Nonlinear gyrokinetic equation

BY YOUJUN HU

Institute of Plasma Physics, Chinese Academy of Sciences
Email: yjhu@ipp.cas.cn

Abstract

The nonlinear gyrokinetic equation in Frieman-Chen's paper[3] is re-derived step by step in this note. Numerical implementation of the gyrokinetic model using PIC method is also discussed.

1 Gyrokinetic?

Most physicists believe that low-frequency (lower than ion cyclotron frequency) electromagnetic perturbations are more important than high-frequency ones in transporting tokamak plasma, based on some non-conclusive observations and analytical theories. This assumption can be verified numerically when we are able to do a full simulation including both low-frequency and high-freqeuncy perturbations. This kind of verification is not possible at present due to the difficulties in doing a full simulation.

If only the low-frequency perturbations are present, the original Vlasov equation can be simplified. Specifically, some kind of symmetry of the perturbed particle distribution function in the phase space can be found if we choose suitable coordinates (independent variables) for the phase space and split the distribution function in a proper way. The symmetry is along the so-called gyro-angle α in the guiding-center coordinates $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$. In obtaining the equation for the gyro-angle independent part of the distribution function, we need to average the coefficients of the equation (particle motion) over the gyro-angle α of the guiding-center coordinates and thus this model is called "gyrokinetic".

In deriving the gyrokinetic equation, the perturbed electromagnetic field is assumed to be known and of low-frequency. To do a kinetic simulation, we need to solve the field equation to obtain the perturbed electromagnetic field. High frequency modes (e.g., compressional Alfven waves and Ω_H modes) can still appear in a gyrokinetic simulation. If the amplitude of high frequency modes is large, then the simulation is not meaningful since the gyrokinetic model is invalid in this case.

1.1 Vlasov equation

The Vlasov equation in terms of particle coordinates $(\mathbf{x}, \mathbf{v})$ is written

$$\frac{\partial f_p}{\partial t} + \mathbf{v} \cdot \frac{\partial f_p}{\partial \mathbf{x}} + \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \frac{\partial f_p}{\partial \mathbf{v}} = 0, \tag{1}$$

where $f_p = f_p(\mathbf{x}, \mathbf{v})$ is the particle distribution function, $\mathbf{x}$ and $\mathbf{v}$ are the location and velocity of particles. The left-hand side of the Vlasov equation describe the total time derivative of f_p observed when we follow particle orbit in the 6D phase-space $(\mathbf{x}, \mathbf{v})$. Using a general 6D phase-space coordinates $(z^{(1)}, z^{(2)}, z^{(3)}, z^{(4)}, z^{(5)}, z^{(6)})$, the Vlasov equation in terms of particle orbit is written as

$$\frac{df_p}{dt} \equiv \frac{\partial f_p}{\partial t} + \sum_{i=1}^{6} \frac{dz^{(i)}}{dt} \frac{\partial f_p}{\partial z^{(i)}} = 0, \tag{2}$$

where $dz^{(i)}/dt$ is given by the equation of particle motion.

1.2 Gyro-angle symmetry of distribution function

Let us consider a simple case: a time-independent and spatial uniform magnetic field ($\mathbf{B} = B_0\hat{\mathbf{z}}$). Describe the velocity space using a right-handed cylindrical coordinates ($v_\perp$, α, $v_\parallel$), where $v_\parallel = \mathbf{v} \cdot \mathbf{e}_\parallel$, $\mathbf{e}_\parallel$ is the unit vector along the magnetic field, α is the azimuthal angle of the velocity. Choosing Cartesian coordinate system (x, y, z) for the configuration space, then the 6D phase-space are described by coordinates (x, y, z, $v_\perp$, α, $v_\parallel$). The distribution function f_p depends on 7 variables, namely $x, y, z, v_\parallel, \alpha, v_\perp$, and t.

Next, let use examine the Vlasov equation in this simple case and look if there is any coordinate transform that can reduce the number of dimensions of the distribution function.

In $(\mathbf{x}, v_\perp, \alpha, v_\parallel)$ coordinates, the Vlasov equation is written

$$\frac{\partial f_p}{\partial t} + \mathbf{v} \cdot \frac{\partial f_p}{\partial \mathbf{x}} + \frac{dv_\perp}{dt}\frac{\partial f_p}{\partial v_\perp} + \frac{dv_\parallel}{dt}\frac{\partial f_p}{\partial v_\parallel} + \frac{d\alpha}{dt}\frac{\partial f_p}{\partial \alpha} = 0. \tag{3}$$

Since $dv_\perp/dt = 0$, $dv_\parallel/dt = 0$, and $d\alpha/dt = -\Omega$, where $\Omega = qB/m$, the above equation is written as

$$\frac{\partial f_p}{\partial t} + \mathbf{v} \cdot \frac{\partial f_p}{\partial \mathbf{x}} - \Omega \frac{\partial f_p}{\partial \alpha} = 0, \tag{4}$$

i.e.,

$$\frac{1}{f_p}\frac{\partial f_p}{\partial t}\frac{2\pi}{\Omega} + \frac{2\pi}{f_p\Omega}\mathbf{v}_\parallel \cdot \frac{\partial f_p}{\partial \mathbf{x}} + \frac{2\pi}{f_p\Omega}\mathbf{v}_\perp \cdot \frac{\partial f_p}{\partial \mathbf{x}} - \frac{2\pi}{f_p}\frac{\partial f_p}{\partial \alpha} = 0. \tag{5}$$

We assume that f_p evolves on a time scale much longer the the gyro-period, i.e.,

$$\frac{1}{f_p}\frac{\partial f_p}{\partial t}\frac{2\pi}{\Omega} \sim O(\varepsilon^1), \tag{6}$$

where ε is a small parameter. Further assume that the spatial variation of f_p along the magnetic field is weak:

$$\frac{1}{f_p}\frac{2\pi}{\Omega}\mathbf{v}_\parallel \cdot \frac{\partial f_p}{\partial \mathbf{x}} \sim O(\varepsilon^1), \tag{7}$$

where $2\pi\mathbf{v}_\parallel/\Omega$ is the parallel distance traveled by a particle within one gyro-period. Dropping terms of $O(\varepsilon^1)$, Eq. (5) is written as

$$\frac{1}{f_p}\frac{2\pi}{\Omega}\mathbf{v}_\perp\cdot\frac{\partial f_p}{\partial \mathbf{x}}-\frac{2\pi}{f_p}\frac{\partial f_p}{\partial \alpha}=0, \tag{8}$$

i.e.,

$$\mathbf{v}_\perp\cdot\frac{\partial f_p}{\partial \mathbf{x}}-\Omega\frac{\partial f_p}{\partial \alpha}=0, \tag{9}$$

In terms of $(x, y, z, v_\perp, \alpha, v_\parallel)$ coordinates, the above equation is written as

$$v_\perp\cos\alpha\frac{\partial f_p}{\partial x}+v_\perp\sin\alpha\frac{\partial f_p}{\partial y}-\Omega\frac{\partial f_p}{\partial \alpha}=0. \tag{10}$$

Define the following coordinates transform (guiding-center transform):

$$x'=x+\frac{v_\perp\sin\alpha}{\Omega}, \tag{11}$$

$$y'=y-\frac{v_\perp\cos\alpha}{\Omega}, \tag{12}$$

$$z'=z, \tag{13}$$

$$\alpha'=\alpha, \tag{14}$$

$$v'_\parallel=v_\parallel, \tag{15}$$

$$v'_\perp=v_\perp. \tag{16}$$

Then, in terms of the new coordinates $(x', y', z', v'_\parallel, v'_\perp, \alpha')$, $\partial f_p/\partial\alpha'$ is written as

$$\begin{aligned}\left.\frac{\partial f_p}{\partial \alpha'}\right|_{x',y',z',v'_\parallel,v'_\perp} &= \frac{\partial f_p}{\partial x}\frac{\partial x}{\partial \alpha'}+\frac{\partial f_p}{\partial y}\frac{\partial y}{\partial \alpha'}+\frac{\partial f_p}{\partial \alpha}\frac{\partial \alpha}{\partial \alpha'}\\ &= -\frac{\partial f_p}{\partial x}\frac{1}{\Omega}v_\perp\cos\alpha-\frac{\partial f_p}{\partial y}\frac{1}{\Omega}v_\perp\sin\alpha+\frac{\partial f_p}{\partial \alpha},\end{aligned} \tag{17}$$

Eq. (10) implies that the above expression is zero. Therefore we obtain

$$\left.\frac{\partial f_p}{\partial \alpha'}\right|_{x',y',z',v'_\parallel,v'_\perp}=0, \tag{18}$$

which indicates that, in the new coordinates (x', y', z', $v'_\parallel$, $v'_\perp$, α'), α' is an ignorable coordinate and thus the phase space dimension of f_p is reduced by one. This is obviously beneficial for numerically solving the Vlasov equation.

Note that the gyro-angle symmetry is valid only in the guiding-center coordinates. In particle coordinates, there is no gyro-angle symmetry.

For the case of static uniform magnetic field considered above, everything is simple. For non-uniform static magnetic field plus an electromagnetic perturbation, things are a little complicate and a great deal of efforts of gyrokinetic theory are devoted to taking into account these effects.

2 Transform Vlasov equation from particle coordinates to guiding-center coordinates

Next, we define the guiding-center transform and then we transform the Vlasov equation from the particle coordinates $(\mathbf{x}, \mathbf{v})$ to the guiding-center coordinates, i.e., express the gradient operators $\partial / \partial \mathbf{x}$ and $\partial / \partial \mathbf{v}$ in terms of the guiding-center coordinates.

2.1 Guiding-center transformation

In a magnetic field, given a particle location and velocity $(\mathbf{x}, \mathbf{v})$, we know how to calculate its guiding-center location $\mathbf{X}$, i.e.,

$$\mathbf{X}(\mathbf{x}, \mathbf{v}) = \mathbf{x} + \mathbf{v} \times \frac{\boldsymbol{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})}, \tag{19}$$

where $\boldsymbol{e}_\parallel = \mathbf{B}_0 / B_0$, $\Omega = qB_0/m$, $\mathbf{B}_0 = \mathbf{B}_0(\mathbf{x})$ is the equilibrium (macroscopic) magnetic field at the particle position. We will consider Eq. (19) as a transform and call it guiding-center transform[1]. Note that the transform (19) involves both position and velocity of particles.

For later use, define $\boldsymbol{\rho} \equiv -\mathbf{v} \times \mathbf{e}_\parallel / \Omega$, which is the vector gyro-radius pointing from the the guiding-center to the particle position.

Given $(\mathbf{x}, \mathbf{v})$, it is straightforward to obtain $\mathbf{X}$ by using Eq. (19). On the other hand, the inverse transform, i.e., given $(\mathbf{X}, \mathbf{v})$, to find $\mathbf{x}$, which is in principle not easy because Ω and $\mathbf{e}_\parallel$ depend on $\mathbf{x}$, which usually requires solving for the root of a nonlinear equation. Numerically, one can use

$$\mathbf{x}_{n+1} = \mathbf{X} - \mathbf{v} \times \frac{\boldsymbol{e}_\parallel(\mathbf{x}_n)}{\Omega(\mathbf{x}_n)}, \tag{20}$$

as an iteration scheme to compute $\mathbf{x}$, with the initial guess chosen as $\mathbf{x}_0 = \mathbf{X}$. The equilibrium magnetic field we will consider has spatial scale length much larger than the thermal gyro radius $\boldsymbol{\rho}$. In this case the difference between the values of $\mathbf{e}_\parallel(\mathbf{x}) / \Omega(\mathbf{x})$ and $\mathbf{e}_\parallel(\mathbf{X}) / \Omega(\mathbf{X})$ is small and thus can be neglected. Then the inverse guiding-center transform can be written as

$$\mathbf{x} \approx \mathbf{X} - \mathbf{v} \times \frac{\boldsymbol{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}, \tag{21}$$

which can also be considered as using the iterative scheme (20) to computer $\mathbf{x}$ with initial guess of $\mathbf{x}$ being $\mathbf{X}$ and stopping at the first iteration. [The difference between equilibrium field values evaluated at $\mathbf{X}$ and $\mathbf{x}$ is always neglected in gyrokinetic theory. Therefore it does not matter whether the above $\mathbf{e}_\parallel / \Omega$ is evaluated at $\mathbf{x}$ or $\mathbf{X}$. What matters is where the perturbed fields are evaluated: at $\mathbf{x}$ or at $\mathbf{X}$. The values of perturbed fields at $\mathbf{x}$ or at $\mathbf{X}$ are different and this is called the finite Larmor radius (FLR) effect, which is almost all that the gyrokinetic theory is about.]

The inverse guiding-center transformation (21) needs to be performed numerically in gyrokinetic PIC simulations when we deposit markers to grids or when we calculate the gyro-averaged field to be used in pushing guiding-centers.

2.2 Choosing velocity coordinates

The guiding-center transformations (19) and (21) involve the particle velocity $\mathbf{v}$. It is the cross product between $\mathbf{v}$ and $\boldsymbol{e}_\parallel(\mathbf{x})$ or $\boldsymbol{e}_\parallel(\mathbf{X})$ that is actually used. Therefore, only the perpendicular velocity (which is defined by $\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}\cdot\mathbf{e}_\parallel$) enters the transform. A natural choice of coordinates for the perpendicular velocity is $(v_\perp, \alpha)$, where $v_\perp = |\mathbf{v}_\perp|$ and α is the azimuthal angle of the perpendicular velocity in the local perpendicular plane.

The parallel direction is fully determined by $\mathbf{B}_0(\mathbf{x})$, but there are degrees of freedom in choosing one of the two perpendicular basis vectors. In order to make the azimuthal angle α fully defined, we need to choose a way to define one of the two perpendicular directions. In `GEM` simulations, one of the perpendicular direction is chosen as the direction perpendicular to the magnetic surface, which is fully determined at each spatial point. (We need to define the perpendicular direction at each spatial location to make $\partial\alpha / \partial\mathbf{x}|_{\mathbf{v}}$ defined, which is needed in the Vlasov differential operators. However, it seems that terms related to $\partial\alpha / \partial\mathbf{x}|_{\mathbf{v}}$ are finally dropped due to that they are of higher order**check.)

In the following, α will be called the "gyro-angle" . Note that the common use of the nomenclature "gyro-angle" is to describe the spatial motion of a particle, but here α is a velocity coordinate rather than a spatial coordinate. How to justify the name "gyro-angle" for α? Consider a series of points in terms of guiding-center coordinates $(\mathbf{X},\ v_\parallel,\ v_\perp,\ \alpha)$ with $(\mathbf{X},\ v_\parallel,\ v_\perp)$ fixed but with α changing. Using the inverse guiding-center transform (21), it is obvious that the above points form a gyro-ring in real space, which justifies calling α by the name "gyro-angle".

The gyro-angle is an important variable we will stick to because we need to directly perform averaging over this variable (with $\mathbf{X}$ fixed) in deriving the gyrokinetic equation. We have more than one choice for the remaining velocity coordinates , such as $(v, v_\parallel)$, or $(v, v_\perp)$, or $(v_\parallel, v_\perp)$. In Frieman-Chen's paper, the velocity coordinates other than α are chosen to be (ε, μ) defined by

$$\varepsilon = \varepsilon(v, \mathbf{x}) \equiv \frac{v^2}{2} + \frac{q\Phi_0(\mathbf{x})}{m}, \tag{22}$$

and

$$\mu = \mu(v_\perp, \mathbf{x}) \equiv \frac{v_\perp^2}{2B_0(\mathbf{x})}, \tag{23}$$

where $\Phi_0(\mathbf{x})$ is the equilibrium (macroscopic) electrical potential.

Note that $(\varepsilon, \mu, \alpha)$ is not sufficient in uniquely determining a velocity vector. An additional parameter $\sigma = \text{sign}(v_\parallel)$ is needed to determine the sign of $v_\parallel = \mathbf{v} \cdot \mathbf{e}_\parallel$. In the following, the dependence of the distribution function on σ is often not explicitly shown in the variable list (i.e., σ is hidden/suppressed), which, however, does not mean that the distribution function is independent of σ.

Another frequently used velocity coordinates are $(\mu, v_\parallel, \alpha)$. In the following, I will derive the gyrokinetic equation in $(\varepsilon,\ \mu,\ \alpha)$ coordinates. After that, I transform it to $(\mu, v_\parallel, \alpha)$ coordinates.

One important thing to note about the above velocity coordinates is that they are defined relative to the local magnetic field. If the field itself is spatially varying (such as in tokamaks), the above velocity coordinates are also spatially varying for a fixed velocity $\mathbf{v}$. Specifically, the following derivatives are nonzero:

$$\frac{\partial \alpha}{\partial \mathbf{x}}|_{\mathbf{v}}, \frac{\partial v_\perp}{\partial \mathbf{x}}|_{\mathbf{v}}, \frac{\partial v_\parallel}{\partial \mathbf{x}}|_{\mathbf{v}}, \frac{\partial \mu}{\partial \mathbf{x}}|_{\mathbf{v}}. \tag{24}$$

2.3 Summary of the phase-space coordinate transform

The coordinate transform from particle coordinates $(\mathbf{x},\ \mathbf{v})$ to guiding-center coordinates $(\mathbf{X}, \varepsilon, \mu,, \sigma)$ is given by

$$\begin{cases} \mathbf{X}(\mathbf{x}, \mathbf{v}) = \mathbf{x} + \mathbf{v} \times \frac{\boldsymbol{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} \\ \varepsilon(\mathbf{x}, \mathbf{v}) = \frac{|\mathbf{v}|^2}{2} + \frac{q\Phi_0(\mathbf{x})}{m} \\ \mu(\mathbf{x}, \mathbf{v}) = \frac{|\mathbf{v} - \mathbf{v} \cdot \mathbf{e}_\parallel(\mathbf{x})|^2}{2B_0(\mathbf{x})} \\ \alpha(\mathbf{x}, \mathbf{v}) = (\text{angle between } \mathbf{v}_\perp \text{ and local } \mathbf{e}_\perp) \\ \sigma(\mathbf{x}, \mathbf{v}) = \text{sign}(\mathbf{v} \cdot \mathbf{e}_\parallel(\mathbf{x})) \end{cases} \tag{25}$$

As mentioned above, the σ dependence of the distribution function is hidden in the following.

2.4 Distribution functions in terms of guiding-center variables

Denote the distribution function in terms of the guiding-center variables $(\mathbf{X}, \varepsilon,\ \mu, \alpha)$ by f_g. Then the relation between f_g and the particle distribution function f_p is given by

$$f_g(\mathbf{X}, \varepsilon, \mu, \alpha) = f_p(\mathbf{x}, \mathbf{v}), \tag{26}$$

where the coordinates $(\mathbf{X}, \varepsilon, \mu, \alpha)$ are related to $(\mathbf{x}, \mathbf{v})$ by Eq. (25) . Equation (26) along with Eq. (25) can be considered as the definition of f_g.

As is conventionally adopted in multi-variables calculus, both f_p and f_g are sometimes simply denoted by f. Which one is actually assumed depends on the context, i.e., depends on which independent variables are actually assumed: particle variables or guiding-center variables. This is one of the subtle (trivial) things needed to be noted for gyrokinetic theory in particular and for multi-variables calculus in general. It seems better to use subscript notation on f to identify which coordinates are assumed. One example where this distinguishing is important is encountered when we try to express the diamagnetic flow in terms of f_g, which is discussed in Appendix D.

In practice, f_g is often called the guiding-center distribution function whereas f_p is called the particle distribution function. However, they are actually the same distribution function expressed in different variables. The name "guiding-center distribution function" is misleading because it may imply that we can count the number of guiding-centers to obtain this distribution function but this implication is wrong.

2.5 Spatial gradient operator in guiding-center coordinates

Using the chain-rule, the spatial gradient $\partial f_p/\partial \mathbf{x}$ is written

$$\frac{\partial f_p}{\partial \mathbf{x}}|_{\mathbf{v}} = \frac{\partial \mathbf{X}}{\partial \mathbf{x}}\cdot\frac{\partial f_g}{\partial \mathbf{X}} + \frac{\partial \varepsilon}{\partial \mathbf{x}}\frac{\partial f_g}{\partial \varepsilon} + \frac{\partial \mu}{\partial \mathbf{x}}\frac{\partial f_g}{\partial \mu} + \frac{\partial \alpha}{\partial \mathbf{x}}\frac{\partial f_g}{\partial \alpha}. \tag{27}$$

From the definition of $\mathbf{X}$, Eq. (19), we obtain

$$\frac{\partial \mathbf{X}}{\partial \mathbf{x}} = \mathbf{I} + \mathbf{v}\times\frac{\partial}{\partial \mathbf{x}}\Big(\frac{\boldsymbol{e}_\parallel}{\Omega}\Big), \tag{28}$$

where $\mathbf{I}$ is the unit dyad. From the definition of ε, we obtain

$$\frac{\partial \varepsilon}{\partial \mathbf{x}} = -\frac{q}{m}\mathbf{E}_0, \tag{29}$$

where $\mathbf{E}_0 = -\partial\Phi_0/\partial\mathbf{x}$. Using the above results, equation (27) is written as

$$\frac{\partial f_p}{\partial \mathbf{x}}|_{\mathbf{v}} = \frac{\partial f_g}{\partial \mathbf{X}} + \left[\mathbf{v}\times\frac{\partial}{\partial \mathbf{x}}\Big(\frac{\boldsymbol{e}_\parallel}{\Omega}\Big)\right]\cdot\frac{\partial f_g}{\partial \mathbf{X}} - \frac{q}{m}\mathbf{E}_0\frac{\partial f_g}{\partial \varepsilon} + \frac{\partial \mu}{\partial \mathbf{x}}\frac{\partial f_g}{\partial \mu} + \frac{\partial \alpha}{\partial \mathbf{x}}\frac{\partial f_g}{\partial \alpha}. \tag{30}$$

As mentioned above, the partial derivative $\partial/\partial\mathbf{x}$ is taken by holding $\mathbf{v}$ constant. Since $\mathbf{B}_0$ is spatially varying, $\mathbf{v}_\perp$ is spatially varying when holding $\mathbf{v}$ constant. Therefore $\frac{\partial \mu}{\partial \mathbf{x}}$ and $\frac{\partial \alpha}{\partial \mathbf{x}}$ are generally nonzero. The explicit expressions of these two derivatives are needed later in the derivation of the gyrokinetic equation and is discussed in Appendix B.

For notation ease, define

$$\boldsymbol{\lambda}_{B1} = \left[\mathbf{v}\times\frac{\partial}{\partial \mathbf{x}}\Big(\frac{\boldsymbol{e}_\parallel}{\Omega}\Big)\right]\cdot\frac{\partial}{\partial \mathbf{X}}, \tag{31}$$

and

$$\boldsymbol{\lambda}_{B2}=\frac{\partial\mu}{\partial\mathbf{x}}\frac{\partial}{\partial\mu}+\frac{\partial\alpha}{\partial\mathbf{x}}\frac{\partial}{\partial\alpha}, \tag{32}$$

then expression (30) is written as

$$\frac{\partial f_p}{\partial\mathbf{x}}|_{\mathbf{v}}=\frac{\partial f_g}{\partial\mathbf{X}}+[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}]f_g-\frac{q}{m}\mathbf{E}_0\frac{\partial f_g}{\partial\varepsilon}. \tag{33}$$

2.6 Velocity gradient operator in guiding-center coordinates

Next, consider the form of the velocity gradient $\partial f/\partial\mathbf{v}$ in terms of the guiding-center variables. Using the chain rule, $\partial f/\partial\mathbf{v}$ is written

$$\frac{\partial f_p}{\partial\mathbf{v}}|_{\mathbf{x}}=\frac{\partial\mathbf{X}}{\partial\mathbf{v}}\cdot\frac{\partial f_g}{\partial\mathbf{X}}+\frac{\partial\varepsilon}{\partial\mathbf{v}}\frac{\partial f_g}{\partial\varepsilon}+\frac{\partial\mu}{\partial\mathbf{v}}\frac{\partial f_g}{\partial\mu}+\frac{\partial\alpha}{\partial\mathbf{v}}\frac{\partial f_g}{\partial\alpha}. \tag{34}$$

From the definition of $\mathbf{X}$, we obtain

$$\begin{aligned}\frac{\partial\mathbf{X}}{\partial\mathbf{v}} &= \frac{\partial}{\partial\mathbf{v}}\left(\frac{\boldsymbol{v}\times\boldsymbol{e}_{\parallel}}{\Omega}\right)\\ &= \frac{\partial\mathbf{v}}{\partial\mathbf{v}}\times\frac{\boldsymbol{e}_{\parallel}}{\Omega}\\ &= \mathbf{I}\times\frac{\boldsymbol{e}_{\parallel}}{\Omega}.\end{aligned} \tag{35}$$

From the definition of ε, we obtain

$$\frac{\partial\varepsilon}{\partial\mathbf{v}}=\mathbf{v}, \tag{36}$$

From the definition of μ, we obtain

$$\frac{\partial\mu}{\partial\mathbf{v}}=\frac{\mathbf{v}_{\perp}}{B_0}, \tag{37}$$

From the definition of α, we obtain

$$\frac{\partial\alpha}{\partial\mathbf{v}}=\frac{1}{v_{\perp}}\left(\boldsymbol{e}_{\parallel}\times\frac{\mathbf{v}_{\perp}}{v_{\perp}}\right)=\frac{\boldsymbol{e}_{\alpha}}{v_{\perp}}, \tag{38}$$

where $\mathbf{e}_{\alpha}$ is defined by

$$\boldsymbol{e}_{\alpha}=\boldsymbol{e}_{\parallel}\times\frac{\mathbf{v}_{\perp}}{v_{\perp}}. \tag{39}$$

Using the above results, expression (34) is written

$$\frac{\partial f_p}{\partial\mathbf{v}}|_{\mathbf{x}}=\frac{\mathbf{I}\times\boldsymbol{e}_{\parallel}}{\Omega}\cdot\frac{\partial f_g}{\partial\mathbf{X}}+\mathbf{v}\frac{\partial f_g}{\partial\varepsilon}+\frac{\mathbf{v}_{\perp}}{B_0}\frac{\partial f_g}{\partial\mu}+\frac{\boldsymbol{e}_{\alpha}}{v_{\perp}}\frac{\partial f_g}{\partial\alpha}. \tag{40}$$

2.7 Time derivatives in guiding-center coordinates

In the guiding-center variables, the time partial derivative $\partial f / \partial t$ appearing in Vlasov equation is written as

$$\frac{\partial f_p}{\partial t}|_{\mathbf{x},\mathbf{v}} = \frac{\partial f_g}{\partial t}|_{\mathbf{X},\mathbf{V}} + \frac{\partial \mathbf{X}}{\partial t} \cdot \frac{\partial f_g}{\partial \mathbf{X}} + \frac{\partial \mathbf{V}}{\partial t} \cdot \frac{\partial f_g}{\partial \mathbf{V}}, \tag{41}$$

where $\mathbf{V} = (\varepsilon, \mu, \alpha)$. Here $\partial \mathbf{X} / \partial t$ and $\partial \mathbf{V} / \partial t$ are not necessarily zero because the equilibrium quantities involved in the definition of the guiding-center transformation are in general time dependent. This time dependence is assumed to be very slow in the gyrokinetic ordering discussed later. In the following, $\partial \mathbf{X} / \partial t$ and $\partial \mathbf{V} / \partial t$ will be dropped.

2.8 Final form of Vlasov equation in guiding-center coordinates

Using the above results, the Vlasov equation in guiding-center coordinates is written

$$\begin{aligned} & \frac{\partial f_g}{\partial t} + \mathbf{v} \cdot \left[\frac{\partial f_g}{\partial \mathbf{X}} + [\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2}] f_g - \frac{q}{m} \mathbf{E}_0 \frac{\partial f_g}{\partial \varepsilon} \right] \\ + \; & \frac{q}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \left(\frac{\mathbf{I} \times \boldsymbol{e}_\parallel}{\Omega} \cdot \frac{\partial f_g}{\partial \mathbf{X}} + \boldsymbol{v} \frac{\partial f_g}{\partial \varepsilon} + \frac{\boldsymbol{v}_\perp}{B_0} \frac{\partial f_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \\ = \; & 0, \end{aligned} \tag{42}$$

Using tensor identity $\mathbf{a} \cdot \boldsymbol{I} \times \mathbf{b} = \mathbf{a} \times \mathbf{b}$, equation (42) is written as

$$\begin{aligned} & \frac{\partial f_g}{\partial t} + \mathbf{v} \cdot \left[\frac{\partial f_g}{\partial \mathbf{X}} + [\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2}] f_g - \frac{q}{m} \mathbf{E}_0 \frac{\partial f_g}{\partial \varepsilon} \right] \\ + \; & \frac{q}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \cdot \frac{\partial f_g}{\partial \mathbf{X}} + \frac{q}{m} (\mathbf{v} \times \mathbf{B}) \cdot \left(\frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \\ + \; & \frac{q}{m} \mathbf{E} \cdot \left(\mathbf{v} \frac{\partial f_g}{\partial \varepsilon} + \frac{\mathbf{v}_\perp}{B_0} \frac{\partial f_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \\ = \; & 0, \end{aligned} \tag{43}$$

This is the Vlasov equation in guiding-center coordinates.

3 Perturbed Vlasov equation in guiding-center variables

Since the definition of the guiding-center variables ($\mathbf{X}$, ε, μ, α) involves the macroscopic (equilibrium) fields $\mathbf{B}_0$ and $\mathbf{E}_0$, to further simplify Eq. (43), we need to separate electromagnetic field into equilibrium and perturbation parts. Writing the electromagnetic field as

$$\mathbf{E} = \mathbf{E}_0 + \delta \mathbf{E} \tag{44}$$

and

$$\mathbf{B} = \mathbf{B}_0 + \delta \mathbf{B}, \tag{45}$$

then substituting these expressions into equation (43) and moving all terms involving the perturbed fields to the right-hand side, we obtain

$$\begin{aligned}
&\frac{\partial f_g}{\partial t}+\mathbf{v}\cdot\frac{\partial f_g}{\partial \mathbf{X}}+\mathbf{v}\cdot\left[[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}]f_g-\frac{q}{m}\mathbf{E}_0\frac{\partial f_g}{\partial \varepsilon}\right]\\
+\;&\frac{q}{m}(\mathbf{E}_0+\mathbf{v}\times\mathbf{B}_0)\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial f_g}{\partial \mathbf{X}}+\frac{q}{m}(\mathbf{v}\times\mathbf{B}_0)\cdot\left(\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial f_g}{\partial \alpha}\right)\\
+\;&\frac{q}{m}\mathbf{E}_0\cdot\left(\mathbf{v}\frac{\partial f_g}{\partial \varepsilon}+\frac{\mathbf{v}_\perp}{B_0}\frac{\partial f_g}{\partial \mu}+\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial f_g}{\partial \alpha}\right)\\
=\;&\delta R f_g,
\end{aligned}\tag{46}$$

where δR is defined by

$$\begin{aligned}
\delta R \;=\;&-\frac{q}{m}(\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B})\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial}{\partial \mathbf{X}}-\frac{q}{m}(\mathbf{v}\times\delta\mathbf{B})\cdot\left(\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial}{\partial \alpha}\right)\\
&-\frac{q}{m}\delta\mathbf{E}\cdot\left(\mathbf{v}\frac{\partial}{\partial \varepsilon}+\frac{\mathbf{v}_\perp}{B_0}\frac{\partial}{\partial \mu}+\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial}{\partial \alpha}\right).
\end{aligned}\tag{47}$$

Next, let us simplify the left-hand side of Eq. (46). Note that

$$\frac{q}{m}\mathbf{E}_0\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial f_g}{\partial \mathbf{X}}=c\left(\frac{\mathbf{E}_0\times\boldsymbol{e}_\parallel}{B_0}\right)\cdot\frac{\partial f_g}{\partial \mathbf{X}}=\mathbf{v}_{\mathbf{E}0}\cdot\frac{\partial f_g}{\partial \mathbf{X}},\tag{48}$$

where $\mathbf{v}_{\mathbf{E}0}$ is defined by $\mathbf{v}_{\mathbf{E}0}=c\mathbf{E}_0\times\mathbf{e}_\parallel/B_0$, which is the macroscopic (equilibrium) flow due to $\mathbf{E}_0\times\mathbf{B}_0$ drift. Further note that

$$\begin{aligned}
\frac{q}{m}\frac{\mathbf{v}\times\mathbf{B}_0}{c}\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial f_g}{\partial \mathbf{X}} \;&=\; [(\mathbf{v}\times\mathbf{e}_\parallel)\times\boldsymbol{e}_\parallel]\cdot\frac{\partial f_g}{\partial \mathbf{X}}\\
&=\; [v_\parallel\mathbf{e}_\parallel-\mathbf{v}]\cdot\frac{\partial f_g}{\partial \mathbf{X}},
\end{aligned}\tag{49}$$

which can be combined with $\mathbf{v}\cdot\partial f_g/\partial\mathbf{X}$ term, yielding $v_\parallel\mathbf{e}_\parallel\cdot\partial f_g/\partial\mathbf{X}$. Finally note that

$$\begin{aligned}
&\frac{q}{m}(\mathbf{v}\times\mathbf{B}_0)\cdot\left(\frac{\mathbf{e}_\alpha}{v_\perp}\frac{\partial f_g}{\partial \alpha}\right)\\
=&\frac{q}{m}(\mathbf{v}\times\mathbf{B}_0)\cdot\frac{\boldsymbol{e}_\parallel\times\mathbf{v}_\perp}{v_\perp^2}\frac{\partial f_g}{\partial \alpha}\\
=&-\Omega\frac{\partial f_g}{\partial \alpha}.
\end{aligned}\tag{50}$$

Using Eqs. (48), (49), and (50), the left-hand side of equation (46) is written as

$$\begin{aligned}
&\frac{\partial f_g}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_{\mathbf{E}0})\cdot\frac{\partial f_g}{\partial \mathbf{X}}+\mathbf{v}\cdot[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}]f_g-\Omega\frac{\partial f_g}{\partial \alpha}\\
+\;&\frac{q}{m}\mathbf{E}_0\cdot\left(\frac{\mathbf{v}_\perp}{B_0}\frac{\partial f_g}{\partial \mu}+\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial f_g}{\partial \alpha}\right)\equiv L_g f_g
\end{aligned}\tag{51}$$

which corresponds to Eq. (7) in Frieman-Chen's paper[3]. (In Frieman-Chen's equation (7), there is a term

$$\frac{q}{m}(\mathbf{E}-\mathbf{E}_0)\cdot\mathbf{v}\frac{\partial}{\partial \varepsilon}$$

where $\mathbf{E}$ is the macroscopic electric field and is in general different from the $\mathbf{E}_0$ introduced when defining the guiding-center transformation. In my derivation $\mathbf{E}_0$ is chosen to be equal to the macroscopic electric field thus the above term does not appear.) In expression (51), L_g is often called the unperturbed Vlasov propagator in guiding-center coordinates $(\mathbf{X}, \varepsilon, \mu, \alpha)$.

Using the above results, Eq. (46) is written as

$$L_g f_g = \delta R f_g, \tag{52}$$

i.e.

$$\begin{aligned}
&\frac{\partial f_g}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_{\mathbf{E}0}) \cdot \frac{\partial f_g}{\partial \mathbf{X}} \\
&+\mathbf{v} \cdot \left[\mathbf{v} \times \frac{\partial}{\partial \mathbf{x}} \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \cdot \frac{\partial f_g}{\partial \mathbf{X}} + \frac{\partial \mu}{\partial \mathbf{x}} \frac{\partial f_g}{\partial \mu} + \frac{\partial \alpha}{\partial \mathbf{x}} \frac{\partial f_g}{\partial \alpha} \right] - \Omega \frac{\partial f_g}{\partial \alpha} \\
&+\frac{q}{m} \mathbf{E}_0 \cdot \left(\frac{\mathbf{v}_\perp}{B_0} \frac{\partial f_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \\
&= -\frac{q}{m} (\delta \mathbf{E} + \mathbf{v} \times \delta \mathbf{B}) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \cdot \frac{\partial f_g}{\partial \mathbf{X}} - \frac{q}{m} (\mathbf{v} \times \delta \mathbf{B}) \cdot \left(\frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \\
&-\frac{q}{m} \delta \mathbf{E} \cdot \left(\mathbf{v} \frac{\partial f_g}{\partial \varepsilon} + \frac{\mathbf{v}_\perp}{B_0} \frac{\partial f_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right).
\end{aligned} \tag{53}$$

It is instructive to consider some special cases of the above complicated equation. Consider the case that the equilibrium magnetic field $\mathbf{B}_0$ is uniform and time-independent, $\mathbf{E}_0 = 0$, and the electrostatic limit $\delta \mathbf{B} = 0$, then equation (53) is simplified as

$$\begin{aligned}
&\frac{\partial f_g}{\partial t} + v_\parallel \mathbf{e}_\parallel \cdot \frac{\partial f_g}{\partial \mathbf{X}} - \Omega \frac{\partial f_g}{\partial \alpha} \\
&= -\frac{q}{m} (\delta \mathbf{E}) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \cdot \frac{\partial f_g}{\partial \mathbf{X}} \qquad (54) \\
&-\frac{q}{m} \delta \mathbf{E} \cdot \left(\mathbf{v} \frac{\partial f_g}{\partial \varepsilon} + \frac{\mathbf{v}_\perp}{B_0} \frac{\partial f_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial f_g}{\partial \alpha} \right) \qquad (55)
\end{aligned}$$

If neglecting all the perturbations and the time and parallel spatial variation of f_g, the above equation reduces to

$$\Omega \frac{\partial f_g}{\partial \alpha} = 0. \tag{56}$$

i.e., the gyro-angle symmetry discussed in Sec. 1.2.

3.1 Distribution function perturbation

Expand the distribution function f_g as

$$f_g = F_g + \delta F_g, \tag{57}$$

where F_g is assumed to be an equilibrium distribution function, i.e.,

$$L_g F_g = 0. \tag{58}$$

Using Eqs. (57) and (58) in Eq. (52), we obtain an equation for δF_g:

$$L_g \delta F_g = \delta R F_g + \delta R \delta F_g. \tag{59}$$

3.2 Gyrokinetic orderings

To facilitate the simplification of the Vlasov equation in the low-frequency regime, we assume the following orderings (some of which are roughly based on experiment measure of fluctuations responsible for tokamak plasma transport, some of which can be invalid in some interesting cases.) These ordering are often called the standard gyrokinetic orderings.

3.2.1 Ordering of macroscopic quantities

Define the spatial scale length L_0 of equilibrium quantities by $L_0 \approx F_g / |\nabla_X F_g|$. Assume that L_0 is much larger than the thermal gyro-radius $\rho_i \equiv v_t/\Omega$, i.e., $\lambda \equiv \rho_i / L_0$ is a small parameter, where $v_t = \sqrt{T/m}$ is the thermal velocity. That is

$$\frac{1}{F_g} \rho_i |\nabla_X F_g| \sim O(\lambda^1). \tag{60}$$

The equilibrium (macroscopic) $\mathbf{E}_0 \times \mathbf{B}_0$ flow, i.e.,

$$\mathbf{v}_{\mathbf{E}0} = \mathbf{E}_0 \times \mathbf{e}_\parallel / B_0 = -\nabla \Phi_0 \times \mathbf{e}_\parallel / B_0, \tag{61}$$

is assumed to be weak with

$$\frac{|\mathbf{v}_{E0}|}{v_t} \sim O(\lambda^1), \tag{62}$$

3.2.2 Orderings of microscopic quantities

We consider low frequency perturbations with $\omega/\Omega \sim O(\lambda^1)$, then

$$\frac{1}{\delta F_g} \frac{1}{\Omega} \frac{\partial \delta F_g}{\partial t} \sim O(\lambda^1). \tag{63}$$

We assume that the amplitudes of perturbations are small. Specifically, we assume

$$\frac{\delta F_g}{F_g} \sim \frac{q\delta\Phi}{T} \sim \frac{|\delta \mathbf{B}|}{B_0} \sim O(\lambda^1), \tag{64}$$

where $\delta\Phi$ is the perturbed scalar potential defined later in Eq. (69).

The perturbation is assumed to have a long wavelength (much longer than ρ_i) in the parallel direction

$$\frac{1}{\delta F_g} |\rho_i \mathbf{e}_\parallel \cdot \nabla_X \delta F_g| \sim O(\lambda^1), \tag{65}$$

and have a short wavelength comparable to the thermal gyro-radius in the perpendicular direction

$$\frac{1}{\delta F_g} |\rho_i \nabla_{X\perp} \delta F_g| \sim O(\lambda^0). \tag{66}$$

Combining Eq. (65) and (66), we obtain

$$\frac{k_\parallel}{k_\perp} \approx \frac{\mathbf{e}_\parallel \cdot \nabla_X}{\nabla_{X\perp}} \sim O(\lambda), \tag{67}$$

i.e., the parallel wave number is one order smaller than the perpendicular wave-number.]

In terms of the scalar and vector potentials $\delta\Phi$ and $\delta\mathbf{A}$, the perturbed electromagnetic field is written as

$$\delta\mathbf{B} = \nabla_x \times \delta\mathbf{A}, \tag{68}$$

and

$$\delta\mathbf{E} = -\nabla_x \delta\Phi - \frac{\partial \delta\mathbf{A}}{\partial t}. \tag{69}$$

Then

$$\delta\mathbf{E}_\parallel = -\nabla_\parallel \delta\Phi - \left(\frac{\partial \delta\mathbf{A}}{\partial t}\right)_\parallel \tag{70}$$

$$\delta\mathbf{E}_\perp = -\nabla_\perp \delta\Phi - \left(\frac{\partial \delta\mathbf{A}}{\partial t}\right)_\perp, \tag{71}$$

Using the above orderings, it is ready see that $\delta E_\parallel$ is one order smaller than $\delta E_\perp$, i.e.,

$$\frac{\delta E_\parallel}{\delta E_\perp} = O(\lambda^1). \tag{72}$$

3.3 Equation for macroscopic distribution function F_g

The evolution of the macroscopic quantity F_g is governed by Eq. (58), i.e.,

$$L_g F_g = 0, \tag{73}$$

where the left-hand side is written as

$$\begin{aligned} L_g F_g &= \frac{\partial F_g}{\partial t} + \frac{\partial \mathbf{X}}{\partial t}\cdot\frac{\partial F_g}{\partial \mathbf{X}} + \frac{\partial \mathbf{V}}{\partial t}\cdot\frac{\partial F_g}{\partial \mathbf{V}} \\ &+ (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_{\mathbf{E}0})\cdot\frac{\partial F_g}{\partial \mathbf{X}} + \mathbf{v}\cdot[(\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2})F_g] - \Omega\frac{\partial F_g}{\partial \alpha} \\ &+ \frac{q}{m}\mathbf{E}_0 \cdot \left(\frac{\mathbf{v}_\perp}{B_0}\frac{\partial F_g}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial F_g}{\partial \alpha}\right) \end{aligned}$$

Expand F_g as $F_g = F_{g0} + F_{g1} + ...$, where $F_{gi} \sim F_{g0} O(\lambda^i)$. Then, the balance on order $O(\lambda^0)$ gives

$$\frac{\partial F_{g0}}{\partial \alpha} = 0 \tag{74}$$

i.e., F_{g0} is independent of the gyro-angle α. The balance on $O(\lambda^1)$ gives

$$v_\parallel \mathbf{e}_\parallel \cdot \frac{\partial F_{g0}}{\partial \mathbf{X}} + \frac{q}{m}\mathbf{E}_0 \cdot \left(\frac{\mathbf{v}_\perp}{B_0}\frac{\partial F_{g0}}{\partial \mu}\right) = \Omega\frac{\partial F_{g1}}{\partial \alpha}. \tag{75}$$

Performing averaging over α, $\int_0^{2\pi}(...)d\alpha$, on the above equation and noting that F_{g0} is independent of α, we obtain

$$\left(\int_0^{2\pi} d\alpha v_\parallel \mathbf{e}_\parallel\right)\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}}+\frac{q}{m}\frac{\partial F_{g0}}{\partial \mu}\int_0^{2\pi} d\alpha \mathbf{E}_0\cdot\left(\frac{\mathbf{v}_\perp}{B_0}\right)=\int_0^{2\pi} d\alpha\Omega\frac{\partial F_{g1}}{\partial\alpha} \tag{76}$$

Note that a quantity $A=A(\mathbf{x})$ that is independent of $\mathbf{v}$ will depend on $\mathbf{v}$ when transformed to guiding-center coordinates, i.e., $A(\mathbf{x})=A_g(\mathbf{X},\mathbf{v})$. Therefore A_g depends on gyro-angle α. However, since $\rho_i/L\ll 1$ for equilibrium quantities, the gyro-angle dependence of the equilibrium quantities can be neglected. Specifically, $\mathbf{e}_\parallel$, B_0 and Ω can be considered to be independent of α. As to $v_\parallel$, we have $v_\parallel=\pm\sqrt{2(\varepsilon-B_0\mu)}$. Since B_0 is considered independent of α, so does $v_\parallel$. Using these results, equation (76) is written

$$v_\parallel\mathbf{e}_\parallel\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}}+\frac{q}{m}\frac{\partial F_{g0}}{\partial \mu}\int_0^{2\pi} d\alpha \mathbf{E}_0\cdot\left(\frac{\mathbf{v}_\perp}{B_0}\right)=0. \tag{77}$$

Using $\mathbf{E}_0=-\nabla\Phi_0$, the above equation is written as

$$v_\parallel\mathbf{e}_\parallel\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}}+\frac{q}{m}\frac{\partial F_{g0}}{\partial \mu}\int_0^{2\pi} d\alpha\left(\frac{-\mathbf{v}_\perp\cdot\nabla\Phi_0}{B_0}\right)=0, \tag{78}$$

Note that

$$\int_0^{2\pi} d\alpha\frac{1}{B_0}\mathbf{v}_\perp\cdot\nabla_X\Phi_0\approx 0, \tag{79}$$

where the error is of $O(\lambda^2)\Phi_0$, and thus, accurate to $O(\lambda)$, the last term of equation (78) is zero. Then equation (78) is written as

$$v_\parallel\mathbf{e}_\parallel\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}}=0, \tag{80}$$

which implies that F_{g0} is constant along a magnetic field line.

3.4 Equation for perturbed distribution function δF_g

Using $F_g\approx F_{g0}$, equation (59) is written as

$$L_g\delta F_g=\delta R F_{g0}+\underbrace{\delta R\delta F_g}_{\mathrm{NonlinearTerm}\sim O(\lambda^2)}, \tag{81}$$

where $\delta R\delta F_g$ is a nonlinear term which is of order $O(\lambda^2)$ or higher, $L_g\delta F_g$ and $\delta R F_{g0}$ are linear terms which are of order $O(\lambda^1)$ or higher. The linear term $\delta R F_{g0}$ is given by

$$\delta R F_{g0} = \underbrace{-\frac{q}{m}\left(\delta\mathbf{E}+\frac{\mathbf{v}\times\delta\mathbf{B}}{c}\right)\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}}}_{O(\lambda^2)} - \underbrace{\frac{q}{m}\delta\mathbf{E}\cdot\left(\mathbf{v}\frac{\partial F_{g0}}{\partial\varepsilon}+\frac{\mathbf{v}_\perp}{B_0}\frac{\partial F_{g0}}{\partial\mu}\right)}_{O(\lambda^1)}, \tag{82}$$

In obtaining (82), use has been made of $\partial F_{g0}/\partial\alpha=0$. Another linear term $L_g\delta F_g$ is written as

$$L_g\delta F_g = \frac{\partial\delta F_g}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_{\mathbf{E}0})\cdot\frac{\partial\delta F_g}{\partial\mathbf{X}}+\mathbf{v}\cdot[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}]\delta F_g-\underbrace{\Omega\frac{\partial\delta F_g}{\partial\alpha}}_{O(\lambda^1)}$$
$$+\frac{q}{m}\mathbf{E}_0\cdot\left(\frac{\mathbf{v}_\perp}{B_0}\frac{\partial\delta F_g}{\partial\mu}+\frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial\delta F_g}{\partial\alpha}\right), \tag{83}$$

where $\Omega\partial\delta F_g/\partial\alpha$ is of order $O(\lambda^1)$ and all the other terms are of order $O(\lambda^2)$.

Next, to reduce the complexity of algebra, we consider the easier case in which $\partial F_{g0}/\partial\mu=0$.

3.4.1 Balance on order $O(\lambda^1)$: adiabatic response

The balance between the leading terms (terms of $O(\lambda)$) in Eq. (81) requires that

$$\Omega\frac{\partial\delta F_a}{\partial\alpha}=\frac{q}{m}\delta\mathbf{E}\cdot\left(\mathbf{v}\frac{\partial F_{g0}}{\partial\varepsilon}\right), \tag{84}$$

where δF_a is a unknown distribution function to be solved from the above equation. It is ready to verify that

$$\delta F_a=\frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial\varepsilon}, \tag{85}$$

is a solution to the above equation, accurate to $O(\lambda)$. [Proof: Substitute expression (85) into the left-hand side of Eq. (84), we obtain

$$\begin{aligned}\Omega\frac{\partial\delta F_a}{\partial\alpha} &= \Omega\frac{\partial}{\partial\alpha}\left(\frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial\varepsilon}\right)\\ &= \frac{q}{m}\frac{\partial F_{g0}}{\partial\varepsilon}\Omega\frac{\partial}{\partial\alpha}(\delta\Phi)\\ &= \frac{q}{m}\frac{\partial F_{g0}}{\partial\varepsilon}\Omega(\nabla_x\delta\Phi)\cdot\frac{\partial\mathbf{x}}{\partial\alpha}\end{aligned} \tag{86}$$

Using

$$\begin{aligned}\frac{\partial\mathbf{x}}{\partial\alpha} &= \frac{\partial}{\partial\alpha}\left[-\mathbf{v}\times\frac{\boldsymbol{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\right]\\ &= \frac{\partial}{\partial\alpha}[-\mathbf{v}]\times\frac{\boldsymbol{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\\ &= -\frac{\mathbf{v}_\perp}{\Omega}\end{aligned} \tag{87}$$

Eq. (86) is written as

$$\begin{aligned}\Omega\frac{\partial\delta F_a}{\partial\alpha} &= -\frac{q}{m}\frac{\partial F_{g0}}{\partial\varepsilon}\Omega(\nabla_x\delta\Phi)\cdot\frac{\mathbf{v}_\perp}{\Omega}\\ &= \frac{q}{m}\frac{\partial F_{g0}}{\partial\varepsilon}\left(\delta\mathbf{E}+\frac{\partial\delta\mathbf{A}}{\partial t}\right)\cdot\mathbf{v}_\perp \qquad (88)\\ &\approx \frac{q}{m}\frac{\partial F_{g0}}{\partial\varepsilon}(\delta\mathbf{E})\cdot\mathbf{v}_\perp, \qquad (89)\end{aligned}$$

where terms of $O(\lambda^2)$ have been dropped. Similarly, dropping the parallel electric field term (which is of $O(\lambda^2)$) on the right-hand side of Eq. (84), we find it is identical to the right-hand side of Eq. (89)]

3.4.2 Separate δF_g into adiabatic and non-adiabatic part

As is discussed above, the terms of $O(\lambda)$ can be eliminated by splitting a so-called adiabatic term form δF_g. Specifically, write δF_g as

$$\delta F_g = \delta F_a + \delta G, \tag{90}$$

where δF_a is given by (85), i.e.,

$$\delta F_a = \frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial \varepsilon}, \tag{91}$$

which depends on the gyro-angle via $\delta\Phi$ and this term is often called adiabatic term. Plugging expression (90) into equation (81), we obtain

$$L_g\delta G = \underbrace{\delta R F_{g0} - L_g\delta F_a}_{\text{LinearTerms}} + \underbrace{\delta R\delta F_g}_{\text{NonlinearTerms}}. \tag{92}$$

Next, let us simplify the linear term on the right-hand side, i.e, $\delta R F_{g0} - L_g\delta F_a$, (which should be of $O(\lambda^2)$ or higher because $\Omega\partial\delta F_a / \partial\alpha$ cancels all the $O(\lambda^1)$ terms in $\delta R F_{g0}$).

$L_g\delta F_a$ is written

$$\begin{aligned} L_g\delta F_a &= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}L_g\delta\Phi + \frac{q}{m}\delta\Phi L_g\frac{\partial F_{g0}}{\partial \varepsilon} \\ &\approx \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}L_g\delta\Phi, \end{aligned} \tag{93}$$

where the error is of order $O(\lambda^3)$. In obtaining the above expression, use has been made of $\mathbf{e}_\parallel \cdot \partial F_{g0} / \partial \mathbf{X} = 0$, $\partial F_{g0} / \partial \mathbf{X} = O(\lambda^1)F_{g0}$, $\partial F_{g0} / \partial\alpha = 0$, $\partial F_{g0} / \partial\mu = 0$, and the definition of $\boldsymbol{\lambda}_{B1}$ and $\boldsymbol{\lambda}_{B2}$ given in expressions (31) and (32). The expression (93) involves $\delta\Phi$ operated by the Vlasov propagator L_g. Since $\delta\Phi$ takes the most simple form when expressed in particle coordinates (if in guiding-center coordinates, $\delta\Phi(\mathbf{x}) = \delta\Phi(\mathbf{X} - \mathbf{v}\times\mathbf{e}_\parallel / \Omega)$, which depends on velocity coordinates and thus more complicated), it is convenient to use the Vlasov propagator L_g expressed in particle coordinates. Transforming L_g back to the particle coordinates, expression (93) is written

$$\begin{aligned} L_g\delta F_a &= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial\delta\Phi}{\partial t}|_{\mathbf{x},\mathbf{v}} + \mathbf{v}\cdot\nabla_{\mathbf{x}}\delta\Phi + \frac{q}{m}(\mathbf{E}_0 + \mathbf{v}\times\mathbf{B}_0)\cdot\frac{\partial\Phi}{\partial\mathbf{v}}|_{\mathbf{x}}\right] \\ &= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial\delta\Phi}{\partial t}|_{\mathbf{x},\mathbf{v}} + \mathbf{v}\cdot\nabla_{\mathbf{x}}\delta\Phi\right] \\ &= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial\delta\Phi}{\partial t}|_{\mathbf{x},\mathbf{v}} + \mathbf{v}\cdot\left(-\delta\mathbf{E} - \frac{\partial\delta\mathbf{A}}{\partial t}|_{\mathbf{x},\mathbf{v}}\right)\right] \end{aligned} \tag{94}$$

$$= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial \delta\Phi}{\partial t}|_{\mathbf{x},\mathbf{v}} - \mathbf{v}\cdot\delta\mathbf{E} - \frac{\partial \mathbf{v}\cdot\delta\mathbf{A}}{\partial t}|_{\mathbf{x},\mathbf{v}}\right]. \tag{95}$$

$$= \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial \delta\Phi}{\partial t} - \mathbf{v}\cdot\delta\mathbf{E} - \frac{\partial \mathbf{v}\cdot\delta\mathbf{A}}{\partial t}\right]. \tag{96}$$

Using this and expression (82), $\delta RF_{g0} - L_g\delta F_a$ is written as

$$\begin{aligned}\delta RF_{g0} - L_g\delta F_a &= -\frac{q}{m}(\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B})\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}} - \frac{q}{m}\delta\mathbf{E}\cdot\left(\mathbf{v}\frac{\partial F_{g0}}{\partial \varepsilon}\right)\\ &\quad - \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial \delta\Phi}{\partial t} - \mathbf{v}\cdot\delta\mathbf{E} - \frac{\partial \mathbf{v}\cdot\delta\mathbf{A}}{\partial t}\right]\\ &= -\frac{q}{m}(\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B})\times\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\cdot\frac{\partial F_{g0}}{\partial \mathbf{X}} - \frac{q}{m}\frac{\partial F_{g0}}{\partial \varepsilon}\left[\frac{\partial \Phi}{\partial t} - \frac{\partial \mathbf{v}\cdot\delta\mathbf{A}}{\partial t}\right],\end{aligned} \tag{97}$$

where the two terms of $O(\lambda^1)$ (the terms in blue and red) cancel each other, with the remain terms being all of $O(\lambda^2)$, i.e, the contribution of the adiabatic term cancels the leading order terms of $O(\lambda^1)$ on the RHS of Eq. (92).

The consequence of this is that, as we will see in Sec. 3.5.1, δG is independent of the gyro-angle, accurate to order $O(\lambda^1)$. Therefore, separating δF into adiabatic and non-adiabatic parts also corresponds to separating δF into gyro-angle dependent and gyro-angle independent parts.

3.4.3 Linear term expressed in terms of $\delta\Phi$ and $\delta\mathbf{A}$

Let us rewrite the linear term (97) in terms of $\delta\Phi$ and $\delta\mathbf{A}$. The $\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B}$ term in expression (97) is written as

$$\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B} = -\nabla_x\delta\Phi - \frac{\partial\delta\mathbf{A}}{\partial t} + \mathbf{v}\times\nabla_x\times\delta\mathbf{A}. \tag{98}$$

Note that this term needs to be accurate to only $O(\lambda)$. Then

$$\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B} \approx -\nabla_x\delta\Phi + \mathbf{v}\times\nabla_x\times\delta\mathbf{A}, \tag{99}$$

where the error is of $O(\lambda^2)$. Using the vector identity $\mathbf{v}\times\nabla_x\times\delta\mathbf{A} = (\nabla\delta\mathbf{A})\cdot\mathbf{v} - (\mathbf{v}\cdot\nabla)\delta\mathbf{A}$ and noting $\mathbf{v}$ is constant for ∇_x operator, the above equation is written

$$\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B} = -\nabla_x\delta\Phi + \nabla_x(\delta\mathbf{A}\cdot\mathbf{v}) - (\mathbf{v}\cdot\nabla_x)\delta\mathbf{A} \tag{100}$$

Note that Eq. (33) indicates that $\nabla_x\delta\Phi \approx \nabla_X\delta\Phi$, where the error is of $O(\lambda^2)$, then the above equation is written

$$\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B} = -\nabla_X\delta\Phi + \nabla_X(\delta\mathbf{A}\cdot\mathbf{v}) - (\mathbf{v}\cdot\nabla_X)\delta\mathbf{A} \tag{101}$$

Further note that the parallel gradients in the above equation are of $O(\lambda^2)$ and thus can be dropped. Then expression (101) is written

$$\begin{aligned}&\delta\mathbf{E}+\mathbf{v}\times\delta\mathbf{B}\\ =&-\nabla_{X\perp}\delta\Phi + \nabla_{X\perp}(\delta\mathbf{A}\cdot\mathbf{v}) - (\mathbf{v}_\perp\cdot\nabla_{X\perp})\delta\mathbf{A}.\\ =&-\nabla_{X\perp}\delta L - \mathbf{v}_\perp\cdot\nabla_{X\perp}\delta\mathbf{A},\end{aligned} \tag{102}$$

where δL is defined by

$$\delta L = \delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}. \tag{103}$$

Using expression (102), equation (97) is written

$$\delta RF_0 - L_g\delta F_a = -\frac{q}{m}\Big[(-\nabla_{X\perp}\delta L - \mathbf{v}_\perp\cdot\nabla_{X\perp}\delta\mathbf{A})\times\frac{\boldsymbol{e}_\parallel}{\Omega}\Big]\cdot\frac{\partial F_{g0}}{\partial\mathbf{X}} - \frac{q}{m}\frac{\partial\delta L}{\partial t}\frac{\partial F_{g0}}{\partial\varepsilon}, \tag{104}$$

where all terms are of $O(\lambda^2)$.

3.5 Equation for the non-adiabatic part δG

Plugging expression (104) into Eq. (92), we obtain

$$L_g\delta G = -\frac{q}{m}\Big[(-\nabla_{X\perp}\delta L - \mathbf{v}_\perp\cdot\nabla_{X\perp}\delta\mathbf{A})\times\frac{\boldsymbol{e}_\parallel}{\Omega}\Big]\cdot\frac{\partial F_{g0}}{\partial\mathbf{X}} - \frac{q}{m}\frac{\partial\delta L}{\partial t}\frac{\partial F_{g0}}{\partial\varepsilon} + \delta R\delta F_g, \tag{105}$$

where L_g is given by Eq. (83), i.e.,

$$\begin{aligned} L_g &= \frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_{\mathbf{E}0})\cdot\frac{\partial}{\partial\mathbf{X}} + \mathbf{v}\cdot[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}] - \Omega\frac{\partial}{\partial\alpha} \\ &+ \frac{q}{m}\mathbf{E}_0\cdot\left(\frac{\mathbf{v}_\perp}{B_0}\frac{\partial}{\partial\mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp}\frac{\partial}{\partial\alpha}\right), \end{aligned} \tag{106}$$

3.5.1 Expansion of δG

Expand δG as

$$\delta G = \delta G_0 + \delta G_1 + ...,$$

where $\delta G_i \sim O(\lambda^{i+1})F_{g0}$, and note that the right-hand side of Eq. (105) is of $O(\lambda^2)$, then, the balance on order $O(\lambda^1)$ requires

$$\frac{\partial\delta G_0}{\partial\alpha} = 0, \tag{107}$$

i.e., δG_0 is gyro-phase independent.

The balance on order $O(\lambda^2)$ requires (for the special case of $\mathbf{E}_0 = 0$):

$$\begin{aligned} &\frac{\partial\delta G_0}{\partial t} + v_\parallel\mathbf{e}_\parallel\cdot\frac{\partial\delta G_0}{\partial\mathbf{X}} + \mathbf{v}\cdot[\boldsymbol{\lambda}_{B1}+\boldsymbol{\lambda}_{B2}]\delta G_0 \\ =\ & -\frac{q}{m}\Big[(-\nabla_{X\perp}\delta L - \mathbf{v}_\perp\cdot\nabla_X\delta\mathbf{A})\times\frac{\boldsymbol{e}_\parallel}{\Omega}\Big]\cdot\frac{\partial F_{g0}}{\partial\mathbf{X}} - \frac{q}{m}\frac{\partial\delta L}{\partial t}\frac{\partial F_{g0}}{\partial\varepsilon} + \delta R\delta F_g. \end{aligned} \tag{108}$$

3.5.2 Gyro-averaging

Define the gyro-average operator $\langle ...\rangle_\alpha$ by

$$\langle h\rangle_\alpha = (2\pi)^{-1}\int_0^{2\pi} h\,d\alpha, \tag{109}$$

where $h = h(\mathbf{X},\alpha,\varepsilon,\mu)$ is an arbitrary function of guiding-center variables. The gyro-averaging is an integration in the velocity space. [For a field quantity, which is independent of the velocity in particle coordinates, i.e., $h = h(\mathbf{x})$, it is ready to see that the above averaging is a spatial averaging over a gyro-ring.]

Gyro-averaging Eq. (108), we obtain

$$\frac{\partial \delta G_0}{\partial t} + \left\langle v_\parallel \mathbf{e}_\parallel \cdot \frac{\partial \delta G_0}{\partial \mathbf{X}} \right\rangle + \langle \mathbf{v} \cdot [\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2}] \delta G_0 \rangle_\alpha$$
$$= -\frac{q}{m}\Big[-\nabla_{X\perp} \langle \delta L \rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega} \Big] \cdot \frac{\partial F_{g0}}{\partial \mathbf{X}} - \frac{q}{m} \frac{\partial \langle \delta L \rangle_\alpha}{\partial t} \frac{\partial F_{g0}}{\partial \varepsilon} + \langle \delta R \delta F_g \rangle_\alpha, \tag{110}$$

where use has been made of $\langle (\mathbf{v}_\perp \cdot \nabla_X) \delta \mathbf{A} \rangle_\alpha \approx 0$, where the error is of order higher than $O(\lambda^2)$. Note that $v_\parallel = \pm\sqrt{2(\varepsilon - B_0\mu)}$. Since B_0 is approximately independent of α, so does $v_\parallel$. Using this, the first gyro-averaging on the left-hand side of the above equation is written

$$\left\langle v_\parallel \mathbf{e}_\parallel \cdot \frac{\partial \delta G_0}{\partial \mathbf{X}} \right\rangle_\alpha = \langle v_\parallel \mathbf{e}_\parallel \rangle \cdot \frac{\partial \delta G_0}{\partial \mathbf{X}} = v_\parallel \mathbf{e}_\parallel \cdot \frac{\partial \delta G_0}{\partial \mathbf{X}} \tag{111}$$

The second gyro-averaging on the left-hand side of Eq. (110) can be written as

$$\langle \mathbf{v} \cdot [\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2}] \delta G_0 \rangle_\alpha = \mathbf{V}_D \cdot \nabla_X \delta G_0, \tag{112}$$

where $\mathbf{V}_D$ is the magnetic curvature and gradient drift (Eq. (112) is derived in Appendix xx, to do later). Then Eq. (110) is written

$$\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D) \cdot \nabla_X \right] \delta G_0$$
$$= -\frac{q}{m}\Big[-\nabla_{X\perp} \langle \delta L \rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega} \Big] \cdot \frac{\partial F_{g0}}{\partial \mathbf{X}} - \frac{q}{m} \frac{\partial \langle \delta L \rangle_\alpha}{\partial t} \frac{\partial F_{g0}}{\partial \varepsilon} + \langle \delta R \delta F_g \rangle_\alpha. \tag{113}$$

3.5.3 Simplification of the nonlinear term

Next, we try to simplify the nonlinear term $\langle \delta R \delta F_g \rangle_\alpha$ appearing in Eq. (113), which is written as

$$\begin{aligned} \langle \delta R \delta F_g \rangle_\alpha &= \left\langle \delta R \left(\frac{q}{m} \delta\Phi \frac{\partial F_{g0}}{\partial \varepsilon} + \delta G_0 \right) \right\rangle_\alpha \\ &= \left\langle \frac{q}{m} \delta R \left(\delta\Phi \frac{\partial F_{g0}}{\partial \varepsilon} \right) \right\rangle_\alpha + \langle \delta R \delta G_0 \rangle_\alpha \end{aligned} \tag{114}$$

First, let us focus on the first term, which can be written as

$$\begin{aligned} \delta R \left(\delta\Phi \frac{\partial F_{g0}}{\partial \varepsilon} \right) &\approx -\frac{q}{m} \frac{\partial F_{g0}}{\partial \varepsilon} \left(\delta\mathbf{E} + \frac{\mathbf{v} \times \delta\mathbf{B}}{c} \right) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \cdot \frac{\partial \delta\Phi}{\partial \mathbf{X}} - \frac{q}{m} \frac{\partial F_{g0}}{\partial \varepsilon} \left(\frac{\mathbf{v} \times \delta\mathbf{B}}{c} \right) \cdot \\ &\quad \left(\frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial \delta\Phi}{\partial \alpha} \right) \\ &\quad - \frac{q}{m} \frac{\partial F_{g0}}{\partial \varepsilon} \delta\mathbf{E} \cdot \left(\mathbf{v} \frac{\partial \delta\Phi}{\partial \varepsilon} + \frac{\mathbf{v}_\perp}{B_0} \frac{\partial \delta\Phi}{\partial \mu} + \frac{\boldsymbol{e}_\alpha}{v_\perp} \frac{\partial \delta\Phi}{\partial \alpha} \right) + \frac{q}{m} \delta\Phi \delta\mathbf{E} \cdot \mathbf{v} \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \\ &= -\frac{q}{m} \left(\delta\mathbf{E} + \frac{\mathbf{v} \times \delta\mathbf{B}}{c} \right) \cdot \nabla_v \delta\Phi + \frac{q}{m} \delta\Phi \delta\mathbf{E} \cdot \mathbf{v} \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \\ &= \frac{q}{m} \delta\Phi \delta\mathbf{E} \cdot \mathbf{v} \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \end{aligned} \tag{115}$$

Using the above results, the nonlinear term $\langle \delta R \delta F \rangle_\alpha$ is written as

$$\langle \delta R \delta F \rangle_\alpha = \frac{q}{m} \left\langle \delta\Phi \delta \mathbf{E} \cdot \mathbf{v} \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \right\rangle_\alpha + \langle \delta R \delta G_0 \rangle_\alpha \tag{116}$$

Accurate to $O(\lambda^2)$, the first term on the right-hand side of the above is zero. [Proof:

$$\begin{aligned} \left\langle \delta\Phi \delta \mathbf{E} \cdot \mathbf{v} \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \right\rangle_\alpha &= \left\langle \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \delta\Phi \nabla \delta\Phi \cdot \mathbf{v} \right\rangle_\alpha \\ &= \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \langle \mathbf{v} \cdot \nabla (\delta\Phi)^2 \rangle_\alpha \\ &\approx \frac{\partial^2 F_{g0}}{\partial \varepsilon^2} \langle \mathbf{v}_\perp \cdot \nabla (\delta\Phi)^2 \rangle_\alpha \\ &\approx 0, \end{aligned} \tag{117}$$

where use has been made of $\langle \mathbf{v}_\perp \cdot \nabla_X \delta\Phi \rangle_\alpha \approx 0$, where the error is of $O(\lambda^2)$. Using the above results, expression (116) is written as

$$\langle \delta R \delta F_g \rangle_\alpha = \langle \delta R \delta G_0 \rangle_\alpha. \tag{118}$$

Using the expression of δR given by Eq. (47), the above expression is written as

$$\begin{aligned} \langle \delta R \delta G_0 \rangle_\alpha &= -\frac{q}{m} \left\langle \left(\delta \mathbf{E} + \frac{\mathbf{v} \times \delta \mathbf{B}}{c} \right) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \right\rangle_\alpha \cdot \frac{\partial \delta G_0}{\partial \mathbf{X}} \\ &\quad - \frac{q}{m} \frac{\partial \delta G_0}{\partial \varepsilon} \langle \delta \mathbf{E} \cdot \mathbf{v} \rangle_\alpha - \frac{q}{m} \frac{\partial \delta G_0}{\partial \mu} \left\langle \delta \mathbf{E} \cdot \frac{\mathbf{v}_\perp}{B_0} \right\rangle_\alpha \end{aligned} \tag{119}$$

where use has been made of $\partial \delta G_0 / \partial \alpha = 0$. Using Eq. (102), we obtain

$$-\frac{q}{m} \left\langle (\delta \mathbf{E} + \mathbf{v} \times \delta \mathbf{B}) \times \left(\frac{\boldsymbol{e}_\parallel}{\Omega} \right) \right\rangle_\alpha = \frac{q}{m} \nabla_{X\perp} \langle \delta L \rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega}. \tag{120}$$

The other two terms in Eq. (119) can be proved to be zero. [Proof:

$$\begin{aligned} -\frac{q}{m} \frac{\partial \delta G_0}{\partial \varepsilon} \langle \delta \mathbf{E} \cdot \mathbf{v} \rangle_\alpha &= \frac{q}{m} \frac{\partial \delta G_0}{\partial \varepsilon} \langle \mathbf{v} \cdot \nabla_x \Phi \rangle_\alpha \\ &\approx \frac{q}{m} \frac{\partial \delta G_0}{\partial \varepsilon} \langle \mathbf{v}_\perp \cdot \nabla_x \Phi \rangle_\alpha \\ &\approx \frac{q}{m} \frac{\partial \delta G_0}{\partial \varepsilon} \langle \mathbf{v}_\perp \cdot \nabla_X \Phi \rangle_\alpha \\ &\approx 0 \end{aligned} \tag{121}$$

$$\begin{aligned} -\frac{q}{m} \frac{\partial \delta G_0}{\partial \mu} \left\langle \delta \mathbf{E} \cdot \frac{\mathbf{v}_\perp}{B_0} \right\rangle_\alpha &= \frac{q}{m} \frac{\partial \delta G_0}{\partial \mu} \left\langle \frac{1}{B_0} \mathbf{v}_\perp \cdot \nabla_x \Phi \right\rangle_\alpha \\ &\approx \frac{q}{m} \frac{\partial \delta G_0}{\partial \mu} \left\langle \frac{1}{B_0} \mathbf{v}_\perp \cdot \nabla_X \Phi \right\rangle_\alpha \\ &\approx 0 \end{aligned} \tag{122}$$

] Using the above results, the nonlinear term is finally written as

$$\langle \delta R \delta G_0 \rangle_\alpha = \frac{q}{m} \left[\nabla_{X\perp} \langle \delta L \rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega} \right] \cdot \nabla_X \delta G_0. \tag{123}$$

Using this in Eq. (118), we obtain

$$\langle \delta R \delta F_g \rangle_\alpha = \frac{q}{m}\Big[\nabla_{X\perp}\langle \delta L\rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega}\Big]\cdot \nabla_X \delta G_0, \tag{124}$$

which is of $O(\lambda^2)$.

3.5.4 Final equation for the non-adiabatic part of the perturbed distribution function

Using the above results, the gyro-averaged kinetic equation for δG_0 is finally written as

$$\begin{aligned}&\frac{\partial \delta G_0}{\partial t} + \left(v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D - \underbrace{\frac{q}{m}\nabla_X \langle \delta L\rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega}}_{\text{nonlinear}} \right)\cdot \nabla_X \delta G_0 \\ =&\underbrace{\Big(\frac{q}{m}\nabla_X \langle \delta L\rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega}\Big)\cdot \nabla_X F_{g0}}_{\text{spatial}-\text{drive}} - \underbrace{\frac{q}{m}\frac{\partial \langle \delta L\rangle_\alpha}{\partial t}\frac{\partial F_{g0}}{\partial \varepsilon}}_{\text{velocit}-\text{space}-\text{damp}}.\end{aligned} \tag{125}$$

where $\mathbf{V}_D$ is the equilibrium guiding-center drift velocity, $\langle ... \rangle_\alpha$ is the gyro-phase averaging operator, $\delta L = \delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}$, and $\delta G_0 = \delta G_0(\mathbf{X}, \varepsilon, \mu, t)$ is gyro-angle independent and is related to the perturbed distribution function δF_g by

$$\delta F_g = \frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial \varepsilon} + \delta G_0, \tag{126}$$

where the first term is called "adiabatic term", which depends on the gyro-phase α via $\delta\Phi$. Equation (125) is the special case ($\partial F_{g0}/\partial\mu|_\varepsilon = 0$) of the Frieman-Chen nonlinear gyrokinetic equation given in Ref. [3]. Note that the nonlinear terms only appear on the left-hand side of Eq. (125) and all the terms on the right-hand side are linear. The term

$$-\frac{q}{m}\nabla_X \langle \delta L\rangle_\alpha \times \frac{\boldsymbol{e}_\parallel}{\Omega}, \tag{127}$$

consists of the perturbed $E\times B$ drift and magnetic fluterring term (refer to Sec. F.3).

4 Characteristic curves of Frieman-Chen nonlinear gyrokinetic equation

The Frieman-Chen nonlinear gyrokinetic equation takes the following form:

$$\begin{aligned}&\frac{\partial \delta G_0}{\partial t} + \Big(v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D - \frac{q}{m}\nabla_X \langle \delta L\rangle_\alpha \times \frac{\mathbf{e}_\parallel}{\Omega}\Big)\cdot \nabla_X \delta G_0 \\ =&\frac{q}{m}\Big[\nabla_X \langle \delta L\rangle_\alpha \times \frac{\mathbf{e}_\parallel}{\Omega}\Big]\cdot \nabla_X F_{g0} - \frac{q}{m}\frac{\partial \langle \delta L\rangle_\alpha}{\partial t}\frac{\partial F_{g0}}{\partial \varepsilon} - \frac{q}{m}S_{l2},\end{aligned} \tag{128}$$

where S_{l2} is due to the μ dependence of F_{g0}, which is not considered in this note. Dropping S_{l2} term, equation (128) agrees with Eq. (125) derived above.

Examining the left-hand side of Eq. (128), it is ready to find that the characteristic curves of this equation are given by the following equations:

$$\frac{d\mathbf{X}}{dt} = v_{\parallel}\mathbf{e}_{\parallel} + \mathbf{V}_D - \frac{q}{m}\nabla_X \langle \delta L \rangle_\alpha \times \frac{\mathbf{e}_{\parallel}}{\Omega}, \tag{129}$$

$$\frac{d\varepsilon}{dt} = 0, \tag{130}$$

$$\frac{d\mu}{dt} = 0. \tag{131}$$

(It is instructive to notice that the kinetic energy ε is conserved along the characteristic curves while the real kinetic energy of a particle is usually not conserved in a perturbed electromagnetic field. This may be an indication that Frieman-Chen equation neglects the velocity space nonlinearity.) For notation ease, we define the perturbed drift $\delta\mathbf{V}_D$ by

$$\delta\mathbf{V}_D = -\frac{q}{m}\nabla_X \langle \delta L \rangle_\alpha \times \frac{\mathbf{e}_{\parallel}}{\Omega}, \tag{132}$$

(this drift can be written as expression (314)), and the total guiding-center velocity $\mathbf{V}_G$ by

$$\mathbf{V}_G = \frac{d\mathbf{X}}{dt} = v_{\parallel}\mathbf{e}_{\parallel} + \mathbf{V}_D + \delta\mathbf{V}_D. \tag{133}$$

The characteristic curve equations (129)-(131), however, is not in the form that can be readily evolved numerically because there is no time evolution equation for $v_{\parallel}$, which appears explicitly in Eq. (129). It is ready to realize that the equation for $v_{\parallel}$ is implicitly contained in the combination of the equations for ε and μ. Next, we derive the equation for $v_{\parallel}$.

4.1 Time evolution equation for $\mathbf{v}_{\parallel}$

Using the definition $\mu = v_{\perp}^2/(2B_0)$, equation (131), i.e., $d\mu/dt = 0$, is written as

$$\frac{d}{dt}\left(\frac{v_{\perp}^2}{2B_0}\right) = 0, \tag{134}$$

which is written as

$$\frac{1}{B}\frac{d}{dt}(v_{\perp}^2) + v_{\perp}^2\frac{d}{dt}(\frac{1}{B_0}) = 0, \tag{135}$$

which can be further written as

$$\frac{d}{dt}(v_{\perp}^2) = 2\mu\frac{d}{dt}(B_0). \tag{136}$$

Using the definition of the characteristics, the right-hand side of the above equation can be expanded, giving

$$\frac{d}{dt}(v_{\perp}^2) = 2\mu\left(\frac{\partial B_0}{\partial t} + \frac{d\mathbf{X}}{dt}\cdot\nabla_X B_0 + \frac{d\varepsilon}{dt}\frac{\partial B_0}{\partial \varepsilon} + \frac{d\mu}{dt}\frac{\partial B_0}{\partial \mu}\right), \tag{137}$$

where $d\mathbf{X}/dt$, $d\varepsilon/dt$, and $d\mu/dt$ are given by Eq. (129), (130), and (131), respectively. Using Eqs. (129)-(131) and $\partial B_0/\partial t = 0$, equation (137) is reduced

to

$$\frac{d}{dt}(v_\perp^2) = 2\mu\left(\frac{d\mathbf{X}}{dt}\cdot\nabla_X B_0\right). \tag{138}$$

On the other hand, equation (130), i.e., $d\varepsilon/dt = 0$, is written as

$$\frac{d}{dt}(v^2) = 0, \tag{139}$$

which can be further written as

$$\frac{d}{dt}(v_\parallel^2) = -\frac{d}{dt}(v_\perp^2). \tag{140}$$

Using Eq. (138), the above equation is written as

$$\frac{d}{dt}(v_\parallel) = -\frac{\mu}{v_\parallel}\left(\frac{d\mathbf{X}}{dt}\cdot\nabla_X B_0\right), \tag{141}$$

which is the equation for the time evolution of $v_\parallel$. This equation involves $d\mathbf{X}/dt$, i,e., the guiding-center drift, which is given by Eq. (129). Equation (141) for $v_\parallel$ can be simplified by noting that the Frieman-Chen equation is correct only to the second order, $O(\lambda^2)$, and thus the characteristics need to be correct only to the first order $O(\lambda)$ and higher order terms can be dropped. Note that, in the guiding-center drift $d\mathbf{X}/dt$ given by Eq. (129), only the $v_\parallel\mathbf{e}_\parallel$ term in is of order $O(\lambda^0)$, all the other terms are of $O(\lambda^1)$. Using this, accurate to order $O(\lambda^1)$, equation (141) is written as

$$\frac{d}{dt}(v_\parallel) = -\mu\mathbf{e}_\parallel\cdot\nabla_X B_0, \tag{142}$$

which is the time evolution equation ready to be used for numerically advancing $v_\parallel$. Note that only the mirror force $-\mu\mathbf{e}_\parallel\cdot\nabla B$ appears in Eq. (142) and there is no parallel acceleration term $qv_\parallel\delta E_\parallel/m$ in Eq. (142). This is because $\delta E_\parallel = -\mathbf{b}\cdot\nabla\delta\Phi - \partial\delta A_\parallel/\partial t$ is of order $O(\lambda^2)$ and (**check** the terms involving $E_\parallel$ are of $O(\lambda^3)$ or higher and thus have been dropped in deriving Frieman-Chen equation.)

5 Gyrokinetic equation in forms amenable to numerical simulation

In the case of F_0 being isotropic, Frieman-Chen's gyrokinetic equation (128) is written as

$$\begin{aligned}&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta G_0\\ &= -\delta\mathbf{V}_D\cdot\nabla_X F_0 - \frac{q}{m}\frac{\partial\langle\delta L\rangle_\alpha}{\partial t}\frac{\partial F_0}{\partial\varepsilon},\end{aligned} \tag{143}$$

where δG_0 is the gyro-phase independent part of the perturbed distribution δF, and is related to δF by

$$\delta F = \frac{q}{m}\delta\Phi\frac{\partial F_0}{\partial\varepsilon} + \delta G_0, \tag{144}$$

where the first term is called "the adiabatic term", which depends on gyro-phase α via $\delta\Phi$. In Eq. (143), $\delta L = \delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}$.

5.1 Eliminate $\partial\langle\delta\phi\rangle_\alpha/\partial t$ term on the right-hand side of Eq. (143)

Note that the coefficient before $\partial F_0/\partial\varepsilon$ in Eq. (143) involves the time derivative of $\langle\delta\phi\rangle_\alpha$, which is problematic if treated by using explicit finite difference in particle simulations (I test the algorithm that treats this term by implicit scheme, the result roughly agrees with the standard method discussed in Sec. 5.6). It turns out that $\partial\langle\delta\Phi\rangle_\alpha/\partial t$ can be eliminated by defining another gyro-phase independent function δf by

$$\delta f = \frac{q}{m}\langle\delta\Phi\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon} + \delta G_0. \tag{145}$$

Then, in terms of δf, the perturbed distribution function δF is written as

$$\delta F = \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon} + \delta f. \tag{146}$$

Using Eq. (145) and Eq. (143), the equation for δf is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta f \\
&-\frac{q}{m}\frac{\partial F_0}{\partial\varepsilon}\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\langle\delta\phi\rangle_\alpha \\
&-\frac{q}{m}\langle\delta\phi\rangle_\alpha\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\frac{\partial F_0}{\partial\varepsilon} \\
&= -\delta\mathbf{V}_D\cdot\nabla_X F_0 - \frac{q}{m}\frac{\partial\langle\delta L\rangle_\alpha}{\partial t}\frac{\partial F_0}{\partial\varepsilon}
\end{aligned} \tag{147}$$

Noting that $\partial F_0/\partial t = 0$, $\mathbf{e}_\parallel\cdot\nabla F_0 = 0$, $\nabla F_0 \sim O(\lambda^1)F_0$, we find that the third line of the above equation is of order $O(\lambda^3)$ and thus can be dropped. Moving the second line to the right-hand side and noting that $\langle\delta L\rangle_\alpha = \langle\delta\phi - \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha$, the above equation is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta f \\
&= -\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}\left[-\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} - (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\langle\delta\Phi\rangle_\alpha\right]\frac{\partial F_0}{\partial\varepsilon},
\end{aligned} \tag{148}$$

where two $\partial\langle\phi\rangle_\alpha/\partial t$ terms cancel each other. Note that the right-hand side of Eq. (148) contains a nonlinear term $\delta\mathbf{V}_D\cdot\nabla_X\langle\delta\Phi\rangle_\alpha$. This is different from the original Frieman-Chen equation, where all nonlinear terms appear on the left-hand side. [Equation (148) corresponds to Eq. (A8) in Yang Chen's paper[2] (where the first minus on the right-hand side is wrong and should be replaced with q/m; one q is missing before $\partial(\mathbf{v}\cdot\delta\mathbf{A})/\partial t$ in A9).]

The blue term in expression (146) gives "the polarization density" when integrated in the velocity space (discussed in Sec. 5.6). The reason for the name "polarization" is that $(\delta\Phi - \langle\delta\Phi\rangle_\alpha)$ is the difference between the local value and the averaged value on a gyro-ring, expressing a kind of "separation".

5.2 Eliminate $\partial\langle\delta\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha/\partial t$ term on the right-hand side of GK equation

Similar to the method of eliminating $\partial\langle\delta\phi\rangle_\alpha/\partial t$, we define another gyro-phase independent function δh by

$$\delta h = \delta f - \frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}. \tag{149}$$

then Eq. (148) is written in terms of δh as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h \\
&+\frac{q}{m}\frac{\partial F_0}{\partial\varepsilon}\left[\left(\frac{\partial}{\partial t} + v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D\right)\cdot\nabla_X\right]\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \\
&+\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\left[\left(\frac{\partial}{\partial t} + v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D\right)\cdot\nabla_X\right]\left(\frac{\partial F_0}{\partial\varepsilon}\right) \\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}\left[-\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} - (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\langle\delta\Phi\rangle_\alpha\right]\frac{\partial F_0}{\partial\varepsilon},
\end{aligned} \tag{150}$$

Noting that $\partial F_0/\partial t = 0$, $\mathbf{e}_\parallel\cdot\nabla F_0 = 0$, $\nabla F_0 \sim O(\lambda^1)F_0$, we find that the third line of the above equation is of order $O(\lambda^3)$ and thus can be dropped. Moving the second line to the right-hand side and noting that $\langle\delta L\rangle_\alpha = \langle\delta\phi - \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha$, the above equation is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h \\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}[(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X(\langle\mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha)]\frac{\partial F_0}{\partial\varepsilon},
\end{aligned} \tag{151}$$

where two $\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha/\partial t$ terms cancel each other and no time derivatives of the perturbed fields appear on the right-hand side. Noting that $\delta\mathbf{V}_D$ given by Eq. (132) is perpendicular to $\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha$ and thus the blue term in Eq. (151) is zero, then Eq. (151) simplifies to

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h \\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}[(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha]\frac{\partial F_0}{\partial\varepsilon}.
\end{aligned} \tag{152}$$

Using $\mathbf{V}_G = v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D$, equation (151) can also be written as

$$\begin{aligned}&\left[\frac{\partial}{\partial t} + \mathbf{V}_G \cdot \nabla_X\right]\delta h \\ &= -\delta\mathbf{V}_D \cdot \nabla_X F_0 - \frac{q}{m}[\mathbf{V}_G \cdot \nabla_X \langle \mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha]\frac{\partial F_0}{\partial \varepsilon}. \end{aligned} \tag{153}$$

5.2.1 For special case $\delta\mathbf{A} \approx \delta A_\parallel \mathbf{e}_\parallel$

Most gyrokinetic simulations approximate the vector potential as $\delta\mathbf{A} \approx \delta A_\parallel \mathbf{e}_\parallel$. Let us simplify Eq. (152) for this case. Then $\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha$ is written as

$$\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \approx \langle v_\parallel \delta A_\parallel\rangle_\alpha. \tag{154}$$

Note that in terms of $(\mathbf{X}, \varepsilon, \mu, \alpha, \sigma)$ coordinates, $v_\parallel$ is written as

$$v_\parallel = \sigma\sqrt{2\varepsilon - 2\mu B_0}, \tag{155}$$

where $B_0(\mathbf{x}) = B_0(\mathbf{X} + \boldsymbol{\rho})$ with $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{X}, \varepsilon, \mu, \alpha)$. Since the scale length of B_0 is much larger than the thermal Larmor radius, $B_0(\mathbf{x}) \approx B_0(\mathbf{X})$ and hence $v_\parallel$ of thermal particles can be approximately considered to be independent of the gyro-angle α. Then $v_\parallel$ can be taken out of the gyro-averaging in expression (154), yielding

$$\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \approx v_\parallel \langle \delta A_\parallel\rangle_\alpha. \tag{156}$$

Using this, the term related to $\delta\mathbf{A}$ in (152) is written as

$$\begin{aligned}(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha &= (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X(v_\parallel\langle\delta A_\parallel\rangle_\alpha) \\ &= \langle\delta A_\parallel\rangle_\alpha(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X(v_\parallel) + v_\parallel(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot \\ &\quad \nabla_X\langle\delta A_\parallel\rangle_\alpha. \end{aligned} \tag{157}$$

Using expression (155), $(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X(v_\parallel)$ is written as

$$\begin{aligned}(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X(v_\parallel) &\approx (v_\parallel\mathbf{e}_\parallel)\cdot\nabla_X(v_\parallel) \\ &= (v_\parallel\mathbf{e}_\parallel)\cdot\nabla_X\big(\sigma\sqrt{2\varepsilon - 2\mu B_0}\big) \\ &= \sigma(v_\parallel\mathbf{e}_\parallel)\cdot\nabla_X\big(\sqrt{2\varepsilon - 2\mu B_0}\big) \\ &= \sigma v_\parallel\frac{-2\mu\mathbf{e}_\parallel\cdot\nabla_X B_0}{2\sqrt{2\varepsilon - 2\mu B_0}} \\ &= v_\parallel\frac{-2\mu\mathbf{e}_\parallel\cdot\nabla_X B_0}{2v_\parallel} \\ &= -\mu\mathbf{e}_\parallel\cdot\nabla_X B_0. \end{aligned} \tag{158}$$

(We can also obtain $\nabla_X(v_\parallel) = -\mu(\nabla B_0)/v_\parallel$ by using Eq. (284)). Using the above results, equation (152) is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h \\
&= -\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}[-(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X\langle\delta\Phi\rangle_\alpha]\frac{\partial F_0}{\partial\varepsilon}, \\
&-\frac{q}{m}[v_\parallel(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X\langle\delta A_\parallel\rangle_\alpha - \langle\delta A_\parallel\rangle_\alpha(\mu\mathbf{e}_\parallel\cdot\nabla B_0)]\frac{\partial F_0}{\partial\varepsilon},
\end{aligned} \tag{159}$$

which agrees with the so-called $p_\parallel$ formulation given in GEM code manual (the first line of Eq. 28), which uses $p_\parallel = v_\parallel + q\langle A_\parallel\rangle_\alpha/m$ as an independent variable.

5.3 Summary of split of the distribution function

In the above, the perturbed part of the distribution function, δF, is split at least three times in order to (1) simplify the gyrokinetic equation by splitting out the adiabatic response and (2) eliminate the time derivatives, $\partial\delta\phi/\partial t$ and (3) $\partial\delta\mathbf{A}/\partial t$, on the right-hand. To avoid confusion, I summarize the split of the distribution function here. The total distribution function F is split as

$$F = F_0 + \delta F, \tag{160}$$

where F_0 is the equilibrium distribution function and δF is the perturbed part of the total distribution function. δF is further split as

$$\delta F = \delta h + \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon} + \frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}, \tag{161}$$

where δh satisfies the gyrokinetic equation (152) or (159). In Eq. (161), the red term gives rise to the so-called polarization density (discussed in Sec. 5.6). The analytic dependence of this term on $\delta\Phi$ is utilized in solving the Poisson equation. The blue term also has an analytic dependence on $\delta\mathbf{A}$, which, however, will cause numerical problems in particle simulations if it is utilized in solving the Ampere equation (the so-called "cancellation problem" in gyrokinetic simulations).

5.3.1 Velocity space moment of $\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}$

Consider the approximation $\delta\mathbf{A}\approx\delta A_\parallel\mathbf{e}_\parallel$, then the blue term in Eq. (161) is written as

$$\frac{q}{m}\langle v_\parallel\delta A_\parallel\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}. \tag{162}$$

Notice that $v_\parallel$ can be taken out of the gyro-averaging. Then the above equation is written

$$\frac{q}{m}v_\parallel\langle\delta A_\parallel\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}. \tag{163}$$

If we neglect the FLR effect, then the above expression is written

$$\frac{q}{m}v_\parallel\delta A_\parallel\frac{\partial F_0}{\partial\varepsilon}. \tag{164}$$

The zeroth order moment (number density) is then written as

$$\delta n = \frac{q}{m}\delta A_{\parallel}\int v_{\parallel}\frac{\partial F_0}{\partial \varepsilon}d\mathbf{v}, \tag{165}$$

which is zero if F_0 is Maxwellian.

Next, consider the parallel current carried by distribution (164), which is written

$$\delta j_{\parallel} = \frac{q^2}{m}\delta A_{\parallel}\int v_{\parallel}^2\frac{\partial F_0}{\partial \varepsilon}d\mathbf{v}. \tag{166}$$

If F_0 is a Maxwellian distribution given by

$$F_0 = n_0\Big(\frac{m}{2\pi T}\Big)^{3/2}\exp\Big(-\frac{mv^2}{2T}\Big). \tag{167}$$

then

$$\frac{\partial F_0}{\partial \varepsilon} = -\frac{m}{T}F_0. \tag{168}$$

Then expression (166) is written

$$\delta j_{\parallel} = -\frac{q^2}{T}n_0\Big(\frac{m}{2\pi T}\Big)^{3/2}\delta A_{\parallel}\int v_{\parallel}^2\exp\Big(-\frac{mv^2}{2T}\Big)d\mathbf{v}. \tag{169}$$

Working in the spherical coordinates, then $v_{\parallel} = v\cos\theta$ and $d\mathbf{v} = v^2\sin\theta dv d\theta d\phi$. Then expression (169) is written

$$\begin{aligned}
\delta j_{\parallel} &= -\frac{q^2}{T}n_0\Big(\frac{m}{2\pi T}\Big)^{3/2}\delta A_{\parallel}\int v^2\cos^2\theta\exp\Big(-\frac{mv^2}{2T}\Big)v^2\sin\theta dv d\theta d\phi \\
&= -\frac{q^2}{T}n_0\Big(\frac{m}{2\pi T}\Big)^{3/2}\delta A_{\parallel}\frac{4\pi}{3}\int v^4\exp\Big(-\frac{mv^2}{2T}\Big)dv \qquad (170)\\
&= -\frac{q^2}{T}n_0\Big(\frac{1}{\pi}\Big)^{3/2}\frac{2T}{m}\delta A_{\parallel}\frac{4\pi}{3}\int x^4\exp(-x^2)dx \\
&= -\frac{q^2}{T}n_0\Big(\frac{1}{\pi}\Big)^{3/2}\frac{2T}{m}\delta A_{\parallel}\frac{4\pi}{3}\frac{3\sqrt{\pi}}{8} \\
&= -\frac{q^2}{m}n_0\delta A_{\parallel}. \qquad (171)
\end{aligned}$$

5.3.2 Ampere's Law

$$-\nabla_{\perp}^2\delta A_{\parallel}^{(n+1)} = \mu_0\,(\delta J_{||i}^{(n+1)} + \delta J_{||e}^{(n+1)}). \tag{172}$$

The parallel currents are given by

$$\delta J_{||i}^{(n+1)} = \delta J_{||i}'\big(\delta\phi^{(n)}, \delta A_{\parallel}^{(n)}\big) + \int\big(v_{\parallel}^{(n+1)}\big)^2\frac{q_i^2}{m_i}\big\langle\delta A_{\parallel}^{(n+1)}\big\rangle_{\alpha}\frac{\partial F_{i0}}{\partial\varepsilon}d\mathbf{v}, \tag{173}$$

$$\delta J_{||e} = \delta J_{||e}'\big(\delta\phi^{(n)}, \delta A_{\parallel}^{(n)}\big) + \int\big(v_{\parallel}^{(n+1)}\big)^2\frac{q_e^2}{m_e}\big\langle\delta A_{\parallel}^{(n+1)}\big\rangle_{\alpha}\frac{\partial F_{e0}}{\partial\varepsilon}d\mathbf{v}, \tag{174}$$

where $\delta J'_{\|i}$ and $\delta J'_{\|e}$ is the parallel current carried by the distribution function δh in Eq. (161), which are updated from the value at the nth time step to the $(n+1)$th time step using an explicit scheme and therefore does not depends on the field at the $(n+1)$th step. The blue terms in Eqs. (173) and (174) depend on the field at the $(n+1)$th step and thus need to be moved to the left-hand side of Ampere's law (172) if we want to solve this equation by direct methods. In this case, equation (172) is written as

$$\begin{aligned}
&-\nabla_\perp^2 \delta A_\|^{(n+1)} - \mu_0 \int \big(v_\|^{(n+1)}\big)^2 \frac{q_i^2}{m_i} \big\langle \delta A_\|^{(n+1)} \big\rangle_\alpha \frac{\partial F_{i0}}{\partial \varepsilon} d\mathbf{v} \\
&-\mu_0 \int \big(v_\|^{(n+1)}\big)^2 \frac{q_e^2}{m_e} \big\langle \delta A_\|^{(n+1)} \big\rangle_\alpha \frac{\partial F_{e0}}{\partial \varepsilon} d\mathbf{v}. \\
&= \mu_0 \left(\delta J'_{\|i}\big(\delta\phi^{(n)}, \delta A_\|^{(n)}\big) + \delta J'_{\|e}\big(\delta\phi^{(n)}, \delta A_\|^{(n)}\big)\right)
\end{aligned} \tag{175}$$

However the gyro-averaging in the blue terms can not be evaluated analytically. If evaluated numerically using markers, the corresponding matrix will depends on markers and thus needs to be re-constructed and inverted each time-step, which is computationally expensive. Therefore we go back to Eq. (172) and try to solve it using iterative methods. However, it is found numerically that directly using Eq. (172) as an iterative scheme is usually divergent. To obtain a convergent iterative scheme, we need to have an approximate form for the blue terms, which is independent of markers and so that it is easy to construct its matrix, and then subtract this approximate form from both sides. After doing this, the iterative scheme has better chance to be convergent. An approximate form is that derived by neglecting the FLR effect given in Sec. 5.3.1. Using this, the iterative scheme for solving Eq. (172) is written as

$$\begin{aligned}
&-\nabla_\perp^2 \delta A_\|^{(n+1)} - \mu_0 \left(-\frac{q_i^2}{m_i} n_{i0} \delta A_\|^{(n+1)} - \frac{q_e^2}{m_e} n_{e0} \delta A_\|^{(n+1)} \right) \\
&= \mu_0 \big[\delta J'_{\|i}\big(\delta\phi^{(n)}, \delta A_\|^{(n)}\big) + \delta J'_{\|e}\big(\delta\phi^{(n)}, \delta A_\|^{(n)}\big) \big] \\
&+\mu_0 \int \big(v_\|^{(n+1)}\big)^2 \frac{q_i^2}{m_i} \big\langle \delta A_\|^{(n+1)} \big\rangle_\alpha \frac{\partial F_{i0}}{\partial \varepsilon} d\mathbf{v} \\
&+\mu_0 \int \big(v_\|^{(n+1)}\big)^2 \frac{q_e^2}{m_e} \big\langle \delta A_\|^{(n+1)} \big\rangle_\alpha \frac{\partial F_{e0}}{\partial \varepsilon} d\mathbf{v} \\
&-\mu_0 \left(-\frac{q_i^2}{m_i} n_{i0} \delta A_\|^{(n+1)} - \frac{q_e^2}{m_e} n_{e0} \delta A_\|^{(n+1)} \right).
\end{aligned} \tag{176}$$

In the drift-kinetic limit (i.e., neglecting the FLR effect), the blue and red terms on the right-hand side of the above equation cancel each other exactly. Even in this case, it is found numerically that these terms need to be retained and the blue terms are evaluated using markers. Otherwise, numerical inaccuracy can give numerical instabilities, which is the so-called cancellation problem. The explanation for this is as follows. The blue terms are part of the current. The remained part of the current is computed by using Monte-Carlo integration over markers. If the blue terms are evaluated analytically, rather than using Monte-

Carlo integration over markers, then the cancellation between this analytical part and Monte-Carlo part can have large error (assume that there are two large contribution that have opposite signs in the two parts) because the two parts are evaluated using different methods and thus have different accuracy, which makes the cancellation less accurate.

The blue terms are sometimes called "adiabatic current". The red terms are approximation to the "adiabatic current" obtained by neglecting the FLR effect. Because the ion adiabatic current is less than the electron adiabatic current by a factor of m_e/m_i, its accuracy is not important, and is approximated by the drift-kinetic limit in `GEM`. And the cancellation error is not a problem and hence can be neglected. In this case, equation (176) is simplified as

$$\begin{aligned}
&-\nabla_\perp^2\delta A_\parallel^{(n+1)}-\mu_0\left(-\frac{q_i^2}{m_i}n_{i0}\delta A_\parallel^{(n+1)}-\frac{q_e^2}{m_e}n_{e0}\delta A_\parallel^{(n+1)}\right)\\
&=\mu_0\left[\delta J'_{\parallel i}\left(\delta\phi^{(n)},\delta A_\parallel^{(n)}\right)+\delta J'_{\parallel e}\left(\delta\phi^{(n)},\delta A_\parallel^{(n)}\right)\right]\\
&+\mu_0\int\left(v_\parallel^{(n+1)}\right)^2\frac{q_e^2}{m_e}\left\langle\delta A_\parallel^{(n+1)}\right\rangle_\alpha\frac{\partial F_{e0}}{\partial\varepsilon}d\mathbf{v}\\
&-\mu_0\left(-\frac{q_e^2}{m_e}n_{e0}\delta A_\parallel^{(n+1)}\right).
\end{aligned}\tag{177}$$

5.4 Generalized split-weight scheme for electrons

The perturbed distribution function is decomposed as given by Eq. (161), i.e.,

$$\delta F=\delta h+\frac{q}{m}(\delta\Phi-\langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon}+\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon},\tag{178}$$

where the term in blue is the so-called adiabatic response, which depends on the gyro-angle. Recall that the red term $\langle\delta\Phi\rangle_\alpha$, which is independent of the gyro-angle, is introduced in order to eliminate the time derivative $\partial\langle\delta\Phi\rangle_\alpha/\partial t$ term on the right-hand side of the original Frieman-Chen gyrokinetic equation.

The so-called generalized split-weight scheme corresponds to going back to the original Frieman-Chen gyrokinetic equation by introducing another $\langle\delta\Phi\rangle_\alpha$ term with a free small parameter ϵ_g. Specifically, δh in the above is split as

$$\delta h=\delta h_s+\epsilon_g\frac{q}{m}\langle\delta\Phi\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}.\tag{179}$$

(If $\varepsilon_g=1$, then the two $\langle\delta\Phi\rangle_\alpha$ terms cancel each other.) Substituting this expression into Eq. (153), we obtain the following equation for δh_s:

$$\begin{aligned}
&\left[\frac{\partial}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h_s\\
&+\epsilon_g\frac{q}{m}\frac{\partial F_0}{\partial\varepsilon}\left[\frac{\partial}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X\right]\langle\delta\Phi\rangle_\alpha\\
&+\epsilon_g\frac{q}{m}\langle\delta\Phi\rangle_\alpha\left[\frac{\partial}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X\right]\frac{\partial F_0}{\partial\varepsilon}
\end{aligned}$$

$$
\begin{aligned}
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0\\
&-\frac{q}{m}[(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X(\langle\mathbf{v}\cdot\delta\mathbf{A}-\delta\Phi\rangle_\alpha)]\frac{\partial F_0}{\partial\varepsilon}.
\end{aligned}
\tag{180}
$$

Noting that $\partial F_0/\partial t=0$, $\mathbf{e}_\parallel\cdot\nabla F_0=0$, $\nabla F_0\sim O(\lambda^1)F_0$, we find that the third line of the above equation is of order $O(\lambda^3)$ and thus can be dropped. Moving the second line to the right-hand side, the above equation is written as

$$
\begin{aligned}
&\left[\frac{\partial}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h_s\\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0\\
&-\frac{q}{m}\Bigg\{(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X[\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha-\langle\delta\Phi\rangle_\alpha]+\epsilon_g\Bigg[\frac{\partial\langle\delta\Phi\rangle_\alpha}{\partial t}+\mathbf{V}_G\cdot\\
&\nabla_X\langle\delta\Phi\rangle_\alpha\Bigg]\Bigg\}\frac{\partial F_0}{\partial\varepsilon}.
\end{aligned}
\tag{181}
$$

5.4.1 special case of $\epsilon_g=1$

For the special case of $\epsilon_g=1$ (the default and most used case in GEM code, Yang Chen said $\epsilon_g<1$ cases are sometimes not accurate, so he gave up using it since 2009), equation (181) can be simplified as:

$$
\begin{aligned}
&\left[\frac{\partial}{\partial t}+(v_\parallel\mathbf{e}_\parallel+\mathbf{V}_D+\delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h_s\\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0\\
&-\frac{q}{m}\left[\mathbf{V}_G\cdot\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha+\frac{\partial\langle\delta\Phi\rangle_\alpha}{\partial t}\right]\frac{\partial F_0}{\partial\varepsilon},
\end{aligned}
\tag{182}
$$

where two $\mathbf{V}_G\cdot\langle\delta\Phi\rangle_\alpha$ terms cancel each other. Because the $v_\parallel E_\parallel$ term is one of the factors that make kinetic electron simulations difficult, eliminating $\mathbf{V}_G\cdot\langle\delta\Phi\rangle_\alpha$ term may be beneficial for obtaining stable algorithms.

For $\epsilon_g=1$, δF is written as

$$
\begin{aligned}
\delta F &= \delta h_s+\varepsilon_g\frac{q}{m}\langle\delta\Phi\rangle_\alpha\frac{\partial F_{g0}}{\partial\varepsilon}+\frac{q}{m}(\delta\Phi-\langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon}+\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}\\
&= \delta h_s+\frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial\varepsilon}+\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon},
\end{aligned}
\tag{183}
$$

where the polarization density term is partially canceled, yielding only the pure adiabatic response. The adiabatic term will be moved to the left-hand side of the Poisson's equation. The descretization of this term is much easier than the polarization density. This term is already in `GEM`.

Equation (182) actually goes back to the original Frieman-Chen equation. The only difference is that $\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}$ is further split from the perturbed distribution function. Considering this, equation (182) can also be obtained from the original Frieman-Chen equation (143) by writing δG_0 as

$$
\delta G_0=\delta h_s+\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon},
\tag{184}
$$

In this case, δF is written as

$$\delta F = \delta h_s + \frac{q}{m}\delta\Phi\frac{\partial F_{g0}}{\partial \varepsilon} + \frac{q}{m}\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \frac{\partial F_0}{\partial \varepsilon}, \tag{185}$$

Substituting expression (184) into equation (143), we obtain the following equation for δh_s:

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h_s \\
&+\frac{q}{m}\frac{\partial F_0}{\partial \varepsilon}\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \\
&+\frac{q}{m}\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\frac{\partial F_0}{\partial \varepsilon} \\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0 - \frac{q}{m}\frac{\partial\langle\delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t}\frac{\partial F_0}{\partial \varepsilon},
\end{aligned} \tag{186}$$

Noting that $\partial F_0/\partial t = 0$, $\mathbf{e}_\parallel\cdot\nabla F_0 = 0$, $\nabla F_0 \sim O(\lambda^1)F_0$, we find that the third line of the above equation is of order $O(\lambda^3)$ and thus can be dropped. Moving the second line to the right-hand side, the above equation is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h_s \\
&=-\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
&-\frac{q}{m}\left[\frac{\partial\langle\delta\Phi\rangle_\alpha}{\partial t} + [(v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X]\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\right]\frac{\partial F_0}{\partial \varepsilon},
\end{aligned} \tag{187}$$

which agrees with Eq. (182).

In GEM, the split weight method is used only for electrons and the $\partial\langle\delta\Phi\rangle_\alpha/\partial t$ is approximated by $\partial\delta\Phi/\partial t$ and this term is obtained from the vorticity equation (rather than from an implicit iteration).

5.5 Comments on how to split the distribution function

In particle simulations, the seemingly trivial thing on how to split the distribution function is often considered to be a big deal. Separating the perturbed part from the equilibrium part is considered to be a big deal and got the famous name "δf particle method", in contrast to the conventional particle method which is now called full-f particle method. Summarizing the above result, we know that the total distribution function F is split in the following form:

$$\begin{aligned}
F &= F_0 + \delta F \\
&= F_0 + \delta h + \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial \varepsilon} + \frac{q}{m}\langle \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial \varepsilon},
\end{aligned} \tag{188}$$

and only δh is actually evolved by using markers and its moment in the phase-space is evaluated via Monte-Carlo integration. The blue and red terms in the above expression explicitly depends on the perturbed field. The velocity integrations of these two terms can be performed analytically. However, in some cases, the phase space integration of the blue terms must be evaluated using markers, i.e., using Monte-Carlo method, to avoid the inaccurate cancellation between the integration of these parts and the integration of δh (the latter is computed using Monte-Carlo method). When will the inaccurate cancellation is significant depends on the problem being investigated and thus can only be determined by actual numerical experiments. Many electromagnetic particle simulation experiments indicate that the parallel current carried by the blue term must be evaluated via Monte-Carlo method, otherwise inaccurate cancellation between this term and δh will give rise to numerical instabilities.

5.6 Poisson's equation and polarization density

Poisson's equation is written as

$$-\varepsilon_0\nabla^2\delta\Phi = q_i\delta n_i + q_e\delta n_e, \tag{189}$$

where $-\varepsilon_0\nabla^2\delta\Phi$ is called the space-charge term. Since we consider modes with $k_\parallel \ll k_\perp$, the space-charge term is approximated as $\nabla^2\delta\Phi \equiv \nabla_\perp^2\delta\Phi + \nabla_\parallel^2\delta\Phi \approx \nabla_\perp^2\delta\Phi$. Then Eq. (189) is written as

$$-\varepsilon_0\nabla_\perp^2\delta\Phi = q_i\delta n_i + q_e\delta n_e. \tag{190}$$

This approximation eliminates the parallel plasma oscillation from the system. The perpendicular plasma oscillations seem to be partially eliminated in system with gyrokinetic ions and drift-kinetic electrons. There are the so-called Ω_H modes (also called electrostatic shear Alfven wave) that appear in the gyrokinetic system which have some similarity with the plasma oscillations but with a much smaller frequency, $\Omega_H \sim (k_\parallel/k_\perp)(\lambda_D/\rho_s)\omega_{pe}$.

Using expression (161), the perturbed number density δn is written as

$$\begin{aligned}\delta n &= \int \delta F d\mathbf{v} \\ &= \int \delta h\, d\mathbf{v} + \int\left[\frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon}\right]d\mathbf{v} + \int\left[\frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}\right]d\mathbf{v},\end{aligned} \tag{191}$$

where the blue term is approximately zero for isotropic F_0 and this term is usually dropped in simulations that assume isotropic F_0 and approximate $\delta\mathbf{A}$ as $\delta A_\parallel\mathbf{e}_\parallel$. The red term in expression (191) is the so-called the polarization density n_p, i.e.,

$$\delta n_p(\mathbf{x}) = \int \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v}, \tag{192}$$

which has an explicit dependence on $\delta\Phi$ and is usually moved to the left hand of Poisson's equation when constructing the numerical solver of the Poisson equation, i.e., equation (190) is written as

$$-\varepsilon_0\nabla_\perp^2\delta\Phi - q_i\int\frac{q_i}{m_i}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_{i0}}{\partial\varepsilon}d\mathbf{v} = q_i\delta n_i' + q_e\delta n_e, \tag{193}$$

where $\delta n_i' = \delta n_i - \delta n_{pi} = \int \delta h_i \, d\mathbf{v}$, which is evaluated by using Monte-Carlo markers. Since some parts depending on $\delta\Phi$ are moved from the right-hand side to the left-hand side of the field equation, numerical solvers (for $\delta\Phi$) based on the left-hand side of Eq. (193) probably behaves better than the one that is based on the left-hand side of Eq. (190), i.e., $-\varepsilon_0\nabla_\perp^2\delta\Phi$.

The polarization density is part of the perturbed density that is extracted from the source term and moved to the left-hand side of the Poisson equation. The polarization density will be evaluated without using Monte-Carlo markers, whereas the remained density on the right-hand side will be evaluated using Mote-Carlo markers. The two different methods of evaluating two parts of the total perturbed density can possibly introduce significant errors if the two terms are expected to cancel each other and give a small quantity that is much smaller than either of the two terms. This is one pitfall for PIC simulations that extract some parts from the source term and move them to the left-hand side. To remedy this, rather than directly moving a part of the distribution function to the left-hand side, we subtract an analytic expression from both sides of Eq. (190). The analytical expression is evaluated based on grid values of perturbed electromagnetic fields and are independent of makers. All the original parts of the distribution functions are kept on the right-hand side and are still evaluated by using markers, which hopefully avoids the possible cancellation problem. This strategy is often called a cancellation scheme.

It turns out the cancellation scheme is not necessary for Eq. (193), but for the field solver for Ampere's equation (discussed later), this cancellation scheme is necessary in order to obtain stable results.

Note that two things appear here: What motivates us to move parts of the distribution function to the left==>hopefully to make the left-hand side matrix more well-behaved (such as good condition number, etc.) Why do we need the cancellation scheme==>to avoid numerical accuracy problem appearing when large terms cancels each other==>iteration is needed when the cancellation scheme is used because the right-hand side explicitly contains unknown electromagnetic fields.

5.7 Polarization density with the velocity integration performed

Since $\delta\Phi$ is independent of the velocity in the particle coordinates, the first term (adiabatic term) in expression (192) is trivial and the velocity integration can

be readily performed (assume F_0 is Maxwellian), giving

$$\begin{aligned}\delta n_{\rm ad} &= \int \frac{q}{m}(\delta\Phi)\frac{\partial F_0}{\partial \varepsilon}d\mathbf{v}.\\ &= \frac{q}{m}(\delta\Phi)\int\Big(-\frac{m}{T}f_M\Big)d\mathbf{v}.\\ &= -\frac{q\delta\Phi}{T}n_0,\end{aligned} \tag{194}$$

which is called adiabatic response. Next, let us perform the gyro-averaging and the velocity integration of the second term in expression (192), i.e.,

$$-\int \frac{q}{m}\langle\delta\Phi\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v}, \tag{195}$$

5.7.1 Gyro-averaging of $\delta\Phi$ in guiding-center coordinates

In order to perform the gyro-averaging of $\delta\Phi$, we expand $\delta\Phi$ in wave-number space as

$$\delta\Phi(\mathbf{x}) = \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x})\frac{d\mathbf{k}}{(2\pi)^3}, \tag{196}$$

and then express $\mathbf{x}$ in terms of the guiding center variables $(\mathbf{X},\mathbf{v})$ since the gyro-averaging is taken by holding $\mathbf{X}$ rather than $\mathbf{x}$ constant. The guiding-center transformation gives

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\rho}(\mathbf{x},\mathbf{v}) \approx \mathbf{X} - \mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}. \tag{197}$$

Using expressions (196) and (197), the gyro-average of $\delta\Phi$ is written as

$$\begin{aligned}\langle\delta\Phi\rangle_\alpha &= \left\langle\int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})\frac{d\mathbf{k}}{(2\pi)^3}\right\rangle_\alpha\\ &= \left\langle\int\delta\Phi_k\exp\left(i\mathbf{k}\cdot\left(\mathbf{X}-\mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\right)\right)\frac{d\mathbf{k}}{(2\pi)^3}\right\rangle_\alpha\\ &= \int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{X})\left\langle\exp\left(-i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\right)\right\rangle_\alpha\frac{d\mathbf{k}}{(2\pi)^3}.\end{aligned} \tag{198}$$

When doing the gyro-averaging, $\mathbf{X}$ is hold constant and thus $\mathbf{e}_\parallel(\mathbf{X})$ is also constant. Then it is straightforward to define the gyro-angle α. Let $\mathbf{k}_\perp$ define one of the perpendicular direction $\hat{\mathbf{e}}_1$, i.e., $\mathbf{k}_\perp = k_\perp\hat{\mathbf{e}}_1$. Then another perpendicular basis vector is defined by $\hat{\mathbf{e}}_2 = \mathbf{e}_\parallel\times\hat{\mathbf{e}}_1$. Then $\mathbf{v}_\perp$ is written as $\mathbf{v}_\perp = v_\perp(\hat{\mathbf{e}}_1\cos\alpha + \hat{\mathbf{e}}_2\sin\alpha)$, which defines the gyro-angle α. Then the blue expression in Eq. (198) is written as

$$\begin{aligned}-i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})} &= -i\mathbf{k}\cdot v_\perp(\hat{\mathbf{e}}_1\cos\alpha+\hat{\mathbf{e}}_2\sin\alpha)\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\\ &= -i\mathbf{k}\cdot\frac{v_\perp}{\Omega(\mathbf{X})}(-\hat{\mathbf{e}}_2\cos\alpha+\hat{\mathbf{e}}_1\sin\alpha)\\ &= -i\frac{k_\perp v_\perp}{\Omega}\sin\alpha.\end{aligned} \tag{199}$$

Then the gyro-averaging in expression (198) is written as

$$\begin{aligned}\left\langle \exp\left(-i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}\right)\right\rangle_\alpha &= \left\langle \exp\left(-i\frac{k_\perp v_\perp}{\Omega}\sin\alpha\right)\right\rangle_\alpha \\ &= \frac{1}{2\pi}\int_0^{2\pi}\exp\left(-i\frac{k_\perp v_\perp}{\Omega}\sin\alpha\right)d\alpha \\ &= J_0\left(\frac{k_\perp v_\perp}{\Omega}\right). \end{aligned} \tag{200}$$

where use has been made of the definition of the zeroth Bessel function of the first kind. Then $\langle\delta\Phi\rangle_\alpha$ in expression (198) is written as

$$\langle\delta\Phi\rangle_\alpha = \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{X}) J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\frac{d\mathbf{k}}{(2\pi)^3}. \tag{201}$$

5.7.2 Gryo-angle integration in particle coordinates

Next, we need to perform the integration in velocity space, which is done by holding $\mathbf{x}$ (rather than $\mathbf{X}$) constant. Therefore, it is convenient to transform back to particle coordinates. Using $\mathbf{X} = \mathbf{x} + \mathbf{v}\times\frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})}$, expression (201) is written as

$$\langle\delta\Phi\rangle_\alpha = \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x}) J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\exp\left(i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel}{\Omega}\right)\frac{d\mathbf{k}}{(2\pi)^3}. \tag{202}$$

Then the velocity integration is written as

$$\begin{aligned}\int \langle\delta\Phi\rangle_\alpha \frac{\partial F_0}{\partial\varepsilon} d\mathbf{v} = \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x})\Bigg[\int J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\exp\Big(i\mathbf{k}\cdot\mathbf{v}\times \\ \frac{\mathbf{e}_\parallel}{\Omega}\Big)\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v}\Bigg]\frac{d\mathbf{k}}{(2\pi)^3}. \end{aligned} \tag{203}$$

Similar to Eq. (199), except for now at $\mathbf{x}$ rather than $\mathbf{X}$, $i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel}{\Omega}$ is written as

$$i\mathbf{k}\cdot\mathbf{v}\times\frac{\mathbf{e}_\parallel}{\Omega} = i\frac{k_\perp v_\perp}{\Omega}\sin\alpha. \tag{204}$$

Since this is at $\mathbf{x}$ rather than $\mathbf{X}$, $k_\perp$, $v_\perp$, and Ω are different from those appearing in expression (199). However, since this difference is due to the variation of the equilibrium quantity $\mathbf{e}_\parallel/\Omega$ in a Larmor radius, and thus is small and is ignored in the following.

Plugging expression (204) into expression (203) and using $d\mathbf{v} = v_\perp dv_\perp dv_\parallel d\alpha$, we get

$$\begin{aligned}\int \langle\delta\Phi\rangle_\alpha \frac{\partial F_0}{\partial\varepsilon} d\mathbf{v} = \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x})\Bigg[\int \\ J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\exp\left(i\frac{k_\perp v_\perp}{\Omega}\sin\alpha\right)\frac{\partial F_0}{\partial\varepsilon}v_\perp dv_\perp dv_\parallel d\alpha\Bigg]\frac{d\mathbf{k}}{(2\pi)^3}. \end{aligned} \tag{205}$$

Note that $\partial F_0/\partial\varepsilon$ is independent of the gyro-angle α in terms of guiding-center variables. When transformed back to particle coordinates, $\mathbf{X}$ contained in $\partial F_0/\partial\varepsilon$ will introduce α dependence via $\mathbf{X}=\mathbf{x}+\mathbf{v}\times\frac{\mathbf{e}_\parallel}{\Omega}$. This dependence on α is weak since the equilibrium quantities can be considered constant over a Larmor radius distance evaluated at the thermal velocity. Therefore this dependence can be ignored when performing the integration over α, i.e., in terms of particle coordinates, $\partial F_0/\partial\varepsilon$ is approximately independent of the gyro-angle α. Then the integration over α in Eq. (205) can be performed, yielding

$$\begin{aligned}\int\langle\delta\Phi\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v} &= \int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})\iint J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\left[\int_0^{2\pi}\exp\left(i\frac{k_\perp v_\perp}{\Omega}\sin\alpha\right)d\alpha\right]\frac{\partial F_0}{\partial\varepsilon}v_\perp dv_\perp dv_\parallel\frac{d\mathbf{k}}{(2\pi)^3}\\ &= \int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})\left[\iint J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)2\pi J_0\left(\frac{k_\perp v_\perp}{\Omega}\right)\frac{\partial F_0}{\partial\varepsilon}v_\perp dv_\perp dv_\parallel\right]\frac{d\mathbf{k}}{(2\pi)^3},\end{aligned}\tag{206}$$

where again use has been made of the definition of the Bessel function.

5.7.3 The remaining velocity integration can be performed analytically if F_0 is Maxwellian

In order to perform the remaining velocity integration in expression (206), we assume that F_0 is a Maxwellian distribution given by

$$F_0 = f_M=\frac{n_0(\mathbf{X})}{(2\pi T(\mathbf{X})/m)^{3/2}}\exp\left(\frac{-mv^2}{2T(\mathbf{X})}\right)\tag{207}$$

$$=\frac{n_0}{(2\pi)^{3/2}v_t^3}\exp\left(\frac{-v^2}{2v_t^2}\right),\tag{208}$$

where $v_t=\sqrt{T/m}$, then

$$\frac{\partial F_0}{\partial\varepsilon}=-\frac{m}{T}f_M.\tag{209}$$

Again we will ignore the weak dependence of $n_0(\mathbf{X})$ and $T(\mathbf{X})$ on $\mathbf{v}$ introduced by $\mathbf{X}=\mathbf{x}+\mathbf{v}\times\mathbf{e}_\parallel/\Omega$ when transformed back to particle coordinates. (For sufficiently large velocity, the corresponding Larmor radius will be large enough to make the equilibrium undergo substantial variation. Since the velocity integration limit is to infinite, this will definitely occur. However, F_0 is exponentially decreasing with velocity, making those particles with velocity much larger than the thermal velocity negligibly few and thus can be neglected.)

Parallel integration

Using Eq. (209), the expression in the square brackets of Eq. (206) is written as

$$
\begin{aligned}
&2\pi\iint J_0^2\left(\frac{k_\perp v_\perp}{\Omega}\right)\frac{\partial F_0}{\partial \varepsilon}v_\perp dv_\perp dv_\parallel \\
&=-\frac{m}{T}\frac{n_0}{(2\pi)^{1/2}}\iint J_0^2\left(\frac{k_\perp v_\perp}{\Omega}\right)\frac{1}{v_t^3}\exp\left(-\frac{v_\parallel^2+v_\perp^2}{2v_t^2}\right)v_\perp dv_\perp dv_\parallel \qquad (210)\\
&=-\frac{m}{T}\frac{n_0}{(2\pi)^{1/2}}\iint J_0^2\left(\frac{k_\perp v_\perp}{\Omega}\right)\exp\left(-\frac{\bar{v}_\parallel^2+\bar{v}_\perp^2}{2}\right)\bar{v}_\perp d\,\bar{v}_\perp d\,\bar{v}_\parallel, \qquad (211)
\end{aligned}
$$

where $\bar{v}_\parallel = v_\parallel / v_t$, $\bar{v}_\perp = v_\perp / v_t$. Using

$$
\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx = \sqrt{2\pi}, \qquad (212)
$$

the integration over $\bar{v}_\parallel$ in expression (211) can be performed, yielding

$$
-\frac{m}{T}n_0\int_0^\infty J_0^2\left(\frac{k_\perp v_\perp}{\Omega}\right)\exp\left(-\frac{\bar{v}_\perp^2}{2}\right)\bar{v}_\perp\, d\,\bar{v}_\perp \qquad (213)
$$

Perpendicular integration

Using (I verified this by using Sympy)

$$
\int_0^\infty J_0^2(a\,x)\exp\left(-\frac{x^2}{2}\right)x\,dx = \exp(-a^2)I_0(a^2), \qquad (214)
$$

where $I_0(a)$ is the zeroth modified Bessel function of the first kind, expression (213) is written

$$
-\frac{m}{T}n_0\exp(-b)I_0(b) \qquad (215)
$$

where $b = k_\perp^2 v_t^2/\Omega^2 = k_\perp^2\rho_t^2$. Then the corresponding density (195) is written as

$$
-\frac{q}{m}\int\langle\delta\Phi\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v} = \frac{qn_0}{T}\int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})\exp(-b)I_0(b)\frac{d\mathbf{k}}{(2\pi)^3}. \qquad (216)
$$

5.7.4 Final form of polarization density

In Fourier space, the adiabatic term in expression (194) is written as

$$
\int\frac{q}{m}(\delta\Phi)\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v} = -\frac{qn_0}{T}\delta\Phi = -\frac{qn_0}{T}\int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})\frac{d\mathbf{k}}{(2\pi)^3}. \qquad (217)
$$

Plugging expression (216) and (217) into expression (192), the polarization density n_p is written as

$$
n_p \;=\; -\frac{qn_0}{T}\int\delta\Phi_k\exp(i\mathbf{k}\cdot\mathbf{x})[1-\exp(-b)I_0(b)]\frac{d\mathbf{k}}{(2\pi)^3}. \qquad (218)
$$

Define

$$\Gamma_0 = \exp(-b) I_0(b), \tag{219}$$

then Eq. (218) is written as

$$n_p = -\frac{q n_0}{T} \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x})[1-\Gamma_0]\frac{d\mathbf{k}}{(2\pi)^3}, \tag{220}$$

Expression (220) agrees with the result given in Yang Chen's notes. Note that the dependence on species mass enters the formula through the Larmor radius ρ_t in Γ_0.

5.8 Pade approximation

Γ_0 defined in Eq. (219) can be approximated by the Pade approximation as

$$\Gamma_0 \approx \frac{1}{1+b}. \tag{221}$$

The comparison between the exact value of Γ_0 and the above Pade approximation is shown in Fig. 1.

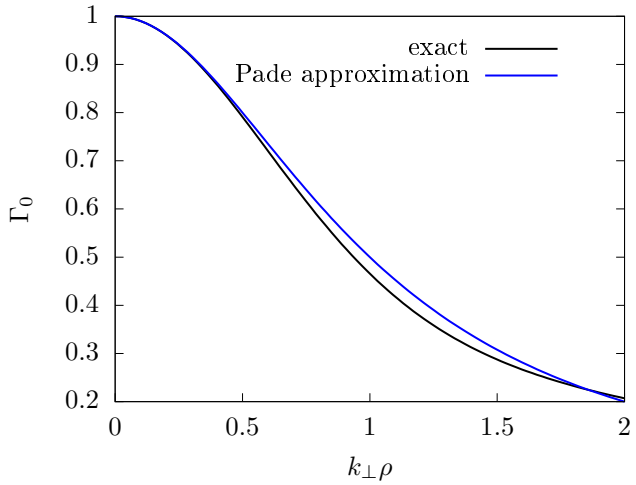

Figure 1. Comparison between the exact value of $\Gamma_0 = \exp(-(k_\perp\rho)^2) I_0((k_\perp\rho)^2)$ and the corresponding Pade approximation $1/(1+(k_\perp\rho)^2)$.

Using the Pade approximation (221), the polarization density n_p in expression (220) can be written as

$$n_p \approx -\frac{q n_0}{T} \int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x}) \frac{k_\perp^2\rho^2}{1+k_\perp^2\rho^2} \frac{d\mathbf{k}}{(2\pi)^3}. \tag{222}$$

(Padé approximant is the "best" approximation of a function by a rational function of given order – under this technique, the approximant's power series agrees with the power series of the function it is approximating.)

5.8.1 Long wavelength approximation of the polarization density

In the long wavelength limit, $k_\perp \rho \ll 1$, expression (222) can be further approximated as

$$\begin{aligned} n_p &\approx -\frac{qn_0}{T}\int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x}) k_\perp^2 \rho^2 \frac{d\mathbf{k}}{(2\pi)^3}, \\ &= \frac{qn_0}{T}\rho^2 \nabla_\perp^2 \delta\Phi. \end{aligned} \tag{223}$$

Then the corresponding term in the Poisson equation is written as

$$\begin{aligned} \frac{q}{\varepsilon_0} n_p &= \frac{q^2 n_0}{\varepsilon_0 T}\rho^2 \nabla_\perp^2 \delta\Phi \\ &= \frac{\rho^2}{\lambda_D^2}\nabla_\perp^2 \delta\Phi, \end{aligned} \tag{224}$$

where λ_D is the Debye length defined by $\lambda_D^2 = T\varepsilon_0/(n_0 q^2)$. For typical tokamak plasmas, the thermal ion gyroradius ρ_i is much larger than λ_D. Therefore the term in expression (224) for ions is much larger than the space charge term $\nabla^2\delta\Phi \equiv \nabla_\perp^2\delta\Phi + \nabla_\parallel^2\delta\Phi \approx \nabla_\perp^2\delta\Phi$ in the Poisson equation. Therefore the space charge term can be neglected in the long wavelenght limit.

Equation (224) also shows that electron polariztion density is smaller than the ion polariztion density by a factor of $\rho_e/\rho_i \approx 1/60$. Note that this conclusion is drawn in the long wavelenth limit. For short wavelength, the electron polariztion and ion polariztion density can be of similar magnitude (to be discussed later).

5.8.2 Polarization density expressed in terms of Laplacian operator

The polarization density expression (223) is for the long wavelength limit, which partially neglects FLR effect. Let us go back to the more general expression (222). The Poisson equation is written

$$-\varepsilon_0 \nabla_\perp^2 \delta\Phi = q_i \delta n_i + q_e \delta n_e. \tag{225}$$

Write $\delta n_i = n_{pi} + \delta n_i'$, where δn_{pi} is the ion polarization density, then the above expression is written

$$-\varepsilon_0 \nabla_\perp^2 \delta\Phi - q_i n_{pi} = q_i \delta n_i' + q_e \delta n_e. \tag{226}$$

Fourier transforming in space, the above equation is written

$$-\varepsilon_0 k_\perp^2 \delta\hat{\Phi} - q_i \hat{n}_{pi} = q_i \delta\hat{n}_i' + q_e \delta\hat{n}_e, \tag{227}$$

where $\hat{n}_{pi}$ is the Fourier transformation (in space) of the polarization density n_{pi} and similar meanings for $\delta\hat{\Phi}$, $\delta\hat{n}_i'$, and $\delta\hat{n}_e$. Expression (222) implies that $\hat{n}_{pi}$ is given by

$$\hat{n}_{pi} = -\frac{q_i n_{i0}}{T_i}\delta\hat{\Phi}\frac{k_\perp^2\rho_i^2}{1+k_\perp^2\rho_i^2}. \tag{228}$$

Using this, equation (227) is written

$$-\varepsilon_0 k_\perp^2 \delta\hat{\Phi} - q_i\left(-\frac{q_i n_{i0}}{T_i}\frac{k_\perp^2\rho_i^2}{1+k_\perp^2\rho_i^2}\delta\hat{\Phi}\right) = q_i\delta\hat{n}_i' + q_e\delta\hat{n}_e, \tag{229}$$

Multiplying both sides by $(1+k_\perp^2\rho_i^2)/\varepsilon_0$, the above equation is written

$$-(1+k_\perp^2\rho_i^2)k_\perp^2\delta\hat{\Phi} - \frac{q_i}{\varepsilon_0}\left(-\frac{q_i n_{i0}}{T_i}(k_\perp^2\rho_i^2)\delta\hat{\Phi}\right) = \frac{1}{\varepsilon_0}(1+k_\perp^2\rho_i^2)(q_i\delta\hat{n}_i' + q_e\delta\hat{n}_e). \tag{230}$$

Next, transforming the above equation back to the real space, we obtain

$$-(1-\rho_i^2\nabla_\perp^2)\nabla_\perp^2\delta\Phi - \frac{q_i}{\varepsilon_0}\left(\frac{q_i n_{i0}}{T_i}\rho_i^2\nabla_\perp^2\delta\Phi\right) = \frac{1}{\varepsilon_0}(1-\rho_i^2\nabla_\perp^2)(q_i\delta n_i' + q_e\delta n_e). \tag{231}$$

Neglecting the Debye shielding term, the above equation is written

$$-\left(\frac{\rho_i^2}{\lambda_{Di}^2}\nabla_\perp^2\delta\Phi\right) = \frac{1}{\varepsilon_0}(1-\rho_i^2\nabla_\perp^2)(q_i\delta n_i' + q_e\delta n_e), \tag{232}$$

which is the equation actually solved in many gyrokinetic codes, where $\lambda_{Di}^2 = \varepsilon_0 T_i/(q_i^2 n_{i0})$.

6 Summary

Equations (152) and (220) are the primary results derived in this note. For reference ease, let us summarize the results. The total distribution function F is split as

$$F = F_0 + \delta F, \tag{233}$$

where F_0 and δF are the equilibrium part and perturbed part of the distribution function, respectively. δF is further split as

$$\delta F = \delta h + \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon} + \frac{q}{m}\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\frac{\partial F_0}{\partial\varepsilon}, \tag{234}$$

where $\delta h = \delta h(\mathbf{X}, \mu, \varepsilon, t)$ satisfies the following nonlinear gyrokinetic equation (152), i.e.,

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\nabla_X\right]\delta h \\
=\ & -\delta\mathbf{V}_D\cdot\nabla_X F_0 \\
-\ & \frac{q}{m}[(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D)\cdot\nabla_X(\langle\mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha)]\frac{\partial F_0}{\partial\varepsilon}.
\end{aligned} \tag{235}$$

where $\mathbf{V}_D$ is the equilibrium guiding-center drift velocity and $\delta\mathbf{V}_D$ is the perturbed drift velocity given by (132), i.e.,

$$\delta\mathbf{V}_D = \frac{q}{m}\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A} - \delta\Phi\rangle_\alpha \times \frac{\mathbf{e}_\parallel}{\Omega}. \tag{236}$$

Here the independent variables for the distribution functions F_0 and δh are $(\mathbf{X}, \mu, \varepsilon)$ with $\mathbf{X}$ being the guiding-center position, $\mu = v_\perp^2/(2B_0)$, and $\varepsilon = v^2/2$. In the

above, $\delta\Phi$ and $\delta\mathbf{A}$ is the perturbed electric potential and vector potential, $\langle ...\rangle_\alpha$ is the gyro-averaging operator over the gyro-phase α, $\mathbf{e}_\parallel = \mathbf{B}_0/B_0$, $\Omega = B_0 q/m$ with q and m being the particle charge and mass, respectively.

When integrated in velocity space, the red term in expression (234) gives rise to the polarization density n_p, i.e.,

$$n_p = \int \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_0}{\partial\varepsilon}d\mathbf{v}, \tag{237}$$

which, for Maxwellian equilibrium distribution, can be written in wave-number space as

$$n_p = -\frac{qn_0}{T}\int \delta\Phi_k \exp(i\mathbf{k}\cdot\mathbf{x})[1-\Gamma_0]\frac{d\mathbf{k}}{(2\pi)^3}, \tag{238}$$

where $\Gamma_0 = \exp(-b)I_0(b)$, $b = k_\perp^2 v_t^2/\Omega^2$, and $I_0(b)$ is the zeroth modified Bessel function of the first kind. This expression is useful for gyrokinetic simulations that use spectral methods in solving the Poisson equation.

A Poloarition density expression using $\delta\Phi$ at gridpoints

In Sec. 5.7, when evaluating the polarization density, the potential $\delta\Phi$ is Fourier expanded in space and then the double gyro-angle integration of each harmonic is expressed as the Bessel function. Are these formulas scaring to beginners? Does this make gyrokinetic theory less accessible to general readers? I think so.

In this section, we avoid using the Fourier expansion at the first place, and directly express the double gyro-angle integral as linear combination of values of $\delta\Phi$ at spatial gridpoints.

The integral we try to evaluate is given by

$$A(\mathbf{x}) = -\int \frac{q}{m}\langle\delta\Phi(\mathbf{x})\rangle_\alpha \frac{\partial F_{i0}}{\partial\varepsilon}d\mathbf{v}, \tag{239}$$

Using $d\mathbf{v} = v_\perp dv_\perp dv_\parallel d\alpha$ and the definition of gyro-averaging, the above integral is written as

$$A(\mathbf{x}) = -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^\infty dv_\perp \frac{\partial F_{i0}}{\partial\varepsilon} v_\perp \int_0^{2\pi} d\alpha \left(\frac{1}{2\pi}\int_0^{2\pi}\delta\Phi(\mathbf{x})d\alpha\right) \tag{240}$$

Note that the gyro-averaging is performed in the guiding-center space and is performed by keeping the value of the guiding-center variable $\mathbf{X}$ constant. So we need to transform $\mathbf{x}$ to $\mathbf{X}$. Using the relation between particle location and

guiding-center location

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\rho}(\mathbf{x}, \mathbf{v}) \approx \mathbf{X} - \mathbf{v} \times \frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})}, \tag{241}$$

$A(\mathbf{x})$ is written as

$$\begin{aligned} A(\mathbf{x}) &= -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^{\infty} dv_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \int_0^{2\pi} d\alpha \left[\frac{1}{2\pi}\int_0^{2\pi} \delta\Phi\left(\mathbf{X} - \mathbf{v} \times \frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})} \right) d\alpha \right] \\ &= -\frac{q}{m}\int_{-\infty}^{\infty} d\, v_\parallel \int_0^{\infty} d\, v_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \int_0^{2\pi} d\alpha \left[\frac{1}{N_2}\sum_{j=1}^{N_2} \delta\Phi\left(\mathbf{X} - \mathbf{v}_{\perp j} \times \frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})} \right) \right] \end{aligned} \tag{242}$$

where the summation is over the gyro-angle of $\mathbf{v}$, and $\mathbf{v}_{\perp j} = \mathbf{v}(v_\perp, \alpha_j)$ denotes a perpendicular velocity corresponding to a discrete gyro-angle.

Next, in order to perform the remaining velocity space integration, transform back to the particle coordinates (because the velocity integration is performed in the particle coordinaes and is performed by keeping the particle coordinate $\mathbf{x}$ constant):

$$\begin{aligned} A(\mathbf{x}) &= -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^{\infty} dv_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \int_0^{2\pi} d\alpha \left[\frac{1}{N_2}\sum_{j=1}^{N_2} \delta\Phi\left(\mathbf{x} + \mathbf{v} \times \frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})} - \mathbf{v}_{\perp j} \times \frac{\mathbf{e}_\parallel(\mathbf{X})}{\Omega(\mathbf{X})} \right) \right] \\ &\approx -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^{\infty} dv_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \int_0^{2\pi} d\alpha \left[\frac{1}{N_2}\sum_{j=1}^{N_2} \delta\Phi\left(\mathbf{x} + \mathbf{v} \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} - \mathbf{v}_{\perp j} \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} \right) \right] \\ &= -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^{\infty} dv_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \\ &\times \frac{2\pi}{N_1}\sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \delta\Phi\left(\mathbf{x} + \mathbf{v}_\perp(v_\perp, \alpha_i) \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} - \mathbf{v}_\perp(v_\perp, \alpha_j) \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} \right). \end{aligned} \tag{243}$$

For notation ease, define

$$\Delta\boldsymbol{\rho}_{ij}(\mathbf{x}, v_\perp, \alpha_i, \alpha_j) = \mathbf{v}_\perp(v_\perp, \alpha_i) \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})} - \mathbf{v}_\perp(v_\perp, \alpha_j) \times \frac{\mathbf{e}_\parallel(\mathbf{x})}{\Omega(\mathbf{x})}, \tag{244}$$

where $\Delta\boldsymbol{\rho}_{ij}$ is a function of $(\mathbf{x}, v_\perp, \alpha_i, \alpha_j)$. Then Eq. (243) is written as

$$A(\mathbf{x}) \;=\; -\frac{q}{m}\int_{-\infty}^{\infty} dv_\parallel \int_0^\infty dv_\perp \frac{\partial F_{i0}}{\partial \varepsilon} v_\perp \frac{2\pi}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\delta\Phi(\mathbf{x}+\Delta\boldsymbol{\rho}_{ij}). \tag{245}$$

The guiding-center transform and the inverse one involved in the above is illustrated in Fig. 2, which also shows how to evaluate the double gyro-angle integration using the discrete values of $\delta\Phi_{ij}$.

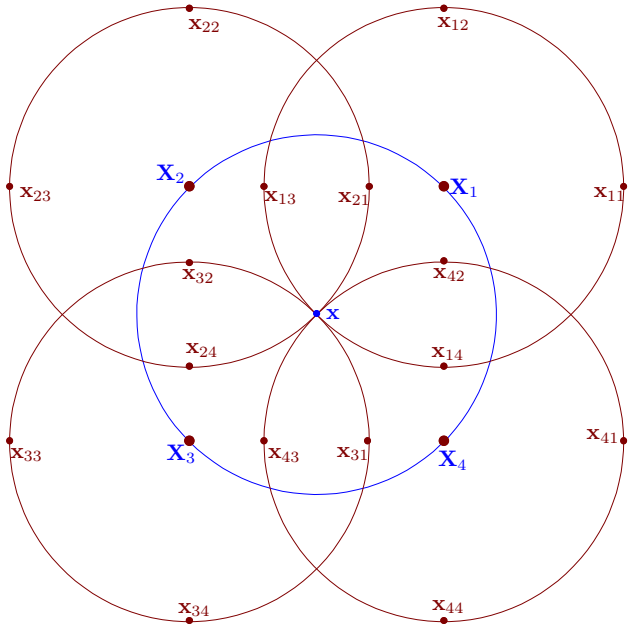

Figure 2. Given a $v_\perp$, then the double gyro-angle integration (240) at a particle location $\mathbf{x}$ is evaluated by the following steps: Use the guiding center transform to get guiding-center locations (four locations are shown in this case, namely $\mathbf{X}_i$ with $i=1,2,3,4$, corresponding gyro-angle α_i with $i=1,2,3,4$; 2.) For each guiding-center location, use the inverse guiding-center transform to calculate points on the corresponding gyro-ring (four points are shown for each guiding-center $\mathbf{X}_i$ in this case, namely $\mathbf{x}_{ij}$ with $j=1,2,3,4$, corresponding gyro angle α_j with $j=1,2,3,4$.) Then the double gyro-angle integral appearing in the polarization density, namely $\int_0^{2\pi} d\alpha\big(\int_0^{2\pi}\delta\Phi(\mathbf{x})d\alpha\big)$, is approximated as

$$\int_0^{2\pi} d\alpha\Big(\int_0^{2\pi}\delta\Phi(\mathbf{x})d\alpha\Big) \approx \frac{(2\pi)^2}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\delta\Phi(\mathbf{x}_{ij}), \tag{247}$$

where $N_1=4, N_2=4$ for the case shown here.

The spatial points $\mathbf{x}_{ij}$ appearing in Eq. (247) are not necessarily grid points. Linear interpolations are used to express $\delta\Phi(\mathbf{x}_{ij})$ as linear combination of values of $\delta\Phi$ on gridpoints.

A.1 Performing the parallel velocity integration

Assume that F_0 is Maxwellian, then

$$\frac{\partial F_0}{\partial \varepsilon} = -\frac{m}{T} f_M = -\frac{m}{T} \frac{n_0}{(2\pi T/m)^{3/2}} \exp\left(\frac{-m v_\parallel^2}{2T}\right) \exp\left(\frac{-m v_\perp^2}{2T}\right). \tag{248}$$

Note that $\Delta\boldsymbol{\rho}$ is independent of $v_\parallel$. Then the integration over $v_\parallel$ can be performed

$$\begin{aligned} A(\mathbf{x}) &= \frac{q}{T} \int_{-\infty}^{\infty} d\,\bar{v}_\parallel \exp\left(\frac{-\bar{v}_\parallel^2}{2}\right) \int_0^{\infty} d\,\bar{v}_\perp \bar{v}_\perp \frac{n_0}{(2\pi)^{3/2}} \exp\left(\frac{-\bar{v}_\perp^2}{2}\right) \\ &\times \frac{2\pi}{N_1} \sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \delta\Phi(\mathbf{x} + \Delta\boldsymbol{\rho}_{ij}) \\ &= n_0 \int_0^{\infty} d\,\bar{v}_\perp \bar{v}_\perp \exp\left(\frac{-\bar{v}_\perp^2}{2}\right) \frac{1}{N_1} \sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \frac{q}{T} \delta\Phi(\mathbf{x} + \Delta\boldsymbol{\rho}_{ij}). \end{aligned} \tag{249}$$

where $\bar{v}_\parallel = v_\perp / v_t$, $v_t = \sqrt{T/m}$, $\bar{v}_\perp = v_\perp / v_t$, and use has been made of

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}. \tag{250}$$

A.2 Toroidal Fourier transform of polarization density in field-aligned coordinates

In field-aligned coordinates (x, y, z), Fourier expansion of $\delta\Phi$ along y is written

$$\delta\Phi(\mathbf{x}) = \sum_{n=-N_t}^{N_t} \exp\left(\iota n \frac{2\pi}{L_y} y\right) \delta\Phi_n(x, z), \tag{251}$$

where N_t is the number of toroidal harmonics included. Use this in Eq. (249), yielding

$$\begin{aligned} A(\mathbf{x}) &= n_0 \int_0^{\infty} d\,\bar{v}_\perp \bar{v}_\perp \exp\left(\frac{-\bar{v}_\perp^2}{2}\right) \\ &\times \frac{1}{N_1} \sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \left[\frac{q}{T} \sum_{n=-N_t}^{N_t} \exp\left(\iota n \frac{2\pi}{L_y}(y + \Delta\rho_{ijy})\right) \delta\Phi_n(x + \Delta\rho_{ijx}, \right. \\ &\left. z + \Delta\rho_{ijz})\right] \end{aligned}$$

$$= n_0 \sum_{n=-N_t}^{N_t} \exp\Big(\iota n \frac{2\pi}{L_y} y\Big) \int_0^\infty d\,\bar{v}_\perp \bar{v}_\perp \exp\Big(\frac{-\bar{v}_\perp^2}{2}\Big)$$
$$\times \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \Big[\frac{q}{T} \exp\Big(\iota\, n \frac{2\pi}{L_y}\Delta\rho_{ijy}\Big)\delta\Phi_n(x + \Delta\rho_{ijx},\ z + \Delta\rho_{ijz})\Big]. \tag{252}$$

From Eq. (252), the Fourier transform of $A(\mathbf{x})$ in y can be identified, which is

$$\begin{aligned} A_n(x,z) &= n_0 \int_0^\infty d\,\bar{v}_\perp \bar{v}_\perp \exp\Big(\frac{-\bar{v}_\perp^2}{2}\Big) \\ &\times \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \Big[\frac{q}{T}\exp\Big(\iota n \frac{2\pi}{L_y}\Delta\rho_{ijy}\Big)\delta\Phi_n(x+\Delta\rho_{ijx}, z+\Delta\rho_{ijz})\Big]. \\ &\approx n_0 \int_0^\infty d\,\bar{v}_\perp \bar{v}_\perp \exp\Big(\frac{-\bar{v}_\perp^2}{2}\Big) \\ &\times \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{1}{N_2} \sum_{j=1}^{N_2} \Big[\exp\Big(\iota n \frac{2\pi}{L_y}\Delta\rho_{ijy}\Big)\frac{q}{T}\delta\Phi_n(x+\Delta\rho_{ijx}, z)\Big]. \end{aligned} \tag{253}$$

B Derivation of Eq. (112)

Using the fact that δG_0 is independent of α, the left-hand side of Eq. (112) is written as

$$\begin{aligned} &\langle \mathbf{v}\cdot[\boldsymbol{\lambda}_{B1} + \boldsymbol{\lambda}_{B2}]\delta G_0\rangle_\alpha \\ &= \frac{1}{2\pi}\int_0^{2\pi} d\alpha\, \mathbf{v}\cdot\Big\{\Big[\mathbf{v}\times\frac{\partial}{\partial\mathbf{x}}\Big(\frac{\boldsymbol{e}_\parallel}{\Omega}\Big)\Big]\cdot\frac{\partial\delta G_0}{\partial\mathbf{X}} + \frac{\partial\mu}{\partial\mathbf{x}}\frac{\partial\delta G_0}{\partial\mu} + \frac{\partial\alpha}{\partial\mathbf{x}}\frac{\partial\delta G_0}{\partial\alpha}\Big\} \\ &= \frac{1}{2\pi}\int_0^{2\pi} d\alpha\, \mathbf{v}\cdot\Big\{\Big[\mathbf{v}\times\frac{\partial}{\partial\mathbf{x}}\Big(\frac{\boldsymbol{e}_\parallel}{\Omega}\Big)\Big]\cdot\frac{\partial\delta G_0}{\partial\mathbf{X}} + \frac{\partial\mu}{\partial\mathbf{x}}\frac{\partial\delta G_0}{\partial\mu}\Big\} \end{aligned}$$

From the definition of μ, we obtain

$$\frac{\partial\mu}{\partial\mathbf{x}} = -\frac{v_\perp^2}{2B_0^2}\frac{\partial B_0}{\partial\mathbf{x}} + \frac{1}{2B_0}\frac{\partial v_\perp^2}{\partial\mathbf{x}} = -\frac{\mu}{B_0}\frac{\partial B_0}{\partial\mathbf{x}} + \frac{\partial\mathbf{v}_\perp}{\partial\mathbf{x}}\cdot\frac{\mathbf{v}_\perp}{B_0} \tag{254}$$

Using

$$\frac{\partial\mathbf{v}_\perp}{\partial\mathbf{x}} = \frac{\partial[\mathbf{v} - v_\parallel\mathbf{e}_\parallel]}{\partial\mathbf{x}} = -v_\parallel\frac{\partial\mathbf{e}_\parallel}{\partial\mathbf{x}} - \frac{\partial v_\parallel}{\partial\mathbf{x}}\mathbf{e}_\parallel, \tag{255}$$

expression (254) is written as

$$\frac{\partial\mu}{\partial\mathbf{x}} = -\frac{\mu}{B_0}\frac{\partial B_0}{\partial\mathbf{x}} - v_\parallel\frac{\partial\mathbf{e}_\parallel}{\partial\mathbf{x}}\cdot\frac{\mathbf{v}_\perp}{B_0}, \tag{256}$$

which agrees with Eq. (10) in Frieman-Chen's paper[3].

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi} d\alpha \mathbf{v}\cdot\frac{\partial \mu}{\partial \mathbf{x}} \\
=&\frac{1}{2\pi}\int_0^{2\pi} d\alpha \mathbf{v}\cdot\left(-\frac{\mu}{B_0}\frac{\partial B_0}{\partial \mathbf{x}} - v_\parallel \frac{\partial \mathbf{e}_\parallel}{\partial \mathbf{x}}\cdot\frac{\mathbf{v}_\perp}{B_0}\right) \\
=&?0
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi} d\alpha \mathbf{v}\cdot\left[\mathbf{v}\times\frac{\partial}{\partial \mathbf{x}}\left(\frac{\boldsymbol{e}_\parallel}{\Omega}\right)\right]\cdot\frac{\partial \delta G_0}{\partial \mathbf{X}} \\
=&
\end{aligned}$$

C ddd

In the above, we assume that $\mathbf{X}$ and $\mathbf{x}$ are always related to each other by the guiding-center transformation (19) or (21) , i.e., $\mathbf{x}$ and $\mathbf{X}$ are not independent. For some cases, it may be convienent to treat $\mathbf{x}$ and $\mathbf{X}$ as independent variables and express the guiding-center transformation via an integral of the Dirac delta function. For example,

$$f_p(\mathbf{x},\mathbf{v}) = \int f_g(\mathbf{X},\mathbf{v})\delta^3(\mathbf{X}-\mathbf{x}+\boldsymbol{\rho})d\mathbf{X}, \tag{257}$$

where $\mathbf{x}$ and $\mathbf{X}$ are considered as independent variables, $\boldsymbol{\rho}$ is evaluated at $\mathbf{x}$, $\delta^3(\mathbf{x}-\mathbf{X}-\boldsymbol{\rho})$ is the three-dimensional Dirac delta function. [In terms of general coordinates (x_1,x_2,x_3), the three-dimensional Dirac delta function is defined via the 1D Dirac delta function as follows:

$$\delta^3(\mathbf{x}) = \frac{1}{|\mathcal{J}|}\delta(x_1)\delta(x_2)\delta(x_3), \tag{258}$$

where $\mathcal{J}$ is the the Jacobian of the general coordinate system. The Jacobian is included in order to make $\delta^3(\mathbf{x})$ satisfy the normalization condition $\int \delta^3(\mathbf{x})d\mathbf{x} = \int \delta^3(\mathbf{x})|\mathcal{J}|dx_1dx_2dx_3 = 1$.]

Expression (257) can be considered as a transformation that transforms an arbitrary function from the guiding-center coordinates to the particle coordinates. Similarly,

$$f_g(\mathbf{X},\mathbf{v}) = \int f_p(\mathbf{x},\mathbf{v})\delta^3(\mathbf{x}-\mathbf{X}-\boldsymbol{\rho})d\mathbf{x}, \tag{259}$$

is a transformation that transforms an arbitrary function from the the particle coordinates to the guiding-center coordinates.

C.1 Moments of distribution function expressed as integration over guiding-center variables

In terms of particle variables $(\mathbf{x}, \mathbf{v})$, it is straightforward to calculate the moment of the distribution function. For example, the number density $n(\mathbf{x})$ is given by

$$n(\mathbf{x}) = \int f_p(\mathbf{x}, \mathbf{v}) d\mathbf{v}. \tag{260}$$

However, it is a little difficult to calculate $n(\mathbf{x})$ at real space location $\mathbf{x}$ by using the guiding-center variables $(\mathbf{X}, \mathbf{v})$. This is because holding $\mathbf{x}$ constant and changing $\mathbf{v}$, which is required by the integration in Eq. (260), means the guiding-center variable $\mathbf{X}$ is changing according to Eq. (19). Using Eq. (26), expression (260) is written as

$$n(\mathbf{x}) = \int f_g(\mathbf{X}(\mathbf{x}, \mathbf{v}), \mathbf{v}) d\mathbf{v}, \tag{261}$$

As is mentioned above, the $d\mathbf{v}$ integration in Eq. (261) should be performed by holding $\mathbf{x}$ constant and changing $\mathbf{v}$, which means the guiding-center variable $\mathbf{X} = \mathbf{X}(\mathbf{x}, \mathbf{v})$ is changing. This means that, in $(\mathbf{X}, \mathbf{v})$ space, the above integration is a (generalized) line integral along the the line $\mathbf{X}(\mathbf{v}) = \mathbf{x} - \boldsymbol{\rho}(\mathbf{x}, \mathbf{v})$ with $\mathbf{x}$ being constant. Treating $\mathbf{X}$ and $\mathbf{x}$ as independent variables and using the Dirac delta function δ, this line integral can be written as the following double integration over the independent variables $\mathbf{X}$ and $\mathbf{v}$:

$$n(\mathbf{x}) = \iint f_g(\mathbf{X}, \mathbf{v}) \delta^3(\mathbf{X} - \mathbf{x} + \boldsymbol{\rho}) d\mathbf{v} d\mathbf{X}. \tag{262}$$

Another perspective of interpreting Eq. (262) is that we are first using the tansformation (257) to transform f_g to f_p and then integrating f_p in the velocity space.

C.2 Numerical evaluation of distribution function moments at grids

Consider the evaluation of integral (262). Actually evaluated in PIC simulation is the average of $n(\mathbf{x})$ in a spatial cell, i.e.,

$$n_{i,j,k} = \frac{1}{V_{i,j,k}} \int_{\Omega_{i,j,k}} \left(\iint f_g(\mathbf{X}, \mathbf{v}) \delta^3(\mathbf{X} - \mathbf{x} + \boldsymbol{\rho}) d\mathbf{v} d\mathbf{X} \right) d\mathbf{x}, \tag{263}$$

where $\Omega_{i,j,k}$ is a spatial cell and $V_{i,j,k}$ is its volume.

In PIC simulations, integral (263) is evaluated by using Monte-Carlo markers. Thinking in terms of full orbit particle simulation, we load an assemble of particles in the phase space $(\mathbf{x}, v_\perp, v_\parallel, \alpha)$. Then we know the phase-space volume occupied by each marker, which is assumed to be constant in time for each marker. For a marker with coordinates $(\mathbf{x}_j, v_{\perp j}, v_{\parallel j}, \alpha_j)$, we can calculate the guiding-center position $\mathbf{X}$ of each marker by using Eq. (19). This is done for only once at the initialization stage of the simulation. Most codes directly set guiding-center localtion $\mathbf{X}$, then the transformation is not needed, but I prefer the former method because particle coordinates are physical and guiding-center coodrinates

$\mathbf{X}$ are only imaginary and auxiliary. Update: However, directly sampling particle position, rather than guiding-center position, is not practical because we need to load a distribution function that is independent of α in $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$ coordinates and this is hard to achieve if we directly sample $(\mathbf{x}, v_\perp, v_\parallel, \alpha)$.

Then $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j})$ can be evolved by using the guiding-center motion equation. But how does the gyro-angle α evolves with time? We do not have an equation for the time evolution of α. The evolution can be calculated by using the local gyro-frequency if we want them. Do we need the gyro angle? Yes, we do. The gyro-angle α is needed in the inverse guiding-center transform when depositing particle weight to grids. But it turns out tracking the evolution of α of a marker is not necessary for PIC simulations because the density and current needed in gyrokinetic particle simulations are those averaged over a gyro-period and during this time interval the gyro-angle of a mark has taken all the possible values.

Assume that we load markers in $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$ with uniform random distribution in the gyro-angle. The gyro-angle of each marker evolves on a very fast time-scale and at later time the distribution in gyro-angle should be still uniform. There is no need to follow the time evolution of α of each marker because any set of random samplings of α are similar to each other in terms of evaluating Monte-Carlo integration. When doing the deposition, the coresponding particle location is calculated by using an arbitrary gyro-angle, which is of cource not the correct value of α taken by that marker at that particular time since we do not follow the time evolution of α and thus the correct value is not available to us. As a result, the particle location calculated by using the incorrect gyro-angle is also incorrect. However, statistically, this phase error of particle location does not matter in terms of evaluating the Monte-Carlo integral because the phase error will cancell each other when summing over particles. (Wrong! no reason to believe that the error will cancel each other) Statistically, we still obtain consistent estimation of the integral. To further reduce the possible error, we select several (rather than just one) possible gyro-angles for each marker and average all these over all markers, yeilding more accurate estimation.

Another way to understand the gyrokinetic simulation is that the density calculated in simulations can be considered as the time average of integral (263) over one typical gyro-period, i.e.,

$$\overline{n}_{i,j,k} = \frac{1}{\delta t}\int_{t_n}^{t_n+\delta t} dt\, n_{i,j,k} = \frac{1}{\delta t}\int_{t_n}^{t_n+\delta t} dt \frac{1}{V_{i,j,k}}\int_{\Omega_{i,j,k}} \left(\iint f_g(\mathbf{X},\, \mathbf{v})\delta^3(\mathbf{X} - \mathbf{x} + \boldsymbol{\rho}) d\mathbf{v} d\mathbf{X} \right) d\mathbf{x}, \tag{264}$$

where δt is a typical gyro-period. In this small time inteval, marker's $(\mathbf{X}, v_\perp, v_\parallel)$ can be considered to be constant and only α changes. Then the time average is equivalent to an average over the gyro-angle. The four point averaging in gyrokinetic simulations can be considered as the time descretization of this time averaging.

Another way of understanding the four point gyro-averaging is that we do 4 separate sampling of the phase space and then average the each estimation to get a bettter estimation.

For a marker with coordinators $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$, where α is the gyro-angle, the corresponding particle position can be calculated by using the inverse guiding-center transformation (21). Then we can deposit the marker weight to grids in the same way that we do in conventional PIC simulations. Looping over all markers, we build the particle density at grids.

Compared with conventional PIC methods, where particle positions are directly sampled, what is the benifit of sampling guiding-center positions and then transform them back to the particle positions? More computations are involved since we need to numerically perform the inverse guiding-center transformation.

to be clean up*The answer lies in the important fact that the distribution function $f_g(\mathbf{X}, v_\perp, v_\parallel, \alpha)$ that needs to be numerically evolved in gyrokinetic simulation is actually independent of the gyro-angle α. Furthermore, we use a probability density function P that is independent of α to sample the 6D phase space $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$. Then the marker weight $w \equiv f_g / (N_p\, P)$ is independent of α, where N_p is number of markers loaded.

Suppose we have a concrete sampling of f_g in the 6D phase-space ($\mathbf{X}$, $v_\perp$, $v_\parallel$, α), i.e., $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j}, \alpha_j, w_j)$ with $j = 1, ..., N_p$, then we can do the inverse guiding-center transform and then deposit particles to the grids to obtain a moment (e.g. density).

Since $P\mathcal{J}_{rv}$, where $\mathcal{J}_{rv} = \mathcal{J}_r v_\perp$ is the Jacobian of $(\mathbf{X}, v_\perp, v_\parallel, \alpha)$, is independent of α, the sampling α_j with $j = 0, 1, ..., N_p$ are uniform distributed random numbers. Therefore we can generate another sampling of α (denoted by α'_j with $j = 1, ..., N_p$) using random number generators and combine α'_j with the old sampling ($\mathbf{X}_j$, $v_{\perp j}$, $v_{\parallel j}$) to obtain a new sampling ($\mathbf{X}_j$, $v_{\perp j}$, $v_{\parallel j}$, α'_j, w'_j). The values of w at the new sampling points are equal to the original values, i.e., $w'_j = w_j$ since the particle weight $w = f_g / (N_p P)$ is independent of α. Using the new sampling and following the same procedures given above, we can estimate values of the moment at grids again, which will differ from the estimation obtained using the old sampling. Taking the average of the two estimations will give a more accurate estimation.

Note that even if $P\mathcal{J}_{rv}$ is dependent on α due to the possible dependence of $\mathcal{J}_{rv}$ on α, we still can easily generate a new set of sampling which differes from the old sampling only in α, i.e., $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j}, \alpha'_j)$. Specifically, in the rejection method, we use the old sampling ($\mathbf{X}_j$, $v_{\perp j}$, $v_{\parallel j}$) for each reject step and only adjust α to satisfy the acceptance criteria.

The independence of P on α is the essential thing, because the combination (f_g / P) must be independent of α, where f_g is already independent of α.

We can also construct a new sampling by replacing α_j by $\alpha_j + \Delta$, where Δ is a constant. Then the new sampling is given by $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j}, \alpha_j + \Delta, w_j')$ with $j = 1, ..., N_p$. Since the original sampling probability function $P\mathcal{J}_r v_\perp$ is independent of α, then $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j}, \alpha_j + \Delta)$ is still a consistent sampling that can be generated by the original probability function $P\mathcal{J}_r v_\perp$. Furthermore, since the particle weight $w = f_g/(N_p P)$ is independent of α, we infer that the values of w at the new sampling points are equal to the original values, i.e., $w_j' = w_j$.

In doing the deposition, each marker has a single gyro-angle (this is the result from the guiding-center transformation, where the gyro-angle must be determined before we can transform a particl location to a guiding-center location). Due to the independence of the weight ($w = f_g/(N_p\, P)$) of the gyro-angle, the resolution over the gyro-angle can be increased in a way that there can be several gyro-angles for a single $(\mathbf{X}_j, v_{\perp j}, v_{\parallel j})$. This gives the impression that the gyro-angle of a guiding-center marker is arbitrary. This is impression is incorrect. The correct understanding is that given above, i.e., we take several times of samplings and average the estimations resulting from each sampling.

In the code, the gyro-angle is defined relative to the direction $\nabla\psi/|\nabla\psi|$ at the guiding-center position. Specificly, the gyro-angle is defined as the included angle between $\nabla\psi/|\nabla\psi|$ and $-\mathbf{v}\times\mathbf{e}_\parallel(\mathbf{x})/\Omega(\mathbf{x})$, where $\mathbf{e}_\parallel(\mathbf{x})/\Omega(\mathbf{x})$ is approximated by the value at the guiding-center location.

C.3 Monte-Carlo sampling of 6D guiding-center phase-space

Suppose that the 6D guiding-center phase-space ($\mathbf{X}$, $\mathbf{v}$) is described by (ψ, θ, ϕ, $v_\parallel$, $v_\perp$, α) coordinates. The Jacobian of the coordinate system is given by $\mathcal{J} = \mathcal{J}_r v_\perp$, where $\mathcal{J}_r = \mathcal{J}(\psi, \theta)$ is the Jacobian of the coordinates (ψ, θ, ϕ). Suppose that we sample the 6D phase-space by using the following probability density function:

$$P(\psi,\theta,\phi,v_\parallel,v_\perp,\alpha) = \frac{1}{V_r}\Big(\frac{m}{2\pi T}\Big)^{3/2}\exp\left[-\frac{m\,(v_\parallel^2+v_\perp^2)}{2T}\right], \tag{265}$$

where V_r is the volume of the spatial simulation box, T is a constant temperature. P given above is independent of ψ, θ, ϕ and α. It is ready to verify that the above P satisfies the following normalization condition:

$$\int_{V_r}\int P d\mathbf{v} d\mathbf{X} = \int_{V_r}\int_{-\infty}^{+\infty}\int_0^\infty\int_0^{2\pi} P v_\perp d\alpha dv_\perp dv_\parallel \mathcal{J}_r d\psi d\theta d\phi = 1. \tag{266}$$

I use the rejection method to numerically generate N_p markers that satisfy the above probability density function. [The effective probability density function actually used in the rejection method is P', which is related to P by

$$P' = |\mathcal{J}_r\mathcal{J}_v|P = |\mathcal{J}_r(\psi, \theta)|v_\perp P = |\mathcal{J}_r(\psi, \theta)|v_\perp \frac{1}{V_r}\Big(\frac{m}{2\pi T}\Big)^{3/2}\exp\left[-\frac{m\,(v_\parallel^2+v_\perp^2)}{2T}\right] \tag{267}$$

Note that P' does not depend on the gyro-angle α.]

Then the weight of a marker is written

$$w = \frac{f_g(\psi, \theta, \phi, v_\parallel, v_\perp)}{N_p P}. \tag{268}$$

Since both f_g and P is independent of the gyro-angle α, w is also independent of α.

The numerical representation of δf_g is written

$$\delta \tilde{f}_g = \frac{1}{\mathcal{J}_r v_\perp} \sum_j^{N_p} w_j \delta(\psi - \psi_j)\delta(\theta - \theta_j)\delta(\phi - \phi_j)\delta(v_\parallel - v_{\parallel j})\delta(v_\perp - v_{\perp j})\delta(\alpha - \alpha_j). \tag{269}$$

Although the distribution function δf_g to be sampled in gyrokinetic simulations is independent of the gyro-angle α, we still need to sample the gyro-angle because we need to use the inverse guiding-center transformation, which needs the gyro-angle. Each marker needs to have a specific gyro-angle value α_j so that we know how to transform its $\mathbf{X}_j$ to $\mathbf{x}_j$ and then do the charge deposition in $\mathbf{x}$ space. To increase the resolution over the gyro-angle, we need to load more markers. However, thanks to the fact that both sampling probability density function P and δf_g are independent of α, the resolution over the gyro-angle can be increased in a simple way. Specifically, for a marker with $(\mathbf{X}_j, v_{\parallel j}, v_{\perp j})$, its gyro-angle value α_j can be adjusted arbitrarily without changing the value of its weight w_j because the weight is independent of α_j. In other words, we have the freedom of choosing the value of α_j (the independence of the weight on α_j guarantees that the weight of the marker is still equal to the original value that this marker takes). Suppose that we choose 4 different values of α_j for the jth marker, denoted by α_{j1}, α_{j2}, α_{j3}, and α_{j4}. Then for each of the four gyro-angle values, we do the inverse guiding-center transformation and then do the charge deposition using their original weight w_j for each marker and then loop over all the markers. Denote a grid quantity (e.g. density) build by the deposition process by $n_1(\mathbf{x})$, $n_2(\mathbf{x})$, $n_3(\mathbf{x})$, and $n_4(\mathbf{x})$, corresponding to using the four gyro-angle values. Then the more accurate estimation of the grid quantity is given by

$$n(\mathbf{x}) = \frac{n_1(\mathbf{x}) + n_2(\mathbf{x}) + n_3(\mathbf{x}) + n_4(\mathbf{x})}{4}. \tag{270}$$

This corresponds to sampling the 6D phase-space 4 separate times (each time with identical sampling points in ($\mathbf{X}$, $v_\parallel$, $v_\perp$) but different sampling points in α) and then using the averaging of the 4 Monte-Carlo integrals to estimate the exact value. This estimation can also be (roughly) considered as a Monte-Carlo estimation using 4 times larger number of markers as that is originally used (the Monte-Carlo estimation using truly 4 time larger number of markers is more accurate than the result we obtained above because the former also increase the resolution of ($\mathbf{X}$, $v_\parallel$, $v_\perp$) while the latter keeps the resolution of ($\mathbf{X}$, $v_\parallel$, $v_\perp$) unchanged.)

In numerical code, we choose N sampling points that are evenly distributed on the gyro-ring (N is usually 4 as a compromise between efficiency and accuracy). Denote the Mote-Carlo weight of the jth marker by w_j. Then the weight is evenly split by the N sub-markers on the gyro-ring. Therefore each sub-marker have a Monte-Carlo weight w_j/N. Then calculating the integration (262) at a grid corresponds to depositing all the N sub-markers associated with each guiding-center marker to the grid, as is illustrated in Fig. 3. However, interpreting in this way is confusing to me because, with a single sampling of the phase-space, the phase-space volume or weight can not be easily split. I prefer the above interpretation that the 6D phase space is sampled 4 separate times and thus we get 4 estimations and finally we take the averaging of these 4 estimations. It took me several days to finally find this way of understanding.

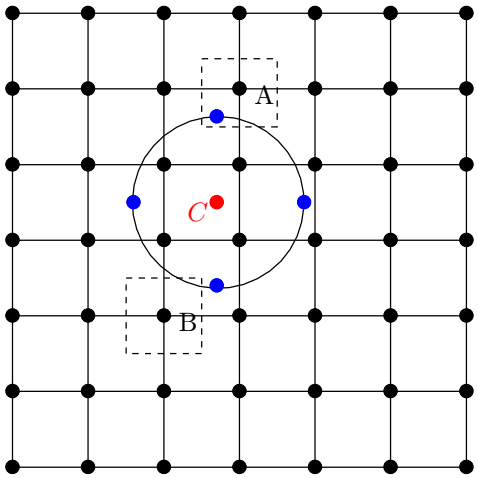

Figure 3. The spatial grids in the plane perpendicular to the equilibrium magnetic field and one gyro-ring with 4 sampling points (sub-markers) on it. For a guiding-center marker C with a Monte-Carlo weight of w_j, the 4 sub-markers are calculated by using the transformation (21) (rebuilding the gyro-ring). The Monte-Carlo weight of each sub-marker is $w_j/4$. The value of integration (262) at a grid point is approximated by $I/\Delta V$, where I is the Monte-Carlo integration of all sub-markers (associated with all guiding-center markers) in the cell, ΔV is the cell volume. The cell associated with a grid-point (e.g., A) is indicated by the dashed rectangle (this is for 2D case, for 3D, it is a cube). If the Dirac delta function is used as the shape function of the sub-markers, then calculating the contribution of a sub-marker to a grid corresponds to the nearest-point interpolation (for example, the 4 sub-markers will contribute noting to grid point B since no-sub marker is located within the cell). In practice, the flat-top shape function with its support equal to the cell size is often used, then the depositing corresponds to linearly interpolating the weight of each sub-marker to the nearby grids.

In summary, the phase-space to be sampled in gyrokinetic simulations are still 6D rather than 5D. In this sense, the statement that gyrokinetic simulation works in a 5D phase space is misleading. We are still working in the 6D phase-

space. The only subtle thing is that the sixth dimension, i.e., gyro-angle, can be sampled in an easy way that is independent of the other 5 variables.

In numerical implementation, the gyro-angle may not be explicitly used. We just try to find 4 arbitrary points on the gyro-ring that are easy to calculate. Some codes (e.g. ORB5) introduces a random variable to rotate these 4 points for different markers so that the gyro-angle can be sampled less biased.

For a marker with coordinators $(\mathbf{X}, v_\perp, v_\parallel)$, we can calculate particle positions by using the transformation (21). There are infinite number of particle positions associated with the marker since the direction of $v_\perp$ (i.e., gyro-angle) is not specified. All these possible particle positions are on a circle around the guiding-center position $\mathbf{X}$. This circle is often called the gyro-ring.

From the view of particle simulations, the gyrokinetic model can be considered as a noise reduction method, where the averaging over the gyro-angle is equivalent to a time averaging over a gyro-period, which reduces the fluctuation level (in both time and space) assosicated with evaluating the Monte-Carlo phase integration. Here the averaging in gyrokinetic particle simulation refers to taking several points on a gyro-ring when depositing markers to spatial grids to obtain the density and current on the grids. (Another gyro-averaging appears in evaluating the guiding-center drift.) In gyrokinetic particle simulation, even a step size smaller than a gyro-period is taken, the quantities used in the model is still the ones averaged over one gyro-period. In this sense, a gyrokinetic simulation is only meaningful when the time step size is larger than one gyro-period. [**Some authors may disagree with that the gyro-averaging is a time-averaging. They may consider the gyro-averaging as the phase-space integration over the gyro-angle coordinate. This view seems to be right in Euler simulations but seems to be wrong in particle simulations. The reason is as follows. For each marker, choose a random gyro-phase and then do the inverse transformation to obtain particle position, and sum over all markers (this corresponds to phase-space Monte-Carlo integration, which include the gyro-angle integration, so no further gyro-angle integration is needed); choose another random gyro-phase and repeat the above procedure (this can be interpreted as do the phase-space Monte-Carlo integration at another time), choose further random gyro-phase for each marker and repeat. Finally averaging all the above values to obtain the final estimation of the phase-space integration. This amounts to a time-averaging over a gyro-motion. In summary, sampling several times with different gyro-phases for each marker and taking the average amounts to the time averaging over gyro-motion**]

When doing the time-average over the ion cylotron motion, the time variation of the low-frequency mode is negilible and only the spatial variation of the modes is important. For the gyro-motion, only the gyro-angle is changing and all the other variables, $(\mathbf{X}, v_\perp, v_\parallel)$, are approximately constant. As a result, this time averaging finally reduces to a gyro-averaging.

I am always reasoning in terms of particle position and velocity, considering the guiding-center location as an image of the particle position in an imaginary space (guiding-center space). When working in the guiding-center space, I am always reasoning by transforming back to the particle position. This reasoning is clear and help me avoid some confusions I used to have.

D Diamagnetic flow

The perturbed distribution function δF given in Eq. (146) contains two terms. The first term is gyro-phase dependent while the second term is gyro-phase independent. The perpendicular velocity moment of the first term will give rise to the so-called $\delta\mathbf{E}\times\mathbf{B}_0$ flow (seems wrong, need checking) and the second term will give rise to the so-called diamagnetic flow. Let us discuss the diamagnetic flow first. For this case, it is crucial to distinguish between the distribution function in terms of the guiding-center variables, $f_g(\mathbf{X},\mathbf{v})$, and that in terms of the particle variables, $f_p(\mathbf{x},\mathbf{v})$. In terms of these denotations, equation (146) is written as

$$\delta F_g = \frac{q}{m}(\delta\Phi - \langle\delta\Phi\rangle_\alpha)\frac{\partial F_{0g}}{\partial\varepsilon} + \delta f_g. \tag{271}$$

Next, consider the perpendicular flow $\mathbf{U}_\perp$ carried by δf_g. This flow is defined by the corresponding distribution function in terms of the particle variables, δf_p, via,

$$n\mathbf{U}_\perp = \int \mathbf{v}_\perp \delta f_p(\mathbf{x},\mathbf{v})d\mathbf{v}, \tag{272}$$

where n is the number density defined by $n=\int\delta f_p d\mathbf{v}$. Using the relation between the particle distribution function and guiding-center distribution function, equation (272) is written as

$$n\mathbf{U}_\perp = \int \mathbf{v}_\perp \delta f_g(\mathbf{x}-\boldsymbol{\rho},\mathbf{v})d\mathbf{v}. \tag{273}$$

Using the Taylor expansion near $\mathbf{x}$, $\delta f_g(\mathbf{x}-\boldsymbol{\rho},\mathbf{v})$ can be approximated as

$$\delta f_g(\mathbf{x}-\boldsymbol{\rho},\mathbf{v}) \approx \delta f_g(\mathbf{x},\mathbf{v}) - \boldsymbol{\rho}\cdot\nabla\delta f_g(\mathbf{x},\mathbf{v}). \tag{274}$$

Plugging this expression into Eq. (273), we obtain

$$n\mathbf{U}_\perp \approx \int \mathbf{v}_\perp \delta f_g(\mathbf{x},\mathbf{v})d\mathbf{v} - \int \mathbf{v}_\perp\boldsymbol{\rho}\cdot\nabla\delta f_g(\mathbf{x},\mathbf{v})d\mathbf{v} \tag{275}$$

As mentioned above, $\delta f_g(\mathbf{x},\mathbf{v})$ is independent of the gyro-angle α. It is obvious that the first integration is zero and thus Eq. (275) is reduced to

$$n\mathbf{U}_\perp = -\int \mathbf{v}_\perp\boldsymbol{\rho}\cdot\nabla\delta f_g(\mathbf{x},\mathbf{v})d\mathbf{v} \tag{276}$$

Using the definition $\boldsymbol{\rho} = -\mathbf{v}\times\mathbf{e}_\parallel/\Omega$, the above equation is written

$$\begin{aligned} n\mathbf{U}_\perp &= \int \mathbf{v}_\perp \frac{\mathbf{v}\times\mathbf{e}_\parallel}{\Omega}\cdot\nabla\delta f_g(\mathbf{x},\mathbf{v})d\mathbf{v} \\ &= \int \mathbf{v}_\perp\Big(\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla\delta f_g(\mathbf{x},\mathbf{v})\Big)\cdot\mathbf{v}_\perp d\mathbf{v}. \\ &= \int \mathbf{v}_\perp\mathbf{H}\cdot\mathbf{v}_\perp d\mathbf{v}, \end{aligned} \tag{277}$$

where $\mathbf{H} = \frac{\mathbf{e}_\parallel}{\Omega}\times\nabla\delta f_g(\mathbf{x},\mathbf{v})$, which is independent of the gyro-angle α because both $\mathbf{e}_\parallel(\mathbf{x})/\Omega(\mathbf{x})$ and $\delta f_g(\mathbf{x},\mathbf{v})$ are independent of α. Next, we try to perform the integration over α in Eq. (277). In terms of velocity space cylindrical coordinates $(v_\parallel, v_\perp, \alpha)$, $\mathbf{v}_\perp$ is written as

$$\mathbf{v}_\perp = v_\perp(\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha), \tag{278}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are two arbitrary unit vectors perpendicular each other and both perpendicular to $\mathbf{B}_0(\mathbf{x})$. $\mathbf{H}$ can be written as

$$\mathbf{H} = H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}}, \tag{279}$$

where H_x and H_y are independent of α. Using these in Eq. (277), we obtain

$$\begin{aligned} n\mathbf{U}_\perp &= \int v_\perp(\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)\, v_\perp(H_x\cos\alpha + H_y\sin\alpha)d\mathbf{v} \\ &= \int v_\perp^2[\hat{\mathbf{x}}(H_x\cos^2\alpha + H_y\sin\alpha\cos\alpha) + \hat{\mathbf{y}}(H_x\cos\alpha\sin\alpha + \\ &\quad H_y\sin^2\alpha)]\, d\mathbf{v}. \end{aligned} \tag{280}$$

Using $d\mathbf{v} = v_\perp dv_\parallel dv_\perp d\alpha$, the above equation is written as

$$\begin{aligned} n\mathbf{U}_\perp &= \int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp \int_0^{2\pi} v_\perp^2[\hat{\mathbf{x}}(H_x\cos^2\alpha + H_y\sin\alpha\cos\alpha) + \hat{\mathbf{y}}(H_x\cos\alpha\sin\alpha + \\ &\quad H_y\sin^2\alpha)]d\alpha \\ &= \int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp \int_0^{2\pi} v_\perp^2(\hat{\mathbf{x}}H_x\cos^2\alpha + \hat{\mathbf{y}}H_y\sin^2\alpha)d\alpha \\ &= \int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp [v_\perp^2(\hat{\mathbf{x}}H_x\pi + \hat{\mathbf{y}}H_y\pi)] \\ &= \int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp [v_\perp^2\mathbf{H}\pi] \\ &= \int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp [v_\perp^2\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla\delta f_g(\mathbf{x},\mathbf{v})\pi] \\ &= \frac{\mathbf{e}_\parallel}{\Omega}\times\nabla\int_{-\infty}^{\infty} dv_\parallel \int_0^\infty v_\perp dv_\perp \delta f_g(\mathbf{x},\mathbf{v})\frac{v_\perp^2}{2}2\pi \\ &= \frac{\mathbf{e}_\parallel}{m\Omega}\times\nabla\delta p_\perp, \end{aligned} \tag{281}$$

where

$$\begin{aligned}\delta p_\perp &\equiv \int_{-\infty}^{\infty} dv_\parallel \int_0^{\infty} v_\perp dv_\perp \delta f_g(\mathbf{x},\mathbf{v}) \frac{m v_\perp^2}{2} 2\pi \\ &= \int \delta f_g(\mathbf{x},\mathbf{v}) \frac{m v_\perp^2}{2} d\mathbf{v}, \end{aligned} \tag{282}$$

is the perpendicular pressure carried by $\delta f_g(\mathbf{x},\mathbf{v})$. The flow given by Eq. (281) is called the diamagnetic flow.

E Transform gyrokinetic equation from $(\mathbf{X}, \mu, \varepsilon, \alpha)$ to $(\mathbf{X}, \mu, v_\parallel, \alpha)$ coordinates

The gyrokinetic equation given above is written in terms of variables $(\mathbf{X}, \mu, \varepsilon, \alpha)$, where α is the gyro-phase. Next, we transform it into coordinates $(\mathbf{X}', \mu', v_\parallel, \alpha')$ which is defined by

$$\begin{cases} \mathbf{X}'(\mathbf{X}, \mu, \varepsilon, \alpha) = \mathbf{X} \\ \mu'(\mathbf{X}, \mu, \varepsilon, \alpha) = \mu \\ \alpha'(\mathbf{X}, \mu, \varepsilon, \alpha) = \alpha \\ v_\parallel(\mathbf{X}, \mu, \varepsilon, \alpha) = \pm\sqrt{2(\varepsilon - \mu B_0(\mathbf{X}))} \end{cases} \tag{283}$$

Use this definition and the chain rule, the gradient operators in $(\mathbf{X},\ \mu,\ \varepsilon,\ \alpha)$ variables are written, in terms of $(\mathbf{X}', \mu', v_\parallel, \alpha')$ variables, as

$$\begin{aligned}\left.\frac{\partial}{\partial \mathbf{X}}\right|_{\mu,\varepsilon,\alpha} &= \frac{\partial \mathbf{X}'}{\partial \mathbf{X}} \cdot \left.\frac{\partial}{\partial \mathbf{X}'}\right|_{\mu',v_\parallel,\alpha'} + \frac{\partial \mu'}{\partial \mathbf{X}} \left.\frac{\partial}{\partial \mu'}\right|_{\mathbf{X}',v_\parallel,\alpha'} + \frac{\partial v_\parallel}{\partial \mathbf{X}} \left.\frac{\partial}{\partial v_\parallel}\right|_{\mathbf{X}',\mu',\alpha'} + \\ &\quad \frac{\partial \alpha'}{\partial \mathbf{X}} \left.\frac{\partial}{\partial \alpha'}\right|_{\mathbf{X}',\mu',v_\parallel} \\ &= \frac{\partial}{\partial \mathbf{X}'} + 0\frac{\partial}{\partial \mu'} - \frac{\mu}{v_\parallel}\frac{\partial B_0}{\partial \mathbf{X}}\frac{\partial}{\partial v_\parallel} + 0\frac{\partial}{\partial \alpha'} \end{aligned} \tag{284}$$

and

$$\begin{aligned}\left.\frac{\partial}{\partial \varepsilon}\right|_{\mathbf{X},\mu,\alpha} &= \frac{\partial \mathbf{X}'}{\partial \varepsilon}\frac{\partial}{\partial \mathbf{X}'} + \frac{\partial \mu'}{\partial \varepsilon}\frac{\partial}{\partial \mu'} + \frac{\partial v_\parallel}{\partial \varepsilon}\frac{\partial}{\partial v_\parallel} + \frac{\partial \alpha'}{\partial \varepsilon}\frac{\partial}{\partial \alpha'} \\ &= 0\frac{\partial}{\partial \mathbf{X}'} + 0\frac{\partial}{\partial \mu'} + \frac{1}{v_\parallel}\frac{\partial}{\partial v_\parallel} + 0\frac{\partial}{\partial \alpha'} \end{aligned} \tag{285}$$

Then, in terms of independent variable $(\mathbf{X}', \mu', v_\parallel, \alpha')$, equation (143) is written

$$\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\partial}{\partial \mathbf{X}'}\right]\delta G_0 - (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\mu}{v_\parallel}\frac{\partial B_0}{d\mathbf{X}}\frac{\partial \delta G_0}{\partial v_\parallel}$$
$$= -\delta\mathbf{V}_D\cdot\left(\frac{\partial F_0}{\partial \mathbf{X}'} - \frac{\mu}{v_\parallel}\frac{\partial B_0}{d\mathbf{X}}\frac{\partial F_0}{\partial v_\parallel}\right) - \frac{q}{m}\frac{\partial \langle \delta L\rangle_\alpha}{\partial t}\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel}, \tag{286}$$

where $\delta\mathbf{V}_D$ and $\langle \delta L\rangle_\alpha$ involve the gyro-averaging operator $\langle ...\rangle_\alpha$. The gyro-averaging operator in $(\mathbf{X}', \mu', v_\parallel, \alpha')$ coordinates is similar to that in the old coordinates since the perpendicular velocity variable μ is identical between the two coordinate systems. Also note that the perturbed guiding-center velocity $\delta\mathbf{V}_D$ is given by

$$\delta\mathbf{V}_D = \frac{\mathbf{e}_\parallel \times \nabla_X \langle \delta\phi\rangle_\alpha}{B_0} + v_\parallel \frac{\langle \delta\mathbf{B}_\perp\rangle_\alpha}{B_0}, \tag{287}$$

where $\partial/\partial\mathbf{X}$ (rather than $\partial/\partial\mathbf{X}'$) is used. Since $\delta\phi(\mathbf{x}) = \delta\phi_g(\mathbf{X}, \mu', \alpha')$, which is independent of $v_\parallel$, then Eq. (284) indicates that $\partial\delta\phi/\partial\mathbf{X} = \partial\delta\phi/\partial\mathbf{X}'$.

Dropping terms of order higher than $O(\lambda^2)$, equation (286) is written as

$$\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\partial}{\partial \mathbf{X}'}\right]\delta G_0 - \mathbf{e}_\parallel\cdot\mu\nabla B_0\frac{\partial \delta G_0}{\partial v_\parallel}$$
$$= -\delta\mathbf{V}_D\cdot\left(\frac{\partial F_0}{\partial \mathbf{X}'}\right) + \left(\delta\mathbf{V}_D\cdot\mu\nabla B_0 - \frac{q}{m}\frac{\partial\langle\delta L\rangle_\alpha}{\partial t}\right)\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel}, \tag{288}$$

Similarly, in terms of independent variable $(\mathbf{X}', \mu', v_\parallel, \alpha')$, equation (148) is written as

$$\left[\frac{\partial}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\partial}{\partial \mathbf{X}'}\right]\delta f - \mathbf{e}_\parallel\cdot\mu\nabla B_0\frac{\partial \delta f}{\partial v_\parallel}$$
$$= -\delta\mathbf{V}_D\cdot\left(\frac{\partial F_0}{\partial \mathbf{X}'}\right) + \delta\mathbf{V}_D\cdot\left(\frac{\mu}{v_\parallel}\nabla B_0\frac{\partial F_0}{\partial v_\parallel}\right)$$
$$-\frac{q}{m}\left[-\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} - \left(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\right)\cdot\nabla_X\langle\delta\phi\rangle_\alpha\right]\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel}, \tag{289}$$

The guiding-center velocity in the macroscopic (equilibrium) field is given by

$$v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D = \frac{\mathbf{B}_0^\star}{B_{\parallel 0}^\star}v_\parallel + \frac{\mu}{\Omega B_{\parallel 0}^\star}\mathbf{B}_0\times\nabla B_0 + \frac{1}{B_0 B_{\parallel 0}^\star}\mathbf{E}_0\times\mathbf{B}_0 \tag{290}$$

where

$$\mathbf{B}_0^\star = \mathbf{B}_0 + B_0\frac{v_\parallel}{\Omega}\nabla\times\mathbf{b}, \tag{291}$$

$$B_\parallel^\star \equiv \mathbf{b}\cdot\mathbf{B}^\star = B\left(1 + \frac{v_\parallel}{\Omega}\mathbf{b}\cdot\nabla\times\mathbf{b}\right), \tag{292}$$

Using $B^\star_{\parallel 0}\approx B_0$, then expression (290) is written as

$$v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D = v_\parallel \mathbf{b} + \underbrace{\frac{v_\parallel^2}{\Omega}\nabla\times\mathbf{b}}_{\text{curvaturedrift}} + \underbrace{\frac{\mu}{\Omega B_0}\mathbf{B}_0\times\nabla B_0}_{\nabla B\text{drift}} + \underbrace{\frac{1}{B_0^2}\mathbf{E}_0\times\mathbf{B}_0}_{E\times B\text{drift}}, \tag{293}$$

where the curvature drift, ∇B drift, and $\mathbf{E}_0\times\mathbf{B}_0$ drift can be identified. Note that the perturbed guiding-center velocity $\delta\mathbf{V}_D$ is given by (refer to Sec. F.3)

$$\delta\mathbf{V}_D = \frac{\mathbf{e}_\parallel\times\nabla_X\langle\delta\phi\rangle_\alpha}{B_0} + v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}. \tag{294}$$

Using the above results, equation (289) is written as

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\partial}{\partial\mathbf{X}'}\right]\delta f - \mathbf{e}_\parallel\cdot\mu\nabla B_0\frac{\partial\delta f}{\partial v_\parallel}\\
&= -\delta\mathbf{V}_D\cdot\left(\frac{\partial F_0}{\partial\mathbf{X}'}\right) + \left(\frac{\mathbf{e}_\parallel\times\nabla_X\langle\delta\phi\rangle_\alpha}{B_0} + v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\right)\cdot\left(\frac{\mu}{v_\parallel}\nabla B_0\frac{\partial F_0}{\partial v_\parallel}\right)\\
&-\frac{q}{m}\left[-\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} - \left(v_\parallel\mathbf{e}_\parallel + \frac{v_\parallel^2}{\Omega}\nabla\times\mathbf{b} + \frac{\mu}{\Omega B_0}\mathbf{B}_0\times\nabla B_0 + \frac{1}{B_0^2}\mathbf{E}_0\times\mathbf{B}_0 + \right.\right.\\
&\left.\left. v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\right)\cdot\nabla_X\langle\delta\phi\rangle_\alpha\right]\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel},
\end{aligned} \tag{295}$$

Collecting coefficients before $\partial F_0/\partial v_\parallel$, we find that the two terms involving ∇B_0 (terms in blue and red) cancel each other, yielding

$$\begin{aligned}
&\left[\frac{\partial}{\partial t} + (v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D)\cdot\frac{\partial}{\partial\mathbf{X}'}\right]\delta f - \mathbf{e}_\parallel\cdot\mu\nabla B_0\frac{\partial\delta f}{\partial v_\parallel}\\
&= -\delta\mathbf{V}_D\cdot\left(\frac{\partial F_0}{\partial\mathbf{X}'}\right)\\
&+\frac{q}{m}\left[\frac{m}{q}v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\cdot(\mu\nabla B_0) + \frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} + \left(v_\parallel\mathbf{b} + \frac{v_\parallel^2}{\Omega}\nabla\times\mathbf{b} + \frac{1}{B_0^2}\mathbf{E}_0\times\mathbf{B}_0 + \right.\right.\\
&\left.\left. v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\right)\cdot\nabla_X\langle\delta\phi\rangle_\alpha\right]\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel},
\end{aligned} \tag{296}$$

This equation agrees with Eq. (8) in I. Holod's 2009 pop paper (gyro-averaging is wrongly omitted in that paper) and W. Deng's 2011 NF paper. Equation (296) drops all terms higher than $O(\lambda^2)$ and as a result the coefficient before $\partial\delta f/\partial v_\parallel$ contains only the mirror force, i.e.,

$$\frac{dv_\parallel}{dt} = -\mathbf{e}_\parallel\cdot\mu\nabla B_0, \tag{297}$$

which is independent of any perturbations.

F Transform gyrokinetic equation from ($\delta\Phi$, $\delta\mathbf{A}$) to ($\delta\mathbf{E}$, $\delta\mathbf{B}$)

F.1 Expression of $\delta B_\perp$ in terms of δA

Note that

$$\begin{aligned}\delta\mathbf{B}_\perp &= \nabla\times\delta\mathbf{A}-(\mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A})\mathbf{e}_\parallel \\ &= \nabla\times(\delta\mathbf{A}_\perp+\delta A_\parallel\mathbf{e}_\parallel)-[\mathbf{e}_\parallel\cdot\nabla\times(\delta\mathbf{A}_\perp+\delta A_\parallel\mathbf{e}_\parallel)]\mathbf{e}_\parallel \end{aligned} \tag{298}$$

Correct to order $O(\lambda)$, $\delta\mathbf{B}_\perp$ in the above equation is written as ($\mathbf{e}_\parallel$ vector can be considered as constant because its spatial gradient combined with $\delta\mathbf{A}$ will give terms of $O(\lambda^2)$, which are neglected)

$$\delta\mathbf{B}_\perp \approx \nabla\times\delta\mathbf{A}_\perp+\nabla\delta A_\parallel\times\mathbf{e}_\parallel-[\mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A}_\perp+\mathbf{e}_\parallel\cdot(\nabla\delta A_\parallel\times\mathbf{e}_\parallel)]\mathbf{e}_\parallel \tag{299}$$

$$= \nabla\times\delta\mathbf{A}_\perp+\nabla\delta A_\parallel\times\mathbf{e}_\parallel-(\mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A}_\perp)\mathbf{e}_\parallel \tag{300}$$

Using local cylindrical coordinates (r,ϕ,z) with z being along the local direction of $\mathbf{B}_0$, and two components of $\mathbf{A}_\perp$ being A_r and A_ϕ, then $\nabla\times\mathbf{A}_\perp$ is written as

$$\nabla\times\delta\mathbf{A}_\perp=\left(-\frac{\partial\delta A_\phi}{\partial z}\right)\mathbf{e}_r+\left(\frac{\partial\delta A_r}{\partial z}\right)\mathbf{e}_\phi+\frac{1}{r}\left[\frac{\partial}{\partial r}(r\delta A_\phi)-\frac{\partial\delta A_r}{\partial\phi}\right]\mathbf{e}_\parallel. \tag{301}$$

Note that the parallel gradient operator $\nabla_\parallel\equiv\mathbf{e}_\parallel\cdot\nabla=\partial/\partial z$ acting on the the perturbed quantities will result in quantities of order $O(\lambda^2)$. Retaining terms of order up to $O(\lambda)$, equation (301) is written as

$$\nabla\times\delta\mathbf{A}_\perp\approx\frac{1}{r}\left[\frac{\partial}{\partial r}(r\delta A_\phi)-\frac{\partial\delta A_r}{\partial\phi}\right]\mathbf{e}_\parallel, \tag{302}$$

i.e., only the parallel component survive, which exactly cancels the last term in Eq. (300), i.e., equation (300) is reduced to

$$\delta\mathbf{B}_\perp=\nabla\delta A_\parallel\times\mathbf{e}_\parallel. \tag{303}$$

F.2 Expression of $\delta B_\parallel$ in terms of δA

$$\begin{aligned}\delta B_\parallel &= \mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A} \\ &= \mathbf{e}_\parallel\cdot\nabla\times(\delta\mathbf{A}_\perp+\delta A_\parallel\mathbf{e}_\parallel)\end{aligned} \tag{304}$$

Accurate to $O(\lambda^1)$, $\delta B_\parallel$ in the above equation is written as ($\mathbf{e}_\parallel$ vector can be considered as constant because its spatial gradient combined with $\delta\mathbf{A}$ will give $O(\lambda^2)$ terms, which are neglected)

$$\begin{aligned}\delta B_\parallel &\approx \mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A}_\perp+\mathbf{e}_\parallel\cdot(\nabla\delta A_\parallel\times\mathbf{e}_\parallel) \\ &= \mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A}_\perp\end{aligned} \tag{305}$$

[Using local cylindrical coordinates (r, ϕ, z) with z being along the local direction of $\mathbf{B}_0$, and two components of $\delta\mathbf{A}_\perp$ being δA_r and δA_ϕ, then $\nabla \times \delta\mathbf{A}_\perp$ is written as

$$\nabla \times \delta\mathbf{A}_\perp = \left(-\frac{\partial \delta A_\phi}{\partial z}\right)\mathbf{e}_r + \left(\frac{\partial \delta A_r}{\partial z}\right)\mathbf{e}_r + \frac{1}{r}\left[\frac{\partial}{\partial r}(r\delta A_\phi) - \frac{\partial \delta A_r}{\partial \phi}\right]\mathbf{e}_\parallel \tag{306}$$

Note that the parallel gradient operator $\nabla_\parallel \equiv \mathbf{e}_\parallel \cdot \nabla = \partial/\partial z$ acting on the the perturbed quantities will result in quantities of order $O(\lambda^2)$. Retaining terms of order up to $O(\lambda)$, equation (301) is written as

$$\nabla \times \delta\mathbf{A}_\perp \approx \frac{1}{r}\left[\frac{\partial}{\partial r}(r\delta A_\phi) - \frac{\partial \delta A_r}{\partial \phi}\right]\mathbf{e}_\parallel, \tag{307}$$

Using this, equation (305) is written as

$$\delta B_\parallel = \frac{1}{r}\left[\frac{\partial}{\partial r}(r\delta A_\phi) - \frac{\partial \delta A_r}{\partial \phi}\right]. \tag{308}$$

However, this expression is not useful for GEM because GEM does not use the local coordinates (r, ϕ, z).]

F.3 Expressing the perturbed drift in terms of δE and δB

The perturbed drift $\delta\mathbf{V}_D$ is given by Eq. (132), i.e.,

$$\delta\mathbf{V}_D = -\frac{q}{m}\nabla_X \langle \delta L\rangle_\alpha \times \frac{\mathbf{e}_\parallel}{\Omega}. \tag{309}$$

Using $\delta L = \delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}$, the above expression can be further written as

$$\begin{aligned}
\delta\mathbf{V}_D &= -\frac{q}{m}\nabla_X \langle \delta\Phi - \mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha \times \frac{\mathbf{e}_\parallel}{\Omega} \\
&= \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega} \times \nabla_X \langle \delta\Phi\rangle_\alpha - \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega} \times \nabla_X \langle v_\parallel \delta A_\parallel\rangle_\alpha \\
&\quad -\frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega} \times \nabla_X \langle \mathbf{v}_\perp \cdot \delta\mathbf{A}_\perp\rangle_\alpha.
\end{aligned} \tag{310}$$

Accurate to order $O(\lambda)$, the term involving $\delta\Phi$ is

$$\begin{aligned}
\frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega} \times \nabla_X \langle \delta\Phi\rangle_\alpha &= \frac{\mathbf{e}_\parallel}{B_0} \times \langle \nabla_X \delta\Phi\rangle_\alpha \\
&\approx \frac{\mathbf{e}_\parallel}{B_0} \times \langle \nabla_x \delta\Phi\rangle_\alpha \\
&\approx \frac{\mathbf{e}_\parallel}{B_0} \times \left\langle -\delta\mathbf{E} - \frac{\partial \delta\mathbf{A}}{\partial t}\right\rangle_\alpha \\
&\approx \frac{\mathbf{e}_\parallel}{B_0} \times \langle -\delta\mathbf{E}\rangle_\alpha \\
&\equiv \delta\mathbf{V}_E,
\end{aligned} \tag{311}$$

which is the $\delta\mathbf{E}\times\mathbf{B}_0$ drift. Accurate to $O(\lambda)$, the $\langle v_\parallel \delta A_\parallel\rangle_\alpha$ term on the right-hand side of Eq. (310) is written

$$\begin{aligned}
-\frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle v_\parallel\delta A_\parallel\rangle_\alpha &\approx -\frac{q}{m}\frac{1}{\Omega}\langle\mathbf{e}_\parallel\times\nabla_{\mathbf{X}}(v_\parallel\delta A_\parallel)\rangle_\alpha\\
&\approx -\frac{q}{m}\frac{1}{\Omega}\langle\mathbf{e}_\parallel\times\nabla_{\mathbf{x}}(v_\parallel\delta A_\parallel)\rangle_\alpha\\
&\approx -\frac{q}{m}\frac{v_\parallel}{\Omega}\langle\mathbf{e}_\parallel\times\nabla_{\mathbf{x}}(\delta A_\parallel)\rangle_\alpha\\
&= v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0},
\end{aligned}\tag{312}$$

which is due to the magnetic fluttering (this is actually not a real drift). In obtaining the last equality, use has been made of Eq. (303), i.e., $\delta\mathbf{B}_\perp=\nabla_{\mathbf{x}}\delta A_\parallel\times\mathbf{e}_\parallel$.

Accurate to $O(\lambda)$, the last term on the right-hand side of expression (310) is written

$$\begin{aligned}
-\frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}_\perp\cdot\delta\mathbf{A}_\perp\rangle_\alpha &\approx -\frac{1}{B_0}\langle\mathbf{e}_\parallel\times\nabla_X(\mathbf{v}_\perp\cdot\delta\mathbf{A}_\perp)\rangle_\alpha\\
&\approx -\frac{1}{B_0}\langle\mathbf{e}_\parallel\times\nabla_x(\mathbf{v}_\perp\cdot\delta\mathbf{A}_\perp)\rangle_\alpha\\
&= -\frac{1}{B_0}\langle\mathbf{e}_\parallel\times(\mathbf{v}_\perp\times\nabla_x\times\delta\mathbf{A}_\perp+\mathbf{v}_\perp\cdot\nabla_x\delta\mathbf{A}_\perp)\rangle_\alpha\\
&= -\frac{1}{B_0}\langle(\mathbf{e}_\parallel\cdot\nabla_x\times\delta\mathbf{A}_\perp)\mathbf{v}_\perp+\mathbf{e}_\parallel\times\mathbf{v}_\perp\cdot\nabla_x\delta\mathbf{A}_\perp\rangle_\alpha
\end{aligned}$$

Using equation (305), i.e., $\delta B_\parallel=\mathbf{e}_\parallel\cdot\nabla\times\delta\mathbf{A}_\perp$, the above expression is written as

$$\begin{aligned}
-\frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}_\perp\cdot\delta\mathbf{A}_\perp\rangle_\alpha &= -\frac{1}{B_0}\langle\delta B_\parallel\mathbf{v}_\perp+\mathbf{e}_\parallel\times\mathbf{v}_\perp\cdot\nabla_x\delta\mathbf{A}_\perp\rangle_\alpha\\
&\approx -\frac{1}{B_0}\langle\delta B_\parallel\mathbf{v}_\perp+\mathbf{e}_\parallel\times\mathbf{v}_\perp\cdot\nabla_X\delta\mathbf{A}_\perp\rangle_\alpha\\
&\approx -\frac{1}{B_0}\langle\delta B_\parallel\mathbf{v}_\perp\rangle_\alpha-\frac{1}{B_0}\mathbf{e}_\parallel\times\langle\mathbf{v}_\perp\cdot\nabla_X\delta\mathbf{A}_\perp\rangle_\alpha.\\
&\approx -\frac{1}{B_0}\langle\delta B_\parallel\mathbf{v}_\perp\rangle_\alpha.
\end{aligned}\tag{313}$$

where use has been made of $\langle\mathbf{v}_\perp\cdot\nabla_X\delta\mathbf{A}_\perp\rangle_\alpha\approx 0$ (**seems wrong**), where the error is of $O(\lambda)\delta\mathbf{A}_\perp$. The term $\langle\delta B_\parallel\mathbf{v}_\perp\rangle_\alpha/B_0$ is of $O(\lambda^2)$ and thus can be neglected (I need to verify this).

Using Eqs. (311), (312), and (313), expression (310) is finally written as

$$\delta\mathbf{V}_D\equiv-\frac{q}{m}\nabla_X\langle\delta L\rangle_\alpha\times\frac{\mathbf{e}_\parallel}{\Omega}=\frac{\langle\delta\mathbf{E}\rangle_\alpha\times\mathbf{e}_\parallel}{B_0}+v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}.\tag{314}$$

Using this, the first equation of the characteristics, equation (129), is written as

$$\begin{aligned}\frac{d\mathbf{X}}{dt} &= v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \delta\mathbf{V}_D \qquad (315)\\ &= v_\parallel \mathbf{e}_\parallel + \mathbf{V}_D + \frac{\langle\delta\mathbf{E}\rangle_\alpha \times \mathbf{e}_\parallel}{B_0} + v_\parallel \frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0}\\ &\equiv \mathbf{V}_G \qquad (316)\end{aligned}$$

F.4 Expressing the coefficient before $\partial F_0 / \partial \varepsilon$ in terms of δE and δB

[Note that

$$\frac{\partial \delta\mathbf{A}_\perp}{\partial t} = -(\delta\mathbf{E}_\perp + \nabla_\perp \delta\Phi), \qquad (317)$$

where $\partial\delta\mathbf{A}_\perp/\partial t$ is of $O(\lambda^2)$. This means that $\delta\mathbf{E}_\perp + \nabla_\perp\delta\phi$ is of $O(\lambda^2)$ although both $\delta\mathbf{E}_\perp$ and $\delta\phi$ are of $O(\lambda)$.]

Note that

$$\begin{aligned}\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} &= v_\parallel \frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t} + \mathbf{v}_\perp \cdot \frac{\partial\langle\delta\mathbf{A}\rangle_\alpha}{\partial t}\\ &= v_\parallel \frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t} + \langle\mathbf{v}_\perp\cdot(-\delta\mathbf{E} - \nabla\delta\Phi)\rangle_\alpha\\ &\approx v_\parallel \frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t} - \langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha \qquad (318)\end{aligned}$$

where use has been made of $\langle\mathbf{v}_\perp\cdot\nabla\delta\phi\rangle \approx 0$, This indicates that $\langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha$ is of $O(\lambda^1)\delta\mathbf{E}$. Using Eq. (318), the coefficient before $\partial F_0/\partial\varepsilon$ in Eq. (148) can be further written as

$$\begin{aligned}&-\frac{q}{m}\Big[-\frac{\partial\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha}{\partial t} - \Big(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D - \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\Big)\cdot\nabla_X\langle\delta\Phi\rangle_\alpha\Big]\\ &=-\frac{q}{m}\Big[-v_\parallel\frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t} + \langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha - \Big(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D - \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\Big)\cdot\\ &\Big\langle -\delta\mathbf{E} - \frac{\partial\delta\mathbf{A}}{\partial t}\Big\rangle_\alpha\Big]\\ &\approx-\frac{q}{m}\Big[-v_\parallel\frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t} + \langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha - \Big(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D - \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\Big)\cdot\\ &\langle-\delta\mathbf{E}\rangle_\alpha + v_\parallel\Big\langle\frac{\partial A_\parallel}{\partial t}\Big\rangle_\alpha\Big]\\ &=-\frac{q}{m}\Big[\langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha + \Big(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D - \frac{q}{m}\frac{\mathbf{e}_\parallel}{\Omega}\times\nabla_X\langle\mathbf{v}\cdot\delta\mathbf{A}\rangle_\alpha\Big)\cdot\langle\delta\mathbf{E}\rangle_\alpha\Big]\\ &\approx-\frac{q}{m}\Big[\langle\mathbf{v}_\perp\cdot\delta\mathbf{E}\rangle_\alpha + \Big(v_\parallel\mathbf{e}_\parallel + \mathbf{V}_D + v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle}{B_0}\Big)\cdot\langle\delta\mathbf{E}\rangle_\alpha\Big]. \qquad (319)\end{aligned}$$

Using Eq. (319) and (), gyrokinetic equation (148) is finally written as

$$\begin{aligned}&\left[\frac{\partial}{\partial t}+\left(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{V}_D+\frac{\langle\delta\mathbf{E}\rangle_\alpha\times\mathbf{e}_{\parallel}}{B_0}+v_{\parallel}\frac{\langle\delta\mathbf{B}_{\perp}\rangle_\alpha}{B_0}\right)\cdot\nabla_X\right]\delta f\\&=-\left(\frac{\langle\delta\mathbf{E}\rangle_\alpha\times\mathbf{e}_{\parallel}}{B_0}+v_{\parallel}\frac{\langle\delta\mathbf{B}_{\perp}\rangle_\alpha}{B_0}\right)\cdot\nabla_X F_0\\&-\frac{q}{m}\left[\langle\mathbf{v}_{\perp}\cdot\delta\mathbf{E}\rangle_\alpha+\left(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{V}_D+v_{\parallel}\frac{\langle\delta\mathbf{B}_{\perp}\rangle_\alpha}{B_0}\right)\cdot\langle\delta\mathbf{E}\rangle_\alpha\right]\frac{\partial F_0}{\partial\varepsilon}.\end{aligned}\tag{320}$$

G Drift-kinetic limit

In the drift-kinetic limit, $\langle\mathbf{v}_{\perp}\cdot\delta\mathbf{E}\rangle_\alpha=0$, $\langle\delta B_{\parallel}\mathbf{v}_{\perp}\rangle_\alpha=0$, and $\langle\delta h\rangle_\alpha=\delta h$, where δh is an arbitrary field quantity. Using these, gyrokinetic equation (320) is written as

$$\begin{aligned}&\left[\frac{\partial}{\partial t}+\left(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D+\mathbf{v}_E+v_{\parallel}\frac{\delta\mathbf{B}_{\perp}}{B_0}\right)\cdot\nabla_X\right]\delta f\\&=-\left(\mathbf{v}_E+v_{\parallel}\frac{\delta\mathbf{B}_{\perp}}{B_0}\right)\cdot\nabla_X F_0-\frac{q}{m}\left[\left(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D+v_{\parallel}\frac{\delta\mathbf{B}_{\perp}}{B_0}\right)\cdot\delta\mathbf{E}\right]\frac{\partial F_0}{\partial\varepsilon}.\end{aligned}\tag{321}$$

G.1 Linear case

Neglecting the nonlinear terms, drift-kinetic equation (321) is written

$$\begin{aligned}&\left[\frac{\partial}{\partial t}+(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D)\cdot\nabla_X\right]\delta f\\&=-\left(\mathbf{v}_E+v_{\parallel}\frac{\delta\mathbf{B}_{\perp}}{B_0}\right)\cdot\nabla_X F_0-\frac{q}{m}[(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D)\cdot\delta\mathbf{E}]\frac{\partial F_0}{\partial\varepsilon}.\end{aligned}\tag{322}$$

Next let us derive the parallel momentum equation from the linear drift kinetic equation (this is needed in my simulation). Multiplying the linear drift kinetic equation (322) by $qv_{\parallel}$ and then integrating over velocity space, we obtain

$$\begin{aligned}\frac{\partial\delta j_{\parallel}}{\partial t}&=-q\int d\mathbf{v}v_{\parallel}(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D)\cdot\nabla_X\delta f\\&-q\int d\mathbf{v}v_{\parallel}\left(\mathbf{v}_E+v_{\parallel}\frac{\delta\mathbf{B}_{\perp}}{B_0}\right)\cdot\nabla_X F_0-\frac{q}{m}q\int d\mathbf{v}v_{\parallel}[(v_{\parallel}\mathbf{e}_{\parallel}+\mathbf{v}_D)\cdot\\&\delta\mathbf{E}]\frac{\partial F_0}{\partial\varepsilon}.\end{aligned}\tag{323}$$

Equation (323) involve $\nabla_X \delta f$ and this should be avoided in particle methods whose goal is to avoid directly evaluating the derivatives of δf over phase-space coordinates. On the other hand, the partial derivatives of velocity moment of δf are allowed. Therefore, we would like to make the velocity integration of δf appear. Note that $\nabla_X \delta f$ here is taken by holding (ε, μ) constant and thus $v_\parallel$ is not a constant and thus can not be moved inside ∇_X. Next, to facilitate performing the integration over $v_\parallel$, we transform the linear drift kinetic equation (322) into variable $(\mathbf{X}, \mu, v_\parallel)$.

G.2 Transform from $(\mathbf{X}, \mu, \varepsilon)$ to $(\mathbf{X}, \mu, v_\parallel)$ coordinates

The kinetic equation given above is written in terms of variable $(\mathbf{X}, \mu, \varepsilon)$. Next, we transform it into coordinates $(\mathbf{X}', \mu', v_\parallel)$ which is defined by

$$\mathbf{X}'(\mathbf{X}, \mu, \varepsilon) = \mathbf{X}, \tag{324}$$

$$\mu'(\mathbf{X}, \mu, \varepsilon) = \mu, \tag{325}$$

and

$$v_\parallel(\mathbf{X}, \mu, \varepsilon) = \sqrt{2(\varepsilon - \mu B_0(\mathbf{X}))}. \tag{326}$$

Use this, we have

$$\begin{aligned} \frac{\partial \delta G_0}{\partial \mathbf{X}}|_{\mu,\varepsilon} &= \frac{\partial \mathbf{X}'}{\partial \mathbf{X}} \frac{\partial \delta G_0}{\partial \mathbf{X}'} + \frac{\partial \mu'}{\partial \mathbf{X}} \frac{\partial \delta G_0}{\partial \mu'} + \frac{\partial v_\parallel}{\partial \mathbf{X}} \frac{\partial \delta G_0}{\partial v_\parallel} \\ &= \frac{\partial \delta G_0}{\partial \mathbf{X}'}|_{\mu, v_\parallel} + 0 \frac{\partial \delta G_0}{\partial \mu'} - \frac{\mu}{v_\parallel} \frac{\partial B_0}{d\mathbf{X}} \frac{\partial \delta G_0}{\partial v_\parallel}, \end{aligned} \tag{327}$$

and

$$\begin{aligned} \frac{\partial F_0}{\partial \varepsilon} &= \frac{\partial F_0}{\partial \mu'} \frac{\partial \mu'}{\partial \varepsilon} + \frac{\partial F_0}{\partial v_\parallel} \frac{\partial v_\parallel}{\partial \varepsilon} \\ &= 0 \frac{\partial F_0}{\partial \mu'} + \frac{\partial F_0}{\partial v_\parallel} \frac{\partial v_\parallel}{\partial \varepsilon} \\ &= \frac{\partial F_0}{\partial v_\parallel} \frac{1}{v_\parallel} \end{aligned} \tag{328}$$

Then, in terms of variable $(\mathbf{X}', \mu, v_\parallel)$, equation (322) is written

$$\begin{aligned} &\frac{\partial \delta f}{\partial t} + (v_\parallel \mathbf{e}_\parallel + \mathbf{v}_D) \cdot \nabla \delta f - \mathbf{e}_\parallel \cdot \mu \nabla B \frac{\partial \delta f}{\partial v_\parallel} \\ &= -\left(\mathbf{v}_E + v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0} \right) \cdot \nabla F_0 + \left(\frac{\mathbf{v}_E}{v_\parallel} + \frac{\delta \mathbf{B}_\perp}{B_0} \right) \cdot \mu \nabla B \frac{\partial F_0}{\partial v_\parallel} - \frac{q}{m} \left[\left(\mathbf{e}_\parallel + \frac{\mathbf{v}_D}{v_\parallel} \right) \cdot \right. \\ &\left. \delta \mathbf{E} \right] \frac{\partial F_0}{\partial v_\parallel}, \end{aligned} \tag{329}$$

where $\nabla \equiv \partial/\partial \mathbf{X}'|_{\mu,v_\parallel}$.

G.3 Parallel momentum equation

Multiplying the linear drift kinetic equation (329) by $qv_\parallel$ and then integrating over velocity space, we obtain

$$\begin{aligned}
&\frac{\partial \delta j_\parallel}{\partial t} + q\int d\mathbf{v} v_\parallel (v_\parallel \mathbf{e}_\parallel + \mathbf{v}_D)\cdot \nabla_X \delta f - q\int d\mathbf{v} v_\parallel \mathbf{e}_\parallel \cdot \mu \nabla B \frac{\partial \delta f}{\partial v_\parallel} \\
=&-q\int d\mathbf{v} v_\parallel \left(\mathbf{v}_E + v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot \nabla_X F_0 + q\int d\mathbf{v} v_\parallel \left(\frac{\mathbf{v}_E}{v_\parallel} + \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot \mu \nabla B \frac{\partial F_0}{\partial v_\parallel} \\
&-\frac{q}{m} q\int d\mathbf{v} v_\parallel \left[\left(\mathbf{e}_\parallel + \frac{\mathbf{v}_D}{v_\parallel}\right)\cdot \delta \mathbf{E}\right]\frac{\partial F_0}{\partial v_\parallel}. \qquad (330)
\end{aligned}$$

Consider the simple case that F_0 does not carry current, i.e., $F_0(\mathbf{X}, \mu, v_\parallel)$ is an even function about $v_\parallel$. Then it is obvious that the integration of the terms in red in Eq. (330) are all zero. Among the rest terms, only the following term

$$-\frac{q}{m} q\int d\mathbf{v} v_\parallel [(v_\parallel \mathbf{e}_\parallel)\cdot \delta \mathbf{E}]\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel} \qquad (331)$$

explicitly depends on $\delta \mathbf{E}$. Using $d\mathbf{v} = 2\pi B dv_\parallel d\mu$, the integration in the above expression can be analytically performed, giving

$$\begin{aligned}
&-\frac{q}{m} q\int d\mathbf{v} v_\parallel [(v_\parallel \mathbf{e}_\parallel)\cdot \delta \mathbf{E}]\frac{\partial F_0}{\partial v_\parallel}\frac{1}{v_\parallel} \\
=&-\frac{q^2}{m}\int 2\pi B dv_\parallel d\mu v_\parallel \delta E_\parallel \frac{\partial F_0}{\partial v_\parallel} \\
=&-\frac{q^2}{m}\int 2\pi B d\mu \delta E_\parallel \int v_\parallel \frac{\partial F_0}{\partial v_\parallel} dv_\parallel \\
=&-\frac{q^2}{m}\int 2\pi B d\mu \delta E_\parallel \left(0 - \int F_0 dv_\parallel\right) \\
=&\frac{q^2}{m}\delta E_\parallel n_0. \qquad (332)
\end{aligned}$$

Using these results, the parallel momentum equation (330) is written

$$\begin{aligned}
\frac{\partial \delta j_\parallel}{\partial t} \;=\; & \frac{q^2}{m}\delta E_\parallel n_0 - q\int d\mathbf{v} v_\parallel (v_\parallel \mathbf{e}_\parallel + \mathbf{v}_D)\cdot \nabla_X \delta f + q\int d\mathbf{v} v_\parallel \mathbf{e}_\parallel \cdot \mu \nabla B \frac{\partial \delta f}{\partial v_\parallel} \\
& -q\int d\mathbf{v} v_\parallel \left(v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot \nabla F_0 + q\int d\mathbf{v} v_\parallel \left(\frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot \mu \nabla B \frac{\partial F_0}{\partial v_\parallel}, \qquad (333)
\end{aligned}$$

where the explicit dependence on $\delta \mathbf{E}$ is via the first term $q^2 n_0 \delta E_\parallel / m$, with all the the other terms being explicitly independent of $\delta \mathbf{E}$ (δf and $\delta \mathbf{B}$ implicitly depend on $\delta \mathbf{E}$).

Equation (333) involve derivatives of δf with respect to space and $v_\parallel$ and these should be avoided in the particle method whose goal is to avoid directly evaluating these derivatives. Using integration by parts, the terms involving $\partial/\partial v_\parallel$ can be simplified, yielding

$$\begin{aligned}\frac{\partial \delta j_\parallel}{\partial t} &= \frac{q^2}{m}\delta E_\parallel n_0 - q\int d\mathbf{v} v_\parallel (v_\parallel \mathbf{e}_\parallel + \mathbf{v}_D)\cdot\nabla_X \delta f - q(\mathbf{e}_\parallel\cdot\nabla B_0)\int \mu \delta f d\mathbf{v} \\ &\quad -q\int d\mathbf{v} v_\parallel \left(v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot\nabla F_0 - q\left(\frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot(\nabla B_0)\int \mu F_0\, d\mathbf{v}, \end{aligned} \tag{334}$$

Define $p_{\perp 0} = \int m v_\perp^2 F_0/2 d\mathbf{v}$ and $\delta p_\perp = \int m v_\perp^2 \delta f/2 d\mathbf{v}$, then the above equation is written

$$\begin{aligned}\frac{\partial \delta j_\parallel}{\partial t} &= \frac{q^2}{m}\delta E_\parallel n_0 - q\int d\mathbf{v} v_\parallel (v_\parallel \mathbf{e}_\parallel + \mathbf{v}_D)\cdot\nabla_X \delta f - q(\mathbf{e}_\parallel\cdot\nabla B_0)\frac{\delta p_\perp}{m B_0} \\ &\quad -q\int d\mathbf{v} v_\parallel \left(v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot\nabla F_0 - q\left(\frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot(\nabla B_0)\frac{p_{\perp 0}}{m B_0}, \end{aligned} \tag{335}$$

Next, we try to eliminate the spatial gradient of δf by changing the order of integration. The second term on the right-hand side of Eq. (335) is written

$$\begin{aligned}&-q\int d\mathbf{v} v_\parallel (v_\parallel \mathbf{e}_\parallel)\cdot\nabla_X \delta f, \\ =&-q\int 2\pi B_0\, dv_\parallel d\mu v_\parallel^2 \mathbf{e}_\parallel\cdot\nabla_X \delta f \\ =&-q 2\pi B_0\, \mathbf{e}_\parallel\cdot\nabla_X \int v_\parallel^2 \delta f dv_\parallel d\mu \\ =&-q B_0\, \mathbf{e}_\parallel\cdot\nabla_X\left(\frac{1}{m B_0}\int m v_\parallel^2 \delta f d\mathbf{v}\right) \\ =&-q B_0\, \mathbf{e}_\parallel\cdot\nabla_X\left(\frac{\delta p_\parallel}{m B_0}\right), \end{aligned} \tag{336}$$

where $\delta p_\parallel = \int m v_\parallel^2 \delta f d\mathbf{v}$. Similarly, the term $-q\int d\mathbf{v} v_\parallel \left(v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot\nabla_X F_0$ is written as

$$\begin{aligned}&-q\int d\mathbf{v} v_\parallel \left(v_\parallel \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot\nabla_X F_0 \\ =&-q\int 2\pi B_0\, dv_\parallel d\mu \left(v_\parallel^2 \frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot\nabla_X F_0 \\ =&-q\left(\frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot B_0 \nabla_X \int (v_\parallel^2 F_0 2\pi dv_\parallel d\mu) \\ =&-q\left(\frac{\delta \mathbf{B}_\perp}{B_0}\right)\cdot B_0\, \nabla_X\left[\frac{1}{B}\int (m v_\parallel^2 F_0 d\mathbf{v})\right] \\ =&-q\delta \mathbf{B}_\perp\cdot\nabla_X\left(\frac{p_{\parallel 0}}{m B_0}\right) \end{aligned} \tag{337}$$

where $p_{\parallel 0}=\int m v_\parallel^2 F_0 d\mathbf{v}$. Similarly, the term $-q\int d\mathbf{v} v_\parallel \mathbf{v}_D\cdot\nabla_X\delta f$ can be written as the gradient of moments of δf. Let us work on this. The drift $\mathbf{v}_D$ is given by

$$\mathbf{v}_D=\frac{B_0\frac{v_\parallel}{\Omega}\nabla\times\mathbf{b}}{B_\parallel^\star}v_\parallel+\frac{\mu}{\Omega B_\parallel^\star}\mathbf{B}_0\times\nabla B_0. \tag{338}$$

where $B_\parallel^\star=B_0\big(1+\frac{v_\parallel}{\Omega}\mathbf{b}\cdot\nabla\times\mathbf{b}\big)$ (refer to my another notes). Using $\mathbf{b}\cdot\nabla\times\mathbf{b}\approx 0$, we obtain $B_\parallel^\star\approx B$. Then $\mathbf{v}_D$ is written

$$\mathbf{v}_D=\frac{v_\parallel^2}{\Omega}\nabla\times\mathbf{b}+\frac{\mu}{\Omega}\mathbf{b}\times\nabla B_0.$$

Using this and $d\mathbf{v}=2\pi B_0\, dv_\parallel d\mu$, the term $-q\int d\mathbf{v} v_\parallel\mathbf{v}_D\cdot\nabla_X\delta f$ is written as

$$\begin{aligned}
-q\int d\mathbf{v} v_\parallel\mathbf{v}_D\cdot\nabla_X\delta f &= -q\int 2\pi B_0\, dv_\parallel d\mu\, v_\parallel\left(\frac{v_\parallel^2}{\Omega}\nabla\times\mathbf{b}+\frac{\mu}{\Omega}\mathbf{b}\times\nabla B_0\right)\cdot\nabla_X\delta f\\
&= -q2\pi B_0\,\frac{1}{\Omega}(\nabla\times\mathbf{b})\cdot\nabla_X\int v_\parallel^3\delta f dv_\parallel d\mu - q2\pi B_0\frac{1}{\Omega}(\mathbf{b}\times\\
&\quad \nabla B_0)\cdot\nabla_X\int v_\parallel\mu\delta f dv_\parallel d\mu\\
&= -qB_0\,\frac{1}{\Omega}(\nabla\times\mathbf{b})\cdot\nabla_X\left(\frac{1}{B_0}\int v_\parallel^3\delta f d\mathbf{v}\right) - qB_0\frac{1}{\Omega}(\mathbf{b}\times\\
&\quad \nabla B_0)\cdot\nabla_X\left(\frac{1}{B_0}\int v_\parallel\mu\delta f d\mathbf{v}\right),\\
&= -m(\nabla\times\mathbf{b})\cdot\nabla_X\left(\frac{1}{B_0}\int v_\parallel^3\delta f d\mathbf{v}\right) - m(\mathbf{b}\times\\
&\quad \nabla B_0)\cdot\nabla_X\left(\frac{1}{B_0}\int v_\parallel\mu\delta f d\mathbf{v}\right),
\end{aligned}\tag{339}$$

which are the third order moments of δf and may be neglect-able (a guess, not verified). Using the above results, the linear parallel momentum equation is finally written

$$\begin{aligned}
\frac{\partial\delta j_\parallel}{\partial t} &= \frac{e^2 n_{e0}}{m}\delta E_\parallel + eB_0\,\mathbf{b}\cdot\nabla_X\left(\frac{\delta p_\parallel}{mB_0}\right)+e(\mathbf{b}\cdot\nabla B_0)\frac{\delta p_\perp}{mB_0}\\
&\quad +e\delta\mathbf{B}_\perp\cdot\nabla_X\left(\frac{p_{\parallel 0}}{mB_0}\right)+e\left(\frac{\delta\mathbf{B}_\perp}{B_0}\right)\cdot(\nabla B_0)\frac{p_{\perp 0}}{mB_0}.\\
&\quad -m(\nabla\times\mathbf{b})\cdot\nabla_X\left(\frac{1}{B_0}\int v_\parallel^3\delta f d\mathbf{v}\right) - m(\mathbf{b}\times\nabla B_0)\cdot\nabla_X\left(\frac{1}{B_0}\int\right.\\
&\quad \left. v_\parallel\mu\delta f d\mathbf{v}\right)
\end{aligned}\tag{340}$$

Define

$$\mathbf{D}_0 = \nabla\left(\frac{p_{\parallel 0}}{m B_0}\right) + \frac{\nabla B_0}{B_0}\frac{p_{\perp 0}}{m B_0}, \tag{341}$$

which, for the isotropic case ($p_{\parallel 0} = p_{\perp 0} = p_0$), is simplified to

$$\mathbf{D}_0 = \frac{\nabla p_0}{m B_0}. \tag{342}$$

then Eq. (340) is written as

$$\begin{aligned}\frac{\partial \delta j_\parallel}{\partial t} &= \frac{e^2 n_0}{m}\delta E_\parallel + e\delta\mathbf{B}_\perp \cdot \mathbf{D}_0 \\ &+ e B_0\, \mathbf{b}\cdot\nabla_X\left(\frac{\delta p_\parallel}{m B_0}\right) + e(\mathbf{b}\cdot\nabla B_0)\frac{\delta p_\perp}{m B_0} \\ &- m(\nabla \times \mathbf{b}) \cdot \nabla_X\left(\frac{1}{B_0}\int v_\parallel^3 \delta f d\mathbf{v}\right) - m(\mathbf{b}\times\nabla B_0)\cdot\nabla_X\left(\frac{1}{B_0}\int \right. \\ &\left. v_\parallel \mu \delta f d\mathbf{v}\right).\end{aligned} \tag{343}$$

G.4 Special case in uniform magnetic field

In the case of uniform magnetic field, the parallel momentum equation (340) is written as

$$\frac{\partial \delta j_\parallel}{\partial t} = \frac{q}{m} q E_\parallel n_{e0} - q\,\mathbf{e}_\parallel \cdot \nabla_X(\delta p_\parallel) - q\frac{\delta \mathbf{B}_\perp}{B_0}\cdot\nabla_X p_{\parallel 0}. \tag{344}$$

G.5 Electron perpendicular flow

Using the gyrokinetic theory and taking the drift-kinetic limit, the perturbed perpendicular electron flow, $\delta\mathbf{V}_{e\perp}$, is written (see Sec. D or Appendix in Yang Chen's paper[2])

$$n_{e0}\delta\mathbf{V}_{e\perp} = \underbrace{\frac{n_{e0}}{B_0}\delta\mathbf{E}\times\mathbf{b}}_{E\times B\,\text{flow}} \underbrace{- \frac{1}{e B_0}\mathbf{b}\times\nabla\delta p_{\perp e}}_{\text{diamagneticflow}} \tag{345}$$

where n_{e0} is the equilibrium electron number density, $\delta p_{e\perp}$ is the perturbed perpendicular pressure of electrons.

G.5.1 Drift kinetic equation

Drift kinetic equation is written

$$\frac{\partial f}{\partial t}+\left(v_{\parallel}\tilde{\mathbf{b}}+\mathbf{v}_D+\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\right)\cdot\nabla f+\left(-\frac{e}{m}\delta E_{\parallel}-\mu\tilde{\mathbf{b}}\cdot\nabla B\right)\frac{\partial f}{\partial v_{\parallel}}=0, \tag{346}$$

where $f=f(\mathbf{x},\mu,v_{\parallel},t)$, $\mu=mv_{\perp}^2/B_0$ is the magnetic moment, $\tilde{\mathbf{b}}=\mathbf{b}+\delta\mathbf{B}_{\perp}/B_0$, $\mathbf{b}=\mathbf{B}_0/B_0$ is the unit vector along the equilibrium magnetic field, $\mathbf{v}_D=\mathbf{v}_D(\mathbf{x},\mu,v_{\parallel})$ is the guiding-center drift in the equilibrium magnetic field. $\delta\mathbf{E}$ and $\delta\mathbf{B}$ are the perturbed electric field and magnetic field, respectively.

G.5.2 Parallel momentum equation

Multiplying the drift kinetic equation () by $v_{\parallel}$ and then integrating over velocity space, we obtain

$$\int\frac{\partial f_e v_{\parallel}}{\partial t}d\mathbf{v}+\int v_{\parallel}\left(v_{\parallel}\tilde{\mathbf{b}}+\mathbf{v}_D+\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\right)\cdot\nabla f_{\ e}d\mathbf{v}+\int v_{\parallel}\left(-\frac{e}{m}\delta E_{\parallel}-\mu\tilde{\mathbf{b}}\cdot\nabla B\right)\frac{\partial f_e}{\partial v_{\parallel}}d\mathbf{v}=0, \tag{347}$$

which can be written as

$$\frac{\partial J_{\parallel e}}{\partial t}+\int v_{\parallel}\left(v_{\parallel}\tilde{\mathbf{b}}+\mathbf{v}_D+\frac{\delta\mathbf{E}\times\mathbf{e}_{\parallel}}{B_0}\right)\cdot\nabla f_e\,d\mathbf{v}+\int v_{\parallel}\left(-\frac{e}{m}\delta E_{\parallel}-\mu\tilde{\mathbf{b}}\cdot\nabla B\right)\frac{\partial f_e}{\partial v_{\parallel}}d\mathbf{v}=0, \tag{348}$$

Using $d\mathbf{v}=B^{-1}2\pi m\,dv_{\parallel}d\mu$, the last term on the RHS of the above equation is written

$$\begin{aligned}
&\iint v_{\parallel}\left(-\frac{e}{m}\delta E_{\parallel}-\mu\tilde{\mathbf{b}}\cdot\nabla B\right)\frac{\partial f_e}{\partial v_{\parallel}}d\mathbf{v}\\
=&\iint v_{\parallel}\left(-\frac{e}{m}\delta E_{\parallel}-\mu\tilde{\mathbf{b}}\cdot\nabla B\right)\frac{\partial f_e}{\partial v_{\parallel}}2\pi\frac{B}{m}\,dv_{\parallel}d\mu\\
=&-\frac{e}{m}\delta E_{\parallel}2\pi\frac{B}{m}\iint v_{\parallel}\frac{\partial f_e}{\partial v_{\parallel}}\,dv_{\parallel}d\mu-(\tilde{\mathbf{b}}\cdot\nabla B)2\pi\frac{B}{m}\int\mu\int v_{\parallel}\frac{\partial f_e}{\partial v_{\parallel}}\,dv_{\parallel}d\mu\\
=&-\frac{e}{m}\delta E_{\parallel}2\pi\frac{B}{m}\int\left(v_{\parallel}f_e|_{-\infty}^{+\infty}-\int f_e dv_{\parallel}\right)d\mu-(\tilde{\mathbf{b}}\cdot\nabla B)2\pi\frac{B}{m}\int\mu\left(v_{\parallel}f_e|_{-\infty}^{+\infty}-\int f_e dv_{\parallel}\right)d\mu\\
=&\frac{e}{m}\delta E_{\parallel}2\pi\frac{B}{m}\iint f_e dv_{\parallel}d\mu+(\tilde{\mathbf{b}}\cdot\nabla B)2\pi\frac{B}{m}\int\mu\int f_e dv_{\parallel}d\mu\\
=&\frac{e}{m}\delta E_{\parallel}n_e+\iint\mu(\tilde{\mathbf{b}}\cdot\nabla B)\,f_e d\mathbf{v}\\
\approx&\frac{e}{m}\delta E_{\parallel}n_e
\end{aligned} \tag{349}$$

$$
\begin{aligned}
&\iint v_\parallel(v_\parallel\tilde{\mathbf{b}})\cdot\nabla f_e d\mathbf{v}\\
=&\iint \tilde{\mathbf{b}}\cdot\nabla(v_\parallel^2 f_e)d\mathbf{v}\\
=&\iint \tilde{\mathbf{b}}\cdot\nabla(v_\parallel^2 f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu\\
=&\tilde{\mathbf{b}}\cdot\nabla\left(\iint v_\parallel^2 f_e 2\pi\frac{1}{m}\,dv_\parallel d\mu\right)B_0\\
=&\tilde{\mathbf{b}}\cdot\nabla\left(\frac{p_\parallel}{B_0}\right)B_0\\
=&\tilde{\mathbf{b}}\cdot\nabla\left(\frac{p_{\parallel 0}+\delta p_\parallel}{B_0}\right)B_0\\
=&\tilde{\mathbf{b}}\cdot\nabla\left(\frac{p_{\parallel 0}}{B_0}\right)B_0+\tilde{\mathbf{b}}\cdot\nabla\left(\frac{\delta p_\parallel}{B_0}\right)B_0\\
\approx&\mathbf{b}\cdot\nabla\left(\frac{p_{\parallel 0}}{B_0}\right)B_0+\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla\left(\frac{p_{\parallel 0}}{B_0}\right)B_0+\mathbf{b}\cdot\nabla\left(\frac{\delta p_\parallel}{B_0}\right)B_0\\
\approx&\mathbf{b}\cdot\nabla(p_{\parallel 0})+\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla(p_{\parallel 0})+\mathbf{b}\cdot\nabla(\delta p_\parallel) \qquad (350)\\
=&\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla(p_{\parallel 0})+\mathbf{b}\cdot\nabla(\delta p_\parallel)
\end{aligned}
$$

where use has been made of $\mathbf{b}\cdot\nabla p_{\parallel 0}=0$.

$$
\frac{\partial\delta J_{e\parallel}}{\partial t}=-\frac{e}{m}\delta E_\parallel n_e-\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla(p_{\parallel 0})-\mathbf{b}\cdot\nabla(\delta p_\parallel) \qquad (351)
$$

Using Eq. () in Eq. (), we obtain

$$
\mu_0 e\frac{e}{m}\delta E_\parallel n_e+\mathbf{b}\cdot\nabla\times\nabla\times\delta\mathbf{E}=-\mu_0 e\left[\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla(p_{e\parallel 0})+\mathbf{b}\cdot\nabla(\delta p_{e\parallel})\right] \qquad (352)
$$

$$
-\mathbf{b}\cdot\nabla\times\nabla\times\delta\mathbf{E}=-\mu_0 e\left(-q\frac{\delta\mathbf{B}_\perp}{B_0}\cdot\nabla_X p_{\parallel 0}+\frac{q}{m}qE_\parallel n_{e0}-q\,\mathbf{e}_\parallel\cdot\nabla_X(\delta p_\parallel)\right). \qquad (353)
$$

ddddd

$$\iint v_\parallel\Big(v_\parallel\tilde{\mathbf{b}}+\mathbf{v}_D+\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\Big)\cdot\nabla f_e d\mathbf{v}$$

$$=\iint\Big(v_\parallel\tilde{\mathbf{b}}+\mathbf{v}_D+\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\Big)\cdot\nabla(v_\parallel f_e)d\mathbf{v}$$

$$=\iint v_\parallel\tilde{\mathbf{b}}\cdot\nabla(v_\parallel f_e)d\mathbf{v}+\iint\mathbf{v}_D\cdot\nabla(v_\parallel f_e)d\mathbf{v}+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)d\mathbf{v}$$

$$=\iint\tilde{\mathbf{b}}\cdot\nabla(v_\parallel^2 f_e)d\mathbf{v}+\iint\mathbf{v}_D\cdot\nabla(v_\parallel f_e)d\mathbf{v}+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)d\mathbf{v}$$

$$=\iint\tilde{\mathbf{b}}\cdot\nabla(v_\parallel^2 f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu+\iint\mathbf{v}_D\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu$$

$$=\int\tilde{\mathbf{b}}\cdot\nabla\Big(v_\parallel^2 f_e 2\pi\frac{1}{m}\,dv_\parallel d\mu\Big)B+\iint\mathbf{v}_D\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu$$

$$=\tilde{\mathbf{b}}\cdot\nabla\Big(\frac{p_\parallel}{B}\Big)B+\iint\mathbf{v}_D\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}dv_\parallel d\mu+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}dv_\parallel d\mu$$

$$=\tilde{\mathbf{b}}\cdot\nabla\Big(\frac{p_\parallel}{B}\Big)B+\iint\frac{1}{m\Omega}\mathbf{b}\times(\mu\nabla B)\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu+\iint\frac{1}{m\Omega}\mathbf{b}\times(mv_\parallel^2\boldsymbol{\kappa})\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu$$

$$+\iint\frac{\delta\mathbf{E}\times\mathbf{b}}{B_0}\cdot\nabla(v_\parallel f_e)2\pi\frac{B}{m}\,dv_\parallel d\mu$$

H Modern view of gyrokinetic equation**wrong**

The modern form of the nonlinear gyrokinetic equation is in the total-f form. The modern way of deriving the gyrokinetic equation is to use transformation methods to eliminate gyro-phase dependence of the total distribution function and thus obtain an equation for the resulting gyro-phase independent distribution function (called gyro-center distribution function).

The resulting equation for the gyro-center distribution function is given by (see Baojian's paper)

$$\Big(\frac{\partial}{\partial t}+\dot{\mathbf{X}}\cdot\nabla+\dot{v}_\parallel\frac{\partial}{\partial v_\parallel}\Big)f(\mathbf{X},v_\parallel,\mu,t)=0, \tag{354}$$

where

$$\dot{\mathbf{R}}=\mathbf{V}_D+\frac{\mathbf{e}_\parallel}{B_0}\times\langle\nabla\delta\Phi\rangle_\alpha+v_\parallel\frac{\langle\delta\mathbf{B}_\perp\rangle_\alpha}{B_0} \tag{355}$$

$$\dot{v}_\parallel = -\frac{1}{m}\frac{\mathbf{B}^\star}{B_\parallel^\star}\cdot(q\nabla\langle\delta\Phi\rangle + \mu\nabla B_0) - \frac{q}{m}\frac{\partial\langle\delta A_\parallel\rangle_\alpha}{\partial t}. \tag{356}$$

Here the independent variables are gyro-center position $\mathbf{X}$, magnetic moment μ and parallel velocity $v_\parallel$.

The gyro-phase dependence of the particle distribution can be recovered by the inverse transformation of the transformation used before. The pull-back transformation gives rise to the polarization density term. (phase-space-Lagrangian Lie perturbation method (Littlejohn, 1982a, 1983), I need to read these two papers.).

These notes were initially written when I visited University of Colorado at Boulder (Sept.-Nov. 2016), where I worked with Dr. Yang Chen, who pointed out that most gyrokinetic simulations essentially employ Frieman-Chen's nonlinear gyrokinetic equation. Therefore a careful re-derivation of the equation to know the gyrokinetic orderings and physics included in the model is highly desirable, which motivates me to write this note.

—to be deleted—

Since $dv_\parallel/dt = 0$ and $dv_\perp/dt = 0$, the phase-space structure of f_p along $v_\parallel$ and $v_\perp$, i.e., $\partial f_p/\partial v_\parallel$ and $\partial f_p/\partial v_\perp$, disappears from Eq. (2).

We assume that f_p evolves at a time scale much longer than typical gyro-periods. Further assume that the spatial variation of f_p along the magnetic field is weak on the space scale of the distance travelled by a particle within one gyro-period. Then, the terms $\partial f_p/\partial t$ and $\partial f_p/\partial z$ can be dropped from the Vlasov equation.

Therefore only 3 of the 7 variables $(x, y, z, v_\perp, \alpha, v_\parallel, t)$, namely x, y, and α, can possibly contribute to the variation of f_p when we follow a particle orbit. Could the number of these variables be further reduced? That is, whether there are more constants of motion can be found (so that $dz^{(i)}/dt = 0$) or weak dependence of f_p on more variables (so that $\partial f/\partial z^{(i)} \sim 0$) can be assumed? It is ready to realise that there are two additional constants of motion, namely the two coordinates (x', y') of the guiding-center. If we use these two variables as new variables replacing x and y. The chain rule indicates that the dependence of f_p on x' and y' will not show up in the expression of df_p/dt because $dx'/dt = 0$ and $dy'/dt = 0$. Then the only variable dependence of f_p remains in the expression of df_p/dt is the gyro-angle α. Then the Vlasov equation implies that f_p is independent of α in the chosen coordinate system, giving a clear information about the phase-space structure of f_p.

Bibliography

[1] P J Catto. Linearized gyro-kinetics. *Plasma Physics*, 20(7):719, 1978.

[2] Yang Chen and Scott E. Parker. Particle-in-cell simulation with vlasov ions and drift kinetic electrons. *Phys. Plasmas*, 16:52305, 2009.

[3] E. A. Frieman and Liu Chen. Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. *Physics of Fluids*, 25(3):502–508, 1982.