

# Long-term analysis of a variational integrator for charged-particle dynamics in a strong magnetic field

Ernst Hairer, Christian Lubich

CRC Preprint 2018/6

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



## Participating universities



**Universität Stuttgart**

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



**Funded by**

**DFG**

ISSN 2365-662X

# Long-term analysis of a variational integrator for charged-particle dynamics in a strong magnetic field

Ernst Hairer<sup>1</sup>, Christian Lubich<sup>2</sup>

**Abstract** The differential equations of motion of a charged particle in a strong non-uniform magnetic field have the magnetic moment as an adiabatic invariant. This quantity is nearly conserved over long time scales covering arbitrary negative powers of the small parameter, which is inversely proportional to the strength of the magnetic field. The numerical discretisation is studied for a variational integrator that is an analogue for charged-particle dynamics of the Störmer–Verlet method. This numerical integrator is shown to yield near-conservation of a modified magnetic moment and a modified energy over similarly long times. The proofs for both the continuous and the discretised equations use modulated Fourier expansions with state-dependent frequencies and eigenvectors.

*Keywords.* Charged particle, magnetic field, adiabatic invariant, modulated Fourier expansion.

*Mathematics Subject Classification (2010):* 65L05, 65P10, 78A35, 78M25

## 1 Introduction

This paper is concerned with the long-time behaviour of a numerical integrator for the equations of motion of a particle in a strong non-uniform magnetic field. In particular we study the long-time near-conservation of the energy and of the magnetic moment, which is an adiabatic invariant of the system of differential equations.

---

<sup>1</sup> Dept. de Mathématiques, Univ. de Genève, CH-1211 Genève 24, Switzerland,  
E-mail: Ernst.Hairer@unige.ch

<sup>2</sup> Mathematisches Institut, Univ. Tübingen, D-72076 Tübingen, Germany,  
E-mail: Lubich@na.uni-tuebingen.de

### 1.1 Motion of charged particles in a strong magnetic field

We consider the differential equation that determines the position  $x(t) \in \mathbf{R}^3$  of a charged particle moving in a strong magnetic field,

$$\ddot{x} = \frac{1}{\varepsilon} \dot{x} \times B(x) + E(x), \quad 0 < \varepsilon \ll 1, \quad (1.1)$$

where  $B(x) = \nabla_x \times A(x)$  with a vector potential  $A(x) \in \mathbf{R}^3$  and  $E(x) = -\nabla_x U(x)$  with a scalar potential  $U(x) \in \mathbf{R}$ . It is assumed that  $A$  and  $U$  are given smooth functions. We are interested in the situation of a small positive scaling parameter  $\varepsilon$ , and we assume  $|B(x)| \geq 1$  in the Euclidean norm  $|\cdot|$ . The differential equations (1.1) are the Euler–Lagrange equations for the Lagrange function  $L(x, \dot{x}) = \frac{1}{2}|\dot{x}|^2 + \frac{1}{\varepsilon}A(x)^\top \dot{x} - U(x)$ .

We consider initial values

$$x(0) = \mathcal{O}(1), \quad \dot{x}(0) = \mathcal{O}(1), \quad (1.2)$$

so that the energy (which is a conserved quantity)

$$H(x, \dot{x}) = \frac{1}{2}|\dot{x}|^2 + U(x) \quad (1.3)$$

is bounded independently of  $\varepsilon$  along the solution.

It is well known [12, 2, 4] that the magnetic moment is an *adiabatic invariant* (see, e.g., [1, 11] for this notion):

$$I(x, \dot{x}) = \frac{1}{2} \frac{|\dot{x} \times B(x)|^2}{|B(x)|^3} = \frac{1}{2} \frac{|\dot{x}_\perp|^2}{|B(x)|}, \quad (1.4)$$

where  $\dot{x}_\perp$  is the component of the velocity that is orthogonal to  $B(x)$ . The particle motion is composed of fast rotations in the plane orthogonal to the magnetic field around a guiding center (Larmor rotation) and slow motion of the guiding center, in such a way that the magnetic moment is approximately conserved.

### 1.2 Numerical integrators

In this paper we are interested in the long-time near-conservation of the energy and the magnetic moment along the numerical solution obtained with a numerical integrator applied to (1.1).

The most widely used integrator for charged-particle dynamics is the *Boris method* [3]

$$\frac{x_{n+1} - 2x_n + x_{n-1}}{h^2} = \frac{1}{\varepsilon} \frac{(x_{n+1} - x_{n-1})}{2h} \times B(x_n) + E(x_n), \quad (1.5)$$

with the velocity approximation

$$v_n = \frac{1}{2h}(x_{n+1} - x_{n-1}). \quad (1.6)$$

The Boris method is known to have good long-time energy behaviour for a fixed  $\varepsilon$  and step size  $h \ll \varepsilon$  in the case of a *constant* magnetic field  $B$ , though not for general magnetic fields [9]. The Boris method is not a variational/symplectic integrator unless the magnetic field  $B$  is constant [5].

A related variational integrator, which coincides with the Boris method for constant  $B$  (or equivalently, linear  $A(x)$ ), is constructed like in the interpretation of the Störmer–Verlet method as a variational integrator: the integral of the Lagrangian  $L(x, \dot{x}) = \frac{1}{2}|\dot{x}|^2 + \frac{1}{\varepsilon}A(x)^\top \dot{x} - U(x)$  over a time step is approximated by replacing  $x(t)$  with the linear interpolant of the endpoint positions and then approximating the integral by the trapezoidal rule, and finally this approximation to the action integral is extremised; see, e.g., [10, Chap. VI, Example 6.2] and [13]. With the derivative matrix  $A'(x) = (\partial_j A_i(x))_{i,j=1}^3$  and its transpose  $A'(x)^\top$ , this variational integrator reads

$$\begin{aligned} \frac{x_{n+1} - 2x_n + x_{n-1}}{h^2} &= \frac{1}{\varepsilon} A'(x_n)^\top \frac{(x_{n+1} - x_{n-1})}{2h} \\ &\quad - \frac{1}{\varepsilon} \frac{A(x_{n+1}) - A(x_{n-1})}{2h} - \nabla U(x_n), \end{aligned} \quad (1.7)$$

again taken with the velocity approximation (1.6). Like the Störmer–Verlet method, this method can be given a one-step formulation that maps  $(x_n, v_n)$  to  $(x_{n+1}, v_{n+1})$ . It is, however, an implicit method. This variational integrator will be shown to have similar behaviour regarding the long-time near-conservation of the energy and the adiabatic invariant as the Störmer–Verlet method has for mechanical systems with high-frequency oscillations that are due to a strong constraining potential [6, 7].

### 1.3 Outline of the paper

In Section 2 we state the main results of this paper: Theorem 2.1 states the long-time near-conservation of the magnetic moment for the continuous system, and Theorem 2.2 states the long-time near-conservation, over the time scale  $t \leq \varepsilon^{-N}$ , of a modified magnetic moment for the discretised system with step size  $h \leq c_N \varepsilon$ . Theorem 2.3 states the long-time near-conservation of a modified energy for the discretised system. Section 3 presents numerical experiments that illustrate the theoretical results.

In Section 4 we derive a modulated Fourier expansion for the exact solution, which is then used to prove Theorem 2.1. In Section 5 we derive, in a similar way, a modulated Fourier expansion for the numerical solution, which is then used to prove Theorems 2.2 and 2.3. In contrast to previous works using modulated Fourier expansions (e.g., [10, Chapter XIII] and [7]), in the problem considered here both the eigenvalues (frequencies) and the eigenvectors of the dominant linear map, which here is  $v \mapsto \frac{1}{\varepsilon}v \times B(x)$ , are state-dependent. The dependence on  $x$  needs to be addressed specifically in the construction of the modulated Fourier expansion. The particular discretization of the magnetic term in (1.7) is of no importance for the existence of the modulated Fourier

expansion, but it is essential for the existence of its almost-invariants, which yield the near-conservation of the modified magnetic moment and energy.

## 2 Main results

### 2.1 Adiabatic invariant of the continuous problem

While it has been common knowledge in physics that the magnetic moment is an adiabatic invariant, at least since the work by Northrop [12], it seems that a rigorous proof of the near-conservation of the magnetic moment *over very long times* was first given by Benettin & Sempio [2], who show the adiabatic invariance over exponentially long times  $t \leq e^{-c/\varepsilon}$  with  $c > 0$  (for real-analytic  $B$  and in the case  $E = 0$ ). Their proof is based on a sequence of nonlinear canonical transformations of Hamiltonian perturbation theory that transform the Hamiltonian to a normal form. Since this technique of proof does not appear to lend itself to studying the long-time behaviour of numerical discretisations such as (1.7), we prove the long-time near-conservation of the magnetic moment with the alternative technique of modulated Fourier expansions (see [10, Chapter XIII] and [7]), which does not use nonlinear coordinate transforms and will later be shown to apply equally to the numerical discretisation. We prove the following result in Section 4.

**Theorem 2.1** *There exists  $\varepsilon_0 > 0$  such that the following holds for  $0 < \varepsilon \leq \varepsilon_0$ . Under the bounded-energy condition (1.2) and provided that the solution  $x(t)$  of (1.1) stays in a compact set  $K$ , the magnetic moment is nearly conserved over long times: for arbitrary  $N \geq 1$ ,*

$$|I(x(t), \dot{x}(t)) - I(x(0), \dot{x}(0))| \leq C_N \varepsilon \quad \text{for } t \leq \varepsilon^{-N}.$$

The constant  $C_N$  is independent of  $\varepsilon \leq \varepsilon_0$  and  $t$  with  $t \leq \varepsilon^{-N}$ , but depends on  $N$ , on bounds of the first  $N + 1$  derivatives of  $B$  and  $E$  on the compact set  $K$ , and on the constants in (1.2).

### 2.2 Near-conservation of the magnetic moment of the variational integrator

We define

$$\xi(x) = 2 \arctan\left(\frac{h}{2\varepsilon}|B(x)|\right) \quad (2.1)$$

and we note that  $\xi(x) \approx (h/\varepsilon)|B(x)|$  for small  $h/\varepsilon$ .

We consider the modified magnetic field

$$B_h(x) = \rho(\xi(x))B(x) \quad \text{with} \quad \rho(\xi) = \cos^4(\xi/2),$$

and we note that  $1 - \rho(\xi) \sim \frac{1}{2}\xi^2$  for small  $\xi$ . We then define the modified magnetic moment as

$$I_h(x, v) = \frac{1}{2} \frac{|v \times B_h(x)|^2}{|B_h(x)|^3} = \frac{1}{\rho(\xi(x))} I(x, v).$$

With the modified velocity  $\tilde{v} = v/\rho(\xi(x))^{1/2}$  we have  $I_h(x, v) = I(x, \tilde{v})$ . The proof of the following theorem is given in Section 5.

**Theorem 2.2** *There exist  $\varepsilon_0 > 0$  and  $h_0 > 0$  such that the following holds for  $0 < \varepsilon \leq \varepsilon_0$  and  $0 < h \leq h_0$ . Let  $N \geq 1$  be an arbitrary integer. Under the bounded-energy condition (1.2) and provided that the numerical solution  $(x_n)$  of the variational integrator (1.7) stays in a compact set  $K$  and is obtained with a step size  $h$  for which  $\frac{h}{\varepsilon} |B(x_n)| \leq 2 \tan(\frac{\pi}{2(N+3)})$  for some  $N \geq 1$ , the modified magnetic moment is nearly conserved over long times:*

$$|I_h(x_n, v_n) - I_h(x_0, v_0)| \leq C_N \varepsilon \quad \text{for } nh \leq \varepsilon^{-N}.$$

The constant  $C_N$  is independent of  $\varepsilon$ ,  $h$ , and  $n$  with  $nh \leq \varepsilon^{-N}$ , but depends on  $N$ , on bounds of the first  $N + 1$  derivatives of  $B$  and  $E$  on the compact set  $K$ , and on the constants in (1.2).

### 2.3 Near-conservation of energy of the variational integrator

We consider the modified energy

$$H_h(x, v) = H(x, v) + \theta(\xi(x))I(x, v) |B(x)|$$

with  $\xi(x)$  from (2.1) and  $\theta(\xi) = \theta_{kin}(\xi) + \theta_{mag}(\xi)$ , where<sup>1</sup>

$$\begin{aligned} \theta_{kin}(\xi) &= \frac{2 \operatorname{sinc}(\xi) - \operatorname{sinc}(\xi/2)^2}{\operatorname{sinc}(\xi)^2} - 1, \\ \theta_{mag}(\xi) &= \frac{\operatorname{sinc}(\xi/2)(\operatorname{sinc}(\xi) - \cos(\xi))}{\cos(\xi/2) \operatorname{sinc}(\xi)^2}. \end{aligned}$$

We note that  $\theta(\xi) \sim \frac{5}{12}\xi^2$  for small  $\xi$ . The proof of the following theorem is given in Section 5.

**Theorem 2.3** *Under the assumptions of Theorem 2.2, the modified energy is nearly conserved over long times:*

$$|H_h(x_n, v_n) - H_h(x_0, v_0)| \leq C_N \varepsilon \quad \text{for } nh \leq \varepsilon^{-N}.$$

The constant  $C_N$  is independent of  $\varepsilon$ ,  $h$ , and  $n$  with  $nh \leq \varepsilon^{-N}$ , but depends on  $N$ , on bounds of the first  $N + 1$  derivatives of  $B$  and  $E$  on the compact set  $K$ , and on the constants in (1.2).

*Remark 2.1* In the case of a confining potential, i.e., such that  $U(x) \rightarrow +\infty$  as  $|x| \rightarrow \infty$ , Theorem 2.3 actually implies by an induction argument that the numerical solution  $x_n$  stays in a compact set  $K$  that depends on bounds of the initial values, but is independent of  $\varepsilon$ ,  $h$ , and  $n$  subject to the conditions of the theorem.

*Remark 2.2* Theorems 2.2 and 2.3 hold also for the Boris method in the special case of a constant magnetic field  $B$ , because the variational integrator coincides with the Boris method in this case.

<sup>1</sup> We use the notation  $\operatorname{sinc}(\xi) = \sin(\xi)/\xi$ .

### 3 Numerical experiments

To illustrate the statements of the preceding section we consider Example 5.2 of [9] with an additional factor  $1/\varepsilon$ . We have the scalar potential

$$U(x) = x_1^3 - x_2^3 + \frac{1}{5}x_1^4 + x_2^4 + x_3^4, \quad (3.1)$$

the magnetic field

$$B(x) = \nabla \times \frac{1}{4} \begin{pmatrix} x_3^2 - x_2^2 \\ x_3^2 - x_1^2 \\ x_2^2 - x_1^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_2 - x_3 \\ x_1 + x_3 \\ x_2 - x_1 \end{pmatrix}, \quad (3.2)$$

and we take the initial values

$$x(0) = (0.0, 1.0, 0.1)^\top, \quad \dot{x}(0) = (0.09, 0.55, 0.30)^\top. \quad (3.3)$$

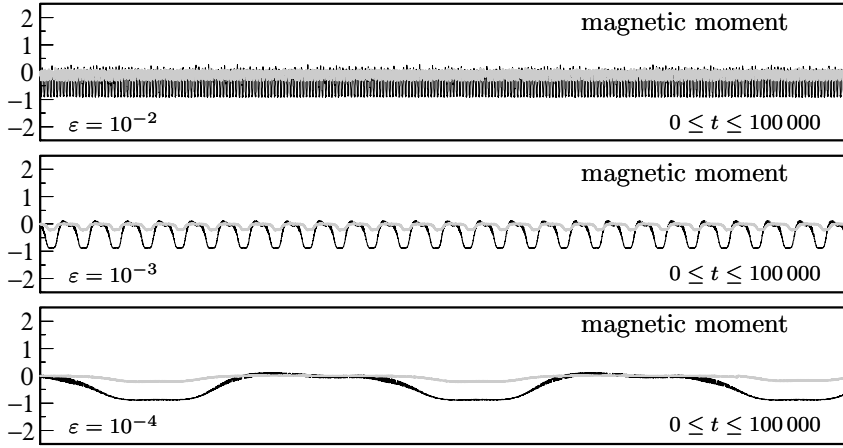
The quartic terms in (3.1) imply compact level sets of the energy, so that the exact solution of the problem exists and remains bounded for all times.

For different values of  $\varepsilon$  and  $\eta$  we apply the variational integrator (1.7) with step size  $h = \eta\varepsilon$ . The nonlinear equation is solved by simplified Newton iterations. Figure 3.1 shows the relative error  $(I(x_n, v_n) - I(x_0, v_0))/I(x_0, v_0)$  of the magnetic moment along the numerical solution. It is drawn in black for  $\eta = 0.8$  and in grey for  $\eta = 0.2$ . We observe that this error is of size  $\mathcal{O}(1)$  for all choices of  $\varepsilon$ . Figure 3.2 shows the same experiment, where the exact magnetic moment is replaced by the modified magnetic moment  $I_h(x, v)$  with  $\rho(\xi) = \cos^4(\xi/2)$ . For this modified magnetic moment the relative error is of size  $\mathcal{O}(\varepsilon)$  (note that the vertical axis is scaled by  $\varepsilon$ ). This is in perfect agreement with the statement of Theorem 2.2.

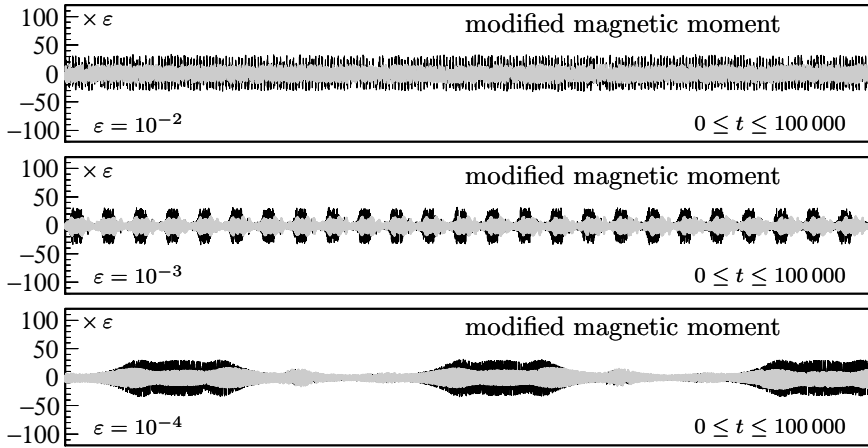
Our next numerical experiment is concerned with the energy  $H(x, v)$  of the system. Figure 3.3 shows the relative error  $(H(x_n, v_n) - H(x_0, v_0))/H(x_0, v_0)$  of the energy along the numerical solution of the variational integrator. We observe that the error is not smaller than  $\mathcal{O}(1)$ . This experiment also demonstrates that, although the variational integrator is symplectic and symmetric, the energy error is small only for step sizes  $h \ll \varepsilon$ . Figure 3.4 shows the relative error  $(H_h(x_n, v_n) - H_h(x_0, v_0))/H_h(x_0, v_0)$  of the modified energy  $H_h(x, v)$ , which is observed to be of size  $\mathcal{O}(\varepsilon)$  in agreement with the statement of Theorem 2.3.

Numerical experiments as in Figures 3.1 and 3.3 with the non-symplectic Boris algorithm are similar to those obtained with the variational integrator. Note, however, that for the Boris algorithm we do not know of a modified magnetic moment or of a modified energy that would be conserved up to errors of size  $\mathcal{O}(\varepsilon)$ .





**Fig. 3.1** Relative error  $(I(x_n, v_n) - I(x_0, v_0))/I(x_0, v_0)$  of the magnetic moment  $I(x, v)$  as a function of time, along the numerical solution of the variational integrator (1.7), obtained with  $h = \eta\varepsilon$ , where  $\eta = 0.8$  (black) and  $\eta = 0.2$  (grey).



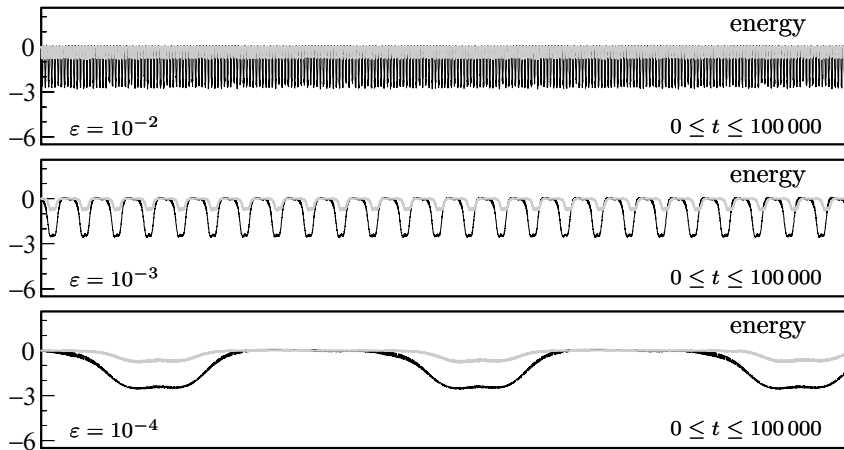
**Fig. 3.2** Relative error  $(I_h(x_n, v_n) - I_h(x_0, v_0))/I_h(x_0, v_0)$  of the modified magnetic moment  $I_h(x, v)$  as a function of time, along the numerical solution of the variational integrator (1.7), obtained with  $h = \eta\varepsilon$ , where  $\eta = 0.8$  (black) and  $\eta = 0.2$  (grey).

#### 4 Modulated Fourier expansion of the exact solution

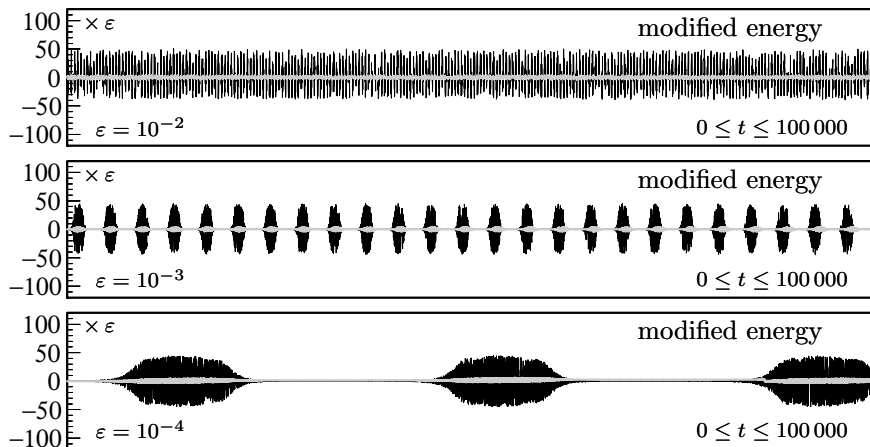
For the solution of (1.1) we consider the *modulated Fourier expansion*

$$x(t) \approx \sum_{k \in \mathbf{Z}} z^k(t) e^{ik\phi(t)/\varepsilon}, \quad (4.1)$$

where  $z^0(t)$  describes the guiding-center motion, and the Larmor rotation is captured in the functions  $z^k(t) e^{ik\phi(t)/\varepsilon}$  for  $k \neq 0$ . We require that the phase function  $\phi(t)$  and the modulation functions  $z^k(t)$  are smooth in the sense that



**Fig. 3.3** Relative error  $(H(x_n, v_n) - H(x_0, v_0))/H(x_0, v_0)$  of the energy  $H(x, v)$  as a function of time, along the numerical solution of the variational integrator (1.7), obtained with  $h = \eta\varepsilon$ , where  $\eta = 0.8$  (black) and  $\eta = 0.2$  (grey).



**Fig. 3.4** Relative error  $(H_h(x_n, v_n) - H_h(x_0, v_0))/H_h(x_0, v_0)$  of the modified energy  $H_h(x, v)$  as a function of time, along the numerical solution of the variational integrator (1.7), obtained with  $h = \eta\varepsilon$ , where  $\eta = 0.8$  (black) and  $\eta = 0.2$  (grey).

all their derivatives are bounded independently of  $\varepsilon$ . Since  $x(t)$  is real, we assume  $z^{-k}(t) = \overline{z^k(t)}$ .

In this section we show how the phase function  $\phi(t)$  and the coefficient functions  $z^k(t)$  have to be chosen to get an accurate approximation of the solution  $x(t)$  of (1.1) over some short time interval. We follow the approach of the varying-frequency modulated Fourier expansion constructed in [7] for a different problem with state-dependent high frequencies. We now encounter the additional difficulties that in the present situation the right-hand side of

(1.1) depends on the velocity and the eigenvectors of the linear map  $v \mapsto \varepsilon^{-1}(v \times B(x))$  depend on  $x$ .

Inserting the ansatz (4.1) and its derivatives into the differential equation (1.1) and comparing the coefficients of  $e^{ik\phi(t)/\varepsilon}$  yields

$$\ddot{z}^k + 2ik\frac{\dot{\phi}}{\varepsilon}z^k + \left(ik\frac{\ddot{\phi}}{\varepsilon} - k^2\frac{\dot{\phi}^2}{\varepsilon^2}\right)z^k = F^k, \quad (4.2)$$

where we use the abbreviation

$$\begin{aligned} F^k &= \frac{1}{\varepsilon} \sum_{k_1+k_2=k} \left( z^{k_1} + ik_1\frac{\dot{\phi}}{\varepsilon}z^{k_1} \right) \times \sum_{\substack{m \geq 0 \\ s(\alpha)=k_2}} \frac{1}{m!} B^{(m)}(z^0) \mathbf{z}^\alpha \\ &\quad - \sum_{\substack{m \geq 0 \\ s(\alpha)=k}} \frac{1}{m!} (\nabla U)^{(m)}(z^0) \mathbf{z}^\alpha. \end{aligned}$$

Here,  $\alpha = (\alpha_1, \dots, \alpha_m)$  is a multi-index with  $\alpha_j \in \mathbf{Z} \setminus \{0\}$ ,  $s(\alpha) = \alpha_1 + \dots + \alpha_m$ , and  $\mathbf{z}^\alpha = (z^{\alpha_1}, \dots, z^{\alpha_m})$ . For  $m = 0$  we interpret  $\alpha$  as the empty set  $\emptyset$ , and we use the convention that  $s(\emptyset) = 0$  and that the argument  $\mathbf{z}^\emptyset$  is not present in the above expressions.

#### 4.1 Construction of the modulated Fourier expansion

Our construction of  $\phi(t)$  and the coefficient functions  $z^k(t)$  in (4.1) is based on asymptotic expansions, which are typically divergent. To obtain rigorous statements we have to truncate the series. We choose a truncation index  $N$  and we consider

$$\tilde{x}(t) = \sum_{|k| \leq N+1} z^k(t) e^{ik\phi(t)/\varepsilon}, \quad (4.3)$$

where  $\phi(t)$  and the coefficient functions  $z^k(t)$  are determined such that the defect when  $\tilde{x}(t)$  is inserted into the differential equation (1.1), is of size  $\mathcal{O}(\varepsilon^N)$ .

We first diagonalize the linear map  $v \mapsto v \times B(x)$ . It has eigenvalues  $\lambda_1 = i|B(x)|$ ,  $\lambda_0 = 0$ , and  $\lambda_{-1} = -i|B(x)|$ . The corresponding normalised eigenvectors are denoted  $v_1(x), v_0(x), v_{-1}(x)$ . Note that  $v_0(x)$  is collinear to  $B(x)$ . We further denote the orthogonal projections onto the eigenspaces by  $P_j(x) = v_j(x)v_j(x)^*$ .

We put

$$\zeta^0 = z^0 - c^0, \quad \zeta^k = \varepsilon^{-|k|} z^k, \quad k \neq 0, \quad (4.4)$$

where  $c^0$  is a constant vector that will be fixed later. We will determine the scaled vectors  $\zeta^k$  such that  $\zeta^k = \mathcal{O}(1)$ . These vectors are written in the time-dependent basis  $v_1(z^0), v_0(z^0), v_{-1}(z^0)$ ,

$$\zeta^k = \zeta_1^k + \zeta_0^k + \zeta_{-1}^k$$

so that  $\zeta_j^k = P_j(z^0)\zeta^k$  is collinear to  $\tilde{v}_j(y^0)$ . As a consequence of  $v_{-1}(y^0) = \overline{v_1(y^0)}$  and the fact that  $v_0(y^0)$  is real, it follows from  $z^{-k} = \overline{z^k}$  that

$$\zeta_{-1}^{-k} = \overline{\zeta_1^k}, \quad \zeta_0^{-k} = \overline{\zeta_0^k}, \quad \zeta_1^{-k} = \overline{\zeta_{-1}^k}. \quad (4.5)$$

Differentiating the identity  $\zeta_j^k = P_j(z^0)\zeta^k$  with respect to time, the relation (4.2), after suitable truncation, yields for  $j \in \{1, 0, -1\}$  (with  $z^0 = c^0 + \zeta^0$ )

$$\begin{aligned} \ddot{\zeta}_j^k - 2\dot{P}_j(z^0)\dot{\zeta}^k - \ddot{P}_j(z^0)(z^0 - c^0) &= P_j(z^0)\tilde{F}^k \\ \left( \ddot{\zeta}_j^k - 2\dot{P}_j(z^0)\dot{\zeta}^k - \ddot{P}_j(z^0)\zeta^k \right) + 2ik\frac{\dot{\phi}}{\varepsilon} \left( \dot{\zeta}_j^k - \dot{P}_j(z^0)\zeta^k \right) \\ &+ \left( ik\frac{\ddot{\phi}}{\varepsilon} - k^2\frac{\dot{\phi}^2}{\varepsilon^2} \right) \zeta_j^k = P_j(z^0)\tilde{F}^k, \end{aligned} \quad (4.6)$$

where we use the notation  $\dot{P}_j(z^0) = \frac{d}{dt}P_j(z^0)$  and similar for  $\ddot{P}_j(z^0)$ . The vector  $\tilde{F}^k$ , written in terms of the scaled variables, is given by

$$\begin{aligned} \tilde{F}^k &= \varepsilon^{-|k|-1} \sum_{k_1+k_2=k} \varepsilon^{|k_1|} \left( \dot{\zeta}^{k_1} + ik_1\frac{\dot{\phi}}{\varepsilon}\zeta^{k_1} \right) \times \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k_2}} \frac{\varepsilon^{|\alpha|}}{m!} B^{(m)}(z^0) \zeta^\alpha \\ &- \varepsilon^{-|k|} \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k}} \frac{\varepsilon^{|\alpha|}}{m!} (\nabla U)^{(m)}(z^0) \zeta^\alpha, \end{aligned}$$

where  $\zeta^\alpha = (\zeta^{\alpha_1}, \dots, \zeta^{\alpha_m})$ ,  $|\alpha| = |\alpha_1| + \dots + |\alpha_m|$ , and all indices  $k, k_1, k_2, \alpha_1, \dots, \alpha_m$  are restricted to the range  $\{-N-1, \dots, N+1\}$  with  $\alpha_j \neq 0$ .

We extract the leading term from  $\tilde{F}^k$ ,

$$\tilde{F}^k = \frac{1}{\varepsilon} \left( \dot{\zeta}^k + ik\frac{\dot{\phi}}{\varepsilon}\zeta^k \right) \times B(z^0) + \widehat{F}^k,$$

and we note that

$$\begin{aligned} P_0(z^0)\tilde{F}^k &= P_0(z^0)\widehat{F}^k, \\ P_{\pm 1}(z^0)\tilde{F}^k &= \mp k \frac{\dot{\phi}|B(z^0)|}{\varepsilon^2} \zeta_{\pm 1}^k \pm i \frac{|B(z^0)|}{\varepsilon} P_{\pm 1}(z^0)\dot{\zeta}^k + P_{\pm 1}(z^0)\widehat{F}^k, \end{aligned}$$

which follows from  $P_0(z^0)(v \times B(z^0)) = 0$  and from  $P_{\pm 1}(z^0)(v \times B(z^0)) = \pm i|B(z^0)|P_{\pm 1}(z^0)v$  for all vectors  $v$ . We further have  $\widehat{F}^0 = \mathcal{O}(1)$ ,  $\widehat{F}^{\pm 1} = \mathcal{O}(\varepsilon^{-1})$ , and  $\widehat{F}^k = \mathcal{O}(\varepsilon^{-2})$  for all other values of  $k$ .

To annihilate the  $\varepsilon^{-2}$  terms in (4.6) for  $j = k = \pm 1$ , we define the phase function  $\phi(t)$  by

$$\phi(t) = \int_0^t |B(z^0(s))| ds, \quad \text{so that} \quad \dot{\phi}(t) = |B(z^0(t))|. \quad (4.7)$$

Our aim is to define smooth coefficient functions  $\zeta_j^k$  of size  $\mathcal{O}(1)$  satisfying (4.6). In each equation we look for the highest derivative appearing in the terms with the lowest power of  $\varepsilon$ .

- $k = 0, j = 0$ : The highest derivative of  $\zeta_0^0$  is the second one and it appears in the term  $\ddot{\zeta}_0^0 - \ddot{P}_0(z^0)(z^0 - c^0)$ .
- $k = 0, j = \pm 1$ : The only  $\varepsilon^{-1}$ -term for  $j = 1$  is  $\varepsilon^{-1}i|B(z^0)|P_1(z^0)\dot{\zeta}^0$  and the highest derivative of  $\zeta_1^0$  is the first one. Using  $P_1(z^0)\dot{\zeta}^0 = \dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0)$ , this term becomes  $\varepsilon^{-1}i|B(z^0)|(\dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0))$ . The case  $j = -1$  is treated in the same way.
- $k = j = \pm 1$ : By (4.7), the  $\varepsilon^{-2}$  terms cancel each other in (4.6). For  $k = j = 1$ , the dominant terms containing  $\dot{\zeta}_1^1$  are  $\varepsilon^{-2}2i\phi\dot{\zeta}_1^1$  (on the left side) and  $\varepsilon^{-1}i|B(z^0)|P_1(z^0)\dot{\zeta}^1$  (on the right side). From  $P_1(z^0)\dot{\zeta}^1 = \dot{\zeta}_1^1 - \dot{P}_1(z^0)\zeta^1$  we thus obtain the first derivative of  $\zeta_1^1$ . The same argument can be repeated for  $k = j = -1$ .
- the remaining values of  $k$  and  $j$ : combining the dominant  $\varepsilon^{-2}$  term from the left and right sides of (4.6) gives  $\varepsilon^{-2}k(j-k)\dot{\phi}^2\zeta_j^k$  for  $j \in \{1, 0, -1\}$ . Hence we obtain algebraic relations for  $\zeta_j^k$ .

This leads to the following system of equations (with  $z^0 = c^0 + \zeta_1^0 + \zeta_0^0 + \zeta_{-1}^0$ ):

$$\ddot{\zeta}_0^0 - \ddot{P}_0(z^0)(z^0 - c^0) = f_0^0(\varepsilon, z^0, \dot{z}^0, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}) \quad (4.8)$$

$$\begin{aligned} \dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0) &= -i\varepsilon|B(z^0)|^{-1}(\ddot{\zeta}_1^0 - \ddot{P}_1(z^0)(z^0 - c^0)) \\ &\quad + \varepsilon f_1^0(\varepsilon, z^0, \dot{z}^0, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}) \end{aligned} \quad (4.9)$$

$$\begin{aligned} \dot{\zeta}_{-1}^0 - \dot{P}_{-1}(z^0)(z^0 - c^0) &= i\varepsilon|B(z^0)|^{-1}(\ddot{\zeta}_{-1}^0 - \ddot{P}_{-1}(z^0)(z^0 - c^0)) \\ &\quad + \varepsilon f_{-1}^0(\varepsilon, z^0, \dot{z}^0, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}) \end{aligned} \quad (4.10)$$

$$\dot{\zeta}_1^1 = f_1^1(\varepsilon, z^0, \dot{z}^0, \ddot{z}^0, \ddot{\zeta}_1^1, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}) \quad (4.11)$$

$$\dot{\zeta}_{-1}^1 = f_{-1}^1(\varepsilon, z^0, \dot{z}^0, \ddot{z}^0, \ddot{\zeta}_{-1}^1, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}) \quad (4.12)$$

$$\zeta_j^k = f_j^k(\varepsilon, z^0, \dot{z}^0, \ddot{z}^0, \ddot{\zeta}_j^k, (\zeta^l, \dot{\zeta}^l)_{l \neq 0}), \quad (4.13)$$

where  $|k| \leq N + 1$  and  $|l| \leq N + 1$ . We note that the functions  $f_j^k$  are polynomials in  $\varepsilon$ . Equation (4.13) is valid for  $|k| \leq N + 1$  and  $j \in \{1, 0, -1\}$  satisfying  $k(j - k) \neq 0$ . For these values of  $k$  and  $j$  the  $\varepsilon$ -independent term of  $f_j^k$  consists of those terms from  $\widehat{F}^k$  for which  $k_1$  and  $\alpha_1, \dots, \alpha_m$  have all the same sign. It thus depends only on  $z^0$  and on  $\zeta^l$  with  $0 < |l| < |k|$ . Moreover,  $f_j^k$  starts with the  $\mathcal{O}(\varepsilon)$ -term for  $|k| = 1$  and  $j \neq k$ .

To obtain a regular differential equation we have to eliminate recursively the algebraic variables and the higher derivatives of the differential variables. We start with the highest index  $|k| = N + 1$ . Using the relation (4.13) for such an index and its differentiated versions, we can recursively push the dependence on  $\zeta_j^{\pm(N+1)}$  (for  $j \in \{1, 0, -1\}$ ) and on its derivatives to higher powers of  $\varepsilon$  until it contains a sufficiently large power of  $\varepsilon$ . For decreasing  $|k|$  we then eliminate all  $\zeta_j^k$  (and their derivatives) with  $k(j - k) \neq 0$  from the right-hand side of the system (4.8)-(4.13).

We next consider the equations (4.11) and (4.12). The  $\varepsilon$ -independent term of  $f_{\pm 1}^{\pm 1}$  depends on  $\zeta_{\pm 1}^{\pm 1}$ , but not on its derivatives. Therefore, these equations permit us to eliminate the first and higher derivatives of  $\zeta_{\pm 1}^{\pm 1}$  from the whole system (4.8)-(4.13).

For the functions  $g_j = \dot{\zeta}_j^0 - \dot{P}_j(z^0)(z^0 - c^0)$  we have  $\dot{g}_j = \ddot{\zeta}_j^0 + \ddot{P}_j(z^0)(z^0 - c^0) + \dot{P}_j(z^0)\dot{z}^0$ . Including  $\dot{P}_j(z^0)\dot{z}^0$  in the expression  $f_j^0$ , the equations (4.9)-(4.10) become  $g_j = \pm i\varepsilon|B(z^0)|^{-1}\dot{g}_j + \dots$ . By differentiation and substitution the first term in the right-hand side of (4.9)-(4.10) can be removed.

Carrying out all these operations, the previous system of equations becomes

$$\ddot{\zeta}_0^0 - \ddot{P}_0(z^0)(z^0 - c^0) = f_0^0(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.14)$$

$$\dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0) = \varepsilon f_1^0(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.15)$$

$$\dot{\zeta}_{-1}^0 - \dot{P}_{-1}(z^0)(z^0 - c^0) = \varepsilon f_{-1}^0(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.16)$$

$$\dot{\zeta}_1^1 = f_1^1(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N+1}) \quad (4.17)$$

$$\dot{\zeta}_{-1}^{-1} = f_{-1}^{-1}(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N+1}) \quad (4.18)$$

$$\zeta_j^k = f_j^k(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N-|k|+2}), \quad (4.19)$$

where the three dots stand for a finite number of derivatives of  $z^0$ . Recalling that  $z^0 = c^0 + \zeta_1^0 + \zeta_0^0 + \zeta_{-1}^0$ , and that  $\ddot{P}_0(z^0) + \ddot{P}_1(z^0) + \ddot{P}_{-1}(z^0) = 0$ , the sum of (4.14) and of the differentiated equations (4.15)-(4.16) gives

$$\ddot{z}^0 = f^0(\varepsilon, z^0, \dot{z}^0, \dots, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N). \quad (4.20)$$

Since the  $\varepsilon$ -independent term of the polynomial  $f^0(\varepsilon, \dots)$  only depends on  $z^0$ ,  $\dot{z}^0$ ,  $\zeta_1^1$ , and  $\zeta_{-1}^{-1}$ , second and higher derivatives of  $z^0$  can be eliminated in the right-hand sides of (4.20) as well as from those of the system (4.14)-(4.19), and we obtain the differential-algebraic system

$$\ddot{\zeta}_0^0 - \ddot{P}_0(z^0)(z^0 - c^0) = f_0^0(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.21)$$

$$\dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0) = \varepsilon f_1^0(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.22)$$

$$\dot{\zeta}_{-1}^0 - \dot{P}_{-1}(z^0)(z^0 - c^0) = \varepsilon f_{-1}^0(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^N) \quad (4.23)$$

$$\dot{\zeta}_1^1 = f_1^1(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N+1}) \quad (4.24)$$

$$\dot{\zeta}_{-1}^{-1} = f_{-1}^{-1}(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N+1}) \quad (4.25)$$

$$\zeta_j^k = f_j^k(\varepsilon, z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}) + \mathcal{O}(\varepsilon^{N-|k|+2}). \quad (4.26)$$

Removing the  $\mathcal{O}(\cdot)$ -terms in (4.21)-(4.26), we obtain a system that is no longer singularly perturbed, i.e., there is no division by  $\varepsilon$ . We still have to prove the solvability of this system. In equations (4.22)-(4.23) the vectors  $\dot{\zeta}_1^0$  and  $\dot{\zeta}_{-1}^0$  appear as  $\dot{\zeta}_j^0 - (P_j'(z^0)(\dot{\zeta}_1^0 + \dot{\zeta}_{-1}^0))(z^0 - c^0)$  in the left-hand side, and multiplied by  $\varepsilon$  in the right-hand side. If  $z^0 - c^0$  is sufficiently small we obtain a unique solution for  $\dot{\zeta}_1^0$ ,  $\dot{\zeta}_{-1}^0$  that is bounded uniformly for  $\varepsilon \rightarrow 0$ . Similarly, the left-hand side of (4.21) contains the second derivative of  $\zeta_0^0$  as  $\ddot{\zeta}_0^0 - (P_0'(z^0)\ddot{\zeta}_0^0)(z^0 - c^0)$ , which permits us to compute  $\ddot{\zeta}_0^0$  provided that  $z^0 - c^0$  is sufficiently small.

This is the correct place to discuss the choice of  $c^0$ , which until now has been treated as an arbitrary vector. Since  $z^0(t)$  approximates the solution  $x(t)$  of the differential equation, a natural choice for  $c^0$  is  $c^0 = x(0)$ . This implies that  $z^0(t)$  is close to  $c^0$  as long as  $|t|$  is not too large.

**Initial values.** For the differential-algebraic system, described above, we need initial values  $\zeta_0^0(0)$ ,  $\dot{\zeta}_0^0(0)$ ,  $\zeta_1^0(0)$ ,  $\zeta_{-1}^0(0)$ ,  $\zeta_1^1(0)$ ,  $\zeta_{-1}^{-1}(0)$  that are bounded independently of  $\varepsilon$ . Assuming  $\phi(0) = 0$ , they have to be computed from

$$\sum_{|k| \leq N+1} z^k(0) = x(0), \quad \sum_{|k| \leq N+1} \left( \dot{z}^k(0) + ik \frac{\dot{\phi}(0)}{\varepsilon} z^k(0) \right) = \dot{x}(0). \quad (4.27)$$

Using the algebraic relations (4.26) with removed remainder term, the first equation of (4.27) can be written as

$$z^0(0) = x(0) + \varepsilon G_0(\varepsilon, z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)). \quad (4.28)$$

Multiplying the second relation with  $P_0(z^0(0)) = P_0(x(0)) + \mathcal{O}(\varepsilon)$  and using the fact that  $\zeta_{\pm 1}^{\pm 1}(0) = \mathcal{O}(\varepsilon)$ , we obtain

$$P_0(z^0(0)) \dot{z}^0(0) = P_0(x(0)) \dot{x}(0) + \varepsilon \widehat{G}_0(\varepsilon, z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)). \quad (4.29)$$

Differentiating  $\zeta_0^0 = P_0(z^0) \zeta^0$  with respect to time yields

$$\dot{z}^0 = \dot{\zeta}^0 = \dot{\zeta}_0^0 + \dot{\zeta}_1^0 + \dot{\zeta}_{-1}^0 = P_0(z^0) \dot{z}^0 + \dot{P}_0(z^0)(z^0 - c^0) + \dot{\zeta}_1^0 + \dot{\zeta}_{-1}^0.$$

Inserting (4.29) and (4.22)-(4.23) with remainder terms removed, and using  $\dot{P}_0(z^0) + \dot{P}_1(z^0) + \dot{P}_{-1}(z^0) = 0$  we obtain the equation

$$\dot{z}^0(0) = P_0(x(0)) \dot{x}(0) + \varepsilon G_1(\varepsilon, z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)). \quad (4.30)$$

Multiplying the second relation of (4.27) with  $P_{\pm 1}(z^0(0))$  gives

$$P_{\pm 1}(z^0(0)) \dot{z}^0(0) \pm i |B(x(0))| \zeta_{\pm 1}^{\pm 1}(0) = P_{\pm 1}(z^0(0)) \dot{x}(0) + \varepsilon \widehat{G}_{\pm}(\varepsilon, z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)). \quad (4.31)$$

Inserting (4.30) and using  $P_{\pm 1}(z^0(0)) P_0(x(0)) = \mathcal{O}(\varepsilon)$  yields

$$\zeta_{\pm 1}^{\pm 1}(0) = \mp i |B(x(0))|^{-1} P_{\pm 1}(x(0)) \dot{x}(0) + \varepsilon G_{\pm}(\varepsilon, z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)). \quad (4.32)$$

Fixed point iteration applied to the system (4.28)-(4.30)-(4.32) gives initial values  $z^0(0)$ ,  $\dot{z}^0(0)$ ,  $\zeta_1^1(0)$ ,  $\zeta_{-1}^{-1}(0)$ . They provide  $\zeta_j^0(0) = P_j(z^0(0))(z^0(0) - c^0)$  and  $\dot{\zeta}_0^0(0) = P_0(z^0(0)) \dot{z}^0(0) + \dot{P}_0(z^0)|_{t=0}(z^0(0) - c^0)$ .

The autonomous differential-algebraic system (4.21)-(4.26), with remainder term removed, together with these initial values define the coefficient functions  $\zeta_j^k(t)$  and hence also  $z_j^k(t)$  of the modulated Fourier expansion on an  $\varepsilon$ -independent non-empty interval.

**Uniqueness of the modulated Fourier expansion.** The above construction shows that, for a given  $c^0$ , the coefficient functions  $\zeta_j^k$  are uniquely determined. Obviously, the functions  $\zeta_j^0$  depend on the choice of  $c^0$ . We show here that in spite of this dependence, the function  $z^0(t)$  and all  $z^k(t)$  of (4.1) are independent of  $c^0$ .

We first note that the functions  $f_j^k$  of (4.8)-(4.13) as well as those of (4.21)-(4.26) do not explicitly depend on the parameter  $c^0$ . This is also true for the function  $f^0$  of (4.20). Therefore, the equation (4.20), where second and higher derivatives of  $z^0$  are recursively removed, together with (4.24) and (4.25) constitute a regular ordinary differential equation for  $z^0, \dot{z}^0, \zeta_1^1, \zeta_{-1}^{-1}$  that does not depend on  $c^0$  as parameter. Concerning the initial values, we note that also the functions  $G_0, G_1$ , and  $G_{\pm}$  are independent of  $c^0$ . Therefore, the values  $z^0(0), \dot{z}^0(0), \zeta_1^1(0), \zeta_{-1}^{-1}(0)$ , obtained from (4.28)-(4.30)-(4.32) do not depend on  $c^0$ .

#### 4.2 Bounds for the coefficient functions and for the remainder

The construction of Section 4.1 yields bounds for the coefficient functions that are collected in the following theorem.

**Theorem 4.1** *Let  $x(t)$  be a solution of (1.1) that satisfies the bounded energy condition (1.2) and stays in a compact set  $K$  for  $0 \leq t \leq T_\varepsilon$  with  $T_\varepsilon = \mathcal{O}(\varepsilon)$ . Then, we have*

$$x(t) = \sum_{|k| \leq N+1} z^k(t) e^{ik\phi(t)/\varepsilon} + R_N(t) \quad (4.33)$$

for arbitrary  $N \geq 1$ , where the phase function is given by (4.7). The functions  $z^k(t)$  together with their derivatives (up to order  $N$ ) are bounded as

$$z^k = \mathcal{O}(\varepsilon^{|k|}) \quad \text{for all } |k| \leq N+1 \quad (4.34)$$

and further satisfy

$$\dot{z}^0 \times B(z^0) = \mathcal{O}(\varepsilon), \quad P_j(z^0)z^k = \mathcal{O}(\varepsilon^2) \quad \text{for } |k| = 1, j \neq k. \quad (4.35)$$

The remainder term and its derivative are bounded by

$$R_N(t) = \mathcal{O}(t^2\varepsilon^N), \quad \dot{R}_N(t) = \mathcal{O}(t\varepsilon^N) \quad \text{for } 0 \leq t \leq T_\varepsilon. \quad (4.36)$$

The functions  $z^k$  are unique up to  $\mathcal{O}(\varepsilon^{N+2})$ . The constants symbolised by the  $\mathcal{O}$ -notation are independent of  $\varepsilon$  and  $t$  with  $0 \leq t \leq T_\varepsilon$ , but they depend on  $N$ , on the constants in (1.2), on bounds of derivatives of  $B(x)$  and  $E(x)$ , and on  $T_\varepsilon$ .

*Proof* The construction of Section 4.1 yields the bounds (4.34) and the uniqueness up to  $\mathcal{O}(\varepsilon^{N+2})$  of the coefficient functions. The first relation of (4.35) follows from (4.2) with  $k = 0$  by multiplying with  $\varepsilon$  and then considering the  $\varepsilon$ -independent term. The improved estimates for  $|k| = 1$  and  $j \neq k$  are a consequence of the fact that the corresponding functions  $f_j^k$  in (4.13) contain a factor  $\varepsilon$ .

The above construction shows that the truncated series (4.3) satisfies the differential equation (1.1) up to a defect of size  $\mathcal{O}(\varepsilon^N)$ . Since the Lipschitz constant of vector field (1.1), when written as a first order system, is of size  $\mathcal{O}(\varepsilon^{-1})$ , an application of Gronwall's lemma proves the bounds (4.36) on intervals of length  $\mathcal{O}(\varepsilon)$ .  $\square$



### 4.3 The modulation system as Euler–Lagrange equation

With the functions  $\phi(t)$  and  $z^k(t)$ , constructed in Section 4.1, we let

$$y^k(t) = z^k(t)e^{ik\phi(t)/\varepsilon} \quad \text{for} \quad |k| \leq N+1,$$

and we put  $y^k(t) = 0$  for  $|k| > N+1$ . For  $|k| \leq N+1$  the equations (4.2) then have the form

$$\begin{aligned} \dot{y}^k &= \frac{1}{\varepsilon} \sum_{k_1+k_2=k} \dot{y}^{k_1} \times \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k_2}} \frac{1}{m!} B^{(m)}(y^0) \mathbf{y}^\alpha \\ &\quad - \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k}} \frac{1}{m!} (\nabla U)^{(m)}(y^0) \mathbf{y}^\alpha + \mathcal{O}(\varepsilon^N), \end{aligned} \quad (4.37)$$

where  $\alpha = (\alpha_1, \dots, \alpha_m)$  with  $\alpha_j \in \mathbf{Z} \setminus \{0\}$ , and  $\mathbf{y}^\alpha = (y^{\alpha_1}, \dots, y^{\alpha_m})$ . In this section it is essential that the magnetic field  $B(x) = \nabla_x \times A(x)$  is given by a vector potential  $A(x)$ . For the vector  $\mathbf{y} = (y^k)_{k \in \mathbf{Z}}$  we then consider the extended potentials

$$\begin{aligned} \mathcal{U}(\mathbf{y}) &= \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=0}} \frac{1}{m!} U^{(m)}(y^0) \mathbf{y}^\alpha \\ \mathcal{A}(\mathbf{y}) &= \left( \sum_{\substack{0 \leq m \leq N+2 \\ s(\alpha)=k}} \frac{1}{m!} A^{(m)}(y^0) \mathbf{y}^\alpha \right)_{k \in \mathbf{Z}} = \left( \mathcal{A}_k(\mathbf{y}) \right)_{k \in \mathbf{Z}}. \end{aligned} \quad (4.38)$$

**Theorem 4.2** *Under the assumptions of Theorem 4.1 the system (4.37) can be written as the Euler–Lagrange equations*

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{y}^k}(\mathbf{y}, \dot{\mathbf{y}}) = \frac{\partial \mathcal{L}}{\partial y^k}(\mathbf{y}, \dot{\mathbf{y}}) + \mathcal{O}(\varepsilon^N) \quad (4.39)$$

for the Lagrangian

$$\mathcal{L}(\mathbf{y}, \dot{\mathbf{y}}) = \frac{1}{2} \dot{\mathbf{y}}^* \dot{\mathbf{y}} + \frac{1}{\varepsilon} \mathcal{A}(\mathbf{y})^* \dot{\mathbf{y}} - \mathcal{U}(\mathbf{y}). \quad (4.40)$$

*Proof* With (4.40) the Euler–Lagrange equations (4.39) become

$$\dot{y}^k + \frac{1}{\varepsilon} \sum_{j \in \mathbf{Z}} \frac{\partial \mathcal{A}_k}{\partial y^j}(\mathbf{y}) \dot{y}^j = \frac{1}{\varepsilon} \sum_{j \in \mathbf{Z}} \left( \frac{\partial \mathcal{A}_j}{\partial y^k}(\mathbf{y}) \right)^* \dot{y}^j - \left( \frac{\partial \mathcal{U}}{\partial y^k}(\mathbf{y}) \right)^* + \mathcal{O}(\varepsilon^N), \quad (4.41)$$

where we have used that  $\mathcal{A}_{-k}(\mathbf{y})$  is the complex conjugate of  $\mathcal{A}_k(\mathbf{y})$ . We note that<sup>2</sup>

$$\frac{\partial \mathcal{A}_k}{\partial y^j}(\mathbf{y}) = \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k-j}} \frac{1}{m!} A^{(m+1)}(y^0) \mathbf{y}^\alpha$$

<sup>2</sup> Actually, for  $j = 0$  the sum is over  $0 \leq m \leq N+2$ . However, the term with  $m = N+2$  is of size  $\mathcal{O}(\varepsilon^{N+2})$  and can be absorbed in the remainder.

where  $\alpha = (\alpha_1, \dots, \alpha_m)$ . This expression only depends on the difference  $k - j$ , so that  $\partial \mathcal{A}_j / \partial y^k$  is the complex conjugate of  $\partial \mathcal{A}_k / \partial y^j$ . Consequently, (4.41) can be written as

$$\begin{aligned} \ddot{y}^k + \frac{1}{\varepsilon} \sum_{j \in \mathbf{Z}} \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha) = k-j}} \frac{1}{m!} \left( A^{(m+1)}(y^0) \mathbf{y}^\alpha - (A^{(m+1)}(y^0) \mathbf{y}^\alpha)^\top \right) \dot{y}^j \\ = - \left( \frac{\partial \mathcal{U}}{\partial y^k}(\mathbf{y}) \right)^* + \mathcal{O}(\varepsilon^N). \end{aligned}$$

Because of  $v \times B(x) = (A'(x)^\top - A'(x))v$ , and consequently

$$v \times B^{(m)}(x) \mathbf{y}^\alpha = \left( (A^{(m+1)}(x) \mathbf{y}^\alpha)^\top - A^{(m+1)}(x) \mathbf{y}^\alpha \right) v, \quad (4.42)$$

the above equation is seen to be equivalent to (4.37).  $\square$

#### 4.4 Almost-invariant of the modulated Fourier expansion

With the mapping  $S(\lambda) \mathbf{y} = (e^{ik\lambda} y^k)_{k \in \mathbf{Z}}$  we have the invariance properties

$$\mathcal{U}(S(\lambda) \mathbf{y}) = \mathcal{U}(\mathbf{y}), \quad \mathcal{A}(S(\lambda) \mathbf{y}) = S(\lambda) \mathcal{A}(\mathbf{y}) \quad \text{for all } \lambda.$$

Differentiation with respect to  $\lambda$  (at  $\lambda = 0$ ) yields

$$\sum_{k \in \mathbf{Z}} ik \frac{\partial \mathcal{U}}{\partial y^k}(\mathbf{y}) y^k = 0 \quad (4.43)$$

$$\sum_{j \in \mathbf{Z}} ij \frac{\partial \mathcal{A}_k}{\partial y^j}(\mathbf{y}) y^j = ik \mathcal{A}_k(\mathbf{y}) \quad \text{for } k \in \mathbf{Z}. \quad (4.44)$$

Here, we assume that only finitely many  $y^k$  are non-zero. The expression (motivated by Noether's theorem)

$$\mathcal{I}(\mathbf{y}, \dot{\mathbf{y}}) = \left( \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{y}}} \right)^* \frac{d}{d\lambda} \Big|_{\lambda=0} S(\lambda) \mathbf{y} = \frac{1}{\varepsilon} \sum_{k \in \mathbf{Z}} \left( \dot{y}^k + \frac{1}{\varepsilon} \mathcal{A}_k(\mathbf{y}) \right)^* ik y^k \quad (4.45)$$

then satisfies

$$\frac{d}{dt} \mathcal{I}(\mathbf{y}, \dot{\mathbf{y}}) = \frac{1}{\varepsilon} \sum_{k \in \mathbf{Z}} ik \left( \left( \dot{y}^k + \frac{1}{\varepsilon} \frac{d}{dt} \mathcal{A}_k(\mathbf{y}) \right)^* y^k + \frac{1}{\varepsilon} \mathcal{A}_k(\mathbf{y})^* \dot{y}^k \right),$$

because the sum of the terms for  $k$  and  $-k$  vanishes in  $\sum_{k \in \mathbf{Z}} ik |\dot{y}^k|^2$ . For the solution of the Euler–Lagrange equations (4.39) (or equivalently (4.41)) this shows that

$$\begin{aligned} \frac{d}{dt} \mathcal{I}(\mathbf{y}, \dot{\mathbf{y}}) &= \frac{1}{\varepsilon} \sum_{k \in \mathbf{Z}} ik \left( \frac{1}{\varepsilon} \sum_{j \in \mathbf{Z}} \left( \frac{\partial \mathcal{A}_j}{\partial y^k}(\mathbf{y}) \right)^* \dot{y}^j - \left( \frac{\partial \mathcal{U}}{\partial y^k}(\mathbf{y}) \right)^* \right)^* y^k \\ &\quad + \frac{1}{\varepsilon^2} \sum_{k \in \mathbf{Z}} ik \mathcal{A}_k(\mathbf{y})^* \dot{y}^k + \mathcal{O}(\varepsilon^N). \end{aligned}$$

The sum containing the derivative of  $\mathcal{U}(\mathbf{y})$  vanishes by (4.43). In the arising double sum we exchange the (finite) summation, we use the identity (4.44) and the fact that  $\overline{y^k} = y^{-k}$  and  $\overline{\mathcal{A}_k} = \mathcal{A}_{-k}$ . This proves that

$$\frac{d}{dt}\mathcal{I}(\mathbf{y}, \dot{\mathbf{y}}) = \mathcal{O}(\varepsilon^N). \quad (4.46)$$

**Theorem 4.3** *Under the assumption of Theorem 4.1 we have*

$$\begin{aligned} \mathcal{I}(\mathbf{y}(t), \dot{\mathbf{y}}(t)) &= \mathcal{I}(\mathbf{y}(0), \dot{\mathbf{y}}(0)) + \mathcal{O}(t\varepsilon^N) \\ \mathcal{I}(\mathbf{y}(t), \dot{\mathbf{y}}(t)) &= I(x(t), \dot{x}(t)) + \mathcal{O}(\varepsilon). \end{aligned}$$

The constants symbolised by  $\mathcal{O}$  are independent of  $\varepsilon$  and  $t$  with  $0 \leq t \leq T_\varepsilon$  (recall that  $T_\varepsilon = \mathcal{O}(\varepsilon)$ ), but depend on the truncation index  $N$ .

*Proof* The first statement follows by integration of (4.46). To prove the second statement we write  $I(x(t), \dot{x}(t))$  in terms of the coefficient functions  $z_j^k$ . The bounds of Theorem 4.1 yield

$$x = z^0 + \mathcal{O}(\varepsilon), \quad \dot{x} = \dot{z}^0 + \mathbf{i} \frac{\dot{\phi}}{\varepsilon} z_1^1 e^{\mathbf{i}\phi/\varepsilon} - \mathbf{i} \frac{\dot{\phi}}{\varepsilon} z_{-1}^{-1} e^{-\mathbf{i}\phi/\varepsilon} + \mathcal{O}(\varepsilon).$$

Consequently, we have

$$\dot{x} \times B(x) = \dot{z}^0 \times B(z^0) - \frac{|B(z^0)|^2}{\varepsilon} \left( z_1^1 e^{\mathbf{i}\phi/\varepsilon} + z_{-1}^{-1} e^{-\mathbf{i}\phi/\varepsilon} \right) + \mathcal{O}(\varepsilon).$$

From (4.35) and from the orthogonality of  $z_1^1$  and  $z_{-1}^{-1}$  it follows that

$$I(x, \dot{x}) = \frac{1}{2} \frac{|\dot{x} \times B(x)|^2}{|B(x)|^3} = |B(z^0)| \frac{|z_1^1|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon).$$

The dominant terms for  $\mathcal{I}(\mathbf{y}(t), \dot{\mathbf{y}}(t))$  are for  $k = \pm 1$ , so that

$$\begin{aligned} \mathcal{I}(\mathbf{y}, \dot{\mathbf{y}}) &= \frac{\mathbf{i}}{\varepsilon} \left( \mathbf{i} \frac{\dot{\phi}}{\varepsilon} z_1^1 + \frac{1}{\varepsilon} A'(z^0) z_1^1 \right)^* z_1^1 - \frac{\mathbf{i}}{\varepsilon} \left( -\mathbf{i} \frac{\dot{\phi}}{\varepsilon} z_{-1}^{-1} + \frac{1}{\varepsilon} A'(z^0) z_{-1}^{-1} \right)^* z_{-1}^{-1} + \mathcal{O}(\varepsilon) \\ &= \frac{|B(z^0)|}{\varepsilon^2} \left( |z_1^1|^2 + |z_{-1}^{-1}|^2 \right) + \frac{\mathbf{i}}{\varepsilon^2} (z_1^1)^* \left( A'(z^0)^\top - A'(z^0) \right) z_1^1 + \mathcal{O}(\varepsilon) \\ &= \frac{2|B(z^0)|}{\varepsilon^2} |z_1^1|^2 + \frac{\mathbf{i}}{\varepsilon^2} (z_1^1)^* \left( z_1^1 \times B(z^0) \right) + \mathcal{O}(\varepsilon) \\ &= |B(z^0)| \frac{|z_1^1|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon). \end{aligned}$$

This implies the second statement of the theorem.  $\square$

#### 4.5 From short to long time intervals

We are now prepared to prove that the magnetic moment (1.4) is an adiabatic invariant (Theorem 2.1). We divide the interval  $[0, T]$  into small intervals of length  $\mathcal{O}(\varepsilon)$ , and we put  $t_n = n\varepsilon$ . On each subinterval  $[t_n, t_{n+1}]$  we consider the modulated Fourier expansion corresponding to initial values at  $t_n$  given by the exact solution of (1.1), and for  $\phi(t_n)$  we take the value at  $t_n$  of the phase function from the previous subinterval. We denote its coefficient functions by  $\mathbf{y}_n(t)$ . The uniqueness of the modulation system up to  $\mathcal{O}(\varepsilon^{N+2})$  implies that  $\mathbf{y}_n(t_n) = \mathbf{y}_{n-1}(t_n) + \mathcal{O}(\varepsilon^{N+2})$ . The first statement of Theorem 4.3 then shows that  $\mathcal{I}(\mathbf{y}_n(t_n), \dot{\mathbf{y}}_n(t_n))$  differs from  $\mathcal{I}(\mathbf{y}_{n-1}(t_n), \dot{\mathbf{y}}_{n-1}(t_n))$  only by  $\mathcal{O}(\varepsilon\varepsilon^N)$ . Summing up these differences yields

$$\mathcal{I}(\mathbf{y}_n(t), \dot{\mathbf{y}}_n(t)) = \mathcal{I}(\mathbf{y}_0(0), \dot{\mathbf{y}}_0(0)) + \mathcal{O}(t\varepsilon^N) \quad (4.47)$$

for  $t \in [t_n, t_{n+1}]$ . The second statement of Theorem 4.3 then gives

$$\begin{aligned} I(x(t), \dot{x}(t)) &= I(x(0), \dot{x}(0)) + \mathcal{O}(\varepsilon) + \mathcal{O}(t\varepsilon^N) \\ &= I(x(0), \dot{x}(0)) + \mathcal{O}(\varepsilon) \quad \text{for } t \leq \varepsilon^{-N+1}. \end{aligned} \quad (4.48)$$

If we replace the arbitrarily chosen integer  $N$  by  $N+1$ , this is the result stated in Theorem 2.1.

### 5 Modulated Fourier expansion of the variational integrator

For the proof of near-conservation of the magnetic moment and the energy (see Theorems 2.2 and 2.3) we extend the modulated Fourier expansion of Section 4 to the numerical solution of the variational integrator.

With the finite difference notation  $\delta_{2h}x_n = (x_{n+1} - x_{n-1})/2h$  and  $\delta_h^2x_n = (x_{n+1} - 2x_n + x_{n-1})/h^2$  the variational integrator (1.7) becomes

$$\delta_h^2x_n = \frac{1}{\varepsilon} \left( A'(x_n)^\top \delta_{2h}x_n - \delta_{2h}A(x_n) \right) - \nabla U(x_n), \quad (5.1)$$

and, because of  $v \times B(x) = (A'(x)^\top - A'(x))v$ , can also be written as

$$\delta_h^2x_n = \frac{1}{\varepsilon} \delta_{2h}x_n \times B(x_n) - \nabla U(x_n) + \frac{1}{\varepsilon} \left( A'(x_n) \delta_{2h}x_n - \delta_{2h}A(x_n) \right). \quad (5.2)$$

For its solution we consider the *modulated Fourier expansion*

$$x_n \approx \sum_{k \in \mathbf{Z}} z^k(t) e^{ik\phi(t)/\varepsilon} = \sum_{k \in \mathbf{Z}} y^k(t) \quad (5.3)$$

where  $y^k(t) = z^k(t) e^{ik\phi(t)/\varepsilon}$  and  $t = t_n$ . We use the same notation  $z^k(t)$  and  $\phi(t)$  as in Section 4, but here these functions depend on the stepsize  $h$  and on

the quotient  $\eta = h/\varepsilon$ . For the finite differences we have

$$\begin{aligned}\delta_{2h}y^k(t) &= \frac{y^k(t+h) - y^k(t-h)}{2h} = e^{ik\phi(t)/\varepsilon} \sum_{l \geq 0} \varepsilon^{l-1} c_l^k(t) \frac{d^l}{dt^l} z^k(t) \\ \delta_h^2 y^k(t) &= \frac{y^k(t+h) - 2y^k(t) + y^k(t-h)}{h^2} = e^{ik\phi(t)/\varepsilon} \sum_{l \geq 0} \varepsilon^{l-2} d_l^k(t) \frac{d^l}{dt^l} z^k(t),\end{aligned}$$

where  $c_{2j}^0 = 0$ ,  $c_{2j+1}^0 = \eta^{2j}/(2j+1)!$ , and  $d_0^0 = 0$ ,  $d_{2j}^0 = 2\eta^{2j-2}/(2j)!$ ,  $d_{2j+1}^0 = 0$ . The first coefficients for  $k \neq 0$  are

$$\begin{aligned}c_0^k(t) &= \frac{i}{\eta} \sin(k\eta\dot{\phi}(t)) - \varepsilon \frac{k\eta}{2} \sin(k\eta\dot{\phi}(t)) \ddot{\phi}(t) + \mathcal{O}(\varepsilon^2) \\ c_1^k(t) &= \cos(k\eta\dot{\phi}(t)) + \mathcal{O}(\varepsilon) \\ d_0^k(t) &= -\frac{4}{\eta^2} \sin^2\left(\frac{k\eta\dot{\phi}(t)}{2}\right) + i\varepsilon k \cos(k\eta\dot{\phi}(t)) \ddot{\phi}(t) + \mathcal{O}(\varepsilon^2) \\ d_1^k(t) &= \frac{2i}{\eta} \sin(k\eta\dot{\phi}(t)) + \mathcal{O}(\varepsilon).\end{aligned}\tag{5.4}$$

Note that these coefficients depend on  $\eta$ ,  $\varepsilon$ , and  $t$  via derivatives of  $\phi(t)$ .

Inserting the modulated Fourier expansion into (5.2) and comparing the coefficients of  $e^{ik\phi(t)/\varepsilon}$  yields (omitting the argument  $t$ )

$$\sum_{l \geq 0} \varepsilon^{l-2} d_l^k \frac{d^l}{dt^l} z^k = F^k\tag{5.5}$$

where the right-hand side is given by

$$\begin{aligned}F^k &= \frac{1}{\varepsilon} \sum_{k_1+k_2=k} \left( \sum_{l \geq 0} \varepsilon^{l-1} c_l^{k_1} \frac{d^l}{dt^l} z^{k_1} \right) \times \sum_{\substack{m \geq 0 \\ s(\alpha)=k_2}} \frac{1}{m!} B^{(m)}(z^0) \mathbf{z}^\alpha \\ &\quad - \sum_{\substack{m \geq 0 \\ s(\alpha)=k}} \frac{1}{m!} (\nabla U)^{(m)}(z^0) \mathbf{z}^\alpha \\ &\quad + \frac{1}{\varepsilon} \sum_{k_1+k_2=k} \left( \sum_{\substack{m \geq 0 \\ s(\alpha)=k_2}} \frac{1}{m!} A^{(m+1)}(z^0) \mathbf{z}^\alpha \right) \left( \sum_{l \geq 0} \varepsilon^{l-1} c_l^{k_1} \frac{d^l}{dt^l} z^{k_1} \right) \\ &\quad - \frac{1}{\varepsilon} \sum_{l \geq 0} \varepsilon^{l-1} c_l^k \frac{d^l}{dt^l} \left( \sum_{\substack{m \geq 0 \\ s(\alpha)=k}} \frac{1}{m!} A^{(m)}(z^0) \mathbf{z}^\alpha \right).\end{aligned}$$

These equations are the analogue of (4.2). The main difference is that there are additional terms with derivatives of  $z^k$  higher than 2 which, however, are multiplied by some power of  $\varepsilon$ .

### 5.1 Construction of the modulated Fourier expansion

We closely follow the construction of Section 4 for the exact solution. We consider a truncated modulated Fourier expansion (4.3), we consider the projections  $P_j(z^0)$  onto the eigenspaces of the linear mapping  $v \mapsto v \times B(z^0)$ , and we introduce scaled vectors  $\zeta^k$  as in (4.4), which are decomposed as  $\zeta^k = \zeta_1^k + \zeta_0^k + \zeta_{-1}^k$ . Written in the scaled variables the above system becomes

$$\begin{aligned} \ddot{\zeta}_j^0 - 2\dot{P}_j(z^0)\dot{z}^0 - \ddot{P}_j(z^0)(z^0 - c^0) + P_j(z^0) \sum_{l \geq 4} \varepsilon^{l-2} d_l^0 \frac{d^l}{dt^l} \zeta^0 &= P_j(z^0) \tilde{F}^0 \\ \frac{1}{\varepsilon^2} d_0^k \zeta_j^k + \frac{1}{\varepsilon} d_1^k (\dot{\zeta}_j^k - \dot{P}_j(z^0) \zeta^k) + P_j(z^0) \sum_{l \geq 2} \varepsilon^{l-2} d_l^k \frac{d^l}{dt^l} \zeta^k &= P_j(z^0) \tilde{F}^k. \end{aligned} \tag{5.6}$$

The vector  $\tilde{F}^k$ , written in terms of the scaled variables, is given by

$$\begin{aligned} \tilde{F}^k &= \varepsilon^{-|k|-1} \sum_{k_1+k_2=k} \varepsilon^{|k_1|} \left( \sum_{l \geq 0} \varepsilon^{l-1} c_l^{k_1} \frac{d^l}{dt^l} \zeta^{k_1} \right) \times \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k_2}} \frac{\varepsilon^{|\alpha|}}{m!} B^{(m)}(z^0) \zeta^\alpha \\ &- \varepsilon^{-|k|} \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k}} \frac{\varepsilon^{|\alpha|}}{m!} (\nabla U)^{(m)}(z^0) \zeta^\alpha \\ &+ \varepsilon^{-|k|-1} \sum_{k_1+k_2=k} \varepsilon^{|k_1|} \left( \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k_2}} \frac{\varepsilon^{|\alpha|}}{m!} A^{(m+1)}(z^0) \zeta^\alpha \right) \left( \sum_{l \geq 0} \varepsilon^{l-1} c_l^{k_1} \frac{d^l}{dt^l} \zeta^{k_1} \right) \\ &- \varepsilon^{-|k|-1} \sum_{l \geq 0} \varepsilon^{l-1} c_l^k \frac{d^l}{dt^l} \left( \sum_{\substack{0 \leq m \leq N+1 \\ s(\alpha)=k}} \frac{\varepsilon^{|\alpha|}}{m!} A^{(m)}(z^0) \zeta^\alpha \right). \end{aligned}$$

This is the analogue of equation (4.6). As in Section 4 we extract the leading term from  $\tilde{F}^k$ ,

$$\tilde{F}^k = \frac{1}{\varepsilon} \left( \frac{c_0^k}{\varepsilon} \zeta^k + c_1^k \dot{\zeta}^k \right) \times B(z^0) + \hat{F}^k,$$

and we note that

$$\begin{aligned} P_0(z^0) \tilde{F}^k &= P_0(z^0) \hat{F}^k, \\ P_{\pm 1}(z^0) \tilde{F}^k &= \pm i c_0^k \frac{|B(z^0)|}{\varepsilon^2} \zeta_{\pm 1}^k \pm i c_1^k \frac{|B(z^0)|}{\varepsilon} P_{\pm 1}(z^0) \dot{\zeta}^k + P_{\pm 1}(z^0) \hat{F}^k, \end{aligned}$$

Despite the additional terms in the formula for  $\tilde{F}^k$  (and hence also in  $\hat{F}^k$ ) we still have  $\hat{F}^0 = \mathcal{O}(1)$ ,  $\hat{F}^{\pm 1} = \mathcal{O}(\varepsilon^{-1})$ , and  $\hat{F}^k = \mathcal{O}(\varepsilon^{-2})$  for all other values of  $k$ . This follows for  $k = 0$  from  $c_0^0 = 0$  and  $\frac{d}{dt} A(z^0) = A'(z^0) \dot{z}^0$ , and for  $k = \pm 1$  from the fact that the only  $\mathcal{O}(\varepsilon^{-2})$ -term in the third line of the definition of  $\tilde{F}^k$  is obtained for  $k_1 = k = \pm 1$  and  $k_2 = 0$  and this term is equal to the  $\mathcal{O}(\varepsilon^{-2})$ -term in the fourth line.

To annihilate the  $\varepsilon^{-2}$  terms in (5.6) for  $j = k = \pm 1$  we require that  $\pm i c_0^{\pm 1} |B(z^0)| = d_0^{\pm 1}$  with coefficients  $c_0^{\pm 1}$  and  $d_0^{\pm 1}$  taken for  $\varepsilon = 0$ . Inserting (5.4) this gives

$$-\frac{1}{\eta} \sin(\eta \dot{\phi}(t)) |B(z^0(t))| = -\frac{4}{\eta^2} \sin^2\left(\frac{\eta \dot{\phi}(t)}{2}\right)$$

which is equivalent to

$$\tan\left(\frac{\eta}{2} \dot{\phi}(t)\right) = \frac{\eta}{2} |B(z^0(t))|. \quad (5.7)$$

This definition of  $\dot{\phi}(t) \in (0, \pi/\eta)$  (and of  $\phi(t)$  by integration) replaces (4.7) for the exact solution.

As in Section 4 we extract the dominant term from (5.6) which is  $\ddot{\zeta}_0^0$  for  $k = j = 0$ ,  $\dot{\zeta}_{\pm 1}^0$  for  $k = 0$  and  $j = \pm 1$ ,  $\dot{\zeta}_{\pm 1}^{\pm 1}$  for  $k = j = \pm 1$ , and  $\zeta_j^k$  for all other choices of  $(k, j)$ . We get the following system of equations

$$\ddot{\zeta}_0^0 - \ddot{P}_0(z^0)(z^0 - c^0) = f_0^0(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}) \quad (5.8)$$

$$\begin{aligned} \dot{\zeta}_1^0 - \dot{P}_1(z^0)(z^0 - c^0) &= -i\varepsilon |B(z^0)|^{-1} (\ddot{\zeta}_1^0 - \ddot{P}_1(z^0)(z^0 - c^0)) \\ &\quad + \varepsilon f_1^0(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}) \end{aligned} \quad (5.9)$$

$$\begin{aligned} \dot{\zeta}_{-1}^0 - \dot{P}_{-1}(z^0)(z^0 - c^0) &= i\varepsilon |B(z^0)|^{-1} (\ddot{\zeta}_{-1}^0 - \ddot{P}_{-1}(z^0)(z^0 - c^0)) \\ &\quad + \varepsilon f_{-1}^0(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}) \end{aligned} \quad (5.10)$$

$$e_1^1 \dot{\zeta}_1^1 = f_1^1(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}) \quad (5.11)$$

$$e_{-1}^{-1} \dot{\zeta}_{-1}^{-1} = f_{-1}^{-1}(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}) \quad (5.12)$$

$$e_j^k \zeta_j^k = f_j^k(\varepsilon, z^0, \dot{z}^0, \dots, (\zeta^l, \dot{\zeta}^l, \dots)_{l \neq 0}), \quad (5.13)$$

where  $z^0 = c^0 + \zeta_1^0 + \zeta_0^0 + \zeta_{-1}^0$ , and

$$e_{\pm 1}^{\pm 1} = d_1^{\pm 1} \mp i c_1^{\pm 1} |B(z^0)|, \quad e_j^k = d_0^k - i j c_0^k |B(z^0)|$$

with coefficients  $d_1^{\pm 1}, c_1^{\pm 1}, d_0^k, c_0^k$  taken at  $\varepsilon = 0$ . The difference to the system (4.8)-(4.13) is that the right-hand side also depends on a finite number of higher derivatives of the variables, and that an additional factor is present in the equations (5.11)-(5.13). Using (5.7), these factors are given by

$$\begin{aligned} e_{\pm 1}^{\pm 1} &= \pm \frac{2i}{\eta} \cos(\eta \dot{\phi}) \left( \tan(\eta \dot{\phi}) - \tan\left(\frac{\eta}{2} \dot{\phi}\right) \right) \\ e_j^k &= -\frac{2}{\eta^2} \sin(k\eta \dot{\phi}) \left( \tan\left(\frac{k\eta}{2} \dot{\phi}\right) - j \tan\left(\frac{\eta}{2} \dot{\phi}\right) \right). \end{aligned}$$

Under the assumption

$$(N+1)\eta |\dot{\phi}(t)| \leq C < \pi, \quad (5.14)$$

which corresponds to a step size restriction for  $h = \eta\varepsilon$ , it follows from the inequalities  $\tan(\alpha + \beta) > \tan \alpha + \tan \beta$  (with  $\alpha > 0, \beta > 0, \alpha + \beta < \pi/2$ ) and

$\tan(k\alpha) > k \tan \alpha$  and from  $|B(x)| \geq 1$  that the absolute values of the factors  $e_{\pm 1}^{\pm 1}, e_j^k$  are bounded from below by a positive constant for  $|k| \leq N+1$ . We can divide the equations (5.11)-(5.13) by these factors and thus obtain a system whose left-hand side is identical to that of (4.8)-(4.13). The right-hand side is again a polynomial in  $\varepsilon$ . We still have that the  $\varepsilon$ -independent term of  $f_j^k$  (for  $k(j-k) \neq 0$ ) only depends on  $z^0$  and  $\zeta^l$  with  $0 < |l| < |k|$ , and that  $f_j^k$  starts with the  $\mathcal{O}(\varepsilon)$ -term for  $|k| = 1$  and  $j \neq k$ . Furthermore, the  $\varepsilon$ -independent term of  $f_{\pm 1}^{\pm 1}$  depends on  $\zeta_{\pm 1}^{\pm 1}$ , but not on its derivatives, and the equations for  $\zeta_{\pm 1}^0$  have the same structure as for the exact solution. Therefore, a regular differential-algebraic system for the coefficient functions  $\zeta_j^k$  is obtained as in Section 4.

**Initial values.** The numerical approximation to the velocity  $\dot{x}(t)$  is given by  $v_n = \delta_{2h} x_n = (x_{n+1} - x_{n-1})/2h$ . It has the expansion (with  $t = nh$ )

$$v_n \approx \sum_{k \in \mathbf{Z}} \delta_{2h} y^k(t) = \sum_{k \in \mathbf{Z}} e^{ik\phi(t)/\varepsilon} \sum_{l \geq 0} \varepsilon^{l-1} c_l^k \frac{d^l}{dt^l} z^k(t). \quad (5.15)$$

Assuming  $\phi(0) = 0$ , the initial values for the differential-algebraic system, constructed above, have to satisfy

$$\sum_{|k| \leq N+1} z^k(0) = x(0), \quad \sum_{|k| \leq N+1} \left( \frac{c_0^k}{\varepsilon} z^k(0) + c_1^k \dot{z}^k(0) + \dots \right) = \dot{x}(0), \quad (5.16)$$

where the coefficient functions  $c_l^k$  are evaluated at  $t = 0$ . The dots indicate  $\mathcal{O}(\varepsilon)$ -terms depending on derivatives of  $z^k$  at  $t = 0$ . The only difference to the construction of Section 4 for the exact solution is formula (5.1), which now becomes

$$P_{\pm 1}(z^0(0)) \dot{z}^0(0) + c_0^{\pm 1} \zeta_{\pm 1}^{\pm 1}(0) = P_{\pm 1}(z^0(0)) \dot{x}(0) + \mathcal{O}(\varepsilon)$$

with  $c_0^{\pm 1} = \pm i \eta^{-1} \sin(\eta \dot{\phi}(0))$ . By assumption (5.14) and  $|B(x)| \geq 1$ , the inverse of  $c_0^{\pm 1}$  exists and is bounded, so that  $\zeta_{\pm 1}^{\pm 1}(0)$  can be extracted from this relation.

**Uniqueness of the modulated Fourier expansion.** The differential-algebraic system and the initial values are obtained from equations that have the same structure as those for the exact solution in Section 4. Therefore the coefficient functions  $z^k(t)$  are independent of the choice of the parameter  $c^0$ .

## 5.2 Bounds for the coefficient functions and for the remainder

We complete the construction of the previous section by giving precise bounds for the coefficient functions of the modulated Fourier expansion and by estimating the remainder caused by truncation of the series.



**Theorem 5.1** *Let  $x_n$  be the numerical solution of the variational integrator (1.7) applied to (1.1) with initial values satisfying (1.2). Suppose that, for  $0 \leq nh \leq T_\varepsilon$  with  $T_\varepsilon = \mathcal{O}(\varepsilon)$ ,  $x_n$  stays in a compact set  $K$  and*

$$\frac{h}{\varepsilon} |B(x_n)| \leq 2 \tan\left(\frac{C}{2(N+1)}\right) \quad \text{with} \quad C < \pi \quad (5.17)$$

for some integer  $N \geq 1$ . Then we have

$$x_n = \sum_{|k| \leq N+1} z^k(t) e^{ik\phi(t)/\varepsilon} + R_N(t), \quad t = nh, \quad (5.18)$$

where the phase function  $\phi(t)$  is given by (5.7). The functions  $z^k(t)$  together with their derivatives (up to order  $N$ ) are bounded as

$$z^k = \mathcal{O}(\varepsilon^{|k|}) \quad \text{for all} \quad |k| \leq N+1 \quad (5.19)$$

and further satisfy

$$z^0 \times B(z^0) = \mathcal{O}(\varepsilon), \quad P_j(z^0)z^k = \mathcal{O}(\varepsilon^2) \quad \text{for} \quad |k| = 1, j \neq k. \quad (5.20)$$

The remainder term is bounded by

$$R_N(t) = \mathcal{O}(t^2\varepsilon^N) \quad \text{for} \quad 0 \leq t \leq T_\varepsilon. \quad (5.21)$$

The functions  $z^k$  are unique up to  $\mathcal{O}(\varepsilon^{N+2})$ . The constants symbolised by the  $\mathcal{O}$ -notation are independent of  $\varepsilon$  and  $n$  with  $0 \leq nh \leq T_\varepsilon$ , but they depend on  $N$ ,  $T_\varepsilon$ , the constants in (1.2), and on bounds of derivatives of  $A(x)$  and  $U(x)$ .

*Proof* The construction of Section 5.1 yields the bounds (5.19) and the uniqueness up to  $\mathcal{O}(\varepsilon^{N+2})$  of the coefficient functions. Over a time interval of length  $\mathcal{O}(\varepsilon)$ , a standard discrete Gronwall argument, similar to the continuous case, allows us to conclude from a defect of size  $\mathcal{O}(\varepsilon^N)$  to an error of size  $\mathcal{O}(t^2\varepsilon^N)$ .  $\square$

### 5.3 The modulation system in terms of $y^k$

With the functions  $z^k(t)$  constructed in Section 5.1, we consider  $y^k(t) = z^k(t)e^{ik\phi(t)/\varepsilon}$  (for  $|k| \leq N+1$ ) with the phase function  $\phi(t)$  given by (5.7). We further put  $y^k(t) = 0$  for  $|k| > N+1$ . Using (4.42) the equation (5.5) for the modulation functions can be written as

$$\delta_h^2 y^k = \frac{1}{\varepsilon} \left( \sum_{j \in \mathbf{Z}} \left( \frac{\partial \mathcal{A}_j}{\partial y^k}(\mathbf{y}) \right)^* \delta_{2h} y^j - \delta_{2h} \mathcal{A}_k(\mathbf{y}) \right) - \left( \frac{\partial \mathcal{U}}{\partial y^k}(\mathbf{y}) \right)^* + \mathcal{O}(\varepsilon^N) \quad (5.22)$$

with  $\mathcal{A}_j(\mathbf{y})$  and  $\mathcal{U}(\mathbf{y})$  given by (4.38). This is equivalent to using the formulation (5.1) of the variational integrator instead of (5.2), which is more convenient for the construction of the modulation functions. In the definitions of  $\mathcal{U}(\mathbf{y})$  and  $\mathcal{A}_k(\mathbf{y})$  we can restrict the sum to  $0 \leq m \leq N+1$  by including further terms in the remainder  $\mathcal{O}(\varepsilon^N)$ . Consequently all appearing sums are finite.

#### 5.4 Near-conservation of a modified magnetic moment

We recall from Section 2.2 the modified magnetic moment

$$I_h(x, v) = \frac{1}{\rho(\xi(x))} I(x, v),$$

where  $\xi(x) = 2 \arctan(\frac{1}{2}\eta|B(x)|)$  with  $\eta = h/\varepsilon$ , and  $\rho(\xi) = \cos^4(\xi/2)$ .

**Theorem 5.2** *Under the assumptions of Theorem 5.1 and with the coefficient functions  $\mathbf{z}(t) = (z^k(t))_{|k| \leq N+1}$  of the modulated Fourier expansion, there exists an almost-invariant  $\mathcal{I}_h[\mathbf{z}](t)$ , such that for  $0 \leq t \leq T_\varepsilon$  (with  $T_\varepsilon = \mathcal{O}(\varepsilon)$ )*

$$\begin{aligned} \mathcal{I}_h[\mathbf{z}](t) &= \mathcal{I}_h[\mathbf{z}](0) + \mathcal{O}(t\varepsilon^N) \\ \mathcal{I}_h[\mathbf{z}](nh) &= I_h(x_n, v_n) + \mathcal{O}(\varepsilon) \quad \text{for } nh \leq T_\varepsilon. \end{aligned}$$

The constants symbolised by  $\mathcal{O}$  are independent of  $\varepsilon$  and  $h$ , but depend on  $N$ .

*Proof* Multiplication of (5.22) with  $-ik(y^k)^*$ , where  $y^k(t) = z^k(t)e^{ik\phi(t)/\varepsilon}$ , and summation over  $k$  yields

$$-\sum_k ik(y^k)^* \delta_h^2 y^k + \frac{1}{\varepsilon} \sum_j ij \mathcal{A}_j(\mathbf{y})^* \delta_{2h} y^j - \frac{1}{\varepsilon} \sum_k ik(y^k)^* \delta_{2h} \mathcal{A}_k(\mathbf{y}) = \mathcal{O}(\varepsilon^{N+1}),$$

where we have used (4.43) and (4.44). This relation can be written as

$$\mathcal{I}_h[\mathbf{z}](t) - \mathcal{I}_h[\mathbf{z}](t-h) = \mathcal{O}(h\varepsilon^N)$$

with<sup>3</sup>

$$\begin{aligned} \mathcal{I}_h[\mathbf{z}](t) &= -\frac{i}{\varepsilon h} \sum_k k y^k(t)^* y^k(t+h) \\ &\quad + \frac{i}{2\varepsilon^2} \sum_k k \left( \mathcal{A}_k(\mathbf{y}(t))^* y^k(t+h) - y^k(t)^* \mathcal{A}_k(\mathbf{y}(t+h)) \right). \end{aligned} \quad (5.23)$$

It defines the almost-invariant of the modulation system and proves the first statement of the theorem.

We shall show that  $\mathcal{I}_h[\mathbf{z}](t)$  is close to the modified magnetic moment. Computing its dominant term, which is for  $k = \pm 1$ , and using the relation (5.7) we obtain

$$\mathcal{I}_h[\mathbf{z}](t) = |B(z^0)| \frac{|z_1^1|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon).$$

On the other hand, inserting the modulated Fourier expansion for  $x_n$  and  $v_n = (x_{n+1} - x_{n-1})/2h$  into (1.4), we obtain

$$I(x_n, v_n) = \text{sinc}^2(\eta\dot{\phi}(t_n)) \frac{\dot{\phi}(t_n)^2}{|B(z^0(t_n))|} \frac{|z_1^1(t_n)|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon). \quad (5.24)$$

<sup>3</sup> We thank Ludwig Gauckler for drawing our attention to this idea in connection with the problem considered in [7].

With  $|B(z^0(t))|$  computed from (5.7) and with the use of the identity  $\sin \xi = 2 \sin \frac{\xi}{2} \cos \frac{\xi}{2}$  this yields

$$I(x_n, v_n) = \cos^4\left(\frac{1}{2}\eta\dot{\phi}(t_n)\right) |B(z^0(t_n))| \frac{|z_1^1(t_n)|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon),$$

and the result follows from the definition of  $I_h(x, v)$ .  $\square$

### 5.5 Near-conservation of a modified energy

We recall the definition of the modified energy

$$H_h(x, v) = H(x, v) + \theta(\xi(x))I(x, v) |B(x)|,$$

where  $\theta(\xi)$  is given in Section 2.3, and again  $\xi(x) = 2 \arctan(\frac{1}{2}\eta|B(x)|)$  with  $\eta = h/\varepsilon$ .

**Theorem 5.3** *Under the assumptions of Theorem 5.1, there exists an almost-invariant  $\mathcal{I}_h[\mathbf{z}](t)$ , such that for  $0 \leq t \leq T_\varepsilon$  (with  $T_\varepsilon = \mathcal{O}(\varepsilon)$ )*

$$\mathcal{I}_h[\mathbf{z}](t) = \mathcal{I}_h[\mathbf{z}](0) + \mathcal{O}(t\varepsilon^N)$$

$$\mathcal{I}_h[\mathbf{z}](nh) = I_h(x_n, v_n) + \mathcal{O}(\varepsilon) \quad \text{for } nh \leq T_\varepsilon.$$

The constants symbolised by  $\mathcal{O}$  are independent of  $\varepsilon$  and  $h$ , but depend on  $N$ .

*Proof* Multiplication of (5.22) with  $(\dot{y}^k)^*$  and summation over  $k$  yields

$$\sum_k (\dot{y}^k)^* \delta_h^2 y^k + \frac{1}{\varepsilon} \sum_k \left( \frac{d}{dt} \mathcal{A}_k(\mathbf{y})^* \delta_{2h} y^k - (\dot{y}^k)^* \delta_{2h} \mathcal{A}_k(\mathbf{y}) \right) + \frac{d}{dt} \mathcal{U}(\mathbf{y}) = \mathcal{O}(\varepsilon^N). \quad (5.25)$$

As in the proof of [7, Theorem 5.1], the first sum is a total differential, and the second sum is a total differential by the proof of [8, Proposition 6.2]. We thus obtain the existence of a smooth function  $\mathcal{H}_h[\mathbf{z}](t)$  such that the above expression equals

$$\frac{d}{dt} \mathcal{H}_h[\mathbf{z}](t) = \mathcal{O}(\varepsilon^N).$$

This proves the first statement of the theorem.

We now determine the dominant part of

$$\mathcal{H}_h[\mathbf{z}](t) = \mathcal{K}_h[\mathbf{z}](t) + \mathcal{M}_h[\mathbf{z}](t) + \mathcal{U}(\mathbf{z}(t)), \quad (5.26)$$

where the time derivatives of the three terms on the right-hand side equal the three corresponding terms on the left-hand side of (5.25). The dominant term in the integral of

$$\begin{aligned} \frac{d}{dt} \mathcal{K}_h[\mathbf{z}] &= \sum_k (\dot{y}^k)^* \delta_h^2 y^k = \sum_k \left( z^k + \frac{ik\dot{\phi}}{\varepsilon} z^k \right)^* \left( \sum_{l \geq 0} \varepsilon^{l-2} d_l^k \frac{d^l}{dt^l} z^k \right) \\ &= (z^0)^* \ddot{z}^0 + \frac{2\dot{\phi}\ddot{\phi}}{\varepsilon^2} \cos(\eta\dot{\phi}) |z^1|^2 \\ &\quad + \frac{2}{\varepsilon^2 \eta^2} \left( \eta\dot{\phi} \sin(\eta\dot{\phi}) - 2 \sin^2\left(\frac{\eta\dot{\phi}}{2}\right) \right) \left( (\dot{z}^1)^* z^1 + (z^1)^* \dot{z}^1 \right) + \mathcal{O}(\varepsilon), \end{aligned}$$

where  $\eta = h/\varepsilon$ , is seen to be

$$\mathcal{K}_h[\mathbf{z}] = \frac{1}{2}|z^0|^2 + \frac{2|z^1|^2}{\varepsilon^2\eta^2} \left( \eta\dot{\phi} \sin(\eta\dot{\phi}) - 2\sin^2\left(\frac{\eta\dot{\phi}}{2}\right) \right) + \mathcal{O}(\varepsilon).$$

We next consider  $\mathcal{M}_h[\mathbf{z}]$ . The dominating terms appear for  $k = 0$  and  $k = \pm 1$ . For  $k = 0$  we deal with smooth functions and hence  $\delta_{2h}y^0 = \dot{y}^0 + \mathcal{O}(h^2)$  and  $\delta_{2h}\mathcal{A}_0(\mathbf{y}) = \frac{d}{dt}\mathcal{A}_0(\mathbf{y}) + \mathcal{O}(h^2)$ . Since these quantities are real, the term for  $k = 0$  is of size  $\mathcal{O}(\varepsilon)$ . For  $k = \pm 1$  we note

$$\begin{aligned} & \frac{d}{dt}\mathcal{A}_k(\mathbf{y})^* \delta_{2h}y^k - (\dot{y}^k)^* \delta_{2h}\mathcal{A}_k(\mathbf{y}) \\ &= \frac{1}{\varepsilon} \left( \frac{d}{dt} \left( A'(z^0)z^k \right) + \frac{ik\dot{\phi}}{\varepsilon} A'(z^0)z^k \right)^* \left( \frac{1}{\varepsilon} c_0^k z^k + c_1^k \dot{z}^k \right) \\ & \quad - \left( \dot{z}^k + \frac{ik\dot{\phi}}{\varepsilon} z^k \right)^* \left( \frac{1}{\varepsilon} c_0^k A'(z^0)z^k + c_1^k \frac{d}{dt} \left( A'(z^0)z^k \right) \right) + \mathcal{O}(\varepsilon) \end{aligned}$$

with  $c_0^k, c_1^k$  given by (5.4). Using the fact that  $z^{-1}, c_l^{-k}$  are the complex conjugate of  $z^1, c_l^k$ , the sum (for  $k = 1$  and  $k = -1$ ) of this expression can be written as

$$\begin{aligned} & -\frac{i\dot{\phi}}{\varepsilon^3} (c_0^1 + c_0^{-1}) \left( (A'(z^0)z^1)^* z^1 - (z^1)^* A'(z^0)z^1 \right) \\ & + \frac{1}{\varepsilon^2} (c_0^1 - i\dot{\phi}c_1^{-1}) \left( \frac{d}{dt} \left( A'(z^0)z^1 \right)^* z^1 + \left( A'(z^0)z^1 \right)^* \dot{z}^1 \right. \\ & \quad \left. - (\dot{z}^1)^* \left( A'(z^0)z^1 \right) - (z^1)^* \frac{d}{dt} \left( A'(z^0)z^1 \right) \right) + \mathcal{O}(\varepsilon). \end{aligned} \tag{5.27}$$

It follows from

$$\begin{aligned} c_0^1 + c_0^{-1} &= -\varepsilon\eta \sin(\eta\dot{\phi})\ddot{\phi} + \mathcal{O}(\varepsilon^2) \\ c_0^1 - i\dot{\phi}c_1^{-1} &= i\dot{\phi}(\text{sinc}(\eta\dot{\phi}) - \cos(\eta\dot{\phi})) + \mathcal{O}(\varepsilon) \end{aligned}$$

and from the relation

$$(A'(z^0)z^1)^* z^1 - (z^1)^* A'(z^0)z^1 = i|B(z^0)| |z^1|^2 + \mathcal{O}(\varepsilon)$$

that the expression (5.27) is the total derivative of

$$\mathcal{M}_h[\mathbf{z}] = \kappa(\eta\dot{\phi}) \dot{\phi} |B(z^0)| \frac{|z^1|^2}{\varepsilon^2} + \mathcal{O}(\varepsilon) \quad \text{with} \quad \kappa(\xi) = \text{sinc}(\xi) - \cos(\xi).$$

Moreover,

$$\mathcal{U}(\mathbf{z}) = U(z^0) + \mathcal{O}(\varepsilon).$$

We now consider the dominant term in  $H(x_n, v_n) = \frac{1}{2}|v_n|^2 + U(x_n)$ . We insert the modulated Fourier expansion for  $x_n$  and  $v_n = (x_{n+1} - x_{n-1})/(2h)$  to obtain  $x_n = z^0(t_n) + \mathcal{O}(\varepsilon)$  and hence

$$U(x_n) = U(z^0(t_n)) + \mathcal{O}(\varepsilon),$$

and (omitting from now on the argument  $t_n$  in the coefficient functions)

$$v_n = \dot{z}^0 + \frac{i}{h} (z_1^1 e^{i\phi/\varepsilon} - z_{-1}^{-1} e^{-i\phi/\varepsilon}) \sin(\eta\dot{\phi}) + \mathcal{O}(\varepsilon),$$

where again  $z_j^k = P_j(z^0)z^k$  and we use the estimates for the coefficient functions in Theorem 5.1. From (5.20) we have that  $\dot{z}^0 = P_0(z^0)\dot{z}^0 + \mathcal{O}(\varepsilon)$ . By orthogonality of the vectors  $P_0(z^0)\dot{z}^0$ ,  $z_1^1$ , and  $z_{-1}^{-1}$ , this implies

$$\frac{1}{2}|v_n|^2 = \frac{1}{2}|\dot{z}^0|^2 + \frac{|z_1^1|^2}{\varepsilon^2\eta^2} \sin^2(\eta\dot{\phi}) + \mathcal{O}(\varepsilon).$$

From (5.24) we note

$$\frac{|z_1^1|^2}{\varepsilon^2\eta^2} \sin^2(\eta\dot{\phi}) = |B(z^0)| I(x_n, v_n) + \mathcal{O}(\varepsilon).$$

Moreover, by (5.7) we have  $\eta\dot{\phi} = \xi(z^0)$  with  $\xi(x) = 2 \arctan(\frac{1}{2}\eta|B(x)|)$ . Subtracting the expressions obtained for  $\mathcal{H}_h[\mathbf{z}](nh)$  and  $H(x_n, v_n)$  from each other, we thus obtain

$$\mathcal{H}_h[\mathbf{z}](nh) - H(x_n, v_n) = \theta(\xi(x_n))|B(x_n)| I(x_n, v_n) + \mathcal{O}(\varepsilon),$$

which yields  $\mathcal{H}_h[\mathbf{z}](nh) - H_h(x_n, v_n) = \mathcal{O}(\varepsilon)$  by the definition of the modified energy  $H_h$ .  $\square$

## 5.6 From short to long time intervals

The statements of Theorems 2.2 and 2.3 can now be obtained by patching together the local near-conservation results of the previous sections in the same way as it was done for the exact solution in Section 4.5.

**Acknowledgement.** The research for this article has been partially supported by the Fonds National Suisse, Project No. 200020\_159856, and by Deutsche Forschungsgemeinschaft, SFB 1173.

## References

1. V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer, Berlin, 1997.
2. G. Benettin and P. Sempio. Adiabatic invariants and trapping of a point charge in a strong nonuniform magnetic field. *Nonlinearity*, 7(1):281–304, 1994.
3. J. P. Boris. Relativistic plasma simulation-optimization of a hybrid code. *Proceeding of Fourth Conference on Numerical Simulations of Plasmas*, pages 3–67, November 1970.
4. J. R. Cary and A. J. Brizard. Hamiltonian theory of guiding-center motion. *Rev. Modern Phys.*, 81(2):693–738, 2009.
5. C. L. Ellison, J. W. Burby, and H. Qin. Comment on “Symplectic integration of magnetic systems”: A proof that the Boris algorithm is not variational. *J. Comput. Phys.*, 301:489–493, 2015.

6. E. Hairer and C. Lubich. Energy conservation by Störmer-type numerical integrators. In D. F. Griffiths G. A. Watson, editor, *Numerical Analysis 1999*, pages 169–190. CRC Press LLC, 2000.
7. E. Hairer and C. Lubich. Long-term analysis of the Störmer-Verlet method for Hamiltonian systems with a solution-dependent high frequency. *Numer. Math.*, 34:119–138, 2016.
8. E. Hairer and C. Lubich. Symmetric multistep methods for charged-particle dynamics. *SMAI J. Comput. Math.*, 3:205–218, 2017.
9. E. Hairer and C. Lubich. Energy behaviour of the Boris method for charged-particle dynamics. *Submitted for publication*, 2018.
10. E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Series in Computational Mathematics 31. Springer-Verlag, Berlin, 2nd edition, 2006.
11. J. Henrard. The adiabatic invariant in classical mechanics. In *Dynamics reported*, volume 2 of *Dynam. Report. Expositions Dynam. Systems (N.S.)*, pages 117–235. Springer, Berlin, 1993.
12. T. G. Northrop. *The adiabatic motion of charged particles*. Interscience Tracts on Physics and Astronomy, Vol. 21. Interscience Publishers John Wiley & Sons New York-London-Sydney, 1963.
13. S. D. Webb. Symplectic integration of magnetic systems. *J. Comput. Phys.*, 270:570–576, 2014.