresponding Bloch waves. We fix some initial wave set and consider generalized Bloch waves as unitary transforms of this set:

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = \sum_m \mathrm{U}^{\mathrm{k}}_{mn} \psi^0_{m,\mathrm{k}}(\mathrm{r}).$$

Remember that the summation over m involves the Bloch waves of just one composite band. In case of a simple band there is no summation and $\mathrm{U}^{\mathrm{k}} = \mathrm{e}^{\mathrm{i}\theta(\mathrm{k})}$ with some real-valued function $\theta(\mathrm{k})$.

Let us switch to the r-space level and see what happens to the Wannier functions under these transformations:

$$w_{n,0}(\mathrm{r}) = \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \sum_m \mathrm{U}^{\mathrm{k}}_{mn} \psi^0_{m,\mathrm{k}}(\mathrm{r}) \, \mathrm{d}\,\mathrm{k}.$$

The matrices U^{k} are unitary and periodic with respect to the reciprocal lattice Γ^* in k-space, i.e. $\mathrm{U}^{\mathrm{k}+\mathrm{K}} = \mathrm{U}^{\mathrm{k}}$, $\mathrm{K} \in \Gamma^*$. This means that U^{k} can be written as a Fourier transform

over the space lattice Γ:

$$\mathrm{U}^{\mathrm{k}} = \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} W^{\mathrm{R}} \tag{7.1.1}$$

with

$$W^{\mathrm{R}} = \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \mathrm{U}^{\mathrm{k}} \,\mathrm{d}\,\mathrm{k}. \tag{7.1.2}$$

For a simple (one-wave) isolated band there will be

$$W^{\mathrm{R}} = \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}(\mathrm{k}\cdot\mathrm{R}+\theta(\mathrm{k}))} \,\mathrm{d}\,\mathrm{k}.$$

The unitarity of U^{k} transforms to the following property of W^{R}:

$$\begin{aligned}
\mathrm{Id} &= \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} (\mathrm{U}^{\mathrm{k}})^{\mathrm{H}} \mathrm{U}^{\mathrm{k}} \,\mathrm{d}\,\mathrm{k} \\
&= \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \sum_{\mathrm{R}} (W^{\mathrm{R}})^{\mathrm{H}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \sum_{\mathrm{R}'} W^{\mathrm{R}'} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}'} \,\mathrm{d}\,\mathrm{k} \\
&= \sum_{\mathrm{R},\mathrm{R}'} (W^{\mathrm{R}})^{\mathrm{H}} W^{\mathrm{R}'} \left(\frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot(\mathrm{R}-\mathrm{R}')} \,\mathrm{d}\,\mathrm{k} \right) \\
&= \sum_{\mathrm{R},\mathrm{R}'} (W^{\mathrm{R}})^{\mathrm{H}} W^{\mathrm{R}'} \delta_{\mathrm{R},\mathrm{R}'} \\
&= \sum_{\mathrm{R}} (W^{\mathrm{R}})^{\mathrm{H}} W^{\mathrm{R}}.
\end{aligned}$$

Therefore, the requirement for W is

$$\sum_{\mathrm{R}} (W^{\mathrm{R}})^{\mathrm{H}} W^{\mathrm{R}} = \mathrm{Id}. \tag{7.1.3}$$

We call this property *sum-unitarity* and the transform W, accordingly, *sum-unitary*. The corresponding generalized Wannier functions will be

$$\begin{aligned}
w_{n,0}(\mathrm{r}) &= \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \sum_{m} \mathrm{U}^{\mathrm{k}}_{mn} \psi^0_{m,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k} \\
&= \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \sum_{m} \sum_{\mathrm{R}\in\Gamma} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} W^{\mathrm{R}}_{mn} \psi^0_{m,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k} \\
&= \sum_{m} \sum_{\mathrm{R}} W^{\mathrm{R}}_{mn} \left(\frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi^0_{m,\mathrm{k}}(\mathrm{r}) \,\mathrm{d}\,\mathrm{k} \right)
\end{aligned}$$

and finally

$$w_{n,0}(\mathrm{r}) = \sum_m \sum_{\mathrm{R}\in\Gamma} W^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}(\mathrm{r}). \tag{7.1.4}$$

Note, here the initial (fixed) Wannier functions correspond to the initial (fixed) Bloch waves:

$$w^0_{n,\mathrm{R}}(\mathrm{r}) \;=\; \frac{1}{(2\pi)^d} \int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \psi^0_{n,\mathrm{k}}(\mathrm{r})\, \mathrm{d}\,\mathrm{k}.$$

One can easily check that such a transform keeps the Wannier functions orthonormal:

$$\begin{aligned}
\langle w_{n,0}, w_{n',0}\rangle_{\mathbb{R}^d,\varepsilon} &= \left\langle \sum_m \sum_{\mathrm{R}} W^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}, \sum_{m'} \sum_{\mathrm{R}'} W^{\mathrm{R}'}_{m'n'} w^0_{m',\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \sum_{m,m'} \sum_{\mathrm{R},\mathrm{R}'} \left(W^{\mathrm{R}}_{mn}\right)^* W^{\mathrm{R}'}_{m'n'} \left\langle w^0_{m,\mathrm{R}}, w^0_{m',\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \sum_{m,m'} \sum_{\mathrm{R},\mathrm{R}'} \left[W^{\mathrm{R}}\right]^{\mathrm{H}}_{nm} W^{\mathrm{R}'}_{m'n'} \delta_{m,m'} \delta_{\mathrm{R},\mathrm{R}'} \\
&= \sum_m \sum_{\mathrm{R}} \left[W^{\mathrm{R}}\right]^{\mathrm{H}}_{nm} W^{\mathrm{R}}_{mn'} \\
&= \left[\sum_{\mathrm{R}} \left[W^{\mathrm{R}}\right]^{\mathrm{H}} W^{\mathrm{R}} \right]_{nn'} = \delta_{n,n'}.
\end{aligned}$$

Let us introduce a manifold of sum-unitary elements W:

$$\mathcal{M} = \left\{ W = \left\{W^{\mathrm{R}}\right\}_{\mathrm{R}\in\Gamma} \quad | \quad \sum_{\mathrm{R}} \left(W^{\mathrm{R}}\right)^{\mathrm{H}} W^{\mathrm{R}} = \mathrm{Id} \right\}. \tag{7.1.5}$$

One can easily check the following properties of $\mathcal{M}$:

1. If $A \in \mathcal{M}$, then $-A \in \mathcal{M}$.

2. If $A \in \mathcal{M}$, then $A^{\mathrm{H}} \in \mathcal{M}$.

3. If $A \in \mathcal{M}$ and $|\alpha|^2 = 1$ (or $\alpha = \mathrm{e}^{\mathrm{i}\theta}$), then $\alpha A \in \mathcal{M}$.

Clearly, the manifold is not linear.

From here on we assume that all further W belong to $\mathcal{M}$. A trivial example of such W is

$$W^0[W^0]^{\mathrm{H}} = \mathrm{Id}; \quad W^{\mathrm{R}} = 0 \text{ for } \mathrm{R} \neq 0.$$

Such a particular case, already considered in Chapter 6, is very restricted. We have shown that it provides a cheap minimization, but the result is not satisfactory. In this chapter we will try to find the minimum of the spread in a wider manifold.

7.2 Ω as a functional over W

Introduce the notations:

$$[P_j^{0;\mathrm{R},\mathrm{R}'}]_{mn} = \left\langle \mathrm{r}^j w^0_{m,\mathrm{R}}, w^0_{n,\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \quad \text{for } j = 1, 2, \tag{7.2.1}$$

$$P_j^{\mathrm{R},\mathrm{R}'} = (W^{\mathrm{R}})^{\mathrm{H}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'} \quad \text{for } j = 1, 2. \tag{7.2.2}$$

One can easily check that $(P_j^{0;\mathrm{R},\mathrm{R}'})^{\mathrm{H}} = P_j^{0;\mathrm{R}',\mathrm{R}}$ and $(P_j^{\mathrm{R},\mathrm{R}'})^{\mathrm{H}} = P_j^{\mathrm{R}',\mathrm{R}}$.

Remark 7.1. $\left[P_1^{0;\mathrm{R},\mathrm{R}'}\right]_{mn}$ *and* $\left[P_1^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ *are vectors of dimensionality* d*, while* $\left[P_2^{0;\mathrm{R},\mathrm{R}'}\right]_{mn}$ *and* $\left[P_2^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ *are scalars.*

Let us find the expressions for the parts of the spread functional $\bar{\mathrm{r}}_n$ and $\langle r^2 \rangle_n$ in terms of the sum-unitary transform of the initial Wannier functions.

$$\begin{aligned}
\left\langle \mathrm{r}^j w_{n,0}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} &= \left\langle \mathrm{r}^j \sum_m \sum_{\mathrm{R}} W^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}, \sum_{m'} \sum_{\mathrm{R}'} W^{\mathrm{R}'}_{m'n} w^0_{m',\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \sum_{\mathrm{R},\mathrm{R}'} \sum_{m,m'} (W^{\mathrm{R}}_{mn})^* W^{\mathrm{R}'}_{m'n} \left\langle \mathrm{r}^j w^0_{m,\mathrm{R}}, w^0_{m',\mathrm{R}'} \right\rangle_{\mathbb{R}^d,\varepsilon} \\
&= \sum_{\mathrm{R},\mathrm{R}'} \sum_{m,m'} (W^{\mathrm{R}})^{\mathrm{H}}_{nm} [P_j^{0;\mathrm{R},\mathrm{R}'}]_{mm'} W^{\mathrm{R}'}_{m'n} \\
&= \sum_{\mathrm{R},\mathrm{R}'} [(W^{\mathrm{R}})^{\mathrm{H}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'}]_{nn} \\
&= \sum_{\mathrm{R},\mathrm{R}'} [P_j^{\mathrm{R},\mathrm{R}'}]_{nn}.
\end{aligned}$$

Therefore,

$$\bar{\mathrm{r}}_n = \sum_{\mathrm{R},\mathrm{R}'} [P_1^{\mathrm{R},\mathrm{R}'}]_{nn}, \qquad \left\langle r^2 \right\rangle_n = \sum_{\mathrm{R},\mathrm{R}'} [P_2^{\mathrm{R},\mathrm{R}'}]_{nn}, \tag{7.2.3}$$

and then

$$\Omega(W) = \sum_{\mathrm{R},\mathrm{R}'} \mathrm{tr}(P_2^{\mathrm{R},\mathrm{R}'}) - \sum_n \left| \sum_{\mathrm{R},\mathrm{R}'} [P_1^{\mathrm{R},\mathrm{R}'}]_{nn} \right|^2 . \tag{7.2.4}$$

Or, equivalently,

$$\Omega(W) = \sum_{\mathrm{R},\mathrm{R}'} \mathrm{tr}\left((W^{\mathrm{R}})^{\mathrm{H}} P_2^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'} \right) - \sum_n \left| \sum_{\mathrm{R},\mathrm{R}'} \left[(W^{\mathrm{R}})^{\mathrm{H}} P_1^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'} \right]_{nn} \right|^2 .$$

7.3 Derivative of Ω

To find a minimum of the spread functional, we need the information about its direction of decay. Without this information the minimization process can be too costly and inefficient since we would have to go over many points, within any discretization. But before we start, we have to find a way how to switch from one sum-unitary W to another — namely, a way to update W keeping it sum-unitary.

Let us perturb W as

$$W \mapsto WA. \tag{7.3.1}$$

The requirement for A then comes from (7.1.3) and reads

$$\sum_{\mathrm{R}} W^{\mathrm{R}} A^{\mathrm{R}} [A^{\mathrm{R}}]^{\mathrm{H}} [W^{\mathrm{R}}]^{\mathrm{H}} = \mathrm{Id},$$

with $\sum_{\mathrm{R}} W^{\mathrm{R}} [W^{\mathrm{R}}]^{\mathrm{H}} = \mathrm{Id}$. Let A have a form

$$A^{\mathrm{R}} = \mathrm{e}^{tX^{\mathrm{R}}} \approx \mathrm{Id} + tX^{\mathrm{R}} \tag{7.3.2}$$

with infinitesimal $t > 0$. Then for X we have the following condition:

$$\begin{aligned} \mathrm{Id} &= \sum_{\mathrm{R}} W^{\mathrm{R}} \left(\mathrm{Id} + tX^{\mathrm{R}} \right) \left(\mathrm{Id} + t[X^{\mathrm{R}}]^{\mathrm{H}} \right) [W^{\mathrm{R}}]^{\mathrm{H}} \\ &= \mathrm{Id} + t \sum_{\mathrm{R}} W^{\mathrm{R}} \left(X^{\mathrm{R}} + [X^{\mathrm{R}}]^{\mathrm{H}} \right) [W^{\mathrm{R}}]^{\mathrm{H}} + O(t^2), \end{aligned}$$

or

$$0 \;=\; \sum_{\mathrm{R}} W^{\mathrm{R}} \left(X^{\mathrm{R}} + [X^{\mathrm{R}}]^{\mathrm{H}}\right) [W^{\mathrm{R}}]^{\mathrm{H}}. \tag{7.3.3}$$

The equation (7.3.3) characterizes $\mathcal{T}_W\mathcal{M}$ — the tangential manifold to $\mathcal{M}$ (7.1.5):

$$\mathcal{T}_W\mathcal{M} = \left\{ X = \left\{X^{\mathrm{R}}\right\}_{\mathrm{R}\in\Gamma} \quad | \quad \sum_{\mathrm{R}} W^{\mathrm{R}} \left(X^{\mathrm{R}} + [X^{\mathrm{R}}]^{\mathrm{H}}\right) [W^{\mathrm{R}}]^{\mathrm{H}} = 0 \right\}. \tag{7.3.4}$$

The update for W corresponding to (7.3.1) reads:

$$W \mapsto W \exp^{tX} \approx W(\mathrm{Id} + tX). \tag{7.3.5}$$

Unfortunately, we cannot derive an explicit formula for such X. One of the possible choices satisfying (7.3.3) is

$$X^{\mathrm{R}} = -[X^{\mathrm{R}}]^{\mathrm{H}} \quad \text{for all R.} \tag{7.3.6}$$

Introduce a notation:

$$\triangle f(W) \;=\; f(W + \triangle W) - f(W).$$

Let us search for the derivative of the spread Ω in a point W in a direction X which we denote as $\mathrm{d}\,\Omega(W)[X]$ and define as

$$\mathrm{d}\,\Omega(W)[X] = \lim_{t\to 0} \frac{\triangle\Omega}{t} = \lim_{t\to 0} \frac{\Omega(W(\mathrm{Id} + tX)) - \Omega(W)}{t}.$$

To find $\triangle\Omega$, we consider $\triangle P$. For $j = 1, 2$

$$\begin{aligned}
\triangle P_j^{\mathrm{R},\mathrm{R}'} &= \left(\mathrm{Id} + t[X^{\mathrm{R}}]^{\mathrm{H}}\right) [W^{\mathrm{R}}]^{\mathrm{H}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'} \left(\mathrm{Id} + tX^{\mathrm{R}'}\right) - P_j^{\mathrm{R},\mathrm{R}'} \\
&= \left(\mathrm{Id} - tX^{\mathrm{R}}\right) P_j^{\mathrm{R},\mathrm{R}'} \left(\mathrm{Id} + tX^{\mathrm{R}'}\right) - P_j^{\mathrm{R},\mathrm{R}'} \\
&= tP_j^{\mathrm{R},\mathrm{R}'} X^{\mathrm{R}'} - tX^{\mathrm{R}} P_j^{\mathrm{R},\mathrm{R}'} \\
&= tP_j^{\mathrm{R},\mathrm{R}'} X^{\mathrm{R}'} - t\left(P_j^{\mathrm{R}',\mathrm{R}} X^{\mathrm{R}}\right)^{\mathrm{H}}.
\end{aligned}$$

From here,

$$\triangle \left(\sum_{\mathrm{R},\mathrm{R}'} \left[P_j^{\mathrm{R},\mathrm{R}'}\right]_{nn} \right) \;=\; 2t \sum_{\mathrm{R},\mathrm{R}'} \mathrm{Re}\left(\left[P_j^{\mathrm{R},\mathrm{R}'} X^{\mathrm{R}'}\right]_{nn} \right).$$

Then

$$\begin{aligned}\triangle\left(\sum_n \langle r^2\rangle_n\right) &= \triangle\left(\sum_{\mathrm{R},\mathrm{R}'} \mathrm{tr}\left(P_2^{\mathrm{R},\mathrm{R}'}\right)\right)\\ &= 2t\sum_{\mathrm{R},\mathrm{R}'} \mathrm{Re}\left(\mathrm{tr}\left(P_2^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right)\right).\end{aligned}$$

To evaluate the second component of the spread functional, first introduce the notation:

$$D = \mathrm{Diag}(\mathrm{Re}(\bar{\mathrm{r}}_n)), \tag{7.3.7}$$

where $\mathrm{Diag}(a)$ is a diagonal matrix with elements of the vector a at its diagonal.

We obtain:

$$\begin{aligned}\triangle\left(\sum_n |\bar{\mathrm{r}}_n|^2\right) &= \sum_n \left(|\bar{\mathrm{r}}_n + \triangle\bar{\mathrm{r}}_n|^2 - |\bar{\mathrm{r}}_n|^2\right)\\ &= 2\sum_n \mathrm{Re}(\bar{\mathrm{r}}_n \cdot \triangle\bar{\mathrm{r}}_n) + O(t^2)\\ &\approx 2\sum_n \mathrm{Re}\left(\bar{\mathrm{r}}_n \cdot \triangle\left(\sum_{\mathrm{R},\mathrm{R}'}[P_1^{\mathrm{R},\mathrm{R}'}]_{nn}\right)\right)\\ &= 2\sum_n \mathrm{Re}(\bar{\mathrm{r}}_n)\cdot 2t\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\left[P_1^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right]_{nn}\right)\\ &= 4t\sum_n D_{nn}\cdot\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\left[P_1^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right]_{nn}\right)\\ &= 4t\sum_n\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(D_{nn}\cdot\left[P_1^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right]_{nn}\right)\\ &= 4t\sum_{\mathrm{R},\mathrm{R}'}\sum_n\mathrm{Re}\left(\left[D\cdot P_1^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right]_{nn}\right)\\ &= 4t\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(\left[D\cdot P_1^{\mathrm{R},\mathrm{R}'}\right]X^{\mathrm{R}'}\right)\right).\end{aligned}$$

Recall that D_{nn} and $\left[P_1^{\mathrm{R},\mathrm{R}'}\right]_{nn}$ are d-dimensional vectors. Therefore,

$$\begin{aligned}\left[D\cdot P_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} &= \left[\sum_{p=1}^{d} D^p \left(P_1^{\mathrm{R},\mathrm{R}'}\right)^p\right]_{mn} \\ &= \sum_{p=1}^{d}\sum_{l} D^p_{ml}\left[P_1^{\mathrm{R},\mathrm{R}'}\right]^p_{ln} \\ &= \sum_{p=1}^{d} D^p_{mm}\left[P_1^{\mathrm{R},\mathrm{R}'}\right]^p_{mn} \\ &= D_{mm}\cdot\left[P_1^{\mathrm{R},\mathrm{R}'}\right]_{mn}.\end{aligned}$$

Finally, for the spread functional we now have:

$$\begin{aligned}\triangle\Omega &= \triangle\left(\sum_n \left\langle r^2\right\rangle_n\right) - \triangle\left(\sum_n |\bar{\mathrm{r}}_n|^2\right) \\ &= 2t\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(P_2^{\mathrm{R},\mathrm{R}'}X^{\mathrm{R}'}\right)\right) - 4t\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(\left[D\cdot P_1^{\mathrm{R},\mathrm{R}'}\right]X^{\mathrm{R}'}\right)\right) \\ &= 2t\sum_{\mathrm{R},\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(\left[P_2^{\mathrm{R},\mathrm{R}'} - 2D\cdot P_1^{\mathrm{R},\mathrm{R}'}\right]X^{\mathrm{R}'}\right)\right) \\ &= 2t\sum_{\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(T^{\mathrm{R}'}X^{\mathrm{R}'}\right)\right),\end{aligned}$$

with

$$T^{\mathrm{R}'} = \sum_{\mathrm{R}}\left[P_2^{\mathrm{R},\mathrm{R}'} - 2D\cdot P_1^{\mathrm{R},\mathrm{R}'}\right]. \tag{7.3.8}$$

Therefore, a derivative of Ω reads:

$$\mathrm{d}\,\Omega(W)[X] = \lim_{t\to 0}\frac{\triangle\Omega}{t} = 2\sum_{\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(T^{\mathrm{R}'}X^{\mathrm{R}'}\right)\right). \tag{7.3.9}$$

Now let us define an inner product of matrix-type structures as

$$\langle A, B\rangle_{\mathcal{F}} := \sum_{\mathrm{R}}\mathrm{tr}\left([A^{\mathrm{R}}]^{\mathrm{H}}B^{\mathrm{R}}\right). \tag{7.3.10}$$

Define an operator

$$\mathcal{A}[B] = \frac{1}{2}\left(B - B^{\mathrm{H}}\right).$$

Then, having in mind our choice $\left[X^{\mathrm{R}}\right]^{\mathrm{H}} = -X^{\mathrm{R}}$,

$$\begin{aligned}
\langle \mathcal{A}[B], X\rangle_{\mathcal{F}} &= \frac{1}{2}\left(\langle B, X\rangle_{\mathcal{F}} - \left\langle B^{\mathrm{H}}, X\right\rangle_{\mathcal{F}}\right) \\
&= \frac{1}{2}\sum_{\mathrm{R}}\left(\mathrm{tr}\left([B^{\mathrm{R}}]^{\mathrm{H}}X^{\mathrm{R}}\right) - \mathrm{tr}\left(B^{\mathrm{R}}X^{\mathrm{R}}\right)\right) \\
&= \frac{1}{2}\sum_{\mathrm{R}}\left(\mathrm{tr}\left([B^{\mathrm{R}}]^{\mathrm{H}}[-X^{\mathrm{R}}]^{\mathrm{H}}\right) - \mathrm{tr}\left(B^{\mathrm{R}}X^{\mathrm{R}}\right)\right) \\
&= -\frac{1}{2}\sum_{\mathrm{R}}\left(\mathrm{tr}^{*}\left(B^{\mathrm{R}}X^{\mathrm{R}}\right) + \mathrm{tr}\left(B^{\mathrm{R}}X^{\mathrm{R}}\right)\right) \\
&= -\sum_{\mathrm{R}}\mathrm{Re}\left(\mathrm{tr}\left(B^{\mathrm{R}}X^{\mathrm{R}}\right)\right).
\end{aligned}$$

Thus we conclude that

$$\mathrm{d}\,\Omega(W)[X] = 2\sum_{\mathrm{R}'}\mathrm{Re}\left(\mathrm{tr}\left(T^{\mathrm{R}'}X^{\mathrm{R}'}\right)\right) = -2\,\langle \mathcal{A}[T], X\rangle_{\mathcal{F}}\,. \tag{7.3.11}$$

7.4 The descent direction

From the expression for the derivative we can try to retrieve the descent direction of the spread in the submanifold of $\mathcal{T}_W\mathcal{M}$ (7.3.4) determined by (7.3.6). Note that since we are restricted to (7.3.6), we cannot talk here about the *gradient* of the spread. Unfortunately, it was not possible to generalize the formulas in a way to consider the whole tangential manifold $\mathcal{T}_W\mathcal{M}$ because of its complicated structure. For this reason, we just search for a *descent direction*. Let us define it as

$$\langle G(W), X\rangle_{\mathcal{F}} \stackrel{!}{=} \mathrm{d}\,\Omega(W)[X].$$

From (7.3.11),

$$\langle G(W), X\rangle_{\mathcal{F}} = -2\,\langle \mathcal{A}[T], X\rangle_{\mathcal{F}}\,,$$

what immediately implies

$$G = -2\mathcal{A}[T], \tag{7.4.1}$$

$$G^{\mathrm{R}'} = \left(T^{\mathrm{R}'}\right)^{\mathrm{H}} - T^{\mathrm{R}'}. \tag{7.4.2}$$

Clearly, $G^{\mathrm{R}'}$ is skew-hermitian. If we take $X^{\mathrm{R}} = -\alpha G^{\mathrm{R}}$ with some $\alpha \geq 0$, this guarantees $d\Omega(W)[X] \leq 0$:

$$\mathrm{d}\,\Omega(W)[X] = \langle G, -\alpha G\rangle_{\mathcal{F}} = -\alpha \sum_{\mathrm{R}} \mathrm{tr}\left([G^{\mathrm{R}}]^{\mathrm{H}} G^{\mathrm{R}}\right) = -\alpha \sum_{\mathrm{R}} \|G^{\mathrm{R}}\|_F^2 \leq 0.$$

Here $\|.\|_F$ is the Frobenius matrix norm. So, if we transform W using such X, we get

$$W^{\mathrm{R}} \mapsto W^{\mathrm{R}} \mathrm{e}^{-\alpha G^{\mathrm{R}}(W)},$$

Ω at least will not increase.

7.5 A weak point in the algorithm

Unfortunately, the numerical computations show that the algorithm described above does not always work well. The main problem is sensitivity of the whole construction to the initial values of $w^0_{n,0}$ and W. To explain this, we recall the formula (7.4.1) for the gradient:

$$\begin{aligned} G^{\mathrm{R}'} &= \left(T^{\mathrm{R}'}\right)^{\mathrm{H}} - T^{\mathrm{R}'} \qquad \text{with} \\ T^{\mathrm{R}'} &= \sum_{\mathrm{R}} \left(P_2^{\mathrm{R},\mathrm{R}'} - 2D \cdot P_1^{\mathrm{R},\mathrm{R}'}\right). \end{aligned}$$

This formula has parts of sort $P_j^{\mathrm{R},\mathrm{R}'} - \left[P_j^{\mathrm{R},\mathrm{R}'}\right]^{\mathrm{H}}$, $j = 1, 2$. What difficulties can they provide? Recall how we update $P_j^{\mathrm{R},\mathrm{R}'}$ by (7.2.1):

$$P_j^{\mathrm{R},\mathrm{R}'} = [W^{\mathrm{R}}]^{\mathrm{H}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'} \text{ for } j = 1, 2,$$

where

$$\left[P_j^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} = \left\langle \mathrm{r}^j w^0_{m,\mathrm{R}}, w^0_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}^d,\varepsilon}.$$

From here we see that if the initial Wannier functions $\left\{w^0_{n,\mathrm{R}}\right\}_{n,\mathrm{R}}$ are real, then the values of $\left\{[P_j^{0;\mathrm{R},\mathrm{R}'}]_{mn}\right\}_{m,n,\mathrm{R},\mathrm{R}'}$ will also be purely real, which means $P_j^{0;\mathrm{R},\mathrm{R}'} = P_j^{0;\mathrm{R}',\mathrm{R}}$. Furthermore, if the matrices $\left\{W^{\mathrm{R}}\right\}_{\mathrm{R}}$ are self-adjoint, this brings us to the following situation:

$$\begin{aligned} P_j^{\mathrm{R},\mathrm{R}'} &= W^{\mathrm{R}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'}, \\ \left[P_j^{\mathrm{R},\mathrm{R}'}\right]^{\mathrm{H}} &= W^{\mathrm{R}'} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}}. \end{aligned}$$

Clearly, for simple isolated bands (where all these matrices are 1×1) the difference $P_j^{\mathrm{R},\mathrm{R}'} - \left[P_j^{\mathrm{R},\mathrm{R}'}\right]^{\mathrm{H}}$ will be zero and as a consequence *the gradient will also be zero.*
An artificial grouping of simple bands into "composites" does not help to cure it since W^{R} is diagonal in this case. Quite generally, the same can happen to composite bands also.

To develop this concept further, we have to change the choice of the descent direction. After (7.3.9), the derivative of Ω looks like

$$\mathrm{d}\,\Omega(W)[X] = 2\sum_{\mathrm{R}} \mathrm{Re}\left(\mathrm{tr}\left(T^{\mathrm{R}}X^{\mathrm{R}}\right)\right).$$

Let us try to find such values of X which make the derivative non-positive. At the moment we forget about the requirement of its skew-hermitivity. For instance, if $X^{\mathrm{R}} = -\alpha[T^{\mathrm{R}}]^{\mathrm{H}}$ with some $\alpha \geq 0$, we have

$$\mathrm{d}\,\Omega(W)[X] = -2\alpha\sum_{\mathrm{R}} \mathrm{Re}\left(\mathrm{tr}\left(T^{\mathrm{R}}[T^{\mathrm{R}}]^{\mathrm{H}}\right)\right) = -2\alpha\sum_{\mathrm{R}} \mathrm{Re}\left(\|T^{\mathrm{R}}\|_F^2\right) \leq 0. \qquad (7.5.1)$$

Unfortunately, on this way we have another weak place: the sum-unitarity of W is not guaranteed. Therefore, we need a trick. Let

$$\tilde{W}^{\mathrm{R}} = W^{\mathrm{R}}\mathrm{e}^{-\alpha[T^{\mathrm{R}}]^{\mathrm{H}}}.$$

With this transform the old W has been shifted from the manifold $\mathcal{M}$. We want to shift $\tilde{W}$ "back" to $\mathcal{M}$ by

$$\tilde{W}^{\mathrm{R}} \mapsto \tilde{W}^{\mathrm{R}}(\mathrm{Id} - \delta)$$

with some small δ. By definition of $\mathcal{M}$, it means that

$$(\mathrm{Id} - \delta)^{\mathrm{H}}\left[\sum_{\mathrm{R}}(\tilde{W}^{\mathrm{R}})^{\mathrm{H}}\tilde{W}^{\mathrm{R}}\right](\mathrm{Id} - \delta) \stackrel{!}{=} \mathrm{Id}.$$

Let

$$A = \sum_{\mathrm{R}}(\tilde{W}^{\mathrm{R}})^{\mathrm{H}}\tilde{W}^{\mathrm{R}}.$$

We look for such δ that for known A

$$(\mathrm{Id}-\delta)^{\mathrm{H}}A(\mathrm{Id}-\delta)=\mathrm{Id},$$

or equivalently

$$A=\left((\mathrm{Id}-\delta)^{-1}\right)^{\mathrm{H}}(\mathrm{Id}-\delta)^{-1}.$$

As soon as A is a hermitian matrix, it permits Cholesky decomposition:

$$A=B^{\mathrm{H}}B$$

with a lower triangular matrix B^{H}. Therefore, we can suppose:

$$\begin{aligned}(\mathrm{Id}-\delta)^{-1} &= B,\\ \delta &= \mathrm{Id}-B^{-1},\end{aligned}$$

and, by composition,

$$\tilde{W}^{\mathrm{R}}(\mathrm{Id}-\delta)\in\mathcal{M}.$$

Rewrite the transform of W in its final form:

$$\begin{aligned}\tilde{W}^{\mathrm{R}} &= W^{\mathrm{R}}\mathrm{e}^{-\alpha[T^{\mathrm{R}}]^{\mathrm{H}}},\\ W^{\mathrm{R}} &\mapsto \tilde{W}^{\mathrm{R}}B^{-1},\end{aligned}$$

where B is a matrix of Cholesky decomposition of $\sum_{\mathrm{R}}\left(\tilde{W}^{\mathrm{R}}\right)^{\mathrm{H}}\tilde{W}^{\mathrm{R}}$.

As far as we apply a shift, or a "projection" of a computed value $\tilde{W}$ to the manifold $\mathcal{M}$, we want to be sure that the new choice is not worse than the one we started from. Namely, we want to know that $\Omega\left(\tilde{W}(\mathrm{Id}-\delta)\right)\leq\Omega(W)$. For this reason we must be careful with $\delta=\delta(\alpha)$ to avoid that the functional increases too much. The non-negativity of the parameter α guarantees that Ω does decay, see (7.5.1). It means that the value of Ω in point $\tilde{W}(\mathrm{Id}-\delta)\in\mathcal{M}$ must be very close to that of the point $\tilde{W}\notin\mathcal{M}$

$$\Omega(W)\quad\mapsto\quad\Omega(\tilde{W})\leq\Omega(W)\quad\mapsto\quad\Omega(\tilde{W}(\mathrm{Id}-\delta))\approx\Omega(\tilde{W}).$$

7.6 On computation of $P_1^{0;\mathrm{R},\mathrm{R}'}$ and $P_2^{0;\mathrm{R},\mathrm{R}'}$

In numerical implementation it is important, how the initial values for $P_1^{0;\mathrm{R},\mathrm{R}'}$ and $P_2^{0;\mathrm{R},\mathrm{R}'}$ are computed. Let us rewrite their constituents as

$$\begin{aligned} w_{n,\mathrm{R}'}(\mathrm{r}) &= w_{n,0}(\mathrm{r}-\mathrm{R}'), \\ w_{m,\mathrm{R}}(\mathrm{r}) &= w_{m,0}(\mathrm{r}-\mathrm{R}) = w_{m,0}((\mathrm{r}-\mathrm{R}')-(\mathrm{R}-\mathrm{R}')) = w_{m,\mathrm{R}-\mathrm{R}'}(\mathrm{r}-\mathrm{R}'). \end{aligned}$$

Consequently,

$$\begin{aligned} \left[P_1^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} &= \left\langle \mathrm{r}w_{m,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \left\langle (\mathrm{r}-\mathrm{R}'+\mathrm{R}')w_{m,\mathrm{R}-\mathrm{R}'}(\cdot-\mathrm{R}'), w_{n,0}(\cdot-\mathrm{R}')\right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \left\langle \mathrm{r}w_{m,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon} + \mathrm{R}'\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'} \end{aligned}$$

and

$$\begin{aligned} \left[P_2^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} &= \left\langle r^2 w_{m,\mathrm{R}}, w_{n,\mathrm{R}'}\right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \left\langle (\mathrm{r}-\mathrm{R}'+\mathrm{R}')^2 w_{m,\mathrm{R}-\mathrm{R}'}(\cdot-\mathrm{R}'), w_{n,0}(\cdot-\mathrm{R}')\right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \left\langle \left((\mathrm{r}-\mathrm{R}')^2 + 2\mathrm{R}'\cdot(\mathrm{r}-\mathrm{R}') + |\mathrm{R}'|^2\right) w_{m,\mathrm{R}-\mathrm{R}'}(\cdot-\mathrm{R}'), w_{n,0}(\cdot-\mathrm{R}')\right\rangle_{\mathbb{R}^d,\varepsilon} \\ &= \left\langle r^2 w_{m,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon} + 2\mathrm{R}'\cdot\left\langle \mathrm{r}w_{m,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon} + |\mathrm{R}'|^2\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}. \end{aligned}$$

We simplify the notations by introducing

$$\begin{aligned} \left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} &= \left\langle \mathrm{r}w_{m,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon}, \\ \left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn} &= \left\langle r^2 w_{m,\mathrm{R}-\mathrm{R}'}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon}. \end{aligned}$$

Note that $\left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ is a d-dimensional vector and $\left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ is a scalar. We can rewrite the formulas:

$$\left[P_1^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} = \left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} + \mathrm{R}\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}, \tag{7.6.1}$$

$$\left[P_2^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} = \left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn} + 2\mathrm{R}'\cdot\left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} + |\mathrm{R}|^2\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}. \tag{7.6.2}$$

With help of Theorem 5.1 we can now calculate $\left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ and $\left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn}$ by integration over the Brillouin zone which is bounded. For further details how to do it numerically see Appendix, Section 10.4. The other values in the computations described in this chapter can be formulated in terms of $P_1^{0;\mathrm{R},\mathrm{R}'}$ and $P_2^{0;\mathrm{R},\mathrm{R}'}$.

Remark 7.2. *As far as we sum up by* R, R′*, in numerical simulations the number of summands is always assumed to be finite. However, we must keep sum-unitarity of* W*. Namely, as far as we fix a quantity of* $\left\{W^{\mathrm{R}}\right\}_{\mathrm{R}}$ *which will be involved into computations (e.g. for* $\mathrm{R} = N_1, \ldots, N_2$*), we "normalize" them to get identity in the sum:*

$$A = \sum_{\mathrm{R}=N_1}^{N_2} \left(W^{\mathrm{R}}\right)^{\mathrm{H}} W^{\mathrm{R}},$$
$$\widetilde{W}^{\mathrm{R}} = W^{\mathrm{R}} A^{-1/2},$$
$$\sum_{\mathrm{R}=N_1}^{N_2} \left(\widetilde{W}^{\mathrm{R}}\right)^{\mathrm{H}} \widetilde{W}^{\mathrm{R}} = A^{-1/2} \left(\sum_{\mathrm{R}=N_1}^{N_2} \left(W^{\mathrm{R}}\right)^{\mathrm{H}} W^{\mathrm{R}} \right) A^{-1/2} = A^{-1/2} A A^{-1/2} = \mathrm{Id}.$$

Then the rest is allowed to be set to zero:

$$\sum_{\mathrm{R}=-\infty}^{\infty} \left(W^{\mathrm{R}}\right)^{\mathrm{H}} W^{\mathrm{R}} = 0 + \sum_{\mathrm{R}=N_1}^{N_2} \left(W^{\mathrm{R}}\right)^{\mathrm{H}} W^{\mathrm{R}} + 0 = \mathrm{Id}.$$

7.7 Starting guess for W

The Marzari-Vanderbilt algorithm has been started with $\mathrm{U}^{\mathrm{k}} = \mathrm{Id}$, for which an analogue would be $W^{\mathrm{R}} = \delta_{\mathrm{R},0}$. But now this ansatz does not work since it would not make the process start. Another idea is to start with Gaussian bell functions. Let us take it up in detail.

Let g_n be a random Gaussian bell centered at 0. Here n corresponds to a band number of the corresponding Wannier function $w_{n,0}^0$ which we want to replace with this bell, having first modified it to keep orthogonality of the basis. The bell have a form

$$g_n(\mathrm{r}) = a_n \mathrm{e}^{-(\mathrm{r}-b_n)^2/(2c_n^2)},$$

where a_n is a peak height, b_n is the location of the peak and c_n determines the full width of the bell at half maximum as $2\sqrt{2\ln 2}c_n$. All parameters are positive and random in range $[\delta, 1-\delta]$, e.g. $[0.1, 0.9]$.

We project the functions $\{g_n\}_n$ onto the set of $\left\{w^0_{n,\mathrm{R}}\right\}_{n,\mathrm{R}}$ by

$$\phi_n(\mathrm{r}) = \sum_m \sum_{\mathrm{R}} \gamma^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}(\mathrm{r}) \quad \text{with} \quad \gamma^{\mathrm{R}}_{mn} = \left\langle w^0_{m,\mathrm{R}}, g_n \right\rangle .$$

Apply orthonormalization:

$$\varphi_n(\mathrm{r}) \;=\; \sum_m \left[S^{-1/2}\right]_{mn} \phi_m(\mathrm{r}), \tag{7.7.1}$$

where

$$\begin{aligned}
S_{mn} &= \langle \phi_m, \phi_n \rangle \\
&= \left\langle \sum_{m'} \sum_{\mathrm{R}} \gamma^{\mathrm{R}}_{m'm} w^0_{m',\mathrm{R}}, \sum_{n'} \sum_{\mathrm{R}'} \gamma^{\mathrm{R}'}_{n'n} w^0_{n',\mathrm{R}'} \right\rangle \\
&= \sum_{m',n'} \sum_{\mathrm{R},\mathrm{R}'} \left(\gamma^{\mathrm{R}}_{m'm}\right)^* \gamma^{\mathrm{R}'}_{n'n} \left\langle w^0_{m',\mathrm{R}}, w^0_{n',\mathrm{R}'} \right\rangle \\
&= \sum_{m'} \sum_{\mathrm{R}} \left(\gamma^{\mathrm{R}}\right)^{\mathrm{H}}_{mm'} \gamma^{\mathrm{R}}_{m'n} \\
&= \sum_{\mathrm{R}} \left[\left(\gamma^{\mathrm{R}}\right)^{\mathrm{H}} \gamma^{\mathrm{R}} \right]_{mn} .
\end{aligned}$$

We take the function φ_n (7.7.1) as a starting value for $w_{n,0}$. In other words, the transform of the initial $w^0_{n,\mathrm{R}}$ such that $\varphi_n = \sum_m \sum_{\mathrm{R}} W^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}$ will be:

$$W^{\mathrm{R}}_{mn} \;=\; \sum_l \gamma^{\mathrm{R}}_{ml} \left[S^{-1/2}\right]_{ln}, \tag{7.7.2}$$

$$W^{\mathrm{R}} \;=\; \gamma^{\mathrm{R}} S^{-1/2}. \tag{7.7.3}$$

Check sum-unitarity:

$$\sum_{\mathrm{R}} (W^{\mathrm{R}})^{\mathrm{H}} W^{\mathrm{R}} = \sum_{\mathrm{R}} (S^{-1/2})^{\mathrm{H}} (\gamma^{\mathrm{R}})^{\mathrm{H}} \gamma^{\mathrm{R}} S^{-1/2} = (S^{-1/2})^{\mathrm{H}} S S^{-1/2} = \mathrm{Id}.$$

7.8 Minimization algorithm

The subroutines

(i) Preparational computations

Compute:

pr1) $u_{n,\mathrm{k}}(\mathrm{r})$ in *Wigner-Seitz cell*

pr2) $\left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} = \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{ik}\cdot(\mathrm{R}-\mathrm{R}')} \langle u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \,\mathrm{d}\,\mathrm{k}$

pr3) $\left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn} = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{ik}\cdot(\mathrm{R}-\mathrm{R}')} \langle \nabla_{\mathrm{k}} u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} \,\mathrm{d}\,\mathrm{k}$

pr4) $\left[P_1^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} = \left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} + \mathrm{R}\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}$

pr5) $\left[P_2^{0;\mathrm{R},\mathrm{R}'}\right]_{mn} = \left[B_2^{\mathrm{R},\mathrm{R}'}\right]_{mn} + 2\mathrm{R}' \cdot \left[B_1^{\mathrm{R},\mathrm{R}'}\right]_{mn} + |\mathrm{R}|^2\delta_{m,n}\delta_{\mathrm{R},\mathrm{R}'}$

pr6) initial Wannier functions $w^0_{n,\mathrm{R}}$

pr7) initial values for W^{R}

(ii) Update P_1, P_2

Input: P_1^0, P_2^0, W.

p1) $P_j^{\mathrm{R},\mathrm{R}'} = (W^{\mathrm{R}})^{\mathrm{H}} P_j^{0;\mathrm{R},\mathrm{R}'} W^{\mathrm{R}'}$, $j = 1, 2$

(iii) The spread Ω

Input: W.

s1) update P_1 and P_2 as in **(ii)**

s2) $\Omega = \sum_{\mathrm{R},\mathrm{R}'} \mathrm{tr}(P_2^{\mathrm{R},\mathrm{R}'}) - \sum_n \left| \sum_{\mathrm{R},\mathrm{R}'} [P_1^{\mathrm{R},\mathrm{R}'}]_{nn} \right|^2$

(iv) descent direction T of the spread

Input: W.

t1) update P_1 and P_2 as in **(ii)**

t2) $\bar{\mathrm{r}}_n = \sum_{\mathrm{R},\mathrm{R}'} \left[P_1^{\mathrm{R},\mathrm{R}'}\right]_{nn}$

t3) $D = \mathrm{Diag}(\mathrm{Re}(\bar{\mathrm{r}}_n))$

t4) $T^{\mathrm{R}'} = \sum_{\mathrm{R}} \left(P_2^{\mathrm{R},\mathrm{R}'} - 2D \cdot P_1^{\mathrm{R},\mathrm{R}'}\right)$

(v) Improved descent direction G of the spread

Input: T.

g1) $G^{\mathrm{R}} = \left(T^{\mathrm{R}}\right)^{\mathrm{H}} - T^{\mathrm{R}}$

(vi) Update W (first version)

Input: α, G, W_{old}.

w1) $W_{\mathrm{new}}^{\mathrm{R}} = W_{\mathrm{old}}^{\mathrm{R}} \mathrm{e}^{-\alpha G^{\mathrm{R}}}$

(vii) Update W (second version)

Input: α, T, W_{old}.

w2.1) $\tilde{W}^{\mathrm{R}} = W_{\mathrm{old}}^{\mathrm{R}} \mathrm{e}^{-\alpha T^{\mathrm{R}}}$

w2.2) $A = \sum_{\mathrm{R}} \left(\tilde{W}^{\mathrm{R}}\right)^{\mathrm{H}} \tilde{W}^{\mathrm{R}}$

w2.3) $B = \mathrm{chol}(A)$ (Cholesky decomposition)

w2.4) $W_{\mathrm{new}}^{\mathrm{R}} = \tilde{W}^{\mathrm{R}} B^{-1}$

Minimization procedure

Input: W, $\alpha_{\max}$, a small value δ (tolerance), m_{iter}.

m1) for known W, compute T as in **(iv)**

m2) for T, compute G as in **(v)**

m3) if $G \neq 0$, define function

$$\tilde{\Omega}(\alpha) = \Omega(W(\alpha, G, W))$$

with W updated by the first version **(vi)** and let $c = 1$;
else define

$$\tilde{\Omega}(\alpha) = \Omega(W(\alpha, T, W))$$

with W updated by the second version **(vii)** and let $c = 2$;
in both cases compute $\Omega(W)$ as in **(iii)**

m4) find $\alpha_0 \in [0, \alpha_{\max}]$ such that $\tilde{\Omega}(\alpha_0) = \min_{\alpha_0 \in [0,\alpha_{\max}]} \tilde{\Omega}(\alpha)$

m5) if $c = 1$, update $W = W(\alpha_0, G, W)$ as in **(vi)**;
if $c = 2$, update $W = W(\alpha_0, T, W)$ as in **(vii)**

Loop m1-m5 till Ω cannot reduce for more than δ_{tol} for 3 consecutive steps, or till $\alpha_0 = 0$, or till the number of iterations achieves the prescribed maximum m_{iter}.

Localized Wannier functions

Input: $w^0_{m,\mathrm{R}}$, W.
For the obtained values of W^{R}, compute the localized Wannier functions:

$$w_{n,0}(\mathrm{r}) = \sum_m \sum_{\mathrm{R}} W^{\mathrm{R}}_{mn} w^0_{m,\mathrm{R}}(\mathrm{r}).$$

7.9 Computational results

1D case

In 1D all bands are isolated and thus can be considered separately. Let us take first four bands.

We take the following input data:

1. The Photonic Crystal: in the Wigner-Seitz cell $[0, 1]$

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, 0.41) \cup (0.59, 1], \\ 11.56, & x \in [0.41, 0.59]; \end{cases}$$

2. Mesh in x-space: $N_x = 51$, $h_x = 0.02$; R $\in \{-5, \ldots, 5\}$;

3. Monkhorst-Pack mesh in k-space as in Section 5.2: $N_{\mathrm{k}} = 100$ $h_{\mathrm{k}} = 2\pi/100$;

4. Stencil (5.2.1), weights $a_1 = a_2 = 1/(2h_{\mathrm{k}}^2)$;

5. Linear Finite Elements to compute $u^0_{n,\mathrm{k}}$ (for details see Appendix, Section 10.2.1);

6. We implemented the first 4 bands ($n = 1, 2, 3, 4$).

n	Ω_0	$\Omega_{\min}$	iterations	time
1	7.527	5.659	3	5 sec
2	0.473	0.147	19	1 min 32 sec
3	6.684	1.426	3	7 sec
4	2.711	1.166	2	5 sec

The corresponding Wannier functions are illustrated on the Figures 7.1. As we see, they do not possess symmetry properties as the maximally localized functions (according to Theorem 3.2 of Kohn). The minimum values of Ω are greater than those of the functions obtained in Chapter 5, but comparable with them and much better as in Chapter 6. The most significant properties are obtained: these functions are well-localized and real.

2D case

In 2D we consider the following Photonic Crystal: In the Wigner-Seitz cell $[0,1]$

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 > 0.18^2, \\ 11.56, & (r_1 - \frac{1}{2})^2 + (r_2 - \frac{1}{2})^2 \leq 0.18^2. \end{cases}$$

The band structure corresponding to this crystal is illustrated in Figure 2.2. The first band is isolated (simple), the following tree bands a gathered together in a composite. Therefore, we consider them in groups: 1) $n = 1$; 2) $n = 2, 3, 4$.
We take the following input data:

1. Mesh in r-space: $N_x = 21$, $h_x = 0.05$; $\mathrm{R} \in \{(-2,-2), \ldots, (2,2)\}$;
2. Monkhorst-Pack mesh in k-space as in Section 5.2 (see Figure 5.1): $N_{\mathrm{k}} = 40$ $h_{\mathrm{k}} = 2\pi/40$;
3. Stencil (5.2.2), weights $a_1 = a_2 = a_3 = a_4 = 1/(2h_{\mathrm{k}}^2)$;
4. Linear Finite Elements to compute $u^0_{n,\mathrm{k}}$ (for details see Appendix, Section 10.2.2);
5. We implemented the first 4 bands ($n = 1, 2, 3, 4$).

n	Ω_0	$\Omega_{\min}$	iterations	time
1	6.864	0.097	3	3 sec
2-4	24.179	11.701	3	2 sec

The corresponding Wannier functions are illustrated on the Figures 7.2.

7.10 Conclusions

As a generalization of the approach presented in Chapter 6, this method provides better localization. This conclusion was expected at the very beginning since the manifold of

sum-unitary transformations (7.1.4) in which we search for a minimizer in this chapter is wider than that of unitary transformations (6.1.1).

However, we have a problem here — the tangential manifold could not be determined, for this reason the presented minimization path is not optimal. Another descent direction is an open question which can be studied in future.

Figure 7.1: Wannier functions in 1D, bands 1-4: Non-localized (left) and localized under sum-unitary transform minimization (right)

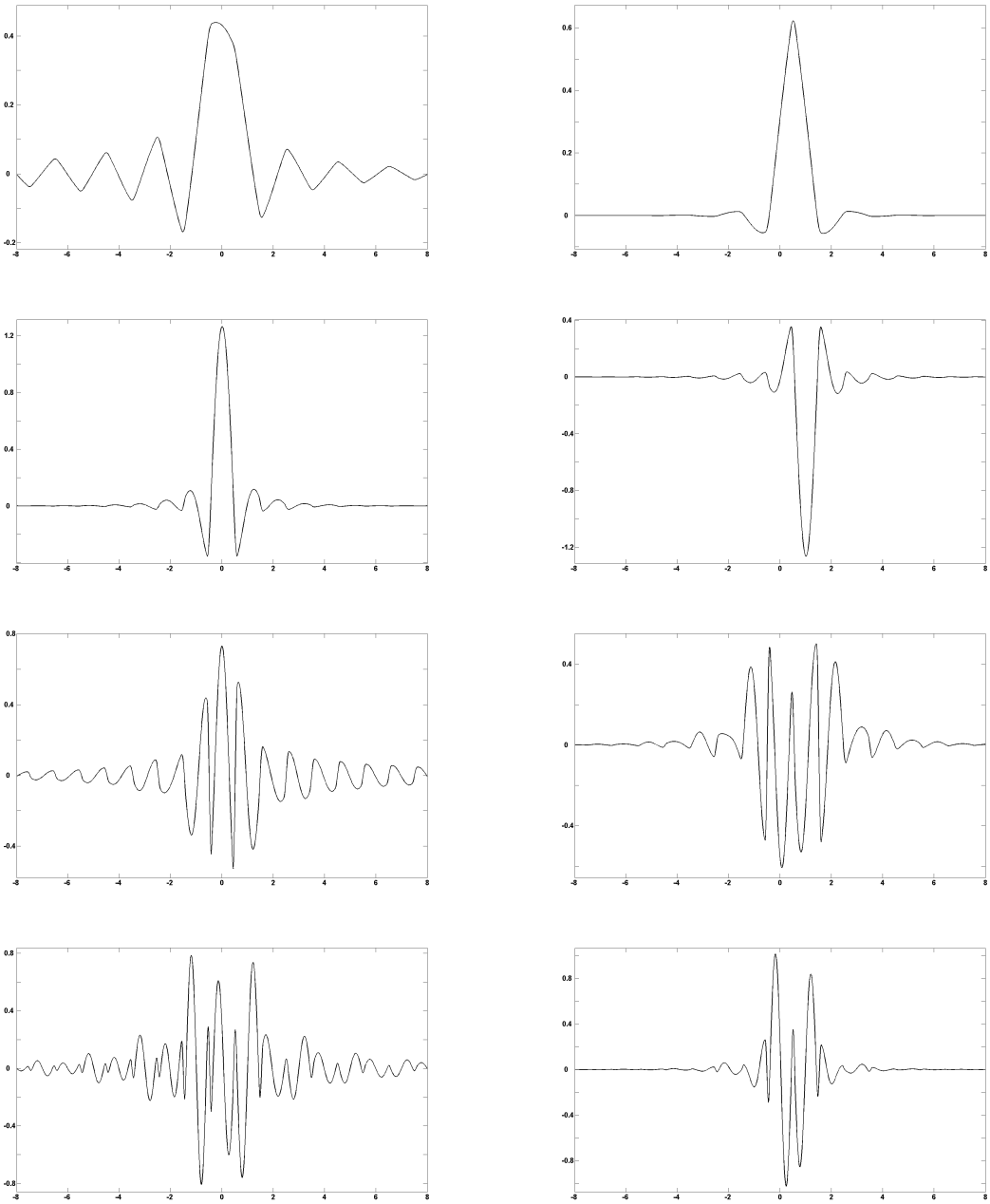
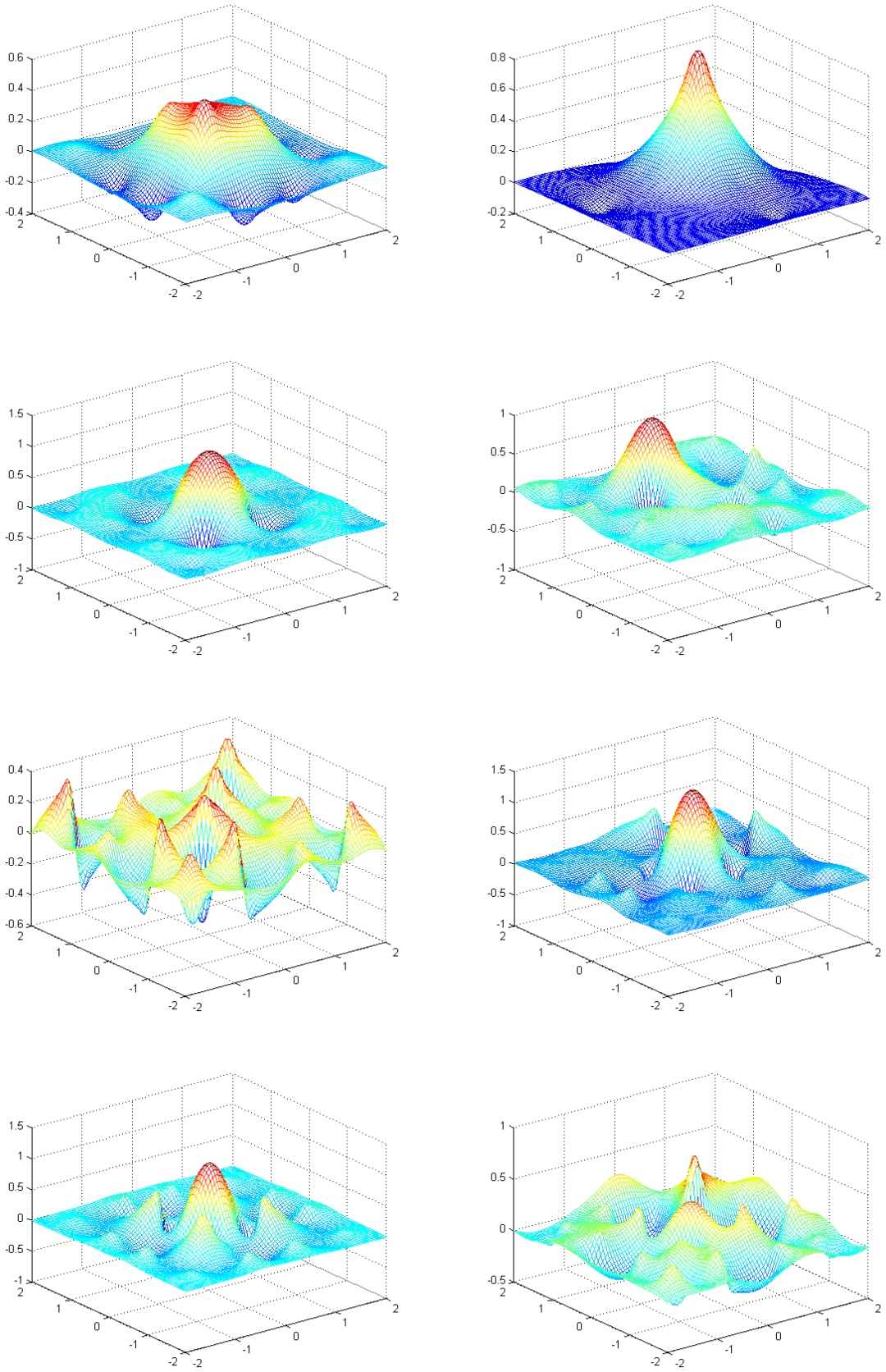


Figure 7.2: Wannier functions in 2D, bands 1-4: Non-localized (left) and localized under sum-unitary transform minimization (right)



Chapter 8

Application of the Wannier functions

The route of light propagation in a Photonic Crystal can be specified by *defects*, properly embedded into a periodic structure [BMGM$^+$03], [Bus03]. An example of such a defect can be an optical *micro-cavity* caused by a modified single rod — this can form a light mode localized inside the PC. A chain of such sense defects can provide a staggering effect — it works as a *waveguide channel* and can form sharp waveguide bends.

Figure 8.1: Single defect rod in a PC

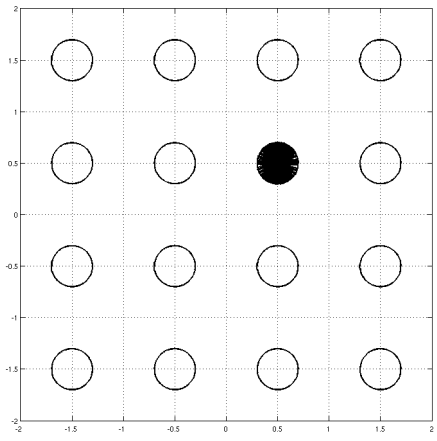
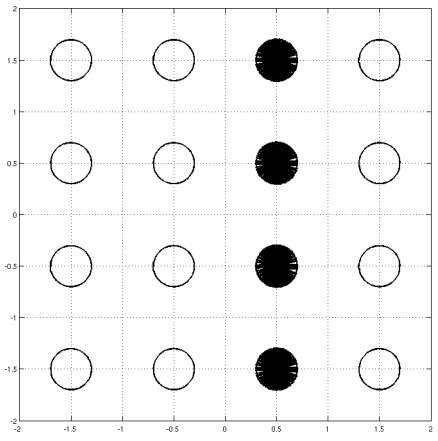


Figure 8.2: Row of defect rods in a PC



Combining these types of basic defects, one can design various optical properties and rule over the propagation of electromagnetic waves in a Photonic Crystal in a way similar to the control over the motion of electrons in electronic circuits. These efforts provide a possibility to create compact all-optical integrated circuits.
In this chapter we will consider an approach to modeling of defect modes using the Wannier function basis.

8.1 Modeling of a defect: Galerkin approach

Let us consider a Photonic Crystal as a dielectric medium with certain electric permittivity ε_1 with periodically located rods of another permittivity ε_2. As a *defect structure embedded into a periodic PC* we define an amount of rods which have permittivity ε_d different from ε_2. Such a crystal with defect can be described, in accordance to (2.1.5), by the following formula [BMGM$^+$03], [SMHB05]:

$$-\Delta E(\mathrm{r}) = \Lambda\left[\varepsilon_p(\mathrm{r}) + \delta\varepsilon(\mathrm{r})\right] E(\mathrm{r}), \tag{8.1.1}$$

where E is the electric field, $\Lambda = (\omega/c)^2$ with c being the speed of light, ε_p is a periodic electric permittivity and $\delta\varepsilon$ is the contributional permittivity of the embedded defect.
Let us expand the electric field E into some complete orthogonal set of functions sufficiently smooth to approximate (8.1.1):

$$E(\mathrm{r}) = \sum_j E_j \varphi_j(\mathrm{r}).$$

Inserting this expression into (8.1.1) gives:

$$\begin{aligned} -\sum_j E_j \Delta\varphi_j(\mathrm{r}) &= \Lambda\left[\varepsilon_p(\mathrm{r}) + \delta\varepsilon(\mathrm{r})\right] \sum_j E_j \varphi_j(\mathrm{r}), \\ -\sum_j E_j \varphi_i^*(\mathrm{r}) \Delta\varphi_j(\mathrm{r}) &= \lambda \sum_j E_j \left[\varphi_i^*(\mathrm{r})\varepsilon_p(\mathrm{r})\varphi_j(\mathrm{r}) + \varphi_i^*(\mathrm{r})\delta\varepsilon(\mathrm{r})\varphi_j(\mathrm{r})\right]. \end{aligned}$$

Integrating both parts by $\mathbb{R}^d$ gives:

$$-\sum_j E_j \int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\Delta\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r} = \Lambda\sum_j E_j \left[\int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\varepsilon_p(\mathrm{r})\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r}\right. \tag{8.1.2}$$

$$\left. + \int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\delta\varepsilon(\mathrm{r})\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r}\right]. \tag{8.1.3}$$

Introduce new notations:

$$\begin{aligned} A_{i,j} &= -\int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\Delta\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r}, \\ C_{i,j} &= \int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\varepsilon(\mathrm{r})\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r}, \\ D_{i,j} &= \int_{\mathbb{R}^d} \varphi_i^*(\mathrm{r})\delta\varepsilon(\mathrm{r})\varphi_j(\mathrm{r})\,\mathrm{d}\,\mathrm{r}. \end{aligned}$$

Therefore, (8.1.2) can be rewritten in matrix form as

$$\sum_j A_{i,j}E_j = \Lambda\sum_j \left(C_{i,j} + D_{i,j}\right) E_j. \tag{8.1.4}$$

This technique of substituting of the partial differential equation (8.1.1) by the linear system (8.1.4) is well known as the *Galerkin approach*. And the quality of this approach depends crucially on a proper choice of the basis set $\{\varphi_j\}_j$.

The Wannier functions $\{w_{n,\mathrm{R}}\}_{n,\mathrm{R}}$, computed for an analogous but perfectly periodic Photonic Crystal ($\delta\varepsilon = 0$), are $\mathrm{L}^2_\varepsilon(\mathbb{R}^d)$-orthogonal and, in contrast to other possible bases, they *already contain* the information about the underlying crystal structure. Moreover, they can be chosen real-valued, well-localized and possessing certain symmetries described in Chapter 3. Apparently, the basis of Wannier functions is an optimal choice for our problem. A confirmation of this statement one can find in [SMHB05], where the authors compare several different bases and make the same conclusion.

Let us insert the Wannier functions instead of φ_j; here $i = \{n, \mathrm{R}\}$, $j = \{n', \mathrm{R}'\}$:

$$A^{\mathrm{R},\mathrm{R}'}_{nn'} = -\int_{\mathbb{R}^d} w^*_{n,\mathrm{R}}(\mathrm{r})\Delta w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r}, \tag{8.1.5}$$

$$C^{\mathrm{R},\mathrm{R}'}_{nn'} = \int_{\mathbb{R}^d} w^*_{n,\mathrm{R}}(\mathrm{r})\varepsilon_p(\mathrm{r})w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} = \delta_{n,n'}\delta_{\mathrm{R},\mathrm{R}'}, \tag{8.1.6}$$

$$D^{\mathrm{R},\mathrm{R}'}_{nn'} = \int_{\mathbb{R}^d} w^*_{n,\mathrm{R}}(\mathrm{r})\delta\varepsilon(\mathrm{r})w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r}. \tag{8.1.7}$$

Now we want to simplify the matrices. Recall the definition (3.1.2) of the Wannier functions via Bloch waves:

$$\begin{aligned} w_{n,\mathrm{R}}(\mathrm{r}) &= \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\psi_{n,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k} \\ &= \frac{1}{V_{\mathrm{BZ}}}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\sum_m \mathrm{U}^{\mathrm{k}}_{mn}\psi^0_{m,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}. \end{aligned}$$

We remember that $\psi^0_{m,\mathrm{k}}$ is an eigenfunction of Transverse Magnetic problem (2.1.5):

$$-\Delta\psi^0_{m,\mathrm{k}}(\mathrm{r}) = \lambda_{m,\mathrm{k}}\varepsilon_p(\mathrm{r})\psi^0_{m,\mathrm{k}}(\mathrm{r}).$$

Therefore,

$$\begin{aligned} -\Delta w_{n',\mathrm{R}'}(\mathrm{r}) &= -\frac{1}{(2\pi)^d}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'}\Delta\psi_{n',k'}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}' \\ &= -\frac{1}{(2\pi)^d}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'}\sum_{m'} \mathrm{U}^{\mathrm{k}}_{m'n'}\Delta\psi^0_{m',k'}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}' \\ &= \frac{\varepsilon_p(\mathrm{r})}{(2\pi)^d}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'}\sum_{m'} \mathrm{U}^{\mathrm{k}}_{m'n'}\lambda_{m',\mathrm{k}'}\psi^0_{m',\mathrm{k}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}'. \end{aligned}$$

This implies:

$$\begin{aligned} A^{\mathrm{R},\mathrm{R}'}_{nn'} &= -\int_{\mathbb{R}^d} w^*_{n,\mathrm{R}}(\mathrm{r})\Delta w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} \\ &= \int_{\mathbb{R}^d}\frac{1}{(2\pi)^d}\int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}}\sum_m [\mathrm{U}^{\mathrm{k}}]^{\mathrm{H}}_{nm}\left(\psi^0_{m,\mathrm{k}}(\mathrm{r})\right)^*\mathrm{d}\,\mathrm{k} \\ &\quad\times\frac{\varepsilon_p(\mathrm{r})}{(2\pi)^d}\int_{\mathrm{BZ}} \mathrm{e}^{-\mathrm{i}\mathrm{k}'\cdot\mathrm{R}'}\sum_{m'} \mathrm{U}^{\mathrm{k}'}_{m'n'}\lambda_{m',\mathrm{k}'}\psi^0_{m',\mathrm{k}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}'\,\mathrm{d}\,\mathrm{r} \\ &= \frac{1}{(2\pi)^{2d}}\int_{\mathrm{BZ}}\int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}(\mathrm{k}\cdot\mathrm{R}-\mathrm{k}'\cdot\mathrm{R}')}\sum_{m,m'}[\mathrm{U}^{\mathrm{k}}]^{\mathrm{H}}_{nm}\lambda_{m',\mathrm{k}'}\mathrm{U}^{\mathrm{k}'}_{m'n'}\times \\ &\quad\times\left[\int_{\mathbb{R}^d}\left(\psi^0_{m,\mathrm{k}}(\mathrm{r})\right)^*\varepsilon_p(\mathrm{r})\psi^0_{m',\mathrm{k}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r}\right]\mathrm{d}\,\mathrm{k}\,\mathrm{d}\,\mathrm{k}'. \end{aligned}$$

By Theorem 2.2, we recall that

$$\int_{\mathbb{R}^d}\left(\psi^0_{m,\mathrm{k}}(\mathrm{r})\right)^*\varepsilon_p(\mathrm{r})\psi^0_{m',\mathrm{k}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} = (2\pi)^d\delta_{mm'}\delta(\mathrm{k}-\mathrm{k}').$$

With this in mind we conclude:

$$A^{\mathrm{R},\mathrm{R}'}_{nn'} = \frac{1}{(2\pi)^d}\int_{\mathrm{BZ}}\sum_m [\mathrm{U}^{\mathrm{k}}]^{\mathrm{H}}_{nm}\lambda_{m,\mathrm{k}}\mathrm{U}^{\mathrm{k}}_{mn'}\mathrm{e}^{\mathrm{i}\mathrm{k}\cdot(\mathrm{R}-\mathrm{R}')}\,\mathrm{d}\,\mathrm{k}. \tag{8.1.8}$$

It is can be shown that this matrix A is positive definite.

As soon as the Wannier functions $w_{n,\mathrm{R}}$ and the defect dielectricity $\delta\varepsilon$ are localized, both matrices A and D are sparse — namely, with increasing distance between the integer-lattice nodes R and R$'$ the values of $A^{\mathrm{R},\mathrm{R}'}$ and $D^{\mathrm{R},\mathrm{R}'}$ vanish. Moreover, since $\{w_{n,\mathrm{R}}\}_{n,\mathrm{R}}$ are real, A and D are real and symmetric.

The equation (8.1.4) can be rewritten as

$$\sum_{n,\mathrm{R}} A^{\mathrm{R},\mathrm{R}'}_{nn'} E^{\mathrm{R}}_{n} \;=\; \Lambda \sum_{n,\mathrm{R}} \left(\delta_{n,n'}\delta_{\mathrm{R},\mathrm{R}'} + D^{\mathrm{R},\mathrm{R}'}_{nn'} \right) E^{\mathrm{R}}_{n}. \tag{8.1.9}$$

We see now very clearly the advantage of the Wannier function basis in defect mode simulations. Instead of the system (8.1.4) with full complex matrices we obtain the system (8.1.9) with sparse real symmetric matrices, one of those even reduced to unity.

Note that the entries of A depend only on the properties of perfectly periodic crystal; the information about defects is contained in D. So, the parameter ε_p and herewith the "initial" periodic PC are excluded from further computations.

8.2 Single defect rod

Let us consider a single embedded rod of electric permittivity ε_d centered at R_d:

$$\delta\varepsilon(\mathrm{r}) = \left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right] \Theta(\mathrm{r} - \mathrm{R}_d)$$

with

$$\Theta(\mathrm{r} - \mathrm{R}_d) = \begin{cases} 1, & \mathrm{r} \text{ inside the rod centered at } \mathrm{R}_d, \\ 0, & \text{otherwise.} \end{cases}$$

Namely, the contribution $\delta\varepsilon$ is zero everywhere except the defect rod where it is equal to $\varepsilon_d - \varepsilon_p$ (see Figure 8.1). Then

$$\begin{aligned} D_{nn'}^{\mathrm{R},\mathrm{R}'} &= \int_{\mathbb{R}^d} w_{n,\mathrm{R}}^*(\mathrm{r})\delta\varepsilon(\mathrm{r})w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} \\ &= \int_{\mathbb{R}^d} w_{n,\mathrm{R}}^*(\mathrm{r})\left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right]\Theta(\mathrm{r}-\mathrm{R}_d)w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} \\ &= \int_{\mathrm{rod}(\mathrm{R}_d)} w_{n,\mathrm{R}}^*(\mathrm{r})\left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right]w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r}. \end{aligned}$$

8.3 Waveguides

A row of defect rods described in the previous section is the most significant type of PC defects. These rows can be constructed in such a way that they form a waveguide which allows to control and guide the propagation of electromagnetic waves with frequencies within a photonic band gap.

A chain of defects is 1D-periodic, therefore its guided modes satisfy the quasi-periodicity condition (Theorem 2.2, Property 2):

$$E(\mathrm{r}+s^w) = \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot s^w}E(\mathrm{r}),$$

where s^w is a *waveguide director* — a d-dimensional vector which specifies the line of defect rods in the Photonic Crystal.

The electric permittivity of the defects is given by

$$\delta\varepsilon(\mathrm{r}) = \sum_{j=-\infty}^{\infty}\left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right]\Theta(\mathrm{r}-\mathrm{R}_j^w),$$

where $\mathrm{R}_j^w = \mathrm{R}_0^w + js^w$ are the positions of the defect rods, R_0^w is the origin of the waveguide. As before, the function Θ defines the area of the defect rods and equals to 1 inside this ares and 0 outside (see Figure 8.2).

Let us compute the matrix D:

$$\begin{aligned} D_{nn'}^{\mathrm{R},\mathrm{R}'} &= \int_{\mathbb{R}^d} w_{n,\mathrm{R}}^*(\mathrm{r})\delta\varepsilon(\mathrm{r})w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} \\ &= \int_{\mathbb{R}^d} w_{n,\mathrm{R}}^*(\mathrm{r}) \sum_{j=-\infty}^{\infty} \left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right] \Theta(\mathrm{r} - \mathrm{R}_j^w) w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r} \\ &= \sum_{j=-\infty}^{\infty} \int_{\mathrm{rod}(\mathrm{R}_j^w)} w_{n,\mathrm{R}}^*(\mathrm{r}) \left[\varepsilon_d(\mathrm{r}) - \varepsilon_p(\mathrm{r})\right] w_{n',\mathrm{R}'}(\mathrm{r})\,\mathrm{d}\,\mathrm{r}. \end{aligned}$$

8.4 The main conclusion

In the framework of this chapter we do not present the numerical computations. Examples of the application of this approach to different types of 2D crystals one can find in physical papers [BMGM$^+$03], [SMHB05]. The most important for us are the conclusions that the authors make: the Galerkin method with the Wannier functions as a basis set allows very efficient calculations which require much less data and time than other techniques such as the Finite Difference Time Domain (FDTD) method or plane-wave supercell approach.

The basis of guided modes of PC waveguides, used instead of plane waves, allows to built a transparent and easy understandable theory of Photonic Crystal circuits. This can be effectively extended to 3D Photonic Crystals as well as to nonlinear and anisotropic materials.

It is straightforward that the Wannier functions possessed of localization and symmetric properties improve the situtation even more by significant decreasing of the data massives required for the computations. For this reason the problem of finding better and better ways to localize the Wannier functions is so asked-for and topical.

Chapter 9

Conclusions and directions for further work

In the framework of this thesis we have explored an important numerical tool in Photonic Crystal simulations — the Wannier functions. We summarized the information about them and unified the notations which differ in various papers denoted to this topic.

We started from the analysis of the Bloch waves as the eigenfunctions of the Transverse Magnetic problem and studied their properties. The non-uniqueness in determination of the Bloch waves allowed us to construct the Wannier functions also non-uniquely and thus to formulate the problem of finding localized Wannier functions.

We reviewed the proofs of existing of the exponentially localized Wannier functions for different types of crystals: we started from a 1D case with inversion symmetry; then turned to a case of a symmetric crystal in arbitrary dimensions; and finally generalized the result to an arbitrary crystal, not necessarily symmetric.

We explored a way to compute the exponentially localized Wannier functions explicitly for a 1D crystal.

We considered the spread functional as a criterion of localization of the Wannier functions and examined its properties. We implemented several different methods of the spread minimization, including the Marzari-Vanderbilt algorithm based on the unitary

transform of the Bloch waves, the Gygi-Fatterbert-Schwegler approach dealing with the unitary transform of the Wannier functions and we also suggested a new method never considered before which minimizes the spread under the sum-unitary transform of the Wannier functions. We researched and explained in detail the features of every method and implemented all of them in Matlab for 1D and 2D Photonic Crystals using the Finite Element method to compute the Bloch waves numerically.

We showed how the Wannier functions can be used in simulation of defect modes in Photonic Crystals and what advantages they have in comparison to other tools. Being well-localized, real and symmetric, they require significantly less computational resources. Moreover, they already contain the information about the periodic structure of a particular Photonic Crystal and perfectly "fit" for modeling of the light propagation problems in it.

At the end we can conclude that a big and fruitful work was done. As in any big research, there remain unfinished problems as well as ways to improve the things. For example, in Chapter 5 we considered two initial guesses for the unitary matrices U^{k}, one of them was taken intuitively as the simplest example of a unitary matrix (identity), the other one was chosen more smart and gave some localization at the very beginning (construction based on the Gaussian bell functions). Nevertheless, we do not exclude the possibility that a better initial value can be found to make the computations faster. In Chapter 7 we were unfortunately not able to suggest an alternative update of the sum-unitary matrices W^{R}, but we believe that this could be a perspective research in this direction. For instance, one could try to search for the minimizer in a space of rotations of the manifold $\mathcal{M}$. Other descent directions of the spread functional as well as other initial values of W^{R} could also be tried. The algorithm from Chapter 6 can be applied to the results of any other method in order to improve them.

The idea to use the Wannier functions in Photonic Crystal simulation is relatively new and very attractive. But at the same time the problem of localization of the Wannier functions becomes of high interest. We hope that the present work will be helpful for

researches of this topic, will give them a good knowledge base and inspire them with new ideas.

Chapter 10

Appendix: Numerical aspects of the presented work

10.1 Explicit computation of the Bloch waves with constant permittivity

Let us consider the simplest case of the Transverse Magnetic problem (2.1.5) — a 1D Photonic crystal with constant electric permittivity ε. Without loss of generality, let $\varepsilon = 1$. The eigenvalue problem in this case reduces to the form

$$-\psi_{\mathrm{k}}''(x) = \frac{\omega_{\mathrm{k}}^2}{c^2}\psi_{\mathrm{k}}(x) \qquad \text{for } x \in [0,1] \tag{10.1.1}$$

with the initial conditions

$$\psi_{\mathrm{k}}(1) = \mathrm{e}^{\mathrm{ik}}\psi_{\mathrm{k}}(0), \quad \psi_{\mathrm{k}}'(1) = \mathrm{e}^{\mathrm{ik}}\psi_{\mathrm{k}}'(0).$$

In further notations let $\lambda_{\mathrm{k}} = \omega_{\mathrm{k}}/c$ so that (10.1.1) turns into

$$-\psi_{\mathrm{k}}''(x) = \lambda_{\mathrm{k}}^2\psi_{\mathrm{k}}(x).$$

Rewrite it for the periodic functions u_{k}:

$$-u_{\mathrm{k}}''(x) - 2\mathrm{ik}u_{\mathrm{k}}'(x) + \mathrm{k}^2 u_{\mathrm{k}}(x) = \lambda_{\mathrm{k}}^2 u_{\mathrm{k}}(x) \tag{10.1.2}$$

with

$$u_{\mathrm{k}}(1) = u_{\mathrm{k}}(0), \quad u'_{\mathrm{k}}(1) = u'_{\mathrm{k}}(0). \tag{10.1.3}$$

The solutions can be found analytically. The dispersion relation reads:

$$\lambda_{\mathrm{k}}^2 - \mathrm{k}^2 + 2\mathrm{i}\mathrm{k}\mu_{\mathrm{k}} + \mu_{\mathrm{k}}^2 = 0.$$

Its roots are

$$\mu_{\mathrm{k}}^+ = \mathrm{i}(-\mathrm{k} + \lambda_{\mathrm{k}}), \quad \mu_{\mathrm{k}}^- = \mathrm{i}(-\mathrm{k} - \lambda_{\mathrm{k}}). \tag{10.1.4}$$

We make the following ansatz: u_{k} can be found in the form

$$u_{\mathrm{k}}(x) \;=\; a\mathrm{e}^{\mu_{\mathrm{k}}^+ x} + b\mathrm{e}^{\mu_{\mathrm{k}}^- x}$$

with some coefficients a and b which we are going to find. The boundary conditions (10.1.3) give:
$u_{\mathrm{k}}(0) = u_{\mathrm{k}}(1)$:

$$\begin{aligned} a + b &= a\mathrm{e}^{\mu_{\mathrm{k}}^+} + b\mathrm{e}^{\mu_{\mathrm{k}}^-}, \\ a\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^+}\right) + b\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^-}\right) &= 0; \end{aligned}$$

$u'_{\mathrm{k}}(0) = u'_{\mathrm{k}}(1)$:

$$\begin{aligned} \mu_{\mathrm{k}}^+ a + \mu_{\mathrm{k}}^- b &= \mu_{\mathrm{k}}^+ a\mathrm{e}^{\mu_{\mathrm{k}}^+} + \mu_{\mathrm{k}}^- b\mathrm{e}^{\mu_{\mathrm{k}}^-}, \\ a\mu_{\mathrm{k}}^+\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^+}\right) + b\mu_{\mathrm{k}}^-\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^-}\right) &= 0. \end{aligned}$$

Finally we obtain the following system for the coefficients a and b:

$$\begin{pmatrix} 1 - \mathrm{e}^{\mu_{\mathrm{k}}^+} & 1 - \mathrm{e}^{\mu_{\mathrm{k}}^-} \\ \mu_{\mathrm{k}}^+\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^+}\right) & \mu_{\mathrm{k}}^-\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^-}\right) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The system has nontrivial solutions only if its determinant vanishes:

$$-2\mathrm{i}\lambda_{\mathrm{k}}\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^+}\right)\left(1 - \mathrm{e}^{\mu_{\mathrm{k}}^-}\right) = 0.$$

Figure 10.1: Eigenvalues of the 1D TM problem with $\varepsilon = 1$

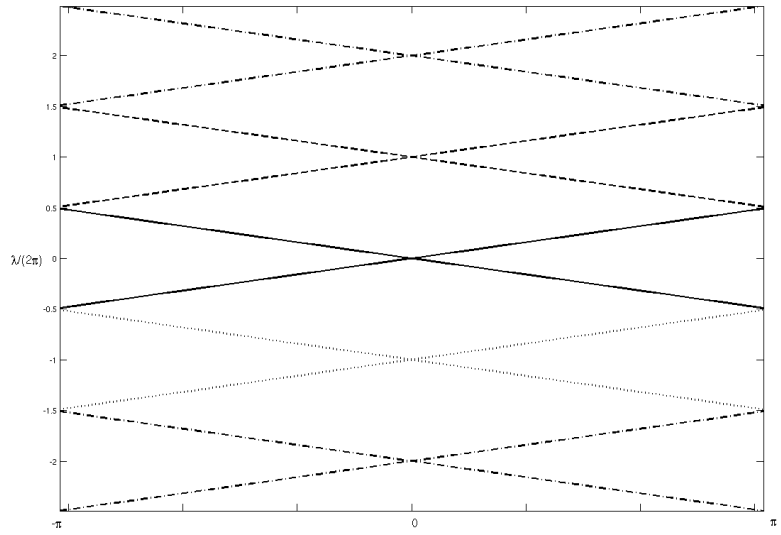
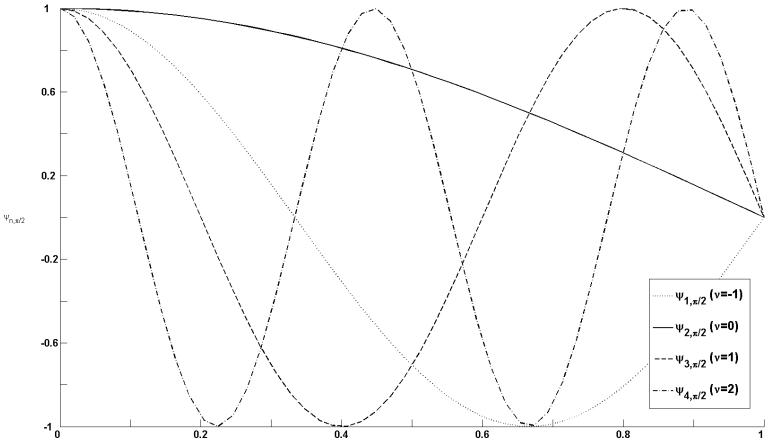


Figure 10.2: Bloch waves of the 1D TM problem with $\varepsilon = 1$



Except the trivial solution $\lambda_{\mathrm{k}} = 0$, this can be achieved in two cases:

$$\begin{aligned} \mathrm{e}^{\mu_{\mathrm{k}}^{+}} = \mathrm{e}^{\mathrm{i}(-\mathrm{k}+\lambda_{\mathrm{k}})} &= 1, \\ \lambda_{n,\mathrm{k}}^{+} &= \mathrm{k} + 2\pi\nu_n,\ \nu_n \in \mathbb{Z}, \end{aligned}$$

and

$$\begin{aligned} \mathrm{e}^{\mu_{\mathrm{k}}^{-}} = \mathrm{e}^{\mathrm{i}(-\mathrm{k}-\lambda_{\mathrm{k}})} &= 1, \\ \lambda_{n,\mathrm{k}}^{-} &= -\mathrm{k} + 2\pi\nu_n,\ \nu_n \in \mathbb{Z}. \end{aligned}$$

Alltogether, the eigenvalues can be rewritten as

$$\lambda_{n,\mathrm{k}}^{2} = (\mathrm{k} + 2\pi\nu_n)^2,\ \nu_n \in \mathbb{Z}. \tag{10.1.5}$$

The eigenfunctions $u_{n,\mathrm{k}}$ satisfy

$$u_{n,\mathrm{k}}(x) = \mathrm{e}^{2\pi\mathrm{i}\nu_n x},\ \nu_n \in \mathbb{Z},$$

and the corresponding Bloch waves are

$$\psi_{n,\mathrm{k}}(x) = \mathrm{e}^{\mathrm{i}(\mathrm{k}+2\pi\nu_n)x},\ \nu_n \in \mathbb{Z}. \tag{10.1.6}$$

The functions $\psi_{n,\mathrm{k}}$ have degeneracy in points $k = \pi m$ with $m \in \mathbb{Z}$.

The graphs $\mathrm{k} \mapsto \lambda(\mathrm{k})$ and $x \mapsto \psi_{n,\pi/2}(x)$ see at Fig.10.1 and Fig.10.2. As we see on the graph, there are no gaps in the spectrum in this case and there is no possibility to single out an isolated composite band.

An analogous situation appears in a 2D Photonic Crystal. The eigenvalue problem read then as

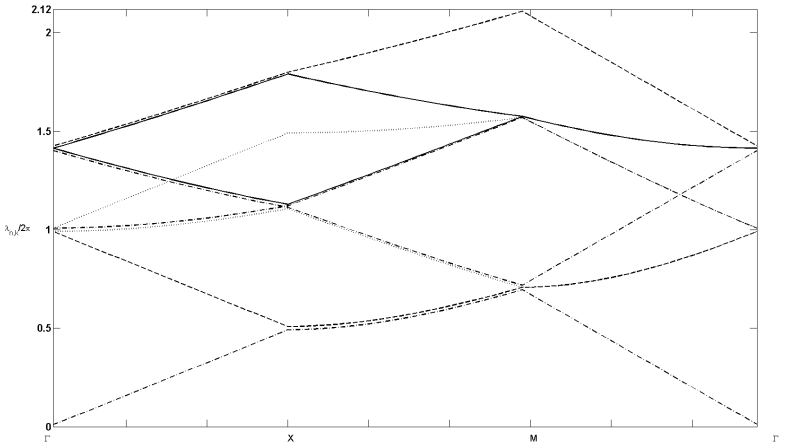
$$-\Delta\psi_{\mathrm{k}}(\mathrm{r}) = \lambda_{\mathrm{k}}\psi_{\mathrm{k}}(\mathrm{r}) \tag{10.1.7}$$

with the boundary conditions

$$\psi_{\mathrm{k}}(\mathrm{r} + \mathrm{R}) = \mathrm{e}^{\mathrm{ik}}\psi_{\mathrm{k}}(\mathrm{r}), \tag{10.1.8}$$

$$\nabla\psi_{\mathrm{k}}(\mathrm{r} + \mathrm{R}) \cdot \mathrm{n} = \mathrm{e}^{\mathrm{ik}}\nabla\psi_{\mathrm{k}}(\mathrm{r}) \cdot \mathrm{n}. \tag{10.1.9}$$

Figure 10.3: Eigenvalues of the 2D TM problem with $\varepsilon = 1$



Here

$$\mathrm{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \in [0,1]^2, \qquad \mathrm{R} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, R_1, R_2 \in \mathbb{Z}, \qquad \mathrm{k} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \in [-\pi, \pi]^2,$$

n is a normal vector to $\partial[0,1]^2$.

The eigenvalues can be found in a similar way and turn out to be

$$\lambda^2_{n,\mathrm{k}} = |\mathrm{k} + 2\pi\nu_n|^2, \ \nu_n \in \mathbb{Z}^2. \tag{10.1.10}$$

As in 1D case, there are no band gaps in the spectrum (see Fig 10.3). The corresponding Bloch waves are

$$\psi_{n,\mathrm{k}}(\mathrm{r}) = \mathrm{e}^{\mathrm{i}(\mathrm{k}+2\pi\nu_n)\cdot\mathrm{r}}, \ \nu_n \in \mathbb{Z}^2. \tag{10.1.11}$$

Again, they have degeneracies in points $\mathrm{k} = \pi m$ with $m \in \mathbb{Z}^2$.

10.2 Numerical computation of the initial functions $u_{n,\mathrm{k}}$: Finite Element method for the eigenvalue problem

In this section we explain shortly the application of the Finite Element method (FEM) to computation of the Bloch waves as the eigenfunctions of the Transverse magnetic problem (2.1.5).

10.2.1 1D case

A Photonic Crystal in 1D is described by the Transverse Magnetic equation:

$$-\psi_{\mathrm{k}}''(x) = \lambda_{\mathrm{k}}\varepsilon(x)\psi_{\mathrm{k}}(x),\ x \in [0,1],\ \mathrm{k} \in [-\pi,\pi]$$

with periodic piecewise constant electric permittivity $\varepsilon(x) = \varepsilon(x+\mathrm{R})$, $\mathrm{R} \in \mathbb{Z}$, given by

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, x_{m_1}), \\ \varepsilon_0, & x \in [x_{m_1}, x_{m_2}], \\ 1, & x \in (x_{m_2}, 1], \end{cases}$$

where $0 < x_{m_1} < x_{m_2} < 1$ and $\varepsilon_0 > 1$.

Floquet ansatz for the solutions:

$$\begin{aligned} \psi_{\mathrm{k}}(x) &= \mathrm{e}^{\mathrm{i}\mathrm{k}x}u_{\mathrm{k}}(x), \\ u_{\mathrm{k}}(x) &= u_{\mathrm{k}}(x+\mathrm{R}),\ \mathrm{R} \in \mathbb{Z}. \end{aligned}$$

Then in terms of the periodic functions u_{k} the problem reads:

$$-u_{\mathrm{k}}''(x) - 2\mathrm{i}\mathrm{k}u_{\mathrm{k}}'(x) + \mathrm{k}^2 u_{\mathrm{k}}(x) = \lambda_{\mathrm{k}}\varepsilon(x)u_{\mathrm{k}}(x),$$

or equivalently:

$$L_{\mathrm{k},\varepsilon}u_{\mathrm{k}}(x) = \lambda_{\mathrm{k}}u_{\mathrm{k}}(x), \tag{10.2.1}$$

where

$$L_{\mathrm{k},\varepsilon} = \frac{1}{\varepsilon}\left(-\,\mathrm{d}_{xx} - 2\mathrm{ik}\,\mathrm{d}_x + \mathrm{k}^2\right) = -\frac{1}{\varepsilon}(\mathrm{d}_x + \mathrm{ik})(\mathrm{d}_x + \mathrm{ik}). \tag{10.2.2}$$

Let us introduce a uniform mesh $x_1..x_{N_x}$ in $[0,1]$ such that $x_1 = 0$ and $x_{N_x} = 1$; mesh step $h_x = x_{j+1} - x_j$. Consider $\varepsilon(x) = \varepsilon(x_j) = \varepsilon_j$ for $x \in [x_j, x_{j+1})$.

Finite elements (FE) φ_j on the mesh are defined as some continuous functions with a support ("hat") on the intervals $[x_{j-1}, x_{j+1}]$ correspondingly and are zero outside [Lay99]. Expansion of u_k in basis of FE:

$$u_\mathrm{k}(x) = \sum_{j=1}^{N_x} \alpha_{\mathrm{k},j}\varphi_j(x).$$

Thus (10.2.1) can be rewritten:

$$\begin{aligned} \sum_{j=1}^{N_x} \alpha_{\mathrm{k},j} L_{\mathrm{k},\varepsilon}\varphi_j(x) &= \lambda_\mathrm{k} \sum_{j=1}^{N_x} \alpha_{\mathrm{k},j}\varphi_j(x), \\ L_{k,\varepsilon}\varphi_j(x) &= \lambda_\mathrm{k}\varphi_j(x). \end{aligned}$$

Multiply every equation by φ_i in sense of the inner product $\langle\cdot,\cdot\rangle_\varepsilon$ in $\mathrm{L}^2_\varepsilon([0,1])$:

$$\begin{aligned} \langle u, v\rangle_\varepsilon &= \int_0^1 u^*(x)\varepsilon(x)v(x)\,\mathrm{d}\,x, \\ \langle\varphi_i, L_{\mathrm{k},\varepsilon}\varphi_j\rangle_\varepsilon &= \langle\varphi_i, \lambda_\mathrm{k}\varphi_j\rangle_\varepsilon. \end{aligned}$$

Insert the operator $L_{\mathrm{k},\varepsilon}$ from (10.2.2):

$$\begin{aligned} \left\langle \varphi_i, -\frac{1}{\varepsilon}(\mathrm{d}_x + \mathrm{ik})(\mathrm{d}_x + \mathrm{ik})\varphi_j \right\rangle_\varepsilon &= \lambda_\mathrm{k}\langle\varphi_i, \varphi_j\rangle_\varepsilon, \\ \left\langle (\mathrm{d}_x + \mathrm{ik})\varphi_i, \frac{1}{\varepsilon}(\mathrm{d}_x + \mathrm{ik})\varphi_j \right\rangle_\varepsilon &= \lambda_\mathrm{k}\langle\varphi_i, \varphi_j\rangle_\varepsilon, \\ \left\langle \varphi_i', \frac{1}{\varepsilon}\varphi_j' \right\rangle_\varepsilon + \mathrm{ik}\left\langle \varphi_i', \frac{1}{\varepsilon}\varphi_j \right\rangle_\varepsilon - \mathrm{ik}\left\langle \varphi_i, \frac{1}{\varepsilon}\varphi_j' \right\rangle_\varepsilon + \mathrm{k}^2\left\langle \varphi_i, \frac{1}{\varepsilon}\varphi_j \right\rangle_\varepsilon &= \lambda_\mathrm{k}\langle\varphi_i, \varphi_j\rangle_\varepsilon. \end{aligned}$$

Introducing the notations

$$\begin{aligned} A^{i,j} &= \left\langle \varphi_i', \frac{1}{\varepsilon}\varphi_j' \right\rangle_\varepsilon, \\ B^{i,j} &= \left\langle \varphi_i', \frac{1}{\varepsilon}\varphi_j \right\rangle_\varepsilon, \\ C^{i,j} &= \left\langle \varphi_i, \frac{1}{\varepsilon}\varphi_j' \right\rangle_\varepsilon, \\ D^{i,j} &= \left\langle \varphi_i, \frac{1}{\varepsilon}\varphi_j \right\rangle_\varepsilon, \\ E^{i,j} &= \langle \varphi_i, \varphi_j \rangle_\varepsilon, \end{aligned}$$

we reformulate the equation:

$$A^{i,j} + \mathrm{i}\mathrm{k}B^{i,j} - \mathrm{i}\mathrm{k}C^{i,j} + \mathrm{k}^2 D^{i,j} = \lambda_{\mathrm{k}} E^{i,j}.$$

From here one can find the matrices A, B, C, D and E.

For every k we reformulate the eigenvalue problem in the form

$$\mathrm{Stif}^{\mathrm{k}} \cdot U^{\mathrm{k}} = \mathrm{Mass} \cdot U^{\mathrm{k}} \cdot \Lambda^{\mathrm{k}},$$

where $\mathrm{Stif}^{\mathrm{k}}$ is a stiffness matrix, Mass is a mass matrix, Λ^{k} is a diagonal matrix of the eigenvalues $\lambda_{n,\mathrm{k}}$ of the operator $L_{\mathrm{k},\varepsilon}$, U^{k} is a desired matrix whose columns are the eigenfunctions $u_{n,\mathrm{k}}$ of $L_{\mathrm{k},\varepsilon}$, orthonormal in sense of the inner product $\langle \cdot, \cdot \rangle_\varepsilon$.
Note the periodic conditions: $\varepsilon(x) = \varepsilon(x + \mathrm{R})$, $u_{\mathrm{k}}(x) = u_{\mathrm{k}}(x + \mathrm{R})$, $\mathrm{R} \in \mathbb{Z}$. In numerical computations, $\varepsilon_1 = \varepsilon_{N_x}$, $u_{\mathrm{k}}(x_1) = u_{\mathrm{k}}(x_{N_x})$, and for this reason the size of the matrices $\mathrm{Stif}^{\mathrm{k}}$, Mass, U^{k}, Λ^{k} is $(N_x - 1) \times (N_x - 1)$ with N_x a mesh size.
One can compute the eigenfunctions for every k numerically, e.g. using Matlab command "eig":
$[U^{\mathrm{k}}, \Lambda^{\mathrm{k}}] = \mathrm{eig}(\mathrm{Stif}^{\mathrm{k}}, \mathrm{Mass})$. The columns of the matrix U^{k} will then give us the eigenvectors $u_{n,\mathrm{k}}$ of the operator $L_{\mathrm{k},\varepsilon}$.

The algorithms for localization of the Wannier functions given in Chapters 5, 6 and 7 are independent on the method of computing the Bloch waves or their periodic parts $u_{n,\mathrm{k}}$. However, to be precise, we give here more details about the initial functions we work with. Let us apply linear finite elements and unitary mesh for x and k with ε defined as

$$\varepsilon(x) = \begin{cases} 1, & x \in [0, 0.41), \\ 11.56, & x \in [0.41, 0.59], \\ 1, & x \in (0.59, 1], \end{cases}$$

The convergence properties of the corresponding functions $u_{n,\mathrm{k}}$ for different mesh sizes N_x and N_k are described in the tables below; consider bands $n = 1, \ldots, 4$.

Table 10.2.1: Convergence in x-mesh, 1D

$N_\mathrm{k} = 100$, fixed point $\mathrm{k}_0 = \pi/2$, norm $\|f_{\mathrm{k}_0}\| = \left(\sum_{j=1}^{N_x} (f_{\mathrm{k}_0})_j^2\right)^{1/2}$

N_x	$\|u_{1,\mathrm{k}_0}(\cdot)\|/\sqrt{N_x}$	$\|u_{2,\mathrm{k}_0}(\cdot)\|/\sqrt{N_x}$	$\|u_{3,\mathrm{k}_0}(\cdot)\|/\sqrt{N_x}$	$\|u_{4,\mathrm{k}_0}(\cdot)\|/\sqrt{N_x}$
51	0.5479	0.8748	0.6879	0.8549
101	0.5381	0.8564	0.6795	0.8510
201	0.5432	0.8610	0.6792	0.8549
401	0.5459	0.8632	0.6799	0.8563

Table 10.2.2: Convergence in k-mesh, 1D

$N_x = 51$, fixed point $x_0 = 0.5$, norm $\|f(x_0)\| = \left(\sum_{\mathrm{k}=1}^{N_\mathrm{k}} f_\mathrm{k}^2(x_0)\right)^{1/2}$

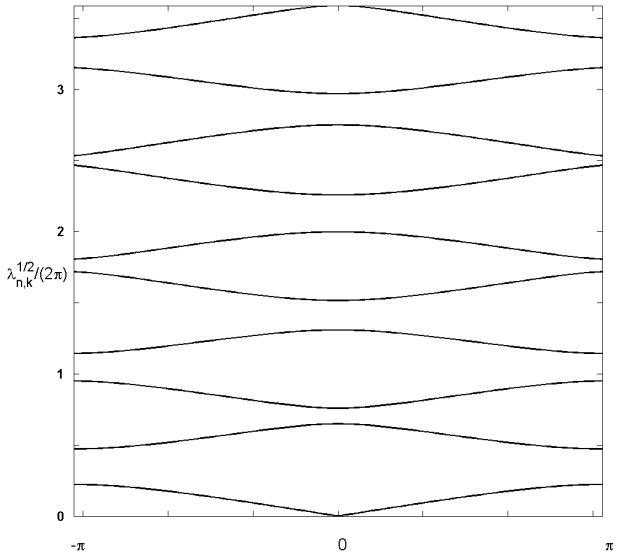
N_k	$\|u_{1,\cdot}(x_0)\|/\sqrt{N_\mathrm{k}}$	$\|u_{2,\cdot}(x_0)\|/\sqrt{N_\mathrm{k}}$	$\|u_{3,\cdot}(x_0)\|/\sqrt{N_\mathrm{k}}$	$\|u_{4,\cdot}(x_0)\|/\sqrt{N_\mathrm{k}}$
50	0.6223	0.3034	0.2129	0.3584
100	0.6223	0.3034	0.2129	0.3584
200	0.6223	0.3034	0.2129	0.3584
400	0.6223	0.3034	0.2129	0.3584

Note that the norms given here are not the same as $\|\cdot\|_{\mathrm{WSC},\varepsilon}$; in this weighted norm the computed functions $u_{n,\mathrm{k}}$ are orthonormal in accordance to Theorem (2.3).

The corresponding eigenvalues $\lambda_{n,\mathrm{k}}$ form a bad gap structure with all simple bands. On

Figure 10.4 we see such a structure computed for $N_x = 51$, $N_{\mathrm{k}} = 100$, $n = 1, \ldots, 10$.

Figure 10.4: Band Gaps in a 1D Photonic Crystal



10.2.2 2D case

Photonic crystal in 2D is described by the equation:

$$-\Delta\psi_{\mathrm{k}}(\mathrm{r}) = \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})\psi_{\mathrm{k}}(\mathrm{r}),\ \mathrm{r} \in [0,1]^2,\ \mathrm{k} \in [-\pi,\pi]^2,$$

with periodic electric permittivity $\varepsilon(\mathrm{r}) = \varepsilon(\mathrm{r}+\mathrm{R}) > 0$, $\mathrm{R} \in \mathbb{Z}^2$:

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & \mathrm{r} \in [0,1]^2, |(x-\frac{1}{2})^2 + (y-\frac{1}{2})^2| > r_1, \\ \varepsilon_0, & |(x-\frac{1}{2})^2 + (y-\frac{1}{2})^2| \le r_1, \end{cases}$$

where $0 < r_1 < \frac{1}{2}$ and $\varepsilon_0 > 1$. Ansatz for the solutions:

$$\begin{aligned} \psi_{\mathrm{k}}(\mathrm{r}) &= \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}} u_{\mathrm{k}}(\mathrm{r}), \\ u_{\mathrm{k}}(\mathrm{r}) &= u_{\mathrm{k}}(\mathrm{r}+\mathrm{R}),\ \mathrm{R} \in \mathbb{Z}^2. \end{aligned}$$

Then in terms of periodic functions u_{k}:

$$-\Delta u_{\mathrm{k}}(\mathrm{r}) - 2\mathrm{i}(\mathrm{k}\cdot\nabla)u_{\mathrm{k}}(\mathrm{r}) + |\mathrm{k}|^2 u_{\mathrm{k}}(\mathrm{r}) = \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})u_{\mathrm{k}}(\mathrm{r}),$$

or

$$L_{\mathrm{k}}u_{\mathrm{k}}(\mathrm{r}) = \lambda_{\mathrm{k}}\varepsilon(\mathrm{r})u_{\mathrm{k}}(\mathrm{r}),$$

where

$$L_{\mathrm{k}} = -(\nabla + \mathrm{ik})\cdot(\nabla + \mathrm{ik})$$

Let us introduce a uniform quadratic mesh $[x_1..x_N]\times[y_1..y_N]$ in $[0,1]^2$ such that $(x_1, y_1) = (0,0)$ and $(x_N, y_N) = (1,1)$; mesh size $h_x = h_y = h$. Renumerating the nodes, we obtain the mesh $\mathrm{r}_s = (x_i, y_j)$: $s = (j-1)N + i$.

Consider $\varepsilon(\mathrm{r}) = \varepsilon(x_i, y_j) = \varepsilon_{i,j}$ for $(x,y) \in [x_i, x_{i+1})\times[y_j, y_{j+1})$. With the same renumeration as before, we get ε_s on the mesh r_s. Introduce finite elements $\Phi_{i,j}$ with a support $[x_{i-1}, x_{i+1}]\times[y_{j-1}, y_{j+1}]$. Ansatz for u_{k} in the FE-basis:

$$u_{\mathrm{k}}(\mathrm{r}) = \sum_{i,j=1}^{N} \zeta_{\mathrm{k},i,j}\Phi_{i,j}(\mathrm{r}).$$

Thus the equations can be rewritten:

$$\begin{aligned}
\sum_{i,j=1}^{N} \zeta_{\mathrm{k},i,j} L_{\mathrm{k}}\Phi_{i,j}(\mathrm{r}) &= \lambda_{\mathrm{k}}^2\varepsilon(\mathrm{r}) \sum_{i,j=1}^{N} \zeta_{\mathrm{k},i,j}\Phi_{i,j}(\mathrm{r}), \\
L_{\mathrm{k}}\Phi_{i,j}(\mathrm{r}) &= \lambda_{\mathrm{k}}^2\varepsilon(\mathrm{r})\Phi_{i,j}(\mathrm{r}).
\end{aligned}$$

Multiply every equation by $\Phi_{\alpha,\beta}(\mathrm{r})$ in sense of inner product $\langle\cdot,\cdot\rangle$ in $\mathrm{L}^2([0,1]^2)$:

$$\begin{aligned}
\langle u, v\rangle &= \int_{[0,1]^2} u^*(\mathrm{r})v(\mathrm{r})\,\mathrm{d}\,\mathrm{r}, \\
\langle \Phi_{\alpha,\beta}, L_{\mathrm{k}}\Phi_{i,j}\rangle &= \lambda_{\mathrm{k}}^2 \langle \Phi_{\alpha,\beta}, \varepsilon(\mathrm{r})\Phi_{i,j}\rangle .
\end{aligned}$$

Insert operator L_{k}:

$$\begin{aligned}
\langle \Phi_{\alpha,\beta}, -(\nabla + \mathrm{ik})\cdot(\nabla + \mathrm{ik})\Phi_{i,j}\rangle &= \lambda_{\mathrm{k}}^2 \langle \Phi_{\alpha,\beta}, \varepsilon\Phi_{i,j}\rangle , \\
\langle (\nabla + \mathrm{ik})\Phi_{\alpha,\beta}, (\nabla + \mathrm{ik})\Phi_{i,j}\rangle &= \lambda_{\mathrm{k}}^2 \langle \Phi_{\alpha,\beta}, \varepsilon\Phi_{i,j}\rangle .
\end{aligned}$$

For vectors this reads as:

$$\left\langle \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \right\rangle = \int_{[0,1]^2} \begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \mathrm{d}\,\mathrm{r} = \langle f_1, g_1 \rangle + \langle f_2, g_2 \rangle .$$

Applying the calculations similar to described for a 1D case, one can find stiffness and mass matrices and then compute the eigenfunctions u_{k} and the eigenvalues λ_{k}.

Let us apply linear finite elements and unitary mesh for x and k with ε defined as

$$\varepsilon(\mathrm{r}) = \begin{cases} 1, & \mathrm{r} \in [0,1]^2, |(x-\frac{1}{2})^2 + (y-\frac{1}{2})^2| > 0.18, \\ 11.56, & |(x-\frac{1}{2})^2 + (y-\frac{1}{2})^2| \le 0.18, \end{cases}$$

The convergence properties of the corresponding functions $u_{n,\mathrm{k}}$ for different mesh sizes N_x and N_{k} are described in the tables below; consider the high symmetry path $\Gamma X M \Gamma$ and bands $n = 1, \ldots, 4$.

Table 10.2.3: Convergence in x-mesh, 2D

$N_{\mathrm{k}} = 40$, fixed point $\mathrm{k}_0 = (\pi/2, \pi/2)$, norm $\|f_{\mathrm{k}_0}\| = \left(\sum_{i=1}^{N_x} \sum_{j=1}^{N_x} (f_{\mathrm{k}_0})_{ij}^2\right)^{1/2}$

N_x	$\|u_{1,\mathrm{k}_0}(\cdot)\|/N_x$	$\|u_{2,\mathrm{k}_0}(\cdot)\|/N_x$	$\|u_{3,\mathrm{k}_0}(\cdot)\|/N_x$	$\|u_{4,\mathrm{k}_0}(\cdot)\|/N_x$
15	0.5741	0.7637	0.5566	0.6676
31	0.5744	0.7379	0.5428	0.6586
61	0.5860	0.7500	0.5575	0.6339

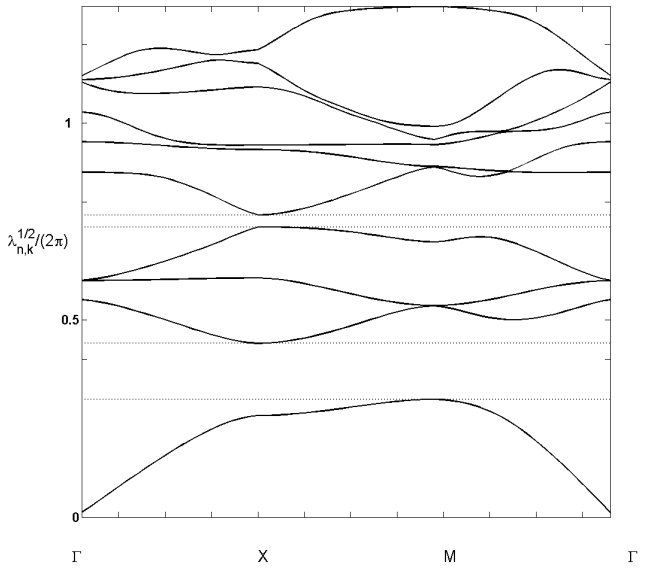
Table 10.2.4: Convergence in k-mesh, 2D

$N_x = 31$, fixed point $x_0 = (0.5, 0.5)$, norm $\|f(x_0)\| = \left(\sum_{k_1=1}^{N_{\mathrm{k}}} \sum_{k_2=1}^{N_{\mathrm{k}}} f_{k_1,k_2}^2(x_0)\right)^{1/2}$

N_{k}	$\|u_{1,\cdot}(x_0)\|/N_{\mathrm{k}}$	$\|u_{2,\cdot}(x_0)\|/N_{\mathrm{k}}$	$\|u_{3,\cdot}(x_0)\|/N_{\mathrm{k}}$	$\|u_{4,\cdot}(x_0)\|/N_{\mathrm{k}}$
40	0.8759	0.5665	0.1994	0.4758
80	0.8780	0.5615	0.2046	0.4673
160	0.8790	0.5591	0.2079	0.4631

The corresponding eigenvalues $\lambda_{n,\mathrm{k}}$ form a bad gap structure with all simple bands. On Figure 10.5 we see such a structure computed for $N_x = 31$, $N_{\mathrm{k}} = 40$, $n = 1, \ldots, 10$.

Figure 10.5: Band Gaps in a 2D Photonic Crystal



10.3 Integration over the Brillouin zone

In this section we describe the numerics used for the computation of the Wannier functions. By their definition,

$$w_{n,0}(\mathrm{r}) = \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}} u_{n,\mathrm{k}}(\mathrm{r})\,\mathrm{d}\,\mathrm{k}.$$

Introduce a Monkhorst-Pack mesh in the Brillouin zone: $k_j = (-2\pi + h_{\mathrm{k}}/2) + jh_{\mathrm{k}}$, $j \in \mathbb{Z}^d$, $h_{\mathrm{k}} = 2\pi/N_{\mathrm{k}}$ (see Chapter 4, section 2). We have the values of the functions $u_{n,\mathrm{k}}$ only in the mesh nodes, since they have been calculated numerically after the section 2 of this chapter.

The simplest intuitive way would be to approximate the Wannier functions as

$$w_{n,0}(\mathrm{r}) \approx \frac{1}{N_{\mathrm{k}}^d} \sum_{j=1}^{N_{\mathrm{k}}^d} \mathrm{e}^{\mathrm{i}k_j\cdot\mathrm{r}} u_{n,k_j}(\mathrm{r}).$$

But, to improve the quality of the result, we apply the following trick.
Divide the Brillouin zone into a $(N_{\mathrm{k}}-1)^d$ smaller areas formed by the mesh nodes:

$$\mathrm{BZ} = \bigcup_{j=1}^{N_{\mathrm{k}}-1} [k_j, k_{j+1}]^d$$

and thus split up the integral:

$$\int_{\mathrm{BZ}} f(\mathrm{k})\,\mathrm{d\,k} = \sum_{j=1}^{N_{\mathrm{k}}-1} \int_{[k_j,k_{j+1}]^d} f(\mathrm{k})\,\mathrm{d\,k}.$$

Let us approximate $u_{n,\mathrm{k}}$ in the k-space by some continuous piecewise polynomial functions:

$$u_{n,\mathrm{k}}(\mathrm{r}) \approx \widetilde{u}_{n,j}(\mathrm{k},\mathrm{r}),\ \mathrm{k} \in [k_j, k_{j+1}]^d.$$

Therefore, on every cube $[k_j, k_{j+1}]^d$ we have a function $\mathrm{k} \mapsto \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}\widetilde{u}_{n,j}(\mathrm{k},\mathrm{r})$ which can be integrated *explicitly*. The Wannier functions can then be calculated as

$$\begin{aligned} w_{n,0}(\mathrm{r}) &\approx \frac{1}{V_{\mathrm{BZ}}} \sum_{j=1}^{N_{\mathrm{k}}-1} \int_{[k_j,k_{j+1}]^d} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{r}}\widetilde{u}_{n,j}(\mathrm{k},\mathrm{r})\,\mathrm{d\,k} \\ &= \frac{1}{V_{\mathrm{BZ}}} \sum_{j=1}^{N_{\mathrm{k}}-1} g_j(\mathrm{r}). \end{aligned}$$

where every g_j is the result of an integration.

10.4 Numerical computation of the inner products of type $\langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}$ and $\left\langle r^2 w_{m,\mathrm{R}}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon}$

In different chapters we use values of the integrals $\langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}$ and $\langle r^2 w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon}$ for further calculations. Here we will describe in detail how to obtain them numerically. Apparently, one cannot integrate over $\mathbb{R}^d$, therefore we need a trick. Recall Blount's theorem 5.1 that we have mentioned in Chapter 5. By this theorem,

$$\begin{aligned} \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0}\rangle_{\mathbb{R}^d,\varepsilon} &= \frac{\mathrm{i}}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \langle u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}}\rangle_{\mathrm{WSC},\varepsilon}\,\mathrm{d\,k}, \\ \left\langle r^2 w_{m,\mathrm{R}}, w_{n,0}\right\rangle_{\mathbb{R}^d,\varepsilon} &= \frac{1}{V_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} \langle \nabla_{\mathrm{k}} u_{m,\mathrm{k}}, \nabla_{\mathrm{k}} u_{n,\mathrm{k}}\rangle_{\mathrm{WSC},\varepsilon}\,\mathrm{d\,k}. \end{aligned}$$

The Brillouin zone is bounded and therefore one can numerically integrate over it. Note that the first expression is a d-dimensional vector and the second one is a scalar. To simplify the notations, we set

$$\begin{aligned} I^1_{\mathrm{R};mn} &= \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon}\,, \\ I^2_{\mathrm{R};mn} &= \left\langle r^2 w_{m,\mathrm{R}}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon}. \end{aligned}$$

Let us shift the Wannier functions by a lattice vector R' as $w_{n,\mathrm{R}} \mapsto w_{n,\mathrm{R}+\mathrm{R}'}$ or equivalently $u_{n,\mathrm{k}} \mapsto \mathrm{e}^{-\mathrm{i}\mathrm{k}\cdot\mathrm{R}'} u_{n,\mathrm{k}}$, which has a meaning of choosing the "home" unit cell. Then the corresponding values of $I^1_{\mathrm{R};mn}$ and $I^2_{\mathrm{R};mn}$ also get shifted:

$$\begin{aligned} \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} &\mapsto \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} + \mathrm{R}'\delta_{m,n}\delta_{\mathrm{R},0}, \\ \left\langle r^2 w_{m,\mathrm{R}}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} &\mapsto \left\langle r^2 w_{m,\mathrm{R}}, w_{n,0} \right\rangle_{\mathbb{R}^d,\varepsilon} + 2\mathrm{R}' \cdot \langle \mathrm{r} w_{m,\mathrm{R}}, w_{n,0} \rangle_{\mathbb{R}^d,\varepsilon} + |\mathrm{R}'|^2 \delta_{m,n}\delta_{\mathrm{R},0}, \end{aligned}$$

and thus

$$I^1_{\mathrm{R};mn} \mapsto I^1_{\mathrm{R};mn} + \mathrm{R}'\delta_{m,n}\delta_{\mathrm{R},0}, \tag{10.4.1}$$

$$I^2_{\mathrm{R};mn} \mapsto I^2_{\mathrm{R};mn} + 2\mathrm{R}' \cdot I^1_{\mathrm{R};mn} + |\mathrm{R}'|^2 \delta_{m,n}\delta_{\mathrm{R},0}. \tag{10.4.2}$$

Let us now follow the notations from Section 5.2 (Monkhorst-Pack mesh in the Brillouin zone, stencil S_k (5.2.1) or (5.2.2), definition (5.2.7) etc.). We approximate the terms with k-gradient by (5.2.3) and (5.2.4):

$$\begin{aligned} \nabla_\mathrm{k} u_{n,\mathrm{k}} &\approx \sum_{\mathrm{s}\in S_\mathrm{k}} a_\mathrm{s} \mathrm{s} \left(u_{n,\mathrm{k}+\mathrm{s}} - u_{n,\mathrm{k}} \right), \\ |\nabla_\mathrm{k} u_{n,\mathrm{k}}|^2 &\approx \sum_{\mathrm{s}\in S_\mathrm{k}} a_\mathrm{s} \left(u_{n,\mathrm{k}+\mathrm{s}} - u_{n,\mathrm{k}} \right)^2 \end{aligned}$$

with s from the stencil S_k and weights a_s such that $\sum_{\mathrm{s}\in S_\mathrm{k}} a_\mathrm{s} \mathrm{s}\otimes\mathrm{s} = \mathrm{Id}$. Recall the definition (5.2.7) of $M^{\mathrm{k},\mathrm{s}}_{mn}$, then

$$I^1_{\mathrm{R};mn} \approx I^{1,\triangle}_{\mathrm{R};mn} = \frac{\mathrm{i}}{N_\mathrm{k}^d} \sum_{\mathrm{k},\mathrm{s}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} a_\mathrm{s} \mathrm{s} \left[M^{\mathrm{k},\mathrm{s}}_{mn} - \delta_{m,n} \right], \tag{10.4.3}$$

$$I^2_{\mathrm{R};mn} \approx I^{2,\triangle}_{\mathrm{R};mn} = \frac{1}{N_\mathrm{k}^d} \sum_{\mathrm{k},\mathrm{s}} \mathrm{e}^{\mathrm{i}\mathrm{k}\cdot\mathrm{R}} a_\mathrm{s} \left[2\delta_{m,n} - 2\,\mathrm{Re}\left(M^{\mathrm{k},\mathrm{s}}_{mn} \right) \right]. \tag{10.4.4}$$

Unfortunately, with a lattice shift $w_{n,\mathrm{R}} \mapsto w_{n,\mathrm{R}+\mathrm{R}'}$ these approximations do not satisfy (10.4.1) and (10.4.2):

$$\begin{aligned} M^{\mathrm{k,s}}_{mn} &\mapsto \mathrm{e}^{-\mathrm{is}\cdot\mathrm{R}'} M^{\mathrm{k,s}}_{mn}, \\ I^{1,\triangle}_{\mathrm{R};mn} &\mapsto \frac{\mathrm{i}}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} a_s \mathrm{s} \left[\mathrm{e}^{-\mathrm{is}\cdot\mathrm{R}'} M^{\mathrm{k,s}}_{mn} - \delta_{m,n} \right], \\ I^{2,\triangle}_{\mathrm{R};mn} &\mapsto \frac{1}{N_{\mathrm{k}}^d} \sum_{\mathrm{k,s}} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} a_s \left[2\delta_{m,n} - 2\,\mathrm{Re}\left(\mathrm{e}^{-\mathrm{is}\cdot\mathrm{R}'} M^{\mathrm{k,s}}_{mn} \right) \right]. \end{aligned}$$

For this reason we want to update the approximation formulas in order to obtain the same shift properties as we have for the continuous analogues. Following [Klo04, Section 4.5], consider small $\beta > 0$ and the Taylor expansion

$$u_{n,\mathrm{k}+\beta\mathrm{s}} = u_{n,\mathrm{k}} + \beta \mathrm{s} \cdot \nabla_k u_{n,\mathrm{k}} + O(\beta^2),$$

where the derivative is understood to be taken in point $\beta = 0$ and s fixed. Suppose

$$M^{\mathrm{k},\beta\mathrm{s}}_{mn} = \delta_{m,n} + \mathrm{i}\mu_1\beta + \frac{1}{2}\mu_2\beta^2 + O(\beta^3),$$

then for the components of $I^{1,\triangle}_{\mathrm{R};mn}$ and $I^{1,\triangle}_{\mathrm{R};mn}$ it holds:

$$\begin{aligned} M^{\mathrm{k},\beta\mathrm{s}}_{mn} - \delta_{m,n} &= \mathrm{i}\mu_1\beta + \frac{1}{2}\mu_2\beta^2 + O(\beta^3), && (10.4.5) \\ 2\delta_{m,n} - 2\,\mathrm{Re}\left(M^{\mathrm{k},\beta\mathrm{s}}_{mn}\right) &= 2\delta_{m,n} - 2\left(\delta_{m,n} + \frac{1}{2}\,\mathrm{Re}(\mu_2)\beta^2\right) + O(\beta^3) && (10.4.6) \\ &= -\,\mathrm{Re}(\mu_2)\beta^2 + O(\beta^3). && (10.4.7) \end{aligned}$$

From the orthonormality of $u_{n,\mathrm{k}}$ we have:

$$\begin{aligned} \delta_{m,n} &= \langle u_{m,\mathrm{k}+\beta\mathrm{s}}, u_{n,\mathrm{k}+\beta\mathrm{s}} \rangle_{\mathrm{WSC},\varepsilon}, \\ \delta_{m,n} &= \langle u_{m,\mathrm{k}} + \beta\mathrm{s}\cdot\nabla_k u_{m,\mathrm{k}}, u_{n,\mathrm{k}} + \beta\mathrm{s}\cdot\nabla_k u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} + O(\beta^2), \\ \delta_{m,n} &= \langle u_{m,\mathrm{k}}, u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} + 2\beta\,\mathrm{Re}\,\langle \mathrm{s}\cdot\nabla_k u_{m,\mathrm{k}}, u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} + O(\beta^2), \\ 0 &= 2\beta\,\mathrm{Re}\,\langle \mathrm{s}\cdot\nabla_k u_{m,\mathrm{k}}, u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon} + O(\beta^2). \end{aligned}$$

But

$$\mathrm{i}\mu_1 = \langle \mathrm{s}\cdot\nabla_k u_{m,\mathrm{k}}, u_{n,\mathrm{k}} \rangle_{\mathrm{WSC},\varepsilon},$$

therefore, $\operatorname{Im}\mu_1 = 0$.

So, we update the way to approximate the integrals in a manner suggested in [MV97] (for detais see also [Klo04, Section 4.5]). Finally,

$$I^{1,\sim}_{\mathrm{R};mn} = -\frac{1}{N^d_{\mathrm{k}}}\sum_{\mathrm{k,s}} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} a_s \mathrm{s} \arg M^{\mathrm{k,s}}_{mn}, \tag{10.4.8}$$

$$I^{2,\sim}_{\mathrm{R};mn} = \frac{1}{N^d_{\mathrm{k}}}\sum_{\mathrm{k,s}} \mathrm{e}^{\mathrm{ik}\cdot\mathrm{R}} a_s \left[\delta_{m,n} - \left|M^{\mathrm{k,s}}_{mn}\right|^2 + \left(\arg M^{\mathrm{k,s}}_{mn}\right)^2\right]. \tag{10.4.9}$$

From (10.4.8) and (10.4.9) one can derive the approximations for the expressions which we need in different algorithms. In particular, with $m = n$ and $\mathrm{R} = 0$ we deduce to the equations (5.2.8) and (5.2.9) of the Section 5.2 and thus find an approximation of the spread functional for Marzari-Vanderbilt algorithm. In Sections 6.2 and 6.3 dedicated to unitary transform of the Wannier functions we introduced matrix A which elements are nothing but $I^{1,\sim}_{0;mn}$; and we used the same approximation for X in Section 3.5 (computation of the eigenfunctions of the projected position operator). In Section 7.6 of the sum-unitary method the values $\left[B_1^{\mathrm{R,R'}}\right]_{mn}$ and $\left[B_2^{\mathrm{R,R'}}\right]_{mn}$ can be obtained by substituting $\left[B_1^{\mathrm{R,R'}}\right]_{mn} = I^{1,\sim}_{\mathrm{R-R';}mn}$, $\left[B_2^{\mathrm{R,R'}}\right]_{mn} = I^{2,\sim}_{\mathrm{R-R';}mn}$.

Bibliography

[Blo62] E.I. Blount. Formalisms of band theory. *Solid State Physics*, 13:305–373, 1962.

[BMGM+03] K. Busch, S. Mingaleev, A. Garcia-Martin, M. Schillinger, and D. Hermann. Wannier function approach to photonic crystal circuits. *J. Phys. Cond. Mat.*, 15, 2003.

[BMR+00] G. Berghold, Ch.J. Mundi, A.H. Romero, J. Hutter, and M. Parinello. General and efficient algorithms for obtaining maximally localized Wannier functions. *Phys. Rev. B*, 62:10040–10048, 2000.

[Bus03] K. Busch. The physics of Photonic Crystals. *Institut für Theorie der Kondensierten Materie, Universität Karlsruhe*, pages 1–30, 2003.

[CNN08] H.D. Cornean, A. Nenciu, and G. Nenciu. Optimally localized Wannier functions for quasi one-dimensional nonperiodic insulators. *J. Phys. A: Math. Theor.*, 41, 2008.

[CS96] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix. Anal. Appl*, 17:161–164, 1996.

[dC63] J. des Cloizeaux. I. Orthogonal orbitals and generalized Wannier functions. *Phys. Rev.*, 129(2):554 – 566, 1963.

[dC64a] J. des Cloizeaux. II. Energy bands and projection operators in a crystal: Analytic and asymptotic properties. *Phys. Rev.*, 135(3A):A685 – A697, 1964.

[dC64b] J. des Cloizeaux. III. Analytical properties of n-dimesional energy bands and Wannier functions. *Phys. Rev.*, 135(3A):A698 – A707, 1964.

[DLP+11] W. Dörfler, A. Lechleiter, M. Plum, G. Schneider, and Ch. Wieners. Mathematics of photonic crystals. 2011.

[FK94] A. Figotin and P. Kuchment. Band-gap structure of the spectrum of periodic Maxwell operators. *Journal of Statistical Physics*, 74:447–455, 1994.

[GFS03] F. Gygi, J.-L. Fattebert, and E. Schwegler. Computation of maximally localized Wannier functions using a simultaneous diagonalization algorithm. *Comput. Phys. Comm.*, 155(1):1–6, 2003.

[Kiv82] S. Kivelson. Wannier functions in one-dimensional disordered systems: Application to fractionally charged solitons. *Phys. Rev. B*, 26(8):4269–4277, 1982.

[Klo04] A. Kloeckner. On the computation of maximally localized Wannier functions. Master's thesis, Universität Karlsruhe (TH), 2004.

[Koh59] W. Kohn. Analytic properties of Bloch waves and Wannier functions. *Phys. Rev.*, 36:809–821, 1959.

[Kra35] H.A. Kramers. *Physica*, volume 2. 1935.

[Kuc01] P. Kuchment. The mathematics of Photonic Crystals. *Mathematical Modelling in Optical Science*, 22:207–272, 2001.

[Kuc08] P. Kuchment. Tight frames of exponentially decaying Wannier functions. *J. Phys. A: Math. Theor.*, 42, 2008.

[Lay99] Y.M. Layevsky. *Finite element method (theory fundamentals, exercises).* Novosibirsk State University, 1999.

[MP76] H.J. Monkhorst and J.D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, 1976.

[MV97] N. Marzari and D. Vanderbilt. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, 1997.

[Nen83] A. Nenciu. Existence of the exponentially localized Wannier functions. *Commun. Math. Phys.*, 91:81–85, 1983.

[NN82] A. Nenciu and G. Nenciu. Dynamics of Bloch electrons in external electric fields: II. The existence of Stark-Wannier resonances. *J. Phys. A: Math. Gen.*, 15:3313–3328, 1982.

[RS78] M. Reed and B. Simon. Analysis of operators. *Methods of modern mathematical physics*, 4, 1978.

[Rud91] W. Rudin. *Functional analysis.* McGraw-Hill Science, 1991.

[She94] J.R. Shewchuck. An introduction to the conjugate gradient method without the agonizing pain. 1994.

[SMHB05] M. Schillinger, S.F. Mingaleev, D. Hermann, and K. Busch. Highly localized Wannier functions for the efficient modeling of Photonic Crystal circuits. *Proc. of SPIE*, 5733:324–335, 2005.

[SMV01] I. Souza, N. Marzari, and D. Vanderbilt. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B*, 65(035101):1–13, 2001.

[Taf88] A. Taflove. Review of the formulation and applications on the finite difference time domain method for numerical modeling of electromagnetic wave interactions with arbitrary structures. *Wave Motion*, 10:547–582, 1988.

[Wan37] G.H. Wannier. The structure of electronic excitation levels in insulating crystals. *Phys. Rev.*, 52(2):191–197, 1937.

[WC03] D.M. Whittaker and M.P. Croucher. Maximally localized Wannier functions for photonic lattices. *Phys. Rev. B*, 67: 085204, 2003.

Curriculum Vitae

Personal data

Name:	Tatiana Bulovyatova
Date of birth:	03.08.1982
Place of birth:	Novosibirsk, Russia
Citizenship:	Russian
Family status:	married
Contact:	bulovyatova@gmail.com

Education

2006-2011	Ph.D. study	GRK1294, Department of Mathematics, Karlsruhe Institute of Technology (KIT) (earlier University Karlsruhe (TH))
1999-2004	Diploma in mathematics	Department of Mechanics and Mathematics, Novosibirsk State University
1988-1999	School	

Working experience

2004-2006	Software tester	Atapy Software, Russia
2002-2004	Research assistant	Institute of petroleum geology and geophysics (SB RAS), Russia