CONFERENCE PRE-PRINT

EXTRACTING THE NEAREST CANONICAL EQUILIBRIUM DISTRIBUTION VIA NATURAL GRADIENT DESCENT METHOD

immediate

October 6, 2025

Chao Li, et al
State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing, China
Beijing, China
Email: 2101110153@stu.pku.edu.cn

Abstract

This paper presents an efficient method for numerically extracting the nearest canonical equilibrium distribution $f_{\rm NE}$ from an arbitrary axisymmetric distribution function of tokamak plasmas by formulating the problem as an optimization task for the discrete form of the gyrokinetic Vlasov equation. An iterative scheme utilizing natural gradient descent is employed to obtain $f_{\rm NE}$ with a specified numerical accuracy. This approach incorporates an enhancement algorithm in order to accelerate the convergence process for the phase space points near the trapped-passing boundary. Possible applications of this algorithm are also discussed.

1. INTRODUCTION

Gyrokinetic simulation is a crucial tool for studying micro-turbulence in magnetized plasma, which plays a dominant role in anomalous particle and heat transport in tokamaks. It is known that the typical frequency spectrum of drift wave turbulence in a tokamak is close to the diamagnetic drift frequency, which is much lower than the ion cyclotron frequency [1]. Based on this fact, the gyrokinetic theory decouples the high-frequency gyromotion of particles from their low-frequency drift motion by averaging over the gyro-angle, enabling the study of low-frequency phenomena in fusion plasma while retaining essential kinetic effects like wave-particle resonances and finite Larmor radius (FLR) effects [2]. This also allows the 6D phase space of particle to be reduced to a 5D gyrocenter phase space, greatly reducing computational costs in simulations. Over the years, many gyrokinetic simulation codes have been developed [3, 4, 5, 6] and applied to investigate turbulence transport properties in tokamaks, effectively providing a detailed understanding of turbulence behavior and transport in plasma.

Gyrokinetic simulations fundamentally require the specification of a initial gyrokinetic equilibrium f_0 —a baseline plasma state perturbed by turbulence. In collisionless plasmas, this initial equilibrium must satisfy the condition

$$\dot{Z}_0^i \frac{\partial}{\partial Z^i} f_0 = 0. \tag{1}$$

Here $\boldsymbol{Z} = \left(\boldsymbol{X}, v_\parallel, \mu\right)$ is the gyro-center coordinate variables, with $\boldsymbol{X}$ the gyro-center position, $v_\parallel$ the parallel velocity along the equilibrium magnetic field and μ the magnetic moment. $\dot{Z}_0^i = \left\{Z^i, H_0\right\}$ is the unperturbed particle motion determined by equilibrium electro-magnetic field with H_0 the unperturbed gyrocenter Hamiltonian and $\{\cdot\}$ the unperturbed Poisson bracket. In toroidal axisymmetric systems like tokamaks, the unperturbed motion is prescribed by three constants of motions (CoMs), the toroidal canonical momentum P_ζ, the particle kinetic energy E and the magnetic momentum μ. Consequently, this equilibrium distribution is naturally an axisymmetric function of constants of motion $f_0(P_\zeta, E, \mu)$, which is also referred as a 'canonical equilibrium distribution' [7].

A well defined canonical equilibrium distribution should be given both initially and dynamically in gyrokinetic simulation and it is essential to develop a reliable method for extracting the canonical equilibrium distribution from a given reference distribution. In principle, there are infinite functions satisfy 1, so that the extracted one should be the nearest one to the given distribution function—a concept referred to as the "Nearest Canonical Equilibrium Distribution" (abbreviated as "NE" for convenience). In this paper, we aim to define NE clearly and propose an algorithm to accurately extract it from any given distribution function. The proposed algorithm is shown to be robust, accurate, and easy to implement. Besides, it is also found that such a definition of NE is consistent with the one given by the orbit average. Therefore, the proposed algorithm can also be regarded as an effective method to calculate orbit average quantities.

In addition, the NE extraction algorithm presented here also performs an effective dimension reduction from 4D function gyrokinetic coordinate down to the 3D CoM space (P_ζ, E, μ), since the extracted NE is already a function of CoM. A direct transformation from f_{NE} to a function of CoM can be done by a simple interpolation along a certain poloidal angle θ. This transformation is highly relevant in EP diagnoses [8, 9] and PSZS related models.

2. EXTRACTING NE VIA NATURAL GRADIENT DESCENT METHOD

As illustrated above, a canonical equilibrium distribution f_{CE} in tokamak configuration should be axisymmetric and invariant along the unperturbed orbit. So it has to satisfy the governing equation:

$$\mathcal{J}\dot{Z}_0^i \frac{\partial}{\partial Z^i} f_{CE}(\psi, \theta, v_\parallel, \mu) = 0. \tag{2}$$

Here we have written equation (1) in magnetic flux coordinates (ψ, θ, ζ). ψ is the flux surface label, θ and ζ are the poloidal and toroidal angle, respectively. Since $\partial_\zeta f_{CE} = 0$ and $\dot{\mu} = 0$, the problem is essentially three dimensional with ψ, θ and $v_\parallel$ in general gyrocenter coordinates. The phase space Jacobian $\mathcal{J} = \frac{1}{\nabla\psi\times\nabla\theta\cdot\nabla\zeta}\frac{B_\parallel^*}{m}$ does not change the solutions of the equation since it is nonzero. Here $B_\parallel^* = B_0 + \frac{m}{e}v_\parallel \mathbf{b_0}\cdot\nabla\times\mathbf{b_0}$ is the Jacobi of the gyrocenter velocity variables, which is assumed essentially unchanged.

It is obvious that the solution to Eq. (2) is not unique. As a matter of fact, any function of the constants of motion, $f(P_\zeta, E, \mu)$, satisfies this equation. The challenge lies in extracting an appropriate part from a given axisymmetric distribution $f_0(\psi, \theta, v_\parallel, \mu)$ in a justifiable manner. Here, we define the NE through basis functions of the unperturbed motion operator $\hat{U} = \mathcal{J}\dot{Z}_0^i \frac{\partial}{\partial Z^i}$.

It can be found that $\hat{U}$ is a skew-adjoint operator:

$$\hat{U}^* = -\hat{U} \tag{3}$$

According to spectral theorem [10], $\hat{U}$ possesses a complete basis (including generalized eigen-functions for the continuous spectrum), and all its eigenvalues are either zero or come in pairs of purely imaginary numbers $i\lambda$. So f_0 can be decomposed into two parts based on whether the eigenvalues are zero,

$$f_0 = \sum_{\lambda=0} c_n\phi_n + \left(\sum c_m\phi_m + \int d\lambda\, c_\lambda\phi_\lambda\right)_{\lambda\neq 0} \tag{4}$$

Here, ϕ_n and c_n are basis functions and expansion coefficients. The expansion is written in a general form. The continuous spectrum may lead to singularity behavior of the system, but it is not the focus of this paper. We only focus on the $\lambda = 0$ part.

So the nearest canonical equilibrium distribution can be properly defined as

$$f_{\mathrm{NE}} = \sum_{\lambda=0} c_n\phi_n, \tag{5}$$

which indicates *the nearest canonical equilibrium* f_{NE} *of a given function* f_0 *is its projection onto the static subspace* $\mathcal{V}_{NE}$.

Directly solving f_{NE} based on $\mathcal{V}_{NE}$ or ϕ_n is computationally impractical due to the high dimensionality of the eigenvalue problem. Fortunately, the specific eigen-functions and eigenvalues are not necessary for the computation of f_{NE} . Instead, the projection process can be reformulated as an optimization task of Eq. (2) with a loss function defined by:

$$S[f] = \int 2\pi\mathcal{J}\, d^5Z^i \left(\dot{Z}_0^i \frac{\partial f}{\partial Z^i}\right)^2. \tag{6}$$

The goal is to optimize S to zero so that $\dot{Z}_0^i\partial_i f = 0$ is satisfied everywhere in phase space. The standard numerical method of this optimization is the natural gradient descent method. The corresponded iterative updating rule for f is:

$$f^{(t+1)} = f^{(t)} - \eta\frac{\delta S[f]}{\delta f} = f^{(t)} + \eta\frac{1}{J}\frac{\partial}{\partial Z^i}\left(\dot{Z}_0^i J \dot{Z}_0^j \frac{\partial f}{\partial Z^j}\right), \tag{7}$$

where η is the step size. This iteration starts from the given distribution, $f^{(t=0)} = f_0$. In each step, $f^{(t)}$ goes to a direction that reduce S.

The operator $\hat{G} = -\frac{1}{J}\frac{\partial}{\partial Z^i}\left(\dot{Z}_0^i J \dot{Z}_0^j \frac{\partial}{\partial Z^j}\right)$ is a positive semi-definite self-adjoint operator, so that (according to the spectral theorem) $\hat{G}$ has a complete basis (self-adjoint) with non-negative eigenvalues (semi-definite). Besides, one can easily prove that:

$$\hat{G}f = 0 \Leftrightarrow \hat{U}f = 0, \tag{8}$$

which indicates the operator $\hat{G}$ and $\hat{U}$ have the same static function subspace $\mathcal{V}_{NE}$. Expanding f in the eigen-basis, we found that the natural gradient descent method effectively shrinks the coefficients associated with non-zero eigenvalues while preserving those with zero eigenvalues:

$$c_\lambda^{(t+1)} = c_\lambda^{(t)}(1 - \eta\lambda) \tag{9}$$

Provided that $\eta < \frac{2}{\lambda_{\max}}$, the algorithm is guaranteed to converge to:

$$f^{(t\to\infty)} = \sum_{\lambda=0} c_n \phi_n = f_{\mathrm{NE}} \tag{10}$$

Thus, this iterative algorithm performs an effective projection of f_0 onto $\mathcal{V}_{NE}$, eliminating all non-equilibrium components.

2.1. Acceleration of the convergence near the T-P boundary

Although the convergence of Eq. (7) is guaranteed, the iterative scheme may still converge slowly for particles near the trapped-passing boundary, where the orbit period $\tau_b \to \infty$. To address this, a positive weighting function $W(Z)$ is introduced to accelerate convergence rate

$$f^{(t+1)} = f^{(t)} + \eta \frac{1}{\mathcal{J}}\frac{\partial}{\partial Z^i}\left(\dot{Z}_0^i W \mathcal{J} W \dot{Z}_0^j \frac{\partial f}{\partial Z^j}\right). \tag{11}$$

The modified operator

$$\hat{G}_w = -\frac{1}{J}\frac{\partial}{\partial Z^i}\left(\dot{Z}_0^i W J W \dot{Z}_0^j \frac{\partial}{\partial Z^j}\right) \tag{12}$$

remains positive semi-definite and self-adjoint, preserving the static function subspace $\mathcal{V}_{\mathrm{NE}}$. The iterative process (11) still converges to the same f_{NE}, provided that $W(Z) > 0$ everywhere. This approach can also be interpreted as redefining the weighted loss function by:

$$S_w[f] = \int 2\pi \mathcal{J}\, d^5 Z^i \left(W(\mathbf{Z}) \dot{Z}_0^i \frac{\partial f}{\partial Z^i}\right)^2, \tag{13}$$

which makes the loss function more sensitive to regions with slow convergence rate, thereby enhancing the overall convergence rate.

3. NUMERICAL IMPLEMENTATION

The proposed algorithm for f_{NE} extraction has been implemented in NLT, a δf semi-Lagrangian gyrokinetic code based on the numerical Lie transform method [11]. Although the algorithm proposed here can be applied to arbitrary distribution functions and equilibrium fields, in this work a local Maxwellian

$$f_0 = f_{LM} = \frac{n(\psi)}{(2\pi)^{3/2} v_T^3} \exp\left(-\frac{\frac{1}{2}m v_\parallel^2 + \mu B}{T(\psi)}\right) \tag{14}$$

is chosen as the test function and no equilibrium electric field is considered for simplicity. The profile parameters employed here are the same with the well-known CBC parameters used in [12]. For comparison, we also explored a direct orbit-averaging method. In this approach, each orbit was discretized into 128 sampling points, determined via Newton-Raphson iteration by inverting the constants of motion. Enhanced sampling density was employed near bounce points to address numerical sensitivities.

3.1. Effectiveness of the optimization algorithm for extracting f_{NE}

As the criterion for the convergence of the iteration, a normalized loss S_{norm} can be defined as

$$S_{norm} = \sqrt{\frac{\sum_{\text{grids}} J(\Delta\beta\alpha)^2}{\sum_{\text{grids}} J\alpha^2}}, \tag{15}$$

which corresponds to the analytical expression

$$S_{norm} = \sqrt{\frac{\int \mathcal{J}\, d^5 Z^i (\dot{Z}_0^i \frac{\partial f}{\partial Z^i})^2}{\int \mathcal{J}\, d^5 Z^i f_0^2}}. \tag{16}$$

The loss function is normalized by f_0 and has the dimension of the inverse of time t^{-1}. In the NLT code, time is normalized by the cyclotron period of ion $\tau_{\mathrm{ci}} = 1/\Omega_{ci}$. Figure 1(a) shows the fast decrease of the normalized loss with iteration round, where one round corresponds to 1000 iterative steps. Different colored curves correspond to various values of μ, with lighter color indicating larger value of μ. The normalized loss S_{norm} can be reduced below 5×10^{-8} after 50 rounds. This convergence result indicates that $\dot{Z}_0^i \frac{\partial f}{\partial Z^i}$ globally approaches zero, validating the effectiveness of the algorithm. Specifically, if the equation

$$\left(\partial_t + \dot{Z}_0^i \frac{\partial}{\partial Z^i}\right) f_{\mathrm{NE}} = 0 \tag{17}$$

is evolved using the same discretization scheme, it takes approximately $T \approx 4\times10^6\tau_{\mathrm{ci}}$ to reach a global variation of $\delta f/f_0 \approx 10^{-4}$, since the numerical error accumulates with time approximately following the scaling $\propto T^{\frac{1}{2}}$. This time significantly exceeds the relevant transport time scale. Therefore, numerically, $\dot{Z}_0^i \frac{\partial f_{\mathrm{NE}}}{\partial Z^i}$ can be considered as zero. Figure 1(b) shows the relative variation in the loss function compared with the initial local Maxwellian. After 50 rounds iterations, all losses are reduced to below 1.1×10^{-4} of their original values, effectively approaching zero. This computation utilizes 16 MPI process (one for each μ) and 128 CPU cores (8 OpenMP threads for each MPI process), requiring only 373 seconds for 50,000 steps, a negligible duration compared to the typical time scales for turbulence transport simulations. The computational efficiency can be further improved by exploiting the equilibrium's up-down symmetry; solving only the upper half-plane within $[0,\pi]$ reduces the computational load by half, as the lower half is simply a mirror image. Additional speed-up could also be achieved with more efficient parallelization methods for sparse matrix multiplication.

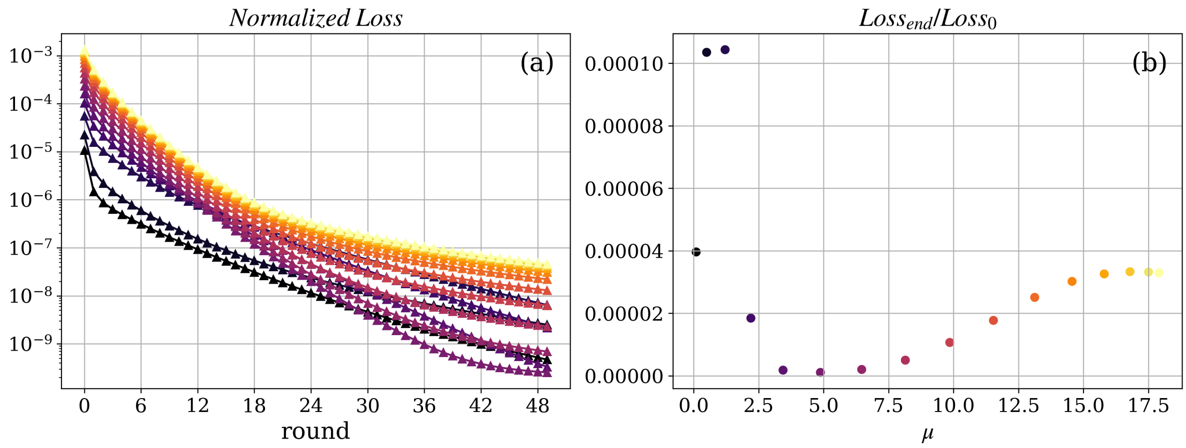

FIG. 1. (a) Decreasing trend of the normalized loss S_{norm} with iteration round, where one round corresponds to 1000 iterative steps. Different colors represent different values of μ (lighter color for larger μ) processed in parallel using MPI. (b) Relative change in loss after iterations, showing convergence below 1.1×10^{-4} of the original value after 50 rounds (50000 steps).

3.2. Transformation to CoM Space

The constants of motion (CoMs) in tokamaks are defined by (P_ζ, E, μ), where:

$$P_\zeta = e_s\left(\psi - \frac{RB_\phi}{\Omega} v_\parallel\right) \tag{18}$$

is the canonical angular momentum, E is the kinetic energy, and μ is the magnetic moment.

A coordinate transformation can be performed from non-canonical gyrocenter coordinates $(\psi, \theta, v_\parallel, \mu)$ to CoM space $(P_\zeta, E, \mu, \theta)_\sigma$. The three constants of motion (P_ζ, E, μ) uniquely define a particle orbit, while θ specify the position along that orbit. Here, σ distinguishes between different Riemann sheets, reflecting the non-uniqueness of the inverse transformation. Specifically, each point in CoM space $(P_\zeta, E, \mu, \theta)$ corresponds to two distinct points in phase space $(\psi, \theta, v_\parallel, \mu)$. If the transformed function is independent of (θ, σ)—meaning it remains constant along the orbit—it effectively depends solely on the CoMs. This transformation is particularly relevant in energetic particle diagnostics [8, 9] and in models involving Phase Space Zonal Structures (PSZS) [13, 14].

As the extracted neighboring equilibrium f_{NE} is already a function of orbit-constant, it inherently depends only on the CoMs. Consequently, converting f_{NE} into CoM space is straightforward, resulting in a function $f_{\mathrm{CoM}}(P_\zeta, E, \mu)$. To evaluate f_{CoM} at any given set of CoMs (P_ζ, E, μ), one can choose an arbitrary θ_0 and then solve for the corresponding $(\psi, \theta_0, v_\parallel, \mu)$ based on the energy relation:

$$v_\parallel^2 = \frac{2(E - \mu B(\psi, \theta_0))}{m}. \tag{19}$$

Substituting this relation into the equation for P_ζ, we have:

$$\psi - \frac{RB_\phi}{\Omega} v_\parallel = \psi - \frac{R(\psi, \theta_0) B_\phi(\psi, \theta_0)}{\Omega(\psi, \theta_0)} \sigma \sqrt{\frac{2(E - \mu B(\psi, \theta_0))}{m}} = \frac{P_\zeta}{e_s}. \tag{20}$$

This equation in ψ is nonlinear and can be efficiently solved using Newton iteration methods. After solving for ψ, $v_\parallel$ is computed directly from the energy relation. Finally, the corresponding value $f_{\mathrm{CoM}}(P_\zeta, E, \mu) = f_{\mathrm{NE}}(\psi, \theta_0, v_\parallel, \mu)$ can be obtained by interpolation. The process, illustrated in Figure 2, efficiently transforms f_{NE} into a CoM-based representation, enabling analysis in EP problems where dependencies on the constants of motions are essential.

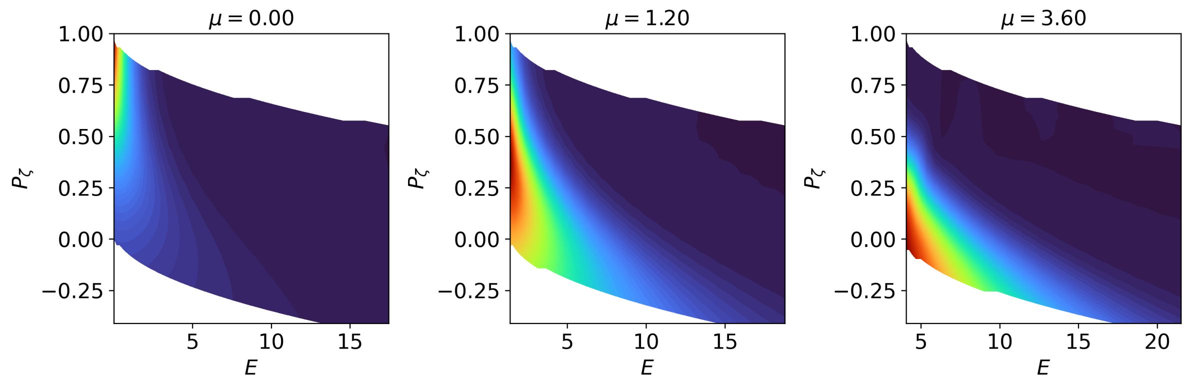

FIG. 2. 2D contour plot of f_{CoM} in (P_ζ, E) with three specified μ. f_{CoM} is obtained from f_{NE} by interpolation.

4. CONCLUSION AND DISCUSSION

In this work, we have developed a method to rigorously define and extract the nearest canonical equilibrium distribution from an arbitrary distribution function and equilibrium fields for gyrokinetic simulation by formulating this task as an optimization problem. This definition aligns consistently with the concept of the orbit average method. An iterative algorithm based on natural gradient descent was proposed, enhanced with a weighted approach to improve convergence rate, especially in numerically challenging regions. Numerical results demonstrate that the equilibrium distribution generated by this algorithm achieves significantly higher accuracy than that attainable through the direct orbit-averaging method. This robust framework effectively isolates macroscopic processes of interest in long-time plasma evolution studies. Furthermore, the computational overhead of this method is negligible compared to typical gyrokinetic simulation times. Besides, the method provides a computationally efficient framework for evaluating orbit-averaged quantities, such as $\langle dP_\zeta/dt \rangle_{OA}$, which are essential inputs for reduced transport models [13].

ACKNOWLEDGMENTS

The authors appreciates the inspiring discussion with Dr. Jian Wang on the convergence of the iteration algorithm. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant Nos. XDB0790201 and XDB0500302, the National MCF Energy R&D Program of China under Grant Nos. 2019YFE03060001 and No. 2024YFE03270400 , and the National Natural Science Foundation of China under Grant No. 12405275.

REFERENCES

[1] W. Horton, "Drift waves and transport," *Reviews of Modern Physics*, vol. 71, no. 3, p. 735, 1999.

[2] A. J. Brizard and T. S. Hahm, "Foundations of nonlinear gyrokinetic theory," *Reviews of modern physics*, vol. 79, no. 2, pp. 421–468, 2007.

[3] Z. Lin, T. S. Hahm, W. Lee, W. M. Tang, and R. B. White, "Turbulent transport reduction by zonal flows: Massively parallel simulations," *Science*, vol. 281, no. 5384, pp. 1835–1837, 1998.

[4] T. Goerler, X. Lapillonne, S. Brunner, T. Dannert, F. Jenko, F. Merz, and D. Told, "The global version of the gyrokinetic turbulence code gene," *Journal of Computational Physics*, vol. 230, no. 18, pp. 7053–7071, 2011.

[5] S. Ku, R. Hager, C.-S. Chang, J. Kwon, and S. E. Parker, "A new hybrid-lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma," *Journal of Computational Physics*, vol. 315, pp. 467–475, 2016.

[6] E. Lanti, N. Ohana, N. Tronko, T. Hayward-Schneider, A. Bottino, B. F. McMillan, A. Mishchenko, A. Scheinberg, A. Biancalani, P. Angelino, *et al.*, "Orb5: a global electromagnetic gyrokinetic code using the pic approach in toroidal geometry," *Computer Physics Communications*, vol. 251, p. 107072, 2020.

[7] Y. Idomura, S. Tokuda, and Y. Kishimoto, "Global gyrokinetic simulation of ion temperature gradient driven turbulence in plasmas using a canonical maxwellian distribution," *Nuclear Fusion*, vol. 43, no. 4, p. 234, 2003.

[8] S. Benjamin, H. Järleblad, M. Salewski, L. Stagner, M. Hole, and D. Pfefferlé, "Distribution transforms for guiding center orbit coordinates in axisymmetric tokamak equilibria," *Computer Physics Communications*, vol. 292, p. 108893, 2023.

[9] A. Bierwage, M. Fitzgerald, P. Lauber, M. Salewski, Y. Kazakov, and Ž. Štancar, "Representation and modeling of charged particle distributions in tokamaks," *Computer Physics Communications*, vol. 275, p. 108305, 2022.

[10] M. Reed, B. Simon, B. Simon, and B. Simon, *Methods of modern mathematical physics*, vol. 1. Academic press New York, 1972.

[11] L. Ye, Y. Xu, X. Xiao, Z. Dai, and S. Wang, "A gyrokinetic continuum code based on the numerical lie transform (nlt) method," *Journal of Computational Physics*, vol. 316, pp. 180–192, 2016.

[12] A. M. Dimits, G. Bateman, M. Beer, B. Cohen, W. Dorland, G. Hammett, C. Kim, J. Kinsey, M. Kotschenreuther, A. Kritz, *et al.*, "Comparisons and physics basis of tokamak transport models and turbulence simulations," *Physics of Plasmas*, vol. 7, no. 3, pp. 969–983, 2000.

[13] M. V. Falessi, L. Chen, Z. Qiu, and F. Zonca, "Nonlinear equilibria and transport processes in burning plasmas," *New Journal of Physics*, vol. 25, no. 12, p. 123035, 2023.

[14] P. Lauber, M. Falessi, G. Meng, T. Hayward-Schneider, V.-A. Popa, F. Zonca, and M. Schneider, "Atep: an advanced transport model for energetic particles," *Nuclear Fusion*, vol. 64, no. 9, p. 096010, 2024.