Second-order nonlinear gyrokinetic theory: from the particle to the gyrocentre

Natalia Tronko\textsuperscript{1,2,†} and Cristel Chandre\textsuperscript{3}

\textsuperscript{1}Max-Planck Institute for Plasma Physics, 85748 Garching, Germany
\textsuperscript{2}TU Munich, Mathematics Center, 85747 Garching, Germany
\textsuperscript{3}Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

(Received 14 September 2017; revised 20 April 2018; accepted 23 April 2018)

A gyrokinetic reduction is based on a specific ordering of the different small parameters characterizing the background magnetic field and the fluctuating electromagnetic fields. In this tutorial, we consider the following ordering of the small parameters: $\epsilon_B = \epsilon_\delta^2$ where $\epsilon_B$ is the small parameter associated with spatial inhomogeneities of the background magnetic field and $\epsilon_\delta$ characterizes the small amplitude of the fluctuating fields. In particular, we do not make any assumption on the amplitude of the background magnetic field. Given this choice of ordering, we describe a self-contained and systematic derivation which is particularly well suited for the gyrokinetic reduction, following a two-step procedure. We follow the approach developed in Sugama (\textit{Phys. Plasmas}, vol. 7, 2000, p. 466): In a first step, using a translation in velocity, we embed the transformation performed on the symplectic part of the gyrocentre reduction in the guiding-centre one. In a second step, using a canonical Lie transform, we eliminate the gyroangle dependence from the Hamiltonian. As a consequence, we explicitly derive the fully electromagnetic gyrokinetic equations at the second order in $\epsilon_\delta$.

\textbf{Key words:} fusion plasma, plasma nonlinear phenomena, plasma simulation

1. Introduction

A strongly magnetized plasma forms a complex multi-scaled system in space and time. Building reduced models allows for the identification of the main physical mechanisms in different regimes and configurations.

After more than three decades of active development, gyrokinetic theory is nowadays one of the important theoretical frameworks for the investigation of strongly magnetized plasmas. The main idea behind the gyrokinetic dynamical reduction consists of a systematic elimination of the smallest scales of motion, which leads to a drastic reduction of computational time.

From a theoretical viewpoint, the gyrokinetic reduction provides access to accurate predictions on long temporal and large spatial scale processes such as transport which is one of the main issues for fusion plasma confinement (Garbet \textit{et al.} 2010). In

\textsuperscript{†} Email address for correspondence: nataliat@ipp.mpg.de
astrophysical plasmas, the gyrokinetic theory is also of interest (Schekochihin et al. 2009): gyrokinetic simulations have been used to access small-scale spectra, in order to fill the gaps when magnetohydrodynamics approximations fail.

The gyrokinetic dynamical reduction exploits the fact that the particle dynamics is decomposed into a fast rotation around the magnetic field lines and a slow drift motion. The temporal scale of the gyromotion is set by the cyclotron frequency \( \Omega = eB/mc \), where \( e \) and \( m \) are, respectively, the charge and mass of the particles, \( B \) is the magnetic field amplitude and \( c \) the speed of light. The gyromotion is described by a fast gyroangle variable \( \theta \) to which the slowly varying magnetic moment \( \mu \) is canonically conjugate. At the lowest order,

\[
\mu = \frac{mv_{\perp}^2}{2B},
\]

where \( v_{\perp} \) is the perpendicular velocity of the charged particle with respect to the magnetic field lines. The magnetic moment is a measure of the magnetic flux through the particle orbit. From this geometrical picture comes the idea of using \( \mu \) as an action variable, canonically conjugated to the fast gyromotion around the magnetic field lines. In the case of a constant and uniform background magnetic field, \( \mu \) is an exact dynamical invariant, regardless of the amplitude of the magnetic field.

The sources for the violation of the conservation of the magnetic moment come from two different origins: first, spatial variations of the background quantities on the gyro-scales such as magnetic field non-uniformities and curvature and, second, the presence of electromagnetic fluctuations generated by the plasma. The goal of the gyrokinetic reduction is to reduce the dynamics taking into account these two sources of perturbations, and restore a conserved quantity, a modified magnetic moment.

A fundamental aspect of the gyrokinetic theory is the assumption on the ordering of the various small parameters present in the system. Each choice of ordering will correspond to a different reduced system and hence a different set of gyrokinetic equations to integrate numerically. The choice of ordering is driven by the specific geometry to be considered and by experimental observations. There are basically two groups of small parameters, each group associated with the two sources of perturbations which break the conservation of the magnetic moment: one group concerns the background magnetic field and another group characterizes the fluctuating fields.

Roughly speaking, in the first group, \( \epsilon_B \) is related to the background magnetic field inhomogeneities. The small parameter \( \epsilon_B \) is given by \( \epsilon_B = \rho_{\text{th}}/L_B \), where \( \rho_{\text{th}} \) is the thermal Larmor radius and \( L_B = \| \nabla B/B \|^{-1} \) defines the spatial scale at which the magnetic field exhibits significant variations.

In the second group of parameters, \( \epsilon_{\delta} \) is related to the amplitude of the fluctuating fields, \( \epsilon_{\parallel} \) to the parallel gradients of these fields, and \( \epsilon_{\omega} \) to their temporal variations: \( \epsilon_{\delta} = (k_{\perp}\rho_{\text{th}})e\phi_1/T_i \), where \( \phi_1 \) is the amplitude of the fluctuating electrostatic potential, \( k_{\perp} \) are perpendicular wavevectors of the fluctuation spectrum and \( T_i \) the ion temperature. Furthermore, we assume that \( \| A_1 \| \) is of the same order as \( \phi_1 c/v_{\text{th}} \), where \( v_{\text{th}} \) is a characteristic thermal velocity. The gyrokinetic theory assumes that \( \epsilon_{\delta} \) is small. In addition, experimental observations report that the most dangerous instabilities occur in the plane perpendicular to the background magnetic field for strongly magnetized plasmas. A small parameter takes this into account: \( \epsilon_{\parallel} = E_{\parallel}/E_{\perp} \sim k_{\parallel}/k_{\perp} \). A usual assumption is that the temporal variations of the fluctuating fields is small, so \( \epsilon_{\omega} = \omega/\Omega \) is taken into account in the expansion, where \( \omega \) is a characteristic frequency.
in the spectrum of the fluctuating fields and $\Omega$ is the ion cyclotron frequency. Other parameters for the fluctuating fields can be defined and we refer to Brizard & Hahm (2007) for a more complete discussion.

The main assumption regarding the small parameters which sustain the derivation of the gyrokinetic model is an assumption which links both groups of parameters, the ones of the background magnetic field and the ones of the fluctuating fields, namely, $\epsilon_B \sim \epsilon^2_\delta$. This assumption is sustained by the majority of global gyrokinetic codes, e.g. ORB5 (Jolliet et al. 2007), GKW (Peeters et al. 2009), GENE (Goerler et al. 2011), GYSELA (Grandgirard et al. 2016) and GYRO (Candy & Waltz 2003). The main constraint behind this choice of ordering is to have simplified models suitable for numerical implementation. It should be noted that recent experimental results for Tore Supra (Casati et al. 2009) as well as numerical results obtained with ORB5 for large systems like ITER and DIII-D (Wersal et al. 2012) indicate the necessity of retaining the maximal ordering $\epsilon_B \sim 100\epsilon_\delta$ for typical turbulence scales in the core of the plasma. The situation is different at the plasma edge, where the difference between both small parameters become more important, i.e. $\epsilon_\delta \sim 100\epsilon_B$, for typical edge turbulence scales. As a consequence, our reduction applies to systems where the magnetic field is not so strong, which is of interest, e.g. in astrophysical plasmas. Concerning the parameters within the group of the fluctuating field parameters, we consider two usual assumptions: $\epsilon_\parallel \sim \epsilon_\delta$ and $\epsilon_\omega \sim \epsilon_\delta$.

Given a particular choice of ordering, there exist several techniques to derive the gyrokinetic theory: direct ones and structural ones. A direct approach akin to Catto (1978, 1981), Frieman & Chen (1982) performs an asymptotic expansion on the Vlasov equation, making the control of orderings and consistency rather cumbersome, for an alternative direct approach, see Abel et al. (2013). Structural derivations of gyrokinetic reductions follow Littlejohn’s seminal work by using the Hamiltonian/Lagrangian framework. The implementation of the Hamiltonian formalism for the particle dynamical reduction towards the guiding-centre motion began with the works of Littlejohn in the Hamiltonian framework (Littlejohn 1979, 1981) and in the Lagrangian one (Littlejohn 1983). The main advantage of these approaches is the consistency of the reduced particle model. An additional step towards gyrokinetic equations considers the coupling between the reduced particle dynamics with electromagnetic fields within the field–particle Lagrangian, providing a framework for the consistent derivation of the gyrokinetic models (Brizard 2000; Sugama 2000); see also Squire et al. (2013), Burby et al. (2015) for a Hamiltonian field formulation. Nowadays this framework is widely used for the derivation of consistently reduced models suitable for the numerical implementation (Brizard & Hahm 2007; Tronko, Bottino & Sonnendrücker 2016; Tronko et al. 2017).

The goal of the systematic gyrokinetic dynamical reduction consists in building a new set of phase-space variables, such that the $\theta$-dependence is completely decoupled from the other variables, and the magnetic moment $\mu$ has trivial dynamics, i.e. $\dot{\mu} = 0$. Therefore, the reduced particle dynamics is described in the five-dimensional phase space with variables $(X, v_\parallel, \mu)$ where $X$ represents the reduced particle position (the gyrocentre), $v_\parallel$ is a variable which mirrors the parallel velocity of the particle and $\mu$ is the conserved magnetic moment. This change of coordinates is constructed via a perturbative series of phase-space transformations. The main advantage of this approach is that these transformations are invertible at each step of the perturbative procedure, allowing one to recover information on the particle dynamics from the averaged one.
The phase-space Lagrangian is the starting point of the gyrokinetic derivation. It is given by

$$L_0 = \left( \frac{e}{c} A(x) + m v \right) \cdot \dot{x} - H(x, v, t), \quad (1.2)$$

where $x$ and $v$ are the position and velocity of the charged particle. The first term, proportional to $\dot{x}$, represents the symplectic (or Liouville) part and the second term the Hamiltonian part.

Electromagnetic perturbations are introduced via the phase-space Lagrangian perturbation:

$$L_1 = \epsilon_\delta \left( \frac{e}{c} A_1(x, \epsilon_\delta t) \cdot \dot{x} - e \phi_1(x, \epsilon_\delta t) \right), \quad (1.3)$$

where we notice that the perturbation associated with the magnetic potential only affects the symplectic part of the phase-space Lagrangian and the electrostatic potential, its Hamiltonian part.

In Hahm (1988), Brizard (1989), Brizard & Hahm (2007), the standard gyrokinetic dynamical reduction is organized in two consecutive steps: the guiding-centre reduction, where the new set of phase-space variables is constructed to evidence the conservation of the magnetic moment $\mu$ with respect to the background magnetic field nonlinearities, and the subsequent gyrocentre reduction which builds a modified set of variables to restore the magnetic moment conservation broken by the introduction of the electromagnetic fluctuations. This gyrokinetic dynamical reduction splits the difficulties in two steps with respect to the small parameter of the system: $\epsilon_B$ for the guiding-centre step and $\epsilon_\delta$ for the gyrocentre step. Each of these steps consists in eliminating the gyrophase dependence from the symplectic and the Hamiltonian part, simultaneously.

In this tutorial, we follow an alternative two-step derivation of the gyrokinetic equations, as proposed in Grebogi, Kaufman & Littlejohn (1979), Sugama (2000), which allows for treating the symplectic and the Hamiltonian part of the Lagrangian in two consecutive steps. This is achieved by an apt shift of the velocity, followed by a modified guiding-centre reduction to move the $\theta$-dependence of the symplectic part of the Lagrangian to the Hamiltonian part. Subsequently, the $\theta$-dependence of the Hamiltonian is removed by some canonical Lie transforms. This approach has also been the one followed in Tronko et al. (2017) in the particular case where the electromagnetic potential does not have any perpendicular component. This method is very well suited for our choice of ordering.

This tutorial is organized as follows: in § 2 we introduce the velocity shift which allows for the application of the guiding-centre theory following Littlejohn (1983). In § 3 we recall the general procedure of canonical Lie transforms and we apply this perturbative procedure to derive the reduced Hamiltonian dynamics at the second order in $\epsilon_\delta$, explicitly providing the new set of variables at order $\epsilon_\delta$. Finally, in § 4, we present the gyrocentre characteristics, from which the gyrokinetic Vlasov equation is reconstructed, and we briefly discuss the variational method to derive the reduced Maxwell equations. For completeness, we add an appendix containing a slightly revisited guiding-centre derivation, with the full details of the derivation, which is suited for our choice of ordering. As a result, the derivation of the gyrokinetic model proposed in this tutorial is self-consistent, and does not rely on previous knowledge of the guiding-centre theory.
2. Gyrocentre as a modified guiding centre

The perturbed one-form associated with the perturbed Lagrangian (1.2)–(1.3) is given by

\[ \gamma_{\text{pert}} = \left( \frac{e}{c} A(x) + \epsilon_\delta \frac{e}{c} A_1(x, \epsilon_\delta t) + m v \right) \cdot dx - H \, dt, \]  

(2.1)

where

\[ H = \frac{1}{2} m v^2 + \epsilon_\delta e \phi_1(x, \epsilon_\delta t). \]  

(2.2)

The part in \( dt \) is referred to as the Hamiltonian part of the one-form, whereas the part in \( dx \) is its symplectic (Liouville) part. We translate the particle velocity \( v \):

\[ \bar{v} = v + \epsilon_\delta \frac{e}{mc} A_1(x, \epsilon_\delta t). \]  

(2.3)

This velocity shift (Grebogi et al. 1979; Sugama 2000) allows us to apply directly and readily Littlejohn’s guiding-centre theory (Littlejohn 1983) without performing the reduction calculations twice (first, the guiding-centre reduction without the perturbation \( A_1 \) and \( \phi_1 \), and then the gyrocentre reduction) in the symplectic part of the one-form. Here we can now apply the guiding-centre results on \( \gamma_{\text{pert}} \) in the variables \((x, \bar{v})\). We decompose \( \bar{v} \) into

\[ \bar{v} = \bar{v}_\parallel \hat{b}(x) + \bar{v}_\perp \hat{\rho}(\bar{\theta}, x), \]  

(2.4)

where \( \hat{b} = B/B \). The above relation defines the two vectors \( \hat{\rho} \) and \( \hat{b} \) of the orthonormal basis \( (\hat{\rho}, \hat{b}) \). Given an orthonormal basis \( (\hat{b}_1(x), \hat{b}_2(x), \hat{b}(x)) \), the following vectors

\[ \hat{\rho}(\bar{\theta}, x) = \hat{b}_1(x) \cos \bar{\theta} - \hat{b}_2(x) \sin \bar{\theta}, \]  

(2.5)

\[ \hat{\rho}(\bar{\theta}, x) = \hat{b}_1(x) \sin \bar{\theta} + \hat{b}_2(x) \cos \bar{\theta}, \]  

(2.6)

define the angle \( \bar{\theta} \). The crucial step in the guiding-centre theory is a shift of the particle position, i.e. \( x = \bar{X} + \rho \) with

\[ \rho = \frac{m \bar{v}_\perp c}{e B(\bar{X})} \hat{\rho}(\bar{\theta}, \bar{X}) + \bar{\rho}_1 + O(\epsilon_B^2), \]  

(2.7)

where the explicit expression of \( \bar{\rho}_1 \) which is of order \( \epsilon_B \) is given in appendix A (for more details on the guiding-centre theory, we refer to Littlejohn (1983), Cary & Brizard (2009), Tronko & Brizard (2015)). Here the expression for \( \bar{\rho}_1 \) is not explicitly needed since we only provide the change of coordinates at order \( \epsilon_\delta \) (see § 3). After the modified guiding-centre reduction, the one-form becomes

\[ \tilde{\gamma}_{\text{pert}} = \left( \frac{e}{c} A(\bar{X}) + m v_\parallel \hat{b}(\bar{X}) - \frac{mc}{e} \mu \hat{R}^* \right) \cdot d\bar{X} + \frac{mc}{e} \bar{\mu} \, d\bar{\theta} - H \, dt + O(\epsilon_B^2), \]  

(2.8)

where \( \bar{\mu} = m \bar{v}_\perp^2/(2B(\bar{X})) \), and \( \hat{R}^* = \nabla \hat{b}_1 \cdot \hat{b}_2 + (\hat{b} \cdot \nabla \times \hat{b}) \hat{b}/2 \). We notice that the symplectic part of \( \tilde{\gamma}_{\text{pert}} \) has no explicit dependence on \( \bar{\theta} \), which was the objective of
the guiding-centre reduction. The Poisson bracket associated with the symplectic part of $\mathcal{H}_\text{pert}$ is given by

$$
\{F, G\}_{\text{gc}} = \frac{e}{mc} \left( \frac{\partial F}{\partial \tilde{\theta}} \frac{\partial G}{\partial \tilde{\mu}} - \frac{\partial F}{\partial \tilde{\mu}} \frac{\partial G}{\partial \tilde{\theta}} \right) + \frac{B^*}{mB^*_\parallel} \cdot \left( \nabla^* F \frac{\partial G}{\partial \tilde{v}_\parallel} - \frac{\partial F}{\partial \tilde{v}_\parallel} \nabla^* G \right)
$$

$$
- \frac{e\hat{b}}{eB^*_\parallel} \cdot (\nabla^* F \times \nabla^* G),
$$

(2.9)

for observables $F$ and $G$, functions of $(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel)$, and where

$$
\nabla^* = \nabla - R^\ast \frac{\partial}{\partial \tilde{\theta}},
$$

(2.10)

$$
B^* = B + \frac{mc}{e} \tilde{v}_\parallel \nabla \times \hat{b} - \frac{mc^2}{e^2} \tilde{\mu} \nabla \times R^*,
$$

(2.11)

and $B^*_\parallel = \hat{b} \cdot B^*$. After the translation in velocity, the Hamiltonian (2.2) becomes

$$
H = \frac{1}{2} m\tilde{v}^2_\parallel + \tilde{\mu} B(\tilde{X}) + \epsilon_\delta e \psi_1(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel, t) + \epsilon^2 \frac{e^2}{2mc^2} \|A_1(\tilde{X} + \rho, \epsilon_\delta t)\|^2,
$$

(2.12)

where there is an explicit $\theta$-dependence at order $\epsilon_\delta$ and $\epsilon^2_\delta$ through the potentials $\psi_1$ and $A_1$. The modified potential $\psi_1$ is given by

$$
\psi_1(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel, t) = \phi_1(\tilde{X} + \rho, \epsilon_\delta t) - \frac{\tilde{v}_\parallel}{c} \hat{b}(\tilde{X} + \rho) \cdot A_1(\tilde{X} + \rho, \epsilon_\delta t)
$$

$$
- \sqrt{\frac{2\tilde{\mu} B(\tilde{X})}{mc^2}} \hat{\mathbf{1}}(\tilde{\theta}, \tilde{X} + \rho) \cdot A_1(\tilde{X} + \rho, \epsilon_\delta t).
$$

(2.13)

It is important to note that all the fluctuating part has been removed from the symplectic part of the one-form and moved to the Hamiltonian part. In this way, the averaging over the fast variable $\tilde{\theta}$ has to be performed only on the Hamiltonian and not on the symplectic part of the one-form. By using canonical transformations, the symplectic part of the one-form is not affected (up to an exact one-form). We notice that Hamiltonian (2.12) has an explicit time dependence, through the fluctuating potentials $\phi_1$ and $A_1$. Therefore in order to perform canonical transformations, it is more convenient to autonomize the system, i.e. by considering that $t$ is an additional dynamical variable and introducing $k$ its canonically conjugate variable. The extended Hamiltonian becomes

$$
\mathcal{H} = \frac{1}{2} m\tilde{v}^2_\parallel + \tilde{\mu} B(\tilde{X}) + \epsilon_\delta e \psi_1(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel, t) + \epsilon^2 \frac{e^2}{2mc^2} \|A_1(\tilde{X} + \rho, \epsilon_\delta t)\|^2 + k,
$$

(2.14)

and the extended Poisson bracket becomes

$$
\{\mathcal{F}, \mathcal{G}\} = \{\mathcal{F}, \mathcal{G}\}_{\text{gc}} + \frac{\partial \mathcal{F}}{\partial \tilde{k}} \frac{\partial \mathcal{G}}{\partial \tilde{k}} - \frac{\partial \mathcal{F}}{\partial \tilde{k}} \frac{\partial \mathcal{G}}{\partial \tilde{k}},
$$

(2.15)

where $\{\cdot, \cdot\}_{\text{gc}}$ is given by (2.9). In what follows, the observables in the extended phase space, i.e. functions of $(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel, t, k)$, are denoted with a calligraphic lettering, whereas functions of $(\tilde{X}, \tilde{\theta}, \tilde{\mu}, \tilde{v}_\parallel, t)$ will be denoted in roman lettering.
As a final remark, we notice that the shift in the position (see (2.7)) contains the Larmor radius from the guiding centre and also the displacement generated by the perturbation field \( A_1 \), and explicitly depends on \( A_1 \). At the leading order we have:

\[
x = \bar{X} + \frac{me}{eb} \hat{b} \times \bar{v} = \bar{X} + \frac{me}{eb} \hat{b} \times v + \epsilon_\delta \frac{1}{B} \hat{b} \times A_1 + O(\epsilon^2_\delta).
\] (2.16)

Furthermore, the averaging procedure performed in the Hamiltonian will modify the position \( \bar{X} \) into the position of the gyrocentre \( X_{gy} \). We will come back to the expressions of the new coordinates after performing the averaging procedure.

3. Averaging procedure of the Hamiltonian

3.1. Canonical Lie transforms

In order to perform the averaging with respect to the fast variable \( \bar{\theta} \) in the Hamiltonian, we use canonical Lie transforms. These transforms are near-identity canonical changes of coordinates which do not modify the expression of the symplectic part of the one-form (up to an exact one-form), or equivalently, do not change the expression of the Poisson bracket. A canonical Lie transform only affects the Hamiltonian, and with an apt choice of generating function eliminates the unwanted part of the Hamiltonian, in our case, its fast-varying part. For more details on Lie transforms, we refer to Cary (1981). The invertible change of coordinates from the old variables \( Z = (X, \bar{\theta}, \bar{\mu}, \bar{v}_p, t, k) \) to the new (gyrokinetic) ones \( Z_{gy} = (X_{gy}, \theta_{gy}, v_{gy}, t, k_{gy}) \) is defined as

\[
Z_{gy} = e^{\xi S} Z,
\] (3.1)

where \( S \) is the scalar generating function of the transformation which is chosen as a function of \( (X, \bar{\theta}, \bar{\mu}, \bar{v}_p, t) \), and the operator \( \xi S \) is defined as \( \xi S = [S, \cdot] \). The bracket \( [\cdot, \cdot] \) is given by (2.9)–(2.15). This change of coordinates transforms any observable \( \mathcal{F}(Z) \) into \( \mathcal{F}(Z_{gy}) \) according to

\[
\mathcal{F}(Z_{gy}) = e^{-\xi S} \mathcal{F}(Z_{gy}) = \mathcal{F} - [S, \mathcal{F}] + \frac{1}{2} [S, [S, \mathcal{F}]] + O(S^3),
\] (3.2)

which is obtained from the scalar invariance \( \mathcal{F}(Z_{gy}) = \mathcal{F}(Z) \) (and the fact that the Poisson bracket satisfies the Leibniz rule) and where the right-hand side of (3.2) is evaluated at \( Z_{gy} \).

We recall that the guiding-centre Poisson bracket (2.9) is decomposed into

\[
[F, G]_{gc} = [F, G]_{-1} + [F, G]_0 + [F, G]_1,
\] (3.3)

and the Hamiltonian \( H = H_0 + \epsilon_\delta H_1 + \epsilon^2_\delta H_2 \) (see (2.12)). In order to remove the \( \bar{\theta} \)-dependence from \( H \), we consider a generating function of the type \( S = \epsilon_\delta S_1 + \epsilon^2_\delta S_2 \). The purpose of \( S_1 \) is to eliminate the fluctuating part of the Hamiltonian \( H \) at order \( \epsilon_\delta \) (i.e. present in \( H_1 \)) and \( S_2 \) eliminates the fluctuating terms at order \( \epsilon^2_\delta \). In order to illustrate the method, we first consider the order \( \epsilon_\delta \). We decompose \( H_1 \) in a fluctuating and an averaged part: \( H_1 = \bar{H}_1 + \langle H_1 \rangle \), where \( \langle H_1 \rangle = (2\pi)^{-1} \int_0^{2\pi} d\bar{\theta} H_1 \). At order \( \epsilon_\delta \), it leads to

\[
\bar{H} = H - \{S, H\}_{gc} - \frac{\partial S}{\partial t} + O(S^2),
\] (3.4)
We extend this analysis to the second order, where the expansion of the Hamiltonian
\[ \tilde{H} = H_0 + \epsilon_{\delta} \left( \langle H_1 \rangle + \tilde{H}_1 - \{S_1, H_0\}_1 - \{S_1, H_0\}_0 - \{S_1, H_0\}_1 - \frac{\partial S_1}{\partial t} \right) + O(\epsilon_{\delta}^2). \] (3.5)

By inspecting the various terms in the above equation, we notice that \( \{S_1, H_0\}_1 \) is of order \( \epsilon_B \) since it involves \( \nabla B \). This term is thus neglected even at the next order since, according to our ordering, \( \epsilon_B \sim \epsilon_{\delta}^2 \). The term \( \{S_1, H_0\}_0 \) involves a term proportional to \( B^* \cdot \nabla S_1 \). Up to order \( \epsilon_B \), this term is the parallel gradient of the generating function. Since the generating function is a function of the fluctuating fields, this term will be of the order of the parallel gradients of the fluctuating fields \( \phi_1 \) and \( A_1 \), which are assumed to be of order \( \epsilon_{\delta} \), so the term \( \{S_1, H_0\}_0 \) is moved to order \( \epsilon_{\delta}^2 \). In addition, we assume that \( \partial S_1/\partial t \) is of order \( \epsilon_{\delta} \) which comes from the ordering \( \epsilon_{\omega} \sim \epsilon_{\delta} \). Therefore the resulting equation which determines the generating function \( S_1 \) is
\[ \{S_1, H_0\}_1 = \frac{eB}{mc} \frac{\partial S_1}{\partial \theta} = \tilde{H}_1. \] (3.6)

At the leading order, the new Hamiltonian becomes
\[ \tilde{H} = H_0 + \epsilon_{\delta} \langle H_1 \rangle + O(\epsilon_{\delta}^2). \] (3.7)

We extend this analysis to the second order, where the expansion of the Hamiltonian is given by (3.2) as
\[ \tilde{H} = H - \{S, H\}_{gc} - \frac{\partial S}{\partial t} + \frac{1}{2} \{S, \{S, H\}_{gc}\}_{gc} + \frac{1}{2} \left\{ S, \frac{\partial S}{\partial t} \right\}_{gc} + O(S^3). \] (3.8)

The terms containing \( \partial S/\partial t \) comes from the fact that the transformation is time dependent (and is taken care by the extended bracket (2.15)). The term \( \{S, \partial S/\partial t\}_{gc} \) in (3.8) is neglected since it is of order \( \epsilon_{\delta}^2 \) given that \( \epsilon_{\omega} \sim \epsilon_{\delta} \). The expansion of the Hamiltonian at the order \( \epsilon_{\delta}^2 \) becomes
\[ \tilde{H} = H_0 + \epsilon_{\delta} \langle H_1 \rangle - \epsilon_{\delta} \left( \frac{\partial S_1}{\partial t} + \{S_1, H_0\}_0 \right) + \epsilon_{\delta}^2 \left( H_2 - \frac{\partial S_2}{\partial t} \right) \]
\[ - \{S_1, H_1\}_{gc} - \{S_2, H_0\}_{gc} + \frac{1}{2} \{S_1, \{S_1, H_0\}_{gc}\}_{gc} + O(\epsilon_{\delta}^3). \] (3.9)

We eliminate the term \( \epsilon_{\delta}^2 \partial S_2/\partial t \) since it is of order \( \epsilon_{\delta}^3 \) given that \( \epsilon_{\omega} \sim \epsilon_{\delta} \). In the term \( \{S_2, H_0\}_{gc} \), only the term \( \{S_2, H_0\}_1 \) matters since \( \{S_2, H_0\}_0 \) and \( \{S_2, H_0\}_1 \) are higher order using the same argument as above for \( S_1 \). Using the same argument, in the term \( \{S_1, \{S_1, H_0\}_{gc}\}_{gc} \), only the terms \( \{S_1, \{S_1, H_0\}_1 \} \) and \( \{S_1, \{S_1, H_0\}_1 \} \) remain, since the term \( \{S_1, \{S_1, H_0\}_1 \} = \{S_1, \tilde{H}_1\}_0 \) involves parallel gradients of \( S_1 \) or \( H_1 \) (of order \( \epsilon_{\delta} \)). This leads to
\[ \tilde{H} = H_0 + \epsilon_{\delta} \langle H_1 \rangle - \epsilon_{\delta} \left( \frac{\partial S_1}{\partial t} + \{S_1, H_0\}_0 \right) + \epsilon_{\delta}^2 \left( \{H_2\} + \tilde{H}_2 - \{S_2, H_0\}_1 - \{S_1, H_1\}_1 \right) \]
\[ - \{S_1, H_1\}_1 + \frac{1}{2} \{S_1, \{S_1, H_0\}_1 \} + \frac{1}{2} \{S_1, \{S_1, H_0\}_1 \}_1 + O(\epsilon_{\delta}^3). \] (3.10)
We rewrite the Poisson bracket
\[
\{ S_2, H_0 \}_{-1} = \tilde{H}_2 - \{ S_1, \langle H_1 \rangle \}_{-1} - \frac{1}{2} \{ S_1, \tilde{H}_1 \}_{-1} - \epsilon_\delta^{-1} \left( \frac{\partial S_1}{\partial t} + \{ S_1, H_0 \}_0 \right)
\]
\[
- \{ S_1, \langle H_1 \rangle \}_1 - \frac{1}{2} \{ S_1, \tilde{H}_1 \}_1,
\]
where we have used (3.6). Consequently, the new Hamiltonian becomes
\[
\tilde{H} = H_0 + \epsilon_\delta \langle H_1 \rangle + \epsilon_\delta^2 (\langle H_2 \rangle - \frac{1}{2} \{ S_1, \tilde{H}_1 \}_{-1} - \frac{1}{2} \{ S_1, \tilde{H}_1 \}_1) + O(\epsilon_\delta^3). \tag{3.12}
\]

3.2. Application to Hamiltonian (2.12)

We rewrite the Poisson bracket \( \{ S_1, \tilde{H}_1 \}_{-1} \) as
\[
\{ S_1, \tilde{H}_1 \}_{-1} = \frac{e}{mc} \frac{\partial}{\partial \hat{\mu}} \left( \tilde{H}_1 \frac{\partial S_1}{\partial \hat{\theta}} \right) - \frac{e}{mc} \frac{\partial}{\partial \hat{\theta}} \left( \tilde{H}_1 \frac{\partial S_1}{\partial \hat{\mu}} \right). \tag{3.13}
\]

From the expression of \( S_1 \), we conclude that the averaged Hamiltonian \( \tilde{H} \) obtained from \( H = H_0 + \epsilon_\delta H_1 + \epsilon_\delta^2 H_2 \) is
\[
\tilde{H} = H_0 + \epsilon_\delta \langle H_1 \rangle + \epsilon_\delta^2 \left( \langle H_2 \rangle - \frac{1}{2B} \frac{\partial}{\partial \hat{\mu}} \langle \tilde{H}_1^{-2} \rangle + \frac{e}{2eB} \hat{b} \cdot (\nabla S_1 \times \nabla \tilde{H}_1) \right) + O(\epsilon_\delta^3), \tag{3.14}
\]
where we have used (3.6) and where we notice the presence of two additional second-order terms, compared to the naïve average of the Hamiltonian. Next, we apply this result to Hamiltonian (2.12) where, in the old coordinates \( Z \),
\[
H_0 = \hat{\mu} B(\bar{X}) + \frac{1}{2} m \bar{v}_||^2, \tag{3.15}
\]
\[
H_1 = e \psi_1 (\bar{X}, \hat{\theta}, \hat{\mu}, \bar{v}_||, t), \tag{3.16}
\]
\[
H_2 = \frac{e^2}{2mc^2} \| A_1 (\bar{X} + \rho, \epsilon_\delta t) \|^2. \tag{3.17}
\]

In the new coordinates \( Z_{gy} \), the reduced Hamiltonian is then
\[
H_{gy} = H_{ge} + \epsilon_\delta e \langle \psi_1 \rangle + \epsilon_\delta^2 \left( \frac{e^2}{2mc^2} \| A_1 (X_{gy} + \rho_{gy}, \epsilon_\delta t) \|^2 \right) - \frac{e^2}{2B(X_{gy})} \frac{\partial}{\partial \mu_{gy}} \langle \tilde{\psi}_1 \rangle^2
\]
\[
- \frac{mc^2}{2B(X_{gy})^2} \hat{b}(X_{gy}) \cdot \left( \nabla \tilde{\psi}_1 \times \int d\theta_{gy} \nabla \tilde{\psi}_1 \right), \tag{3.18}
\]
where
\[
\rho_{gy} = \frac{mc}{eB(X_{gy})} \sqrt{\frac{2\mu_{gy} B(X_{gy})}{m}} \hat{\rho}(\theta_{gy}, X_{gy}). \tag{3.19}
\]
In (3.18), the function $\psi_1$ is evaluated at $(X_{gy}, \theta_{gy}, \mu_{gy}, v_{\parallel gy}, t)$, and $H_{gc}$ is given by

$$H_{gc} = \mu_{gy} B(X_{gy}) + \frac{1}{2} m v_{\parallel gy}^2.$$  

(3.20)

The averaging has been performed using the generating function $S_1$ given by

$$S_1(X_{gy}, \theta_{gy}, \mu_{gy}, v_{\parallel gy}, t) = \frac{mc}{B} \int d\theta_{gy} \tilde{\psi}_1(X_{gy}, \theta_{gy}, \mu_{gy}, v_{\parallel gy}, t),$$  

(3.21)

which is essential in order to determine, at the leading order, the change of coordinates which has realized the reduction.

We recover the expression of the Hamiltonian obtained in (Brizard 1989; Sugama 2000; Brizard & Hahm 2007).

3.3. Changes of coordinates

Next, we look at the expression of the change of coordinates which links the particle dynamics with the gyrocentre dynamics. We recall that, at the leading order, this change is a result of two steps: a translation of the velocity by the perturbation fields and of the position by a modified Larmor radius, and an averaging performed at the Hamiltonian level using a canonical Lie transform. Below, we provide the explicit expressions at order $\epsilon\delta$.

Given our choice of generating function and the ordering of the bracket, the old coordinates $Z$ as functions of the new (gyrokinetic) ones $Z_{gy}$ are written as

$$Z = Z_{gy} - \epsilon\delta \{ S_1, Z_{gy} \}_{gc} + O(\epsilon^2\delta).$$  

(3.22)

We note that there was a first step (a modified guiding-centre step) which mapped $(x, v)$ into $(\bar{X}, \bar{\theta}, \bar{\mu}, \bar{v}_{\parallel})$:

$$x = \bar{X} + \frac{mc}{eB(\bar{X})} \sqrt{\frac{2\bar{\mu}B(\bar{X})}{m}} \hat{\rho}(\bar{\theta}, \bar{X}),$$  

$$v = \bar{v}_{\parallel} \hat{\rho}(\bar{X}) + \sqrt{\frac{2\bar{\mu}B(\bar{X})}{m}} \hat{\omega}(\bar{\theta}, \bar{X}) - \epsilon\delta e \frac{e}{mc} A_1(\bar{X} + \rho, \epsilon\delta t).$$  

(3.23, 3.24)

The second step performed the averaging of the Hamiltonian using a canonical Lie transform. It mapped $(\bar{X}, \bar{\theta}, \bar{\mu}, \bar{v}_{\parallel})$ into $(X_{gy}, \theta_{gy}, \mu_{gy}, v_{\parallel gy})$. Up to order $O(\epsilon^2\delta)$, the expressions for this second change of coordinates are

$$X_{gy} = \bar{X} + \epsilon\delta \{ S_1, \bar{X} \}_0 + \epsilon\delta \{ S_1, \bar{X} \}_1,$$

$$= \bar{X} - \epsilon\delta \left( \frac{1}{m} \hat{b} \frac{\partial S_1}{\partial v_{\parallel}} + \frac{c}{eB} \hat{b} \times \nabla S_1 \right),$$  

(3.25a)

$$\theta_{gy} = \bar{\theta} + \epsilon\delta \{ S_1, \bar{\theta} \}_1 = \bar{\theta} - \epsilon\delta \frac{e}{mc} \frac{\partial S_1}{\partial \mu} ,$$  

(3.25b)

$$\mu_{gy} = \bar{\mu} + \epsilon\delta \{ S_1, \bar{\mu} \}_1 = \bar{\mu} + \epsilon\delta \frac{e}{B} \tilde{\psi}_1,$$

(3.25c)

$$v_{\parallel gy} = \bar{v}_{\parallel}.$$

(3.25d)
By combining the two changes of coordinates, we obtain

\[ x = X_{gy} + \rho_{gy} - \epsilon \delta \{ S_1, X_{gy} + \rho_{gy} \}_g \]

\[ = X_{gy} + \rho_{gy} + \epsilon \delta \left[ -\frac{1}{B} \hat{b} \int d\theta_{gy} \widetilde{A}_1 + \frac{c}{eB} \hat{b} \times \nabla S_1 \right. \]

\[ \left. - \frac{1}{B} \sqrt{\frac{m^2 c^2}{2\mu_{gy} B}} \left( \widetilde{\psi}_1 \hat{\rho} - 2\mu_{gy} \hat{\Lambda} \int d\theta_{gy} \frac{\partial \widetilde{\psi}_1}{\partial \mu_{gy}} \right) \right], \quad (3.26a) \]

\[ v = v_{\parallel gy} \hat{b} - \frac{eB}{mc} \hat{b} \times \rho_{gy} + \epsilon \delta \frac{eB}{mc} \hat{b} \times \{ S_1, \rho_{gy} \}_g - \epsilon \delta \frac{e}{mc} A_1, \quad (3.26b) \]

where the functions \( \hat{b}, B, \hat{\rho} \) and \( \hat{\Lambda} \) are taken at \( X_{gy} \), and the functions \( \widetilde{\psi}_1, S_1 \) and its derivatives, at \( (X_{gy}, \theta_{gy}, \mu_{gy}, v_{\parallel gy}) \). The fields \( A_1 \) and \( \phi_1 \) are evaluated at \( X_{gy} + \rho_{gy} \), where \( \rho_{gy} \) is given by (3.19).

We introduce the variables \((\theta, \mu, v_{\parallel})\) associated with the velocity \( v \):

\[ v = v_{\parallel} \hat{b}(x) + \sqrt{\frac{2\mu B(x)}{m}} \hat{\Lambda}(\theta, x). \quad (3.27) \]

The expressions of the variables \((\theta, \mu, v_{\parallel})\) are given by

\[ \theta = \theta_{gy} + \epsilon \delta \frac{e}{mc} \left( \frac{\partial S_1}{\partial \mu_{gy}} + \sqrt{\frac{m}{2\mu_{gy} B}} \hat{\rho} \cdot A_1 \right), \quad (3.28a) \]

\[ \mu = \mu_{gy} - \epsilon \delta \frac{e}{B} \left( \widetilde{\psi}_1 \frac{1}{c} \sqrt{\frac{2\mu_{gy} B}{m}} \hat{\Lambda} \cdot A_1 \right), \quad (3.28b) \]

\[ v_{\parallel} = v_{\parallel gy} - \epsilon \delta \frac{e}{mc} A_1. \quad (3.28c) \]

From these expressions, we deduce the relation between \( \rho_{gy} \) and the Larmor radius:

\[ \frac{mc}{eB} \hat{b} \times v = \rho_{gy} - \epsilon \delta \{ S_1, \rho_{gy} \}_g - \epsilon \delta \frac{1}{B} \hat{b} \times A_1. \quad (3.29) \]

In the two-step reduction procedure we presented, the complexity of the derivation is shared between the Hamiltonian and the symplectic parts of the phase-space Lagrangian: First, the velocity shift allows one to move all the gyroangle dependencies from the symplectic part to the Hamiltonian, so as to readily apply the guiding-centre transform. Then a series of canonical Lie transforms are applied to the Hamiltonian to finalize the dynamical gyrocentre reduction. An alternative derivation uses a general reduction method applied to the Hamiltonian and the symplectic part at the same time, which requires applying non-canonical Lie transforms to the differential forms (see e.g. Brizard & Hahm (2007)).

4. Gyrokinetic Vlasov–Maxwell equations

We derive the gyrokinetic Maxwell–Vlasov equations following Sugama (2000) which provides the following Lagrangian:
\[ L = \sum_{sp} \int dV_0 dW_0 F(Z_0, t_0) L_p(Z_{gy}(Z_0, t_0; t), \dot{Z}_{gy}(Z_0, t_0; t), t) \]
\[ + \int dV \frac{|E|^2 - |B + \epsilon_\delta \nabla \times A_1|^2}{8\pi}, \quad (4.1) \]

where
\[ L_p = \left( \frac{e}{c} A + mc_{gy} \hat{b} \right) \cdot \dot{X}_{gy} + \frac{mc}{e} \mu_{gy} \dot{\theta}_{gy} - H_{gy}, \quad (4.2) \]

and where the gyrocentre distribution function of the species \( sp \) \( F(Z_0, t_0) \) is defined at the arbitrary initial gyrocentre phase-space position \( Z_0 \) and arbitrary initial time \( t_0 \). We will not use a specific notation to distinguish the distribution functions of the different species for simplicity. The reduced phase-space variables are \( Z_{gy} = (X_{gy}, v_{gy}, \mu_{gy}) \) and the phase-space volume element is given by \( d\Omega = dV_0 dW_0 \) with \( dV_0 \) denoting the volume element in physical space, i.e. \( dV_0 = d^3X_{gy} \) for the gyrocentre part and \( dV = d^3x \) for the electromagnetic part; the velocity gyrocentre phase-space volume is \( dW_0 = B^*_1(Z_0) dv_{gy} d\mu_{gy} d\theta_{gy} \).

We perform the change of variables \( Z_{gy} = Z_{gs}(Z_0, t_0; t) \), such that the first term of (4.1) becomes:
\[ \sum_{sp} \int d\Omega \; F(Z_{gy}, t) L_p(Z_{gy}, \dot{Z}_{gy}, t). \quad (4.3) \]

The gyrokinetic Vlasov equation is obtained using the gyrocentre characteristics, from the conservation of the distribution function along the trajectories, i.e.
\[ \frac{d}{dt} F(Z_{gy}(Z_0, t_0; t), t) = \frac{\partial}{\partial t} F(Z_{gy}, t) + \dot{Z}_{gy} \cdot \frac{\partial}{\partial Z_{gy}} F(Z_{gy}, t) = 0. \quad (4.4) \]

For the electromagnetic part of the Lagrangian, we use the quasi-neutrality and Darwin approximation (see, e.g. Krause, Apte & Morrison (2007), Bottino & Sonnendrücker (2015) for more details) which boils down to neglecting the \( E^2 \) term in the Lagrangian. The resulting expression for the gyrokinetic Lagrangian used for the derivation of the gyrokinetic Poisson and Ampère equations is:
\[ L = \sum_{sp} \int d\Omega \; F(Z_{gy}, t) L_p(Z_{gy}, \dot{Z}_{gy}, t) - \int dV \frac{|\nabla \times (A_0 + \epsilon_\delta A_1)|^2}{8\pi} + O(\epsilon_\delta^3). \quad (4.5) \]

### 4.1. Gyrokinetic Vlasov–Maxwell equations

In this section, we provide the gyrokinetic Vlasov–Maxwell equations in the weak form, since this form is well suited to the finite-element discretization, necessary for Particle-In-Cell Monte Carlo simulations as performed in ORB5.

The gyrokinetic quasi-neutrality equation is obtained in the weak form using the functional derivatives of the action:
\[ 0 = \frac{\delta L}{\delta \phi_1} \circ \hat{\phi}_1 = \epsilon_\delta \sum_{sp} \int d\Omega \; F \left( -e(\hat{\phi}_1) + \epsilon_\delta \frac{e^2}{B} \frac{\partial}{\partial \mu_{gy}} \left( \langle \Psi_1 \hat{\phi}_1 \rangle - \langle \Psi_1 \rangle \langle \hat{\phi}_1 \rangle \right) \right. \]
\[ + \epsilon_\delta \frac{mc^2}{2B^2} \hat{b} \left. \left\langle \nabla \hat{\phi}_1 \times \left( \int d\theta_{gy} \nabla \tilde{\Psi}_1 + \nabla \tilde{\Psi}_1 \times \left( \int d\theta_{gy} \nabla \tilde{\phi}_1 \right) \right) \right. \right), \quad (4.6) \]

where \( \hat{\phi}_1 \) is a test function, evaluated at \( X_{gy} + \rho_{gy} \).
The parallel component of the gyrokinetic Ampère equation is given by

\[
0 = \frac{\partial L}{\partial A_{1\parallel}} \circ \hat{A}_{1\parallel} = -\frac{\epsilon_\delta}{4\pi} \int dV \hat{A}_{1\parallel} \cdot \mathbf{b} \times \mathbf{B} + \epsilon_\delta \sum_{sp} e \int d\Omega \ F \frac{v_{\parallel\rm gy}}{c} (\hat{A}_{1\parallel})
- \frac{\epsilon_\delta^2}{4\pi} \int dV \nabla \hat{A}_{1\parallel} \cdot [\mathbf{b} \times (\nabla \times \mathbf{A}_1)] - \epsilon_\delta^2 \sum_{sp} \frac{e^2}{mc^2} \int d\Omega \ F (\hat{A}_{1\parallel} \hat{A}_{1\parallel})
- \epsilon_\delta^2 \sum_{sp} \int d\Omega \ F \frac{v_{\parallel\rm gy}}{2B^2} \cdot \left\langle \nabla \hat{\psi}_1 \times \int d\theta_{\rm gy} \nabla \hat{\psi}_1 + \nabla \hat{\psi}_1 \times \int d\theta_{\rm gy} \nabla \hat{A}_{1\parallel}\right\rangle,
\]

(4.7)

where \( \hat{A}_{1\parallel} \) is a test function, evaluated at \( X_{\rm gy} + \rho_{\rm gy} \) when the integral is over \( d\Omega \), and evaluated at \( X_{\rm gy} \) when the integral is over \( dV \). The perpendicular component of the gyrokinetic Ampère equation is given by:

\[
0 = \frac{\delta L}{\delta A_{1\perp}} \circ \hat{A}_{1\perp}
= -\frac{\epsilon_\delta}{4\pi} \int dV \ A_{1\perp} \cdot (\nabla \times \mathbf{B}) + \epsilon_\delta \sum_{sp} e \int d\Omega \ F \sqrt{\frac{2\mu_{\rm gy}B}{mc^2}} (\hat{\mathbf{A}} \cdot \hat{A}_{1\perp})
- \frac{\epsilon_\delta^2}{4\pi} \int dV (\nabla \times \mathbf{A}_1) \cdot (\nabla \times \hat{A}_{1\perp}) - \epsilon_\delta^2 \sum_{sp} \frac{e^2}{mc^2} \int d\Omega \ F (\hat{A}_1 \hat{A}_{1\perp})
- \epsilon_\delta^2 \sum_{sp} \int d\Omega \ F \frac{mc^2}{B^2} \sqrt{\frac{2\mu_{\rm gy}B}{mc^2}} \cdot \left\langle \nabla (\hat{\mathbf{A}} \times \hat{A}_{1\perp}) \times \int d\theta_{\rm gy} \nabla \hat{\psi}_1 + \nabla \hat{\psi}_1 \times \int d\theta_{\rm gy} \nabla (\hat{\mathbf{A}} \cdot \hat{A}_{1\perp})\right\rangle,
\]

(4.8)

where \( \hat{A}_{1\perp} \) is a test function, evaluated at \( X_{\rm gy} + \rho_{\rm gy} \) when the integral is over \( d\Omega \), and evaluated at \( X_{\rm gy} \) when the integral is over \( dV \).

We remark that both Ampère’s laws (4.7) and (4.8) are identical to the results presented in Sugama (2000) and Brizard & Hahm (2007).

4.2. Gyrokinetic particle characteristics and the gyrokinetic Vlasov equation

The gyrokinetic Vlasov equation is obtained from the gyrocentre particle characteristics following (4.4) and taking into account that \( d\mathbf{Z}/dt = \{\mathbf{Z}, H_{\rm gy}\}_{\rm gc} \). The gyrocentre characteristics are given by:

\[
\dot{X}_{\rm gy} = \left\{X_{\rm gy}, H_{\rm gy}\right\}_{\rm gc} = \frac{\mathbf{c}}{\mathbf{eB}_\parallel} \times \nabla H_{\rm gy} + \frac{\mathbf{B}_\times}{mB^2} \frac{\partial H_{\rm gy}}{\partial v_{\parallel\rm gy}},
\]

(4.9a)

\[
\dot{v}_{\parallel\rm gy} = \left\{v_{\parallel\rm gy}, H_{\rm gy}\right\}_{\rm gc} = -\frac{\mathbf{B}_\times}{mB^2} \cdot \nabla H_{\rm gy},
\]

(4.9b)
where \( H_{gy} \) is given by (3.18). The fully nonlinear gyrokinetic Vlasov equation for the Vlasov distribution \( F(X_{gy}, v_{\parallel gy}, \mu_{gy}, t) \) is therefore given by:

\[
0 = \frac{dF}{dt} = \frac{\partial F}{\partial t} + \{X_{gy}, H_{gy}\}_{gc} \cdot \nabla F + \{v_{\parallel gy}, H_{gy}\}_{gc} \frac{\partial F}{\partial v_{\parallel gy}},
\]

where we notice that the term in \( \frac{\partial F}{\partial \mu_{gy}} \) is absent since \( \{\mu_{gy}, H_{gy}\}_{gc} = 0 \). The Vlasov equation can be rewritten with the help of the Poisson bracket (2.9):

\[
\frac{\partial F}{\partial t} = -\ F, H_{gy}_{gc}.
\]

Equation (4.11) for the Vlasov equation and equations (4.6), (4.7) and (4.8) for the Maxwell equations constitute the second-order gyrokinetic Vlasov–Maxwell equations associated with the second-order Hamiltonian for the gyrocentres given by (3.18), consistent with the ordering \( \epsilon B \sim \epsilon^2 \). We notice that due to the quasi-neutrality and the Darwin approximations, the gyrokinetic Poisson and Ampère equations do not contain explicit time derivatives.

Acknowledgements

The authors would like to thank C. Angioni, A. Bottino, A. J. Brizard, J. W. Burby, A. Mishchenko, B. D. Scott and E. Sonnendrücker for helpful discussions. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training Programme No. 2014–2018 under grant agreement no. 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Appendix A. Littlejohn’s guiding-centre theory

In this appendix, we revisit Littlejohn’s guiding-centre theory, following Littlejohn (1983). We begin with the one-form for the motion of the particle:

\[
\gamma = \left( \frac{e}{c} A(x) + m v \right) \cdot dx - H dt.
\]

The revisit comes from the fact that we do not consider \( \epsilon \) as a small parameter as it was done in Littlejohn (1979, 1981, 1983). Here the small parameter is \( \epsilon_B \) which relates to the spatial variation of the external magnetic field.

The idea of the guiding centre is to find a change of coordinates which removes the fluctuating part from the one-form. The guiding-centre transformation is based on two different changes of coordinates, a far-from-identity transformation which consists of a shift of the position by the Larmor radius, and a near-identity transformation which eliminates the fluctuating terms of the one-form at order \( \epsilon_B \).

The first step is a translation by the Larmor radius, inspired by the situation where the magnetic field is constant and uniform:

\[
x = \bar{x} + \rho_0.
\]

We translate the velocity, by defining

\[
w = v + \frac{e}{mc} \left[ A(\bar{x} + \rho_0) - A(\bar{x}) - (\rho_0 \cdot \nabla)A(\bar{x}) - \frac{1}{2}(\rho_0 \rho_0 : \nabla \nabla)A(\bar{x}) \right].
\]
where the last term written in indices is $-(1/2)\rho_i\rho_j \partial^2 A/\partial \tilde{x}_i \partial \tilde{x}_j$. We notice that the quantity with which the velocity has been translated is of order $\epsilon^2_B$ (since it involves third derivatives of the vector potential, i.e. second derivatives of the magnetic field). We decompose the velocity $w$ in the following way:

$$w = w \hat{b}(x) + w \hat{1}(\theta, x), \quad (A4)$$

where the orthonormal basis $(\hat{b}, \hat{1}, \hat{\rho})$ is defined in § 2. We decompose the one-form as

$$\gamma = \gamma_0 + \gamma_1, \quad (A5)$$

where

$$\gamma_0 = \left( \frac{e}{c} A(\tilde{x}) + mw \hat{b}(\tilde{x}) \right) \cdot d\tilde{x} - H \, dt, \quad (A6)$$

and

$$\gamma_1 = \left( \frac{e}{c} (\rho_0 \cdot \nabla)A + \frac{e}{2c} (\rho_0 \rho_0 : \nabla \nabla)A + mw \hat{1}(\theta, \tilde{x} + \rho_0) + mw \hat{b}(\tilde{x} + \rho_0) \right) \cdot d\rho_0. \quad (A7)$$

We use a gauge invariance to simplify the one-form: $\gamma$ can be replaced by $\gamma + d\sigma$ where $\sigma$ is any scalar function of $(\tilde{x}, w, w \bot)$. Looking at the shape of $\gamma_1$, especially the terms in $d\rho_0$, some terms are removed by considering

$$\sigma_1 = -\frac{e}{c} A \cdot \rho_0 - \frac{e}{2c} (\rho_0 \cdot \nabla)A \cdot \rho_0 - \frac{e}{6c} (\rho_0 \rho_0 : \nabla \nabla)A \cdot \rho_0. \quad (A8)$$

In what follows unless otherwise specified, the dependence of the functions on the variables is omitted when these functions are unambiguously expressed in the current set of variables. The one-form becomes

$$\gamma_1 + d\sigma_1 = \left( \frac{eB}{c} \hat{b} \times \rho_0 + \frac{e}{2c} (\rho_0 \cdot \nabla)(B\hat{b}) \times \rho_0 + mw \hat{1} + mw \hat{b}(\rho_0 \cdot \nabla)\hat{b} + \rho_0 \rho_0 : \nabla \nabla) \hat{b} \right) \cdot d\rho_0, \quad (A9)$$

where we have neglected the contributions of order $\epsilon^2_B$. In the above expression we have used the identity

$$(\rho_0 \cdot \nabla)A - \nabla A \cdot \rho_0 = B\hat{b} \times \rho_0. \quad (A10)$$
The condition for $\rho$ is
\[
\frac{eB}{c} \hat{b} \times \rho_0 + mw_\perp \hat{\rho} = 0, \tag{A11}
\]
i.e.
\[
\rho_0 = \frac{mw_\perp c}{eB} \hat{\rho}. \tag{A12}
\]
The one-form becomes
\[
\gamma_1 + d\sigma_1 = \frac{m^2 c}{eB} \left( \frac{w^2_\perp}{2B} (\hat{\rho} \cdot \nabla)(\hat{b} \hat{b}) \times \hat{\rho} + w^2_\perp (\hat{\rho} \cdot \nabla) \hat{\rho} \right) + m^2 c \left[ \frac{w_\perp^2}{2} \hat{\rho} + \frac{1}{2} \frac{mc}{eB} \left( \frac{w^2_\perp}{3B} (\hat{\rho} \cdot \nabla)(\hat{b} \hat{b}) \times \hat{\rho} + w^2_\perp (\hat{\rho} \cdot \nabla) \hat{\rho} \right) \right] \cdot d\hat{x}
\]
\[
+ w_\parallel w_\perp (\hat{\rho} \cdot \nabla) \hat{b} \right] \cdot d(w_\perp \hat{\rho}). \tag{A13}
\]
The leading-order term (in the small parameter $\epsilon_B$) is
\[
\frac{m^2 c}{eB} \left( \frac{w^2_\perp}{2B} \hat{b} + \frac{1}{2} w_\perp \hat{\theta} \right) \cdot d(w_\perp \hat{\rho}) = \frac{m^2 c}{eB} \left[ \frac{w^2_\perp}{2} d\theta - \left( \frac{w^2_\perp}{2} \nabla \cdot \hat{\rho} - w_\parallel w_\perp \nabla \hat{\rho} \hat{b} \right) \right] \cdot d\hat{x}, \tag{A14}
\]
since $\hat{\Lambda} = \partial \hat{\rho} / \partial \theta$ and where we have used $\nabla \hat{\rho} \cdot \hat{\rho} = -\nabla \hat{\rho} \cdot \hat{\rho}$. We notice that the terms in $d\hat{x}$ are of order $\epsilon_B$. We rewrite the one-form $\gamma$ as
\[
\gamma + d\sigma_1 = \tilde{\gamma}_0 + \tilde{\gamma}_1, \tag{A15}
\]
where
\[
\tilde{\gamma}_0 = \left( \frac{e}{c} A(\tilde{x}) + mw_\parallel \hat{b}(\tilde{x}) \right) \cdot d\tilde{x} + \frac{m^2 w_\perp^2 c}{2eB} d\theta - H dt, \tag{A16}
\]
is the leading order and
\[
\tilde{\gamma}_1 = \frac{m^2 c}{eB} \left( \frac{w^2_\perp}{2B} (\hat{\rho} \cdot \nabla)(\hat{b} \hat{b}) \times \hat{\rho} + w^2_\perp (\hat{\rho} \cdot \nabla) \hat{\rho} \right) \cdot d\hat{x}
\]
\[
+ \frac{m^2 c^2}{eB^2} \left( \frac{w^2_\perp}{3B} (\hat{\rho} \cdot \nabla)(\hat{b} \hat{b}) \times \hat{\rho} + w^2_\perp (\hat{\rho} \cdot \nabla) \hat{\rho} \right)
\]
\[
+ w_\parallel w_\perp (\hat{\rho} \cdot \nabla) \hat{b} \right) \cdot (\hat{\rho} dw_\perp + w_\perp \hat{\theta} d\theta), \tag{A17}
\]
is of order $\epsilon_B$. Here we have used the identity $(\hat{\rho} \cdot \nabla) \hat{b} + \nabla \hat{\rho} \cdot \hat{b} = (\nabla \times \hat{b}) \times \hat{\rho}$. We notice that the order $\epsilon_B$ of the one-form, i.e. $\tilde{\gamma}_1$, contains fluctuating terms in $\theta$. These terms are eliminated by a near-identity transformation which we consider at order one in $\epsilon_B$:
\[
\tilde{x} = X + \xi(X, W_\parallel, W_\perp, \Theta), \tag{A18a}
\]
Second-order nonlinear gyrokinetic theory

\[ w_\parallel = W_\parallel + \mathcal{W}_\parallel(X, W_\parallel, W_\perp, \Theta), \]  
\[ w_\perp = W_\perp + \mathcal{W}_\perp(X, W_\parallel, W_\perp, \Theta), \]  
\[ \theta = \Theta + \mathcal{T}(X, W_\parallel, W_\perp, \Theta), \]

where the unknown functions \( \xi \) and \( \mathcal{W}_\parallel \), \( \mathcal{W}_\perp \) and \( \mathcal{T} \) are of order \( \epsilon_B \).

Since \( \tilde{\gamma}_1 \) is of order \( \epsilon_B \), its expression in the new coordinates \((X, W_\parallel, W_\perp, \Theta)\) is exactly the same as in the old coordinates \((\hat{x}, w_\parallel, w_\perp, \theta)\) up to \( \epsilon^2_B \) terms. It remains to expand \( \tilde{\gamma}_0 \) at the first order in \( \epsilon_B \). The expansion leads to

\[ \tilde{\gamma}_0 = \left( \frac{e}{c} A(X) + mW_\parallel \hat{b}(X) \right) \cdot dX + \frac{m^2 W^2_\parallel c}{2eB} d\Theta - H dt + \tilde{\gamma}_2, \]  
(A 19)

where

\[ \tilde{\gamma}_2 = \left( \frac{e}{c} (\xi \cdot \nabla) A + m\mathcal{W}_\parallel \hat{b} \right) \cdot dX + \left( \frac{e}{c} A + mW_\parallel \hat{b} \right) \cdot d\xi \]
\[ + \frac{m^2 c}{eB} W_\perp \mathcal{W}_\perp d\Theta + \frac{m^2 W^2_\parallel c}{2eB} d\mathcal{T}. \]  
(A 20)

We notice that we have used the approximation \( B(X + \xi) \approx B(X) \) since the gradients of \( B \) are of order \( \epsilon_B \) and \( \xi \) is also of order \( \epsilon_B \), so the difference \( B(X + \xi) - B(X) \) is of order \( \epsilon^2_B \). The same approximation holds for \( \hat{b} \). Using a gauge transformation similar to (A 8), we simplify \( \tilde{\gamma}_2 \) into

\[ \tilde{\gamma}_2 + d\sigma_2 = \left( \frac{eB}{c} \hat{b} \times \xi + m\mathcal{W}_\parallel \hat{b} \right) \cdot dX + mW_\parallel \hat{b} \cdot d\xi \]
\[ + \frac{m^2 c}{eB} W_\perp \mathcal{W}_\perp d\Theta + \frac{m^2 W^2_\parallel c}{2eB} d\mathcal{T}, \]  
(A 21)

with

\[ \sigma_2 = - \frac{e}{c} A \cdot \xi. \]  
(A 22)

We look at the one-form \( \tilde{\gamma}_1 + \tilde{\gamma}_2 + d\sigma_2 \) and determine the unknown functions \( \xi \), \( \mathcal{W}_\parallel \), \( \mathcal{W}_\perp \) and \( \mathcal{T} \) such that this one-form no longer possesses \( \Theta \)-dependent terms. We notice that the spatial derivatives of \( \mathcal{T} \) and \( \xi \) are of order \( \epsilon_B \); therefore, the terms in \( d\xi \) and \( d\mathcal{T} \) only contain terms in \( dW_\parallel \), \( dW_\perp \) and \( d\Theta \). For the same reason, the terms in \( d\rho_0 \) in \( \tilde{\gamma}_1 \) only involve terms in \( dW_\perp \) and \( d\Theta \).

Combining (A 17) and (A 21), the one-form \( \tilde{\gamma}_1 + \tilde{\gamma}_2 + d\sigma_2 \) can be written as

\[ \tilde{\gamma}_1 + \tilde{\gamma}_2 + d\sigma_2 = \Gamma_X \cdot dX + \Gamma_\theta d\Theta + \Gamma_{W_\parallel} dW_\parallel + \Gamma_{W_\perp} dW_\perp. \]  
(A 23)

The terms in \( dX \) are

\[ \Gamma_X = \frac{eB}{c} \hat{b} \times \xi + m\mathcal{W}_\parallel \hat{b} + \frac{m^2 c}{eB} \left[ \frac{W^2_\parallel}{2B} (\hat{\rho} \cdot \nabla)(B\hat{b}) \times \hat{\rho} + W^2_\perp (\hat{\rho} \cdot \nabla) \hat{\mathcal{X}} - \frac{W^2_\parallel}{2} \nabla \hat{\mathcal{X}} \cdot \hat{\rho} \right] \]
\[ + W_\parallel W_\perp (\nabla \times \hat{b}) \times \hat{\rho}. \]  
(A 24)
The terms in $\mathrm{d}\Theta$ are
\[
\Gamma_\Theta = mW_\parallel \hat{b} \cdot \frac{\partial \xi}{\partial \Theta} + \frac{m^2 W_\perp^2 c}{2eB} \frac{\partial \mathcal{T}}{\partial \Theta} + \frac{m^2 c}{eB} W_\perp \mathcal{W}_\perp 
+ \frac{m^2 c^2}{e^2 B^2} \left[ -\frac{W_\perp^2}{3B} \hat{\rho} \cdot \nabla B + W_\parallel W_\perp^2 (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\lambda} \right], \tag{A 25}
\]
where we have used the identities $\nabla \hat{b} \cdot \hat{b} = \nabla \hat{\lambda} \cdot \hat{\lambda} = 0$ since $\hat{b}$ and $\hat{\lambda}$ are of unit norm. The terms in $\mathrm{d}W_\parallel$ are
\[
\Gamma_\parallel = mW_\parallel \hat{b} \cdot \frac{\partial \xi}{\partial W_\parallel} + \frac{m^2 W_\perp^2 c}{2eB} \frac{\partial \mathcal{T}}{\partial W_\parallel}, \tag{A 26}
\]
and the terms in $\mathrm{d}W_\perp$ are
\[
\Gamma_\perp = mW_\parallel \hat{b} \cdot \frac{\partial \xi}{\partial W_\perp} + \frac{m^2 W_\perp^2 c}{2eB} \frac{\partial \mathcal{T}}{\partial W_\perp} + \frac{m^2 c^2}{e^2 B^2} \left[ W_\perp^2 (\hat{\rho} \cdot \nabla) \hat{\lambda} \cdot \hat{\rho} + W_\parallel W_\perp^2 (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\rho} \right], \tag{A 27}
\]
The terms in $\Gamma_X$ perpendicular to $\hat{b}$ determine the perpendicular component of $\xi$, whereas the terms parallel to $\hat{b}$ determine $\mathcal{W}_\parallel$. The perpendicular component of $\xi$ is obtained from the cross-product of $\Gamma_X$ with $\hat{b}$, and is given by
\[
\xi_\perp = \frac{m^2 c^2}{e^2 B^2} \left( -\frac{W_\perp^2}{2B} (\hat{\rho} \cdot \nabla B) \hat{\rho} + W_\perp^2 \hat{b} \times (\hat{\rho} \cdot \nabla) \hat{\lambda} - W_\parallel W_\perp (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\rho} \right), \tag{A 28}
\]
where we have used the fact that $\hat{b} \times [(\hat{\rho} \cdot \nabla) \hat{b} \times \hat{\rho}] = 0$ which comes from $\nabla \hat{b} \cdot \hat{b} = 0$. The scalar product between $\Gamma_X$ and $\hat{b}$ leads to
\[
\mathcal{W}_\parallel = \frac{mc}{eB} \left( W_\perp^2 \frac{1}{2} (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\lambda} - W_\parallel W_\perp (\nabla \hat{b}) \cdot \hat{\lambda} \right), \tag{A 29}
\]
where we have used the identity $\nabla \hat{\lambda} \cdot \hat{b} = -\nabla \hat{b} \cdot \hat{\lambda}$. In $\Gamma_X$, which this choice of functions, it remains
\[
\Gamma_X = -\frac{mc}{e} \mu R, \tag{A 30}
\]
where $R = \nabla \hat{\lambda} \cdot \hat{\rho} = \nabla \hat{b}_\parallel \cdot \hat{b}_\perp$ which is independent of the gyroangle $\Theta$.

The first two terms in $\Gamma_\Theta$, $\Gamma_\parallel$ and $\Gamma_\perp$ calls for a gauge transformation with
\[
\sigma_3 = -mW_\parallel \hat{b} \cdot \xi - \frac{m^2 W_\perp^2 c}{2eB} \mathcal{T}. \tag{A 31}
\]
The terms in $\Gamma_\Theta$, $\Gamma_\parallel$ and $\Gamma_\perp$ become
\[
\tilde{\Gamma}_\Theta = \frac{m^2 c}{eB} W_\perp \mathcal{W}_\perp + \frac{m^2 c^2}{e^2 B^2} \left[ -\frac{W_\perp^2}{3B} \hat{\rho} \cdot \nabla B + W_\parallel W_\perp^2 (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\lambda} \right], \tag{A 32a}
\]
\[
\tilde{\Gamma}_\parallel = -m\hat{b} \cdot \xi, \tag{A 32b}
\]
With the change of coordinates \((A \ 18)\), the Hamiltonian has been changed into
\[
\mathcal{H} = -\frac{m^2 c}{eB} W_\perp \mathcal{T} + \frac{m^2 c^2}{e^2 B^2} \left[ W_\perp^2 (\hat{\rho} \cdot \nabla) \mathcal{\hat{\rho}} \cdot \mathcal{\hat{\rho}} + W_\parallel W_\perp (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right]. \quad (A 32c)
\]

Using this gauge transformation, the equations which determine the unknown functions \(W_\perp, W_\parallel\) and the parallel component of \(\xi\) have been decoupled. We choose \(\mathcal{T}\) such that it only eliminates the fluctuating terms. Using the following expression for the fluctuating part of \((\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}}\):
\[
(\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} = \frac{1}{2} \left( (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} - (\mathcal{\hat{\rho}} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right), \quad (A 33)
\]
we obtain the following expression for \(\mathcal{T}\):
\[
\mathcal{T} = \frac{mc}{eB} \left[ W_\perp (\hat{\rho} \cdot \nabla) \mathcal{\hat{\rho}} \cdot \mathcal{\hat{\rho}} + \frac{W_\parallel}{2} \left( (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} - (\mathcal{\hat{\rho}} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right) \right]. \quad (A 34)
\]

In the one-form there remains a term in \(dW_\perp\):
\[
\frac{m^3 c^2}{2e^2 B^2} W_\parallel W_\perp \left[ (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} + (\mathcal{\hat{\rho}} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right] dW_\perp. \quad (A 35)
\]

Using a gauge transformation with
\[
\sigma_4 = -\frac{m^2 c^2}{4e^2 B^2} W_\parallel W_\perp^2 \left[ (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} + (\mathcal{\hat{\rho}} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right], \quad (A 36)
\]
the term in \(dW_\perp\) is eliminated and a new term appears in \(dW_\parallel\), which then determines the parallel component of \(\xi\):
\[
\mathcal{\hat{b}} \cdot \mathcal{\hat{\xi}} = -\frac{m^3 c^2}{4e^2 B^2} W_\perp^2 \left[ (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} + (\mathcal{\hat{\rho}} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right]. \quad (A 37)
\]

For \(\mathcal{W}_\perp\), we choose
\[
\mathcal{W}_\perp = \frac{mc}{eB} \left[ \frac{W_\perp^2}{3B} \mathcal{\hat{\rho}} \cdot \nabla B - W_\parallel W_\perp (\hat{\rho} \cdot \nabla) \mathcal{\hat{b}} \cdot \mathcal{\hat{b}} \right], \quad (A 38)
\]
in order to eliminate terms in \(d\Theta\). Other choices of functions \(\xi, \mathcal{W}_\parallel, \mathcal{W}_\perp\) and \(\mathcal{T}\) are possible and lead to equivalent theories, also when pushed to high orders (see Parra & Calvo (2011), Burby, Squire & Qin (2013), Parra et al. (2014), Tronko & Brizard (2015) for more details). Using the change of variables defined by (A 18) with the functions defined above, the \(\Theta\)-dependence has been removed from the symplectic part of the one-form. However there is still some dependence in \(\Theta\) in the Hamiltonian.

Now, we proceed with the last step which is a canonical Lie transform of the Hamiltonian as in § 3. This transformation does not affect the symplectic part of the one-form and is only designed to eliminate the \(\Theta\)-dependence in the Hamiltonian. With the change of coordinates (A 18), the Hamiltonian has been changed into
\[
H = \frac{1}{2} m W_\parallel^2 + \frac{1}{2} m W_\perp^2 + m (\mathcal{W}_\parallel \mathcal{W}_\parallel + W_\perp \mathcal{W}_\perp) + O(\epsilon_B^2). \quad (A 39)
\]

We consider a canonical Lie transform with the following generating function
\[
\frac{\partial S}{\partial \Theta} + \frac{m^2 c}{eB} (W_\parallel \mathcal{\tilde{W}}_\parallel + W_\perp \mathcal{\tilde{W}}_\perp) = 0, \quad (A 40)
\]
where $\tilde{\mathbf{W}}_\parallel$ and $\tilde{\mathbf{W}}_\perp$ are the fluctuating part of $\mathbf{W}_\parallel$ and $\mathbf{W}_\perp$. We notice that the gradient of the generating function, i.e., $\nabla S$, is of order $\epsilon_2^2$. The way to determine the generating function $S$ follows from the same principle as briefly explained in § 3. The Hamiltonian becomes

$$H = \frac{1}{2}mW^2_\parallel + \frac{1}{2}mW^2_\perp + m(W_\parallel \langle \mathbf{W}_\parallel \rangle + W_\perp \langle \mathbf{W}_\perp \rangle) = \frac{1}{2}mW^2_\parallel + \mu B + \frac{mc}{2e} \mathbf{W}_\parallel \mu \hat{\mathbf{b}} \cdot (\nabla \times \hat{\mathbf{b}}),$$

(A 41)

where $\mu = mW^2_\perp/(2B)$. Here we have used the identity $\langle (\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}} \rangle = - (\hat{\mathbf{b}} \cdot (\nabla \times \hat{\mathbf{b}}))/2$. In order to complete the guiding-centre derivation, we need to specify the full change of coordinates, which is a combination of the change given by (A 18a)–(A 18d) and the canonical Lie transform, which is, up to order $\epsilon_2^2$ terms:

$$\hat{x} = X + \xi + [S, X]_{gc} = X + \xi + (\hat{\mathbf{b}} \cdot \xi) \hat{\mathbf{b}} - \frac{\hat{\mathbf{b}}}{m} \frac{\partial S}{\partial \mathbf{W}_\parallel},$$

(A 42a)

$$w_\parallel = W_\parallel + W_\parallel + [S, W_\parallel]_{gc} = W_\parallel + W_\parallel,$$

(A 42b)

$$w_\perp = W_\perp + \mathbf{W}_\perp + [S, W_\perp]_{gc} = W_\perp + \mathbf{W}_\perp + \frac{eB}{m^2 \omega_{\perp c}} \frac{\partial S}{\partial \Theta},$$

(A 42c)

$$\theta = \Theta + \mathcal{T} + [S, \Theta]_{gc} = \Theta + \mathcal{T} - \frac{eB}{m^2 \omega_{\perp c}} \frac{\partial S}{\partial \mathbf{W}_\perp},$$

(A 42d)

where the Poisson bracket is obtained from the symplectic part of the one-form, and is given by

$$\{F, G\}_{gc} = \frac{e}{mc} \left( \frac{\partial F}{\partial \mu} \frac{\partial G}{\partial \mu} - \frac{\partial F}{\partial \mu} \frac{\partial G}{\partial \Theta} \right) + \frac{B^*}{mb_\parallel} \cdot \left( \nabla^* F \frac{\partial G}{\partial W_\parallel} - \frac{\partial F}{\partial W_\parallel} \nabla^* G \right) - \frac{c\hat{\mathbf{b}}}{eB_\parallel} \cdot (\nabla^* F \times \nabla^* G),$$

(A 43)

where

$$\nabla^* = \nabla - R \frac{\partial}{\partial \theta},$$

(A 44)

$$B^* = B + \frac{mc}{e} \mathbf{W}_\parallel \nabla \times \hat{\mathbf{b}} - \frac{mc^2}{e^2} \mu \nabla \times R,$$

(A 45)

and $B^* = B^* \cdot \hat{\mathbf{b}}$. In order to obtain explicit expressions for the change of coordinates, we specify the fluctuating part of the functions $\mathbf{W}_\parallel$ and $\mathbf{W}_\perp$:

$$\tilde{\mathbf{W}}_\perp = \frac{mc}{eB} \left( \frac{W^2_\perp}{3B} \hat{\rho} \cdot \nabla B - \frac{W_\perp W_\parallel}{2} \left( (\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}} + (\hat{\mathbf{\nabla}} \cdot \hat{\mathbf{b}}) \hat{\rho} \right) \right),$$

(A 46)

$$\tilde{\mathbf{W}}_\parallel = \frac{mc}{eB} \left( \frac{W^2_\parallel}{4} \left( (\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}} + (\hat{\mathbf{\nabla}} \cdot \hat{\mathbf{b}}) \hat{\rho} \right) - W_\parallel W_\perp (\nabla \times \hat{\mathbf{b}}) \cdot \hat{\mathbf{\nabla}} \right),$$

(A 47)

where we have used the fact that the fluctuating part of $(\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}}$ is given by

$$\langle (\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}} \rangle = \frac{1}{2} \left( (\hat{\rho} \cdot \nabla) \hat{\mathbf{b}} \cdot \hat{\mathbf{\nabla}} + (\hat{\mathbf{\nabla}} \cdot \hat{\mathbf{b}}) \hat{\rho} \right).$$

(A 48)
The generating function $S$ is chosen to be purely fluctuating:

$$S = \frac{m^3 c^2}{e^2 B^2} \left[ \frac{W^3}{3B} \mathbf{\hat{\lambda}} \cdot \nabla B + \frac{W_\perp W^2}{8} \left( (\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\rho} - (\mathbf{\hat{\lambda}} \cdot \nabla) \hat{b} \cdot \mathbf{\hat{\lambda}} \right) + W_\perp^2 (\nabla \times \hat{b}) \cdot \hat{\rho} \right],$$

(A 49)

where we have used the fluctuating part of $(\hat{\rho} \cdot \nabla) \hat{b} \cdot \hat{\rho}$ given by (A 33). Therefore the change of coordinates from $(x, v)$ to $(X, W_\|, W_\perp, \Theta)$ is given by

$$x = X + \frac{m W_\perp c}{e B} \hat{\rho} + \frac{m^2 c^2}{e^2 B^2} \left( -\frac{W_\perp^2}{2B} (\hat{\rho} \cdot \nabla B) \hat{\rho} + 2(\mathbf{\hat{\lambda}} \cdot \nabla B) \mathbf{\hat{\lambda}} \right)$$

$$- W_\| W_\perp \left( \hat{b} \cdot \nabla \times \hat{b} \right) \hat{\rho} + \frac{W_\perp^2}{8} \left( 3\hat{b} \cdot \nabla \hat{b} \cdot \hat{\rho} + \mathbf{\hat{\lambda}} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} \right) \hat{b} + W_\perp^2 \left( (\mathbf{\hat{\lambda}} \cdot \nabla \times \hat{b}) \hat{\rho} - (\hat{\rho} \cdot \nabla \times \hat{b}) \mathbf{\hat{\lambda}} \right)$$

$$+ \frac{W_\perp^2}{4} \left( -3\hat{\rho} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} + \mathbf{\hat{\lambda}} \cdot \nabla \hat{b} \cdot \hat{\rho} \right) \hat{\rho}$$

$$+ \frac{W_\perp^2}{4} \left( \hat{\rho} \cdot \nabla \hat{b} \cdot \hat{\rho} - \mathbf{\hat{\lambda}} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} \right) \mathbf{\hat{\lambda}},$$

(A 50a)

$$v_\| = W_\| + \frac{mc}{e B} \left( \frac{W_\perp^2}{2} \hat{\rho} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} - W_\| W_\perp \nabla \times \hat{b} \cdot \mathbf{\hat{\lambda}} \right),$$

(A 50b)

$$v_\perp = W_\perp + \frac{mc}{e B} \left( -\frac{3W_\perp W_\|}{4} \hat{\rho} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} + \frac{W_\perp W_\|}{4} \mathbf{\hat{\lambda}} \cdot \nabla \hat{b} \cdot \hat{\rho} + W_\perp^2 \nabla \times \hat{b} \cdot \mathbf{\hat{\lambda}} \right),$$

(A 50c)

$$\theta = \Theta + \frac{mc}{e B} \left( -\frac{W_\perp}{B} \mathbf{\hat{\lambda}} \cdot \nabla B + W_\perp \hat{\rho} \cdot \mathbf{\hat{\lambda}} \cdot \hat{\rho} + \frac{W_\perp}{4} \left( \hat{\rho} \cdot \nabla \hat{b} \cdot \hat{\rho} - \mathbf{\hat{\lambda}} \cdot \nabla \hat{b} \cdot \mathbf{\hat{\lambda}} \right) \right)$$

$$- \frac{W_\perp^2}{W_\perp} \nabla \times \hat{b} \cdot \hat{\rho},$$

(A 50d)

where the right-hand side is evaluated at $(X, W_\|, W_\perp, \Theta)$ and where

$$v = v_\| \hat{b}(x) + v_\perp \mathbf{\hat{\lambda}}(\theta, x).$$

(A 51)

The inversion of the change of variables given by (A 50), i.e. providing $(X, W_\|, W_\perp, \Theta)$ as functions of $(x, v_\|, v_\perp, \theta)$, gives the exact same equations given in Littlejohn (1983).

We note that the term $W_\| \mu \hat{b} \cdot (\nabla \times \hat{b})$ is usually moved to the symplectic part of the one-form by a translation in $W_\|$. In the new coordinates, it leads to the Hamiltonian

$$H_{gc} = \mu B(X) + \frac{1}{2} mW_\|^2,$$

(A 52)

and the one-form

$$\gamma_{gc} = \left[ \frac{e}{c} A(X) + mW_\| \hat{b}(X) - \frac{mc}{e} \mu \mathbf{R} \right] \cdot dX + \frac{mc}{e} \mu d\Theta - H_{gc} \, dt,$$

(A 53)

where $\mathbf{R}^* = \nabla \hat{b}_1 \cdot \hat{b}_2 + (\hat{b} \cdot \nabla \times \hat{b}) \hat{b}/2$. 

---

*Second-order nonlinear gyrokinetic theory* 

21
REFERENCES

ABEL, I. G., PLUNK, G. G., WANG, E., BARNES, M., COWLEY, S. C., DORLAND, W. & SCHekoCHIHN, A. A. 2013 Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. Rep. Progr. Phys. 76, 116201.

Bottino, A. & SonnenDrucker, E. 2015 Monte Carlo particle-in-cell methods for the simulation of the Vlasov-Maxwell gyrokinetic equations. J. Plasma Phys. 81, 435810501.

Brizard, A. J. 1989 Nonlinear gyrokinetic Maxwell-Vlasov equations using magnetic co-ordinates. J. Plasma Phys. 41, 541.

Brizard, A. J. 2000 New variational principle for the Vlasov–Maxwell equations. Phys. Rev. Lett. 84, 5768.

Brizard, A. J. & Hahm, T. S. 2007 Foundations of nonlinear gyrokinetic theory. Rev. Mod. Phys. 79, 421.

Burby, J. W., Brizard, A. J., Morrison, P. J. & Qin, H. 2015 Hamiltonian gyrokinetic Vlasov–Maxwell system. Phys. Lett. A 379, 2073.

Burby, J. W., Squire, J. & Qin, H 2013 Automation of the guiding center expansion. Phys. Plasmas 20, 072105.

Candy, J. & Waltz, R. E. 2003 An Eulearian gyrokinetic-Maxwell solver. J. Comput. Phys. 186, 545.

Cary, J. R. 1981 Lie transform perturbation theory for Hamiltonian systems. Phys. Rep. 186, 435811027.

Cary, J. R. & Brizard, A. J. 2009 Hamiltonian theory of guiding center motion. Rev. Mod. Phys. 81, 693.

Casati, A., Gerbaud, T., Hennequin, P., Bourdelle, C., Candy, J., Clairet, F., Garbet, X., Grandgirard, V., Gurcan, O. D., Heuraux, S. et al. 2009 Turbulence in the torus supra tokamak: measurements and validation of nonlinear simulations. Phys. Rev. Lett. 102 (16), 165005–165005.

Catto, P. J. 1978 Linearized gyro-kinetics. Plasma Phys. Control. Fusion 20 (7), 719–722.

Catto, P. J. 1981 Generalized gyro-kinetics. Plasma Phys. Control. Fusion 23 (7), 639–650.

Frieman, E. A. & Chen, L. 1982 Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. Phys. Fluids 25, 502.

Garbet, X., Idomura, Y., Villard, L. & Watanabe, T. H. 2010 Gyrokinetic simulations of turbulent transport. Nuclear Fusion 50, 043002.

Goerler, T., Lapillonne, X., Brunner, S., Dannert, T., Jenko, F., Merz, F. & Told, D. 2011 The global version of the gyrokinetic turbulence code GENE. J. Computat. Phys. 230, 7053–7071.

Grandgirard, V., Abiteboul, J., Bigot, J., Cartier-Michaud, J., Crouseille, N., Dif-Pradalier, G., Ehrlicher, C., Esteve, D., Garbet, X., Ghendrih, P. et al. 2016 A 5d gyrokinetic full-f global semi-Lagrangian code for flux-driven ion turbulence simulations. Comput. Phys. Commun. 207, 35–68.

Greboji, C., Kaufman, A. N. & Littlejohn, R. G. 1979 Hamiltonian theory of ponderomotive effects of an electromagnetic wave in a nonuniform magnetic field. Phys. Rev. Lett. 43, 22.

Hahm, T. S. 1988 Nonlinear gyrokinetic equations for tokamak microturbulence. Phys. Fluids 31, 2670.

Jolliet, S., Bottino, A., Angelino, P., Hatzky, R., Tran, T. M., McMillan, B. F., Sauter, O., Appert, K., Idomura, Y. & Villard, L. 2007 A global collisionless PIC code in magnetic coordinates. Comput. Phys. Commun. 177, 409.

Krause, T. B., Apte, A. & Morrison, P. J. 2007 A unified approach to the darwin approximation. Phys. Plasmas 14, 102112.

Littlejohn, R. G. 1979 A guiding center Hamiltonian: a new approach. J. Math. Phys. 20, 2445.

Littlejohn, R. G. 1981 Hamiltonian formulation of guiding center motion. Phys. Fluids 29, 1730.

Littlejohn, R. G. 1983 Variational principles of guiding centre motion. J. Plasma Phys. 29, 111.

Parra, F. I. & Calvo, I. 2011 Phase-space Lagrangian derivation of electrostatic gyrokinetics in general geometry. Plasma Phys. Control. Fusion 53, 045001.

Parra, F. I., Calvo, I., Burby, J. W. & Qin, H 2014 Equivalence of two independent calculations of the higher order guiding center Lagrangian. Phys. Plasmas 21, 104506.
PEETERS, A. G., CAMENEN, Y., CASSON, F. J., HORNSBY, W. A., SNODIN, A. P., STRINTZI, D. & SZEPESI, G. 2009 The nonlinear gyro-kinetic flux tube code gkw. *Comput. Phys. Commun.* **180** (12), 2650–2672.

SCHKOCIHIN, A. A., COWLEY, S. C., DORLAND, W., HAMMETT, G. W., HOWES, G. G., QUATAERT, E. & TATSUNO, T. 2009 Astrophysical gyrokinetics: kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas. *Astrophys. J. Suppl.* **182**, 310.

SQUIRE, J., QIN, H., TANG, W. M. & CHANDRE, C. 2013 The Hamiltonian structure and Euler–Poincaré formulation of the Vlasov–Maxwell and gyrokinetic systems. *Phys. Plasmas* **20**, 022501.

SUGAMA, H. 2000 Gyrokinetic field theory. *Phys. Plasmas* **7**, 466.

TRONKO, N., BOTTINO, A., CHANDRE, C. & SONNENDRÜCKER, E. 2017 Hierarchy of second order gyrokinetic Hamiltonian models for particle-in-cell codes. *Plasma Phys. Control. Fusion* **59**, 064008.

TRONKO, N., BOTTINO, A. & SONNENDRÜCKER, E. 2016 Second order gyrokinetic theory for particle-in-cell codes. *Phys. Plasmas* **23**, 082505.

TRONKO, N. & BRIZARD, A. J. 2015 Lagrangian and Hamiltonian constraints for guiding center Hamiltonian theories. *Phys. Plasmas* **22**, 112507.

WERSAL, C., BOTTINO, A., ANGELINO, P. & SCOTT, B. D. 2012 Fluid moments and spectral diagnostics in global particle-in-cell simulations. *J. Phys. Conf. Ser.* **401**, 012025.