J. Plasma Phys. (2024), *vol*. 90, 905900406 © The Author(s), 2024.
Published by Cambridge University Press
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.
doi:10.1017/S0022377824000709

Optimisation of gyrokinetic microstability using adjoint methods

G.O. Acton [1,2,†], M. Barnes [1], S. Newton [2] and H. Thienpondt [3]

[1]Rudolf Peierls Centre For Theoretical Physics, University of Oxford, Oxford OX1 3PU, UK

[2]United Kingdom Atomic Energy Authority, Culham Campus, Abingdon, Oxfordshire OX14 3DB, UK

[3]Laboratorio Nacional de Fusión, CIEMAT, 28040 Madrid, Spain

(Received 13 March 2024; revised 10 May 2024; accepted 13 May 2024)

Microinstabilities drive turbulent fluctuations in inhomogeneous, magnetised plasmas. In the context of magnetic confinement fusion devices, this leads to an enhanced transport of particles, momentum and energy, thereby degrading confinement. In this work, we describe an application of the adjoint method to efficiently determine variations of gyrokinetic linear growth rates on a general set of external parameters in the local δf-gyrokinetic model. We then offer numerical verification of this approach. When coupled with gradient-based techniques, this methodology can facilitate the optimisation process for the microstability of the confined plasmas across a high-dimensional parameter space. We present a numerical demonstration wherein the ion-temperature-gradient instability growth rate in a tokamak plasma is minimised with respect to flux surface shaping parameters. The adjoint method approach demonstrates a significant computational speed-up compared with a finite-difference gradient calculation.

Key words: fusion plasma, plasma confinement, plasma instabilities

1. Introduction

Within the plasma core of magnetic confinement fusion (MCF) devices, gradients of temperature and density act as sources of free energy that are capable of driving microscale instabilities, should they exceed critical values (see e.g. Rudakov & Sagdeev 1961; Drake & Lee 1977 or Tang, Connor & Hastie 1980 for ion-temperature-gradient (ITG), microtearing mode or kinetic ballooning mode descriptions, respectively). Beyond such critical values the transport fluxes increase rapidly, thus requiring a large additional power input to maintain the temperature gradient. One mechanism behind these large fluxes is turbulent mixing, which results in the increased transport of particles, momentum and energy out of the device. To properly predict plasma profiles, one should thus solve the transport equations in which nonlinear, turbulent fluxes determine profile evolution. However, due to this 'stiffness' of the transport (Dimits *et al.* 2000), the critical gradients obtained from a linear instability analysis are often a first reasonable approximation to

† Email address for correspondence: georgia.acton@merton.ox.ac.uk

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

the experimental outcomes[1] (Doyle, Houlberg & Kamada 2007). Hence, it is worthwhile to determine if externally controlled experimental parameters can be chosen to optimise linear stability.

The growth rate of linear microinstabilities can be influenced by a large number of parameters, including plasma density, temperature, flow profiles and the magnetic geometry. In an idealised situation, one would determine how such growth rates, and by extension critical gradients, depend on all parameters governing the system, and then design MCF devices that optimise temperature and/or density gradients. However, analytical searches are intractable, and, because the number of tuneable parameters in modern MCF devices is large, full numerical parameter scans are often prohibitively expensive.

In this paper we address this issue by employing an adjoint approach (cf. Pironneau 1974) to enable efficient calculation of local variations to the linear growth rate with respect to all parameters of interest within the local, δf-gyrokinetic model. In contrast to a finite-difference calculation, the adjoint method is essentially independent of the dimension of the parameter space, and can be used to optimise over a large number of variables at once without incurring additional computational cost beyond solving the system equations; the associated cost of the adjoint method is roughly equivalent to solving the linearised gyrokinetic system of equations twice.

The archetypal microinstability in MCF plasmas is the ITG instability (see e.g. Rudakov & Sagdeev 1961; Cowley, Kulsrud & Sudan 1991), as it has been identified as the main source of heat transport in the core of many tokamaks (Mantica *et al.* 2011). Therefore, for clarity, we shall present the optimisation process with ITG instabilities in mind. We shall apply the adjoint method to the linear gyrokinetic equation and demonstrate its utility by calculating the sensitivity of the growth rate to geometrical parameters in a tokamak, with the aim of maximising the critical ITG through shaping considerations only. However, the technique and equations derived below allow for calculations of the gradient of the linear growth rate with respect to a general set of unspecified parameters, $\boldsymbol{p}$, and could thus be applied more broadly, and to other instabilities. In general, the adjoint method we present here is agnostic to the mechanism behind the instability, and gives the gradient of the linear growth rate with respect to a set of external parameters independent of the mode type. Given that the growth rate is a continuous function this approach could thus also be applied to linear growth rates for which the instability mechanism changes.

This paper is organised as follows: in § 2 we outline the gyrokinetic-Maxwell system, including the governing equations for the evolution of the plasma distribution function and electromagnetic fields. In § 3, we present details of the gyrokinetic-adjoint system, and derive the adjoint equations as applied to a linear, low-flow, electromagnetic δf-gyrokinetic system with collisions. The aim in mind is to minimise the linear growth rate in a parameter space defined by a generic set of independent variables, $\boldsymbol{p}$, that influence the plasma evolution. We also extract the electrostatic, collisionless limit of the adjoint equations. The second portion of the paper then focuses on applying the model derived, and discussing how this method may be used to vary the plasma profile (such as temperature or density gradients) in directions favourable for fusion whilst retaining microstability. In § 4 we

[1] An important caveat to mention is that true critical gradient values can be shifted by nonlinear effects. Linear instabilities grow until nonlinear coupling to 'zonal' modes removes energy from the linearly growing modes. This energy injection into the zonal modes causes them to grow and can suppress turbulence completely in a region beyond the linear critical gradients (e.g. see Dimits *et al.* 2000; Rogers, Dorland & Kotschenreuther 2000). We will neglect this 'Dimits shift' of the critical gradient in the following analysis, and use only the linear gradient as a predictor for experimental profiles. This provides a lower bound for critical gradient values and can be a good approximation for the full nonlinear threshold value.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

briefly discuss how the process is implemented using the δf-gyrokinetic code `stella`, with § 4.2 specialising to the case where $\boldsymbol{p}$ consists of the parameters needed to specify the local magnetic geometry within the Miller formalism (Miller *et al.* 1998); we describe the motivation behind this choice, and describe how it is implemented numerically. Section 5 covers the numerical methods employed in solving the adjoint system. Finally, § 6 presents the numerical results, including tests for an example application in which the adjoint method is utilised to vary the magnetic geometry with the aim to minimise the linear ITG growth rate.

2. Plasma evolution equations

We model the evolution of plasma fluctuations with the δf-gyrokinetic equation (Catto 1978; Antonsen & Lane 1980; Abel *et al.* 2013), which one derives by exploiting spatial and temporal scale separation. Our starting point is the Vlasov–Maxwell system of equations including collisions

$$\frac{\mathrm{d}f_\nu}{\mathrm{d}t} = \sum_{\nu'} C_{\nu\nu'}[f_\nu, f_{\nu'}], \tag{2.1}$$

$$\nabla \cdot \boldsymbol{E} = 4\pi\varrho, \tag{2.2}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{2.3}$$

$$\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t} = -\nabla \times \boldsymbol{E}, \tag{2.4}$$

$$\nabla \times \boldsymbol{B} = \frac{1}{c}\left(4\pi\boldsymbol{j} + \frac{\partial \boldsymbol{E}}{\partial t}\right), \tag{2.5}$$

where ϱ is the electric charge density, $\boldsymbol{j}$ is the current density, ν is the particle species index, c the speed of light, t the time, $\boldsymbol{E}$ and $\boldsymbol{B}$ are the electric and magnetic fields, respectively, and $C_{\nu\nu'}[f_\nu, f_{\nu'}]$ accounts for the effects on species ν from Coulomb collisions with species ν'. We relate the charge density and current density to the particle distribution function, f_ν, via the velocity space integrals

$$\varrho = \sum_\nu Z_\nu e \int \mathrm{d}^3 v f_\nu, \tag{2.6}$$

$$\boldsymbol{j} = \sum_\nu Z_\nu e \int \mathrm{d}^3 v \boldsymbol{v} f_\nu, \tag{2.7}$$

with $Z_\nu e$ the species charge, and e the proton charge. We take the frequency of fluctuations, ω, to be much less than the gyrofrequency of particles, Ω_ν, defined as $\Omega_\nu = Z_\nu e B/m_\nu c$, with B the magnitude of the magnetic field, and m_ν the species mass.

We decompose quantities into their mean and fluctuating parts. The mean components determine the evolution of the background plasma, and are found by averaging over all fluctuations. We represent the average of a given quantity, h, over all fluctuations by $\langle h \rangle_{\mathrm{turb}}$ and define this average as

$$\langle h(t) \rangle_{\mathrm{turb}} = \frac{1}{T}\int_{t-T/2}^{t+T/2} \mathrm{d}t' \langle h(t') \rangle_\perp, \tag{2.8}$$

where T is some intermediate time shorter than the (transport) time scale associated with mean profile evolution, and longer than time scales associated with typical fluctuations,

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

such that $\omega^{-1} \ll T \ll L/v_{\mathrm{th},i}$, where L is the system size, and $v_{\mathrm{th},\nu} \doteq \sqrt{2T_\nu/m_\nu}$ the species thermal velocity, with T_ν the species temperature. Here, $\langle\cdot\rangle_\perp$ is an appropriately defined spatial average over a length l satisfying $\rho_{\mathrm{ref}} \ll l \ll L$ for a surface perpendicular to the magnetic field (Abel *et al.* 2013). The distribution function is decomposed as $f_\nu = F_\nu + \delta f_\nu$ with $F_\nu = \langle f_\nu \rangle_{\mathrm{turb}}$ determining the profile of the equilibrium plasma, and δf_ν the contribution from plasma fluctuations.

We take the Larmor radius of particles, $\rho_\nu = v_\perp/\Omega_\nu$, to be much smaller than the system size, L, where $v_\perp$ is the magnitude of velocity in the plane perpendicular to the magnetic field. We henceforth take the reference scale to be ordered the same as the ion scale, with $\rho_{\mathrm{ref}} = v_{\mathrm{th,ref}}/\Omega_{\mathrm{ref}} \sim \rho_i$. We expand the Vlasov–Maxwell equations including collisions, (2.1)–(2.5), in the small parameter $\epsilon \sim \rho_\star \ll 1$, with $\rho_\star = \rho_i/L$ the ratio of the reference ion gyroradius to the system size, and equate terms of equivalent order. We order terms as follows:

$$\epsilon \sim \rho_\star \doteq \frac{\rho_i}{L} \sim \frac{\omega}{\Omega_i} \sim \frac{k_\parallel}{k_\perp} \sim \frac{e\phi}{T_\nu} \sim \frac{\delta B}{B_0} \sim \frac{\delta f_\nu}{F_{0,\nu}} \ll 1 \quad \text{and} \quad k_\perp \rho_i \sim 1. \tag{2.9}$$

In the above orderings we have decomposed the magnetic field into the equilibrium contribution, $\boldsymbol{B}_0$, and the magnetic fluctuations, $\delta\boldsymbol{B}$, such that the full magnetic field is $\boldsymbol{B} = \boldsymbol{B}_0 + \delta\boldsymbol{B}$. We have also introduced the parallel and perpendicular wavenumbers, $k_\parallel = \boldsymbol{k}\cdot\hat{\boldsymbol{b}}$ and $\boldsymbol{k}_\perp = (\boldsymbol{I} - \hat{\boldsymbol{b}}\hat{\boldsymbol{b}})\cdot\boldsymbol{k}$, with $\hat{\boldsymbol{b}}$ the unit vector in the direction of the magnetic field, and $\boldsymbol{I}$ the identity matrix. The perturbed electric potential has been introduced as ϕ. We also expand the mean and fluctuating components of the distribution function

$$\left.\begin{aligned} F_\nu &= F_{0,\nu} + F_{1,\nu} + F_{2,\nu} \cdots, \\ \delta f_\nu &= \delta f_{1,\nu} + \delta f_{2,\nu} + \cdots, \end{aligned}\right\} \tag{2.10}$$

with $F_{0,\nu} \sim f_\nu$, $F_{1,\nu} \sim \delta f_{1,\nu} \sim \epsilon f_\nu$, $F_{2,\nu} \sim \delta f_{2,\nu} \sim \epsilon^2 f_\nu$ and so on. Equilibrium quantities are taken to have characteristic length scales of order L, and evolve slowly on the long transport time scale $\tau_E^{-1} \sim \epsilon^3\Omega_i$; they are thus understood to be static during our considerations. Small-scale fluctuations, captured in δf_ν, have characteristic length scales of the order $\rho_i \sim \epsilon L$ and frequencies $\omega \sim \epsilon\Omega_i$.

The charged particles follow magnetic field lines and perform rapid gyration in the plane perpendicular to the field. We introduce $\boldsymbol{R}_\nu$ as the gyro-centre for species ν, and $\boldsymbol{r}$ to indicate the spatial position vector for any given particle. These descriptions of particle location are related through $\boldsymbol{R}_\nu = \boldsymbol{r} - \boldsymbol{\rho}_\nu(\vartheta)$, with $\boldsymbol{\rho}_\nu(\vartheta) = \hat{\boldsymbol{b}}_0 \times \boldsymbol{v}/\Omega_\nu$ the velocity-dependent vector gyroradius. Our analysis need only consider $\hat{\boldsymbol{b}}_0$; hence, we shall set $\hat{\boldsymbol{b}} \equiv \hat{\boldsymbol{b}}_0$ in the remainder of this paper for notational brevity. The gyrophase, ϑ, characterises this gyro-motion, and has a large associated frequency $|\dot{\vartheta}| \approx \Omega_\nu$. A natural approach is to average over these fast oscillations by introducing a gyroaverage, defined through

$$\langle h(\boldsymbol{r})\rangle_{\boldsymbol{R}_\nu} = \langle h(\boldsymbol{R}_\nu + \boldsymbol{\rho}_\nu(\vartheta))\rangle_{\boldsymbol{R}_\nu} = \frac{1}{2\pi}\int_0^{2\pi} h(\boldsymbol{R}_\nu + \boldsymbol{\rho}_\nu(\vartheta))\,\mathrm{d}\vartheta, \tag{2.11}$$

$$\langle h(\boldsymbol{R}_\nu)\rangle_{\boldsymbol{r}} = \langle h(\boldsymbol{r} - \boldsymbol{\rho}_\nu(\vartheta))\rangle_{\boldsymbol{r}} = \frac{1}{2\pi}\int_0^{2\pi} h(\boldsymbol{r} - \boldsymbol{\rho}_\nu(\vartheta))\,\mathrm{d}\vartheta, \tag{2.12}$$

where $\boldsymbol{R}_\nu$ and $\boldsymbol{r}$ are held constant when performing the ϑ integrations in (2.11) and (2.12), respectively.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

Expanding the Vlasov–Maxwell equations, then equating terms of order ϵ^{-1} with respect to $v_{\mathrm{th},i}F_{0,i}/L$, we obtain the equation $\partial F_{0,\nu}/\partial\vartheta = 0$, which demands $F_{0,\nu}$ be independent of gyrophase. In the presence of modest collisionality[2] zeroth-order terms provide the further constraint that the equilibrium component is a Maxwellian in velocities

$$F_{0,\nu} := \frac{n_\nu}{(\pi v_{\mathrm{th},\nu}^2)^{3/2}} \mathrm{e}^{-v^2/v_{\mathrm{th},\nu}^2}, \tag{2.13}$$

where n_ν represents the species density.

Equating terms ordered ϵ^1 gives the evolution equation for first-order perturbations. This exists within a six-dimensional phase space with coordinate choice $\{\boldsymbol{R}_\nu, v_\parallel, \mu_\nu, \vartheta\}$. Here, μ_ν is the magnetic moment defined as $\mu_\nu = m_\nu v_\perp^2/2B$ and is a conserved quantity to the order of consideration. Gyrophase dependence is removed by gyroaveraging the full equation, reducing the phase-space dimensionality by one. As a result the gyroaveraged fluctuating distribution function of guiding centres arises, denoted using $g_\nu(\boldsymbol{R}_\nu, v_\parallel, \mu_\nu) \doteq \langle \delta f_\nu \rangle_{\boldsymbol{R}_\nu} = \delta f_\nu + (Z_\nu e/T_\nu)F_{0,\nu}[\phi - \langle\phi\rangle_{\boldsymbol{R}_\nu}]$, such that g_ν is gyrophase independent.

We expand the total time derivative in (2.1) in terms of partial derivatives in $\{t, v_\parallel, \mu_\nu, \vartheta, \boldsymbol{R}_\nu\}$ – with each partial derivative taken assuming all other variables are held fixed, unless explicitly stated otherwise – and then gyroaveraging over ϑ to reduce the dimensionality of our equations from 6 to 5. We define a 'low-flow', or 'drift', ordering to be when the flow speed is ordered as ρ_* small compared with the thermal speed. In this ordering the resulting linear, electromagnetic gyrokinetic equation including collisions is

$$\begin{aligned}
&\frac{\partial g_\nu}{\partial t} + v_\parallel \left[\hat{\boldsymbol{b}}\cdot\nabla g_\nu + \frac{Z_\nu e}{T_\nu}F_{0,\nu}\hat{\boldsymbol{b}}\cdot\nabla\langle\chi\rangle_{\boldsymbol{R}_\nu}\right] + \boldsymbol{v}_{M,\nu}\cdot\left[\nabla_\perp g_\nu + \frac{Z_\nu e}{T_\nu}F_{0,\nu}\nabla_\perp\langle\chi\rangle_{\boldsymbol{R}_\nu}\right] \\
&-\frac{\mu_\nu}{m_\nu}\hat{\boldsymbol{b}}\cdot\nabla B_0\frac{\partial g_\nu}{\partial v_\parallel} + \langle\boldsymbol{v}_\chi\rangle_{\boldsymbol{R}_\nu}\cdot\nabla|_{\mathcal{E}}\,F_{0,\nu} + \frac{Z_\nu e}{T_\nu}\frac{\mu_\nu}{m_\nu c}\hat{\boldsymbol{b}}\cdot\nabla B_0 F_{0,\nu}\langle A_\parallel\rangle_{\boldsymbol{R}_\nu} = C_\nu^{(l)},
\end{aligned} \tag{2.14}$$

where $C_\nu^{(l)} = \sum_{\nu'} C_{\nu,\nu'}^l$ is the linearised collision operator taken to be self-adjoint such that $C_\nu^{(l)\dagger} = C_\nu^{(l)}$.[3] We have introduced $\chi = \phi - \boldsymbol{v}\cdot\boldsymbol{A}/c$ as the gyrokinetic potential, with $\boldsymbol{A} = A_\parallel\hat{\boldsymbol{b}} + \boldsymbol{A}_\perp$ the fluctuating magnetic vector potential, ($\delta\boldsymbol{B} = \nabla\times\boldsymbol{A}$), which has been decomposed into components parallel and perpendicular to the equilibrium magnetic field. The gradient acting on the Maxwellian appears as $\nabla|_{\mathcal{E}}$. This indicates that the derivative has been taken at constant kinetic energy, $\mathcal{E} = m_\nu v^2/2$, rather than at fixed $\{v_\parallel, \mu_\nu\}$ variables, in contrast to the other spatial gradients appearing in equation (2.14). Finally, $\boldsymbol{v}_{M,\nu}$ and $\boldsymbol{v}_\chi$ are the magnetic and generalised $\boldsymbol{E}\times\boldsymbol{B}$ drifts defined through

$$\boldsymbol{v}_{M,\nu} = \frac{1}{\Omega_\nu}\hat{\boldsymbol{b}}\times\left(\frac{\mu_\nu}{m_\nu}\nabla B + v_\parallel^2\boldsymbol{\kappa}\right), \tag{2.15}$$

$$\boldsymbol{v}_\chi = \frac{c}{B}\hat{\boldsymbol{b}}\times\nabla_\perp\chi, \tag{2.16}$$

with $\boldsymbol{\kappa} = \hat{\boldsymbol{b}}\cdot\nabla\hat{\boldsymbol{b}}$ the equilibrium magnetic field curvature.

[2] This requires the collisionality $\nu_* \gtrsim \rho_*\omega$ – a regime within which we work.

[3] It should be noted that this encapsulates a broad range of collision operators including the linearised Landau, and the linearised Fokker–Planck collision operators.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

The system is closed by the field equations consisting of quasineutrality, $\sum_\nu Z_\nu \delta n_\nu = 0$, with δn_ν the perturbed density, and the low-frequency Ampère's law, $\nabla \times \delta \boldsymbol{B} = (4\pi/c)\delta \boldsymbol{J}$, with $\delta \boldsymbol{B}$ and $\delta \boldsymbol{J}$ the fluctuating magnetic field and current density, respectively. When written in terms of the distribution function these relations become

$$\sum_\nu Z_\nu \int \mathrm{d}^3 v \left[\langle g_\nu \rangle_{\boldsymbol{R}_\nu} + \frac{Z_\nu e}{T_\nu} F_{0,\nu} \left(\left\langle \langle \chi \rangle_{\boldsymbol{R}_\nu} \right\rangle_{\boldsymbol{r}} - \phi \right) \right] = 0, \tag{2.17}$$

$$\nabla_\perp^2 A_\parallel - \frac{4\pi}{c} \sum_\nu Z_\nu e \int \mathrm{d}^3 v v_\parallel \left[g_\nu + \frac{Z_\nu e}{T_\nu c} F_{0,\nu} v_\parallel \left\langle \langle A_\parallel \rangle_{\boldsymbol{R}_\nu} \right\rangle_{\boldsymbol{r}} \right] = 0, \tag{2.18}$$

$$\nabla_\perp \delta B_\parallel - \frac{4\pi}{c} \sum_\nu Z_\nu e \int \mathrm{d}^3 v \nabla \cdot \left[\left(g_\nu + \frac{Z_\nu e}{T_\nu} F_{0,\nu} \left\langle \langle \chi \rangle_{\boldsymbol{R}_\nu} \right\rangle_{\boldsymbol{r}} \right) (\hat{\boldsymbol{b}} \times \boldsymbol{v}_\perp) \right] = 0, \tag{2.19}$$

where $\delta B_\parallel = \delta \boldsymbol{B} \cdot \hat{\boldsymbol{b}} = (\nabla \times \boldsymbol{A}_\perp) \cdot \hat{\boldsymbol{b}}$ is the parallel component of the perturbed magnetic field.

2.1. *Magnetic coordinates*

We can express the magnetic field using the Clebsch representation

$$\boldsymbol{B} = \nabla \alpha \times \nabla \psi, \tag{2.20}$$

where ψ and α are the flux surface and field line labels, respectively. We choose to work in field-aligned coordinates (Beer, Cowley & Hammett 1995) denoted by (x, y, z), with z the position along the magnetic field line and (x, y) the position in the plane perpendicular to $\hat{\boldsymbol{b}}$. Together, the unit vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{b}}\}$ form a left-handed, orthonormal basis. We relate the coordinates (x, y) to (ψ, α) using

$$\left.\begin{aligned} x &= \frac{\mathrm{d}x}{\mathrm{d}\psi}(\psi - \psi_0), \\ y &= \frac{\mathrm{d}y}{\mathrm{d}\alpha}(\alpha - \alpha_0), \end{aligned}\right\} \tag{2.21}$$

with (ψ_0, α_0) denoting the values at the centre of the domain.

We take the discrete Fourier transform in $\{x, y\}$ whilst retaining the real-space coordinate for the parallel direction, and evaluate nonlinear terms pseudo-spectrally in order to retain spectral accuracy in spatial derivatives when implementing the equations numerically.[4] The discrete Fourier transform in $\{x, y\}$

$$g_\nu(x, y, z, v_\parallel, \mu_\nu, t) = \sum_{k_x, k_y} \hat{g}_{\boldsymbol{k},\nu}(z, v_\parallel, \mu_\nu, t) e^{(\mathrm{i}k_x x + \mathrm{i}k_y y)}, \tag{2.22}$$

is justified provided the condition $k_x \sim k_y \gg 1/L$ is satisfied. This translates to a requirement that the turbulent fluctuations at the edges of our domain are decorrelated, such that they may be considered statistically identical, and periodic boundary conditions can be enforced in (x, y) (Beer *et al.* 1995).

[4] Spatial derivatives are exact in the Fourier representation and are hence more accurate than using a finite-difference scheme in configuration space.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

2.2. *System equations*

We next transform our evolution equations (2.14), and (2.17)–(2.19). The definition of χ is used to find the Fourier transform of the quantity $\langle\chi\rangle_{\boldsymbol{R}_\nu} = \langle\phi\rangle_{\boldsymbol{R}_\nu} - v_\parallel\langle A_\parallel\rangle_{\boldsymbol{R}_\nu}/c - \langle\boldsymbol{v}_\perp\cdot\boldsymbol{A}_\perp\rangle_{\boldsymbol{R}_\nu}/c$

$$\mathcal{F}_{\boldsymbol{k}}\left[\langle\chi\rangle_{\boldsymbol{R}_\nu}\right] \doteq \hat{\chi}_{\boldsymbol{k},\nu} = \left[\mathrm{J}_{0,\boldsymbol{k},\nu}\hat{\phi}_{\boldsymbol{k}} - \frac{v_\parallel}{c}\mathrm{J}_{0,\boldsymbol{k},\nu}\hat{A}_{\parallel,\boldsymbol{k}} + 2\frac{\mu_\nu}{Z_\nu e}\frac{\mathrm{J}_{1,\boldsymbol{k},\nu}}{a_{\boldsymbol{k},\nu}}\delta\hat{B}_{\parallel,\boldsymbol{k}}\right], \tag{2.23}$$

where we define $\mathcal{F}_{\boldsymbol{k}}[\langle\chi\rangle_{\boldsymbol{R}_\nu}] \doteq \hat{\chi}_{\boldsymbol{k},\nu}$ to be the Fourier components of $\langle\chi\rangle_{\boldsymbol{R}_\nu}$. The variables $\mathrm{J}_{n,\boldsymbol{k},\nu}$ above are the nth-order Bessel functions of the first kind for species ν. The Bessel functions arise naturally as a result of the gyroaverages that appear in (2.14), and have argument $a_{\boldsymbol{k},\nu} = k_\perp v_\perp/\Omega_\nu$. Utilising this, we can write each of our functional operators in terms of $\{\hat{g}_{\boldsymbol{k},\nu}, \hat{\phi}_{\boldsymbol{k}}, \hat{A}_{\parallel,\boldsymbol{k}}, \delta\hat{B}_{\parallel,\boldsymbol{k}}\}$. The resulting gyrokinetic equation is

$$\begin{aligned}\hat{G}_{\boldsymbol{k},\nu} = {} & \frac{\partial\hat{g}_{\boldsymbol{k},\nu}}{\partial t} + v_\parallel\hat{\boldsymbol{b}}\cdot\boldsymbol{\nabla}z\left[\frac{\partial\hat{g}_{\boldsymbol{k},\nu}}{\partial z} + \frac{Z_\nu e}{T_\nu}\frac{\partial\hat{\chi}_{\boldsymbol{k},\nu}}{\partial z}F_{0,\nu}\right] \\ & + \mathrm{i}\omega_{*,\boldsymbol{k},\nu}F_{0,\nu}\hat{\chi}_{\boldsymbol{k},\nu} + \mathrm{i}\omega_{d,\boldsymbol{k},\nu}\left[\hat{g}_{\boldsymbol{k},\nu} + \frac{Z_\nu e}{T_\nu}\hat{\chi}_{\boldsymbol{k},\nu}F_{0,\nu}\right] \\ & - \frac{\mu_\nu}{m_\nu}(\hat{\boldsymbol{b}}\cdot\boldsymbol{\nabla}B_0)\frac{\partial\hat{g}_{\boldsymbol{k},\nu}}{\partial v_\parallel} + \frac{Z_\nu e}{T_\nu}\frac{\mu_\nu}{m_\nu c}(\hat{\boldsymbol{b}}\cdot\boldsymbol{\nabla}B_0)F_{0,\nu}\mathrm{J}_{0,\boldsymbol{k},\nu}\hat{A}_{\parallel,\boldsymbol{k}} - \hat{C}_{\boldsymbol{k},\nu}[\hat{g}_{\boldsymbol{k},\nu}] \\ = {} & 0.\end{aligned} \tag{2.24}$$

The drift frequencies, $\omega_{d,\nu}$ and $\omega_{*,\nu}$, correspond to the magnetic drift frequencies resulting from the gradient and curvature of the magnetic field and the diamagnetic drift respectively, and are given by

$$\omega_{d,\boldsymbol{k},\nu} = \frac{1}{\Omega_\nu}(v_\parallel^2\boldsymbol{v}_\kappa + \mu_\nu\boldsymbol{v}_{\nabla B})\cdot(k_x\boldsymbol{\nabla}x + k_y\boldsymbol{\nabla}y), \tag{2.25}$$

$$\omega_{*,\boldsymbol{k},\nu} = \frac{ck_y}{B_0}\frac{\mathrm{d}y}{\mathrm{d}\alpha}\frac{\mathrm{d}\ln F_{0,\nu}}{\mathrm{d}\psi}, \tag{2.26}$$

with $\boldsymbol{v}_\kappa = \hat{\boldsymbol{b}}\times(\hat{\boldsymbol{b}}\cdot\boldsymbol{\nabla}\hat{\boldsymbol{b}})$, and $\boldsymbol{v}_{\nabla B} = \hat{\boldsymbol{b}}\times\boldsymbol{\nabla}B$.

The distribution function and field quantities are time-dependent functions, and their evolution equations admit normal mode solutions. For a generic set of initial conditions, the fastest growing (or slowest decaying) mode is likely to have an initial amplitude that is comparable to or smaller than other normal modes in the system. This results in an initial transient period due to the competition between these different normal modes, followed by exponential growth or decay once the fastest growing (or slowest decaying) mode has a much larger amplitude than all others. Therefore, we decompose $\hat{f}_k = \sum_j \bar{f}_{k,j}\mathrm{e}^{\gamma_{k,j}t}$[5] with $\gamma_{k,j}\in\mathbb{C}$ the complex frequencies, and $\bar{f}_{k,j}$ the amplitude for the jth normal mode.[6] Quasineutrality and Ampère's law, along with the linearity of the gyrokinetic equation, ensure that all fluctuating quantities for an equivalent normal mode share the same complex frequency and will thus exhibit the same time-dependent behaviour during the

[5]We remind the reader that, here, $\boldsymbol{k}$ denotes the Fourier mode from the spatial decomposition, and j is the subscript denoting the temporal normal mode.

[6]It is worth noting that there are instances when the linear growth rates of two independent modes may be very similar, and a single dominant growth rate may require long run times to become evident.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

period of exponential growth or decay. After a sufficiently long period of time (the exact length of which will depend on both the relative growth rates and starting amplitudes of each mode) the fastest growing mode will dominate, meaning we can approximate the time dependent behaviour after large times using a single temporal mode; $\hat{f}_{\boldsymbol{k}} \approx \bar{f}_{\boldsymbol{k},0}\mathrm{e}^{\gamma_{\boldsymbol{k},0}\tilde{t}}$, with $\mathrm{Re}(\gamma_{\boldsymbol{k},0}) > \mathrm{Re}(\gamma_{\boldsymbol{k},j})$, $\forall j,\ j \neq 0$. Using this normal mode decomposition in the gyrokinetic equation and considering behaviour beyond the transient period gives

$$\begin{aligned}\hat{G}_{\boldsymbol{k},\nu} = {} & \gamma_{\boldsymbol{k},0}\bar{g}_{\boldsymbol{k},\nu,0} + v_\parallel \hat{\boldsymbol{b}}\cdot\nabla z\left[\frac{\partial \bar{g}_{\boldsymbol{k},\nu,0}}{\partial z} + \frac{Z_\nu e}{T_\nu}\frac{\partial \bar{\chi}_{\boldsymbol{k},0}}{\partial z}F_{0,\nu}\right] \\ & + \mathrm{i}\omega_{*,\boldsymbol{k},\nu}F_{0,\nu}\bar{\chi}_{\boldsymbol{k},0} + \mathrm{i}\omega_{d,\boldsymbol{k},\nu}\left[\bar{g}_{\boldsymbol{k},\nu,0} + \frac{Z_\nu e}{T_\nu}\bar{\chi}_{\boldsymbol{k},0}F_{0,\nu}\right] \\ & - \mu_\nu \hat{\boldsymbol{b}}\cdot\nabla B_0\frac{\partial \bar{g}_{\boldsymbol{k},\nu,0}}{\partial v_\parallel} + \frac{Z_\nu e}{T_\nu}\frac{\mu_\nu}{m_\nu c}(\hat{\boldsymbol{b}}\cdot\nabla B_0)F_{0,\nu}\mathrm{J}_{0,\boldsymbol{k},\nu}\bar{A}_{\parallel,\boldsymbol{k},0} - \hat{C}_{\boldsymbol{k},\nu}[\bar{g}_{\boldsymbol{k},\nu,0}]. \end{aligned} \tag{2.27}$$

The corresponding transformed, normalised field equations are $\hat{Q}_{\boldsymbol{k}} = \hat{M}_{\boldsymbol{k}} = \hat{N}_{\boldsymbol{k}} = 0$, where

$$\hat{Q}_{\boldsymbol{k}} = \sum_\nu Z_\nu e\left\{\frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2 v\, \mathrm{J}_{0,\boldsymbol{k},\nu}\bar{g}_{\boldsymbol{k},\nu,0} + \frac{Z_\nu e n_\nu}{T_\nu}\left(\Gamma_{0,\boldsymbol{k},\nu} - 1\right)\bar{\phi}_{\boldsymbol{k},0} + \frac{n_\nu}{B_0}\Gamma_{1,\boldsymbol{k},\nu}\delta\bar{B}_{\parallel,\boldsymbol{k},0}\right\}, \tag{2.28}$$

$$\begin{aligned}\hat{M}_{\boldsymbol{k}} = {} & -\frac{4\pi}{k_\perp^2 c}\sum_\nu Z_\nu e\frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2 v v_\parallel \mathrm{J}_{0,\boldsymbol{k},\nu}\bar{g}_{\boldsymbol{k},\nu,0} \\ & + \left[1 + \frac{4\pi}{k_\perp^2 c^2}\sum_\nu \frac{(Z_\nu e)^2 n_\nu}{m_\nu}\Gamma_{0,\boldsymbol{k},\nu}\right]\bar{A}_{\parallel,\boldsymbol{k},0}, \end{aligned} \tag{2.29}$$

and

$$\begin{aligned}\hat{N}_{\boldsymbol{k}} = {} & 8\pi\sum_\nu \frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2 v \frac{\mathrm{J}_{1,\boldsymbol{k},\nu}}{a_{\boldsymbol{k},\nu}}\mu_\nu \bar{g}_{\boldsymbol{k},\nu,0} + \left[4\pi\sum_\nu \frac{Z_\nu e n_\nu}{B_0}\Gamma_{1,\boldsymbol{k},\nu}\right]\bar{\phi}_{\boldsymbol{k},0} \\ & + \left[1 + 16\pi\sum_\nu \frac{n_\nu T_\nu}{B_0^2}\Gamma_{2,\boldsymbol{k},\nu}\right]\delta\bar{B}_{\parallel,\boldsymbol{k},0}, \end{aligned} \tag{2.30}$$

with $\int \mathrm{d}^2 v \doteq \int \mathrm{d}\mu \int \mathrm{d}v_\parallel$. In the above we have defined

$$\Gamma_{0,\boldsymbol{k},\nu} = \mathrm{I}_0(\alpha_{\boldsymbol{k},\nu})\mathrm{e}^{-\alpha_{\boldsymbol{k},\nu}}, \tag{2.31}$$

$$\Gamma_{1,\boldsymbol{k},\nu} = [\mathrm{I}_0(\alpha_{\boldsymbol{k},\nu}) - \mathrm{I}_1(\alpha_{\boldsymbol{k},\nu})]\mathrm{e}^{-\alpha_{\boldsymbol{k},\nu}}, \tag{2.32}$$

$$\Gamma_{2,\boldsymbol{k},\nu} = \mathrm{I}_1(\alpha_{\boldsymbol{k},\nu})\mathrm{e}^{-\alpha_{\boldsymbol{k},\nu}}, \tag{2.33}$$

where I_0 and I_1 are modified Bessel functions of the first kind, and $\alpha_{\boldsymbol{k},\nu} = k_\perp^2\rho_\nu^2/2$.

3. Adjoint method for gyrokinetics

The adjoint-based optimisation method is a tool that, at its heart, efficiently calculates derivatives of a desired quantity with respect to a potentially large number of parameters. The associated computational cost depends only on the expense of solving both the

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

objective functional and adjoint system of equations, and is essentially independent of the dimension of the parameter space. Adjoint methods have already been successfully applied to certain other aspects of MCF devices, such as optimising coil configurations for stellarator geometries (see, e.g. Paul *et al.* 2018; Geraldini, Landreman & Paul 2021; Nies *et al.* 2022). The novelty here is to apply the adjoint method to geometric optimisation and plasma microstability, which includes the complexity of the full linearised gyrokinetic system.

We now turn to the gyrokinetic equation and the objective of minimising the dominant linear growth rate, $\gamma_{\boldsymbol{k},0}$, with respect to a set of currently unspecified parameters $\{p_i\}$, which we take to be the components of vector $\boldsymbol{p}$. This section outlines how one can take advantage of the adjoint method to efficiently obtain the gradient $\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma_{\boldsymbol{k},0}$. Although a finite-difference scheme could be used to obtain such a gradient, this becomes computationally expensive when the dimension of $\boldsymbol{p}$, $\mathcal{N}_p$, is large: for each gradient a finite-difference scheme demands we solve the system equations $\mathcal{N}_p + 1$ times. In contrast, the adjoint method allows us to solve the system equations only once, and in exchange one must solve the set of adjoint equations, for which the cost is computationally equivalent to the original system equations.

We start by considering the general case of low-flow, linear, electromagnetic, δf-gyrokinetics including collisions, with equations (2.27)–(2.30) as the functional operators defining our system in the long time limit.

3.1. *General formalism*

Since our equations are limited to the linear regime, there is no coupling of different Fourier modes and it is thus possible to consider each perpendicular wavenumber individually. Given that we are also considering the post-transient limit with only one dominant growth rate, it will henceforth be assumed that only a single perpendicular wavenumber is being considered, and we will thus drop the $\boldsymbol{k}$ subscript, along with the subscript that denotes the dominant mode for the distribution function and field quantities. For notational brevity we shall also drop the overbars that appear on the distribution function and fields that denoted normal mode decomposition. Hence, everywhere g_v, ϕ, $A_\parallel$ and $\delta B_\parallel$ appear it shall be assumed that they contain suppressed spatial and temporal Fourier subscripts.

Consider now a set of variables that influences our system and which exists within a parameter space spanned by all possible $\boldsymbol{p}$. We take the set $\{p_i\}$, $i \in [1, \mathcal{N}_p]$, to be linearly independent, with no time variation, and explore how variations within this space affect the linear growth rate. At present we need not specify which variables are denoted by $\boldsymbol{p}$, and thus derive a general set of adjoint equations for the gyrokinetic-Maxwell system above, (2.27)–(2.30).

Our objective functional, $\hat{G}_v$, is a linear functional of $\{g_v, \phi, A_\parallel, \delta B_\parallel\}$, which are coupled to the fields through the field equations, (2.28)–(2.30). When taking derivatives of G_v we invariably end up with derivatives acting on all four of these variables. This is undesirable because it requires we calculate the gradients of $\hat{g}_v$ and the fields, which in turn requires us to solve the gyrokinetic system $\mathcal{N}_p + 1$ times, as discussed above. In order to eliminate these four derivatives we use something akin to the method of Lagrange multipliers and introduce four adjoint variables which multiply the corresponding constraint equations. The optimisation Lagrangian is thus

$$\mathcal{L} := \left\langle \hat{G}_v, \lambda_v \right\rangle_{z,v,v} + \left\langle \hat{Q}, \xi \right\rangle_z + \left\langle \hat{M}, \zeta \right\rangle_z + \left\langle \hat{N}, \sigma \right\rangle_z , \tag{3.1}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

with the angle brackets representing inner products defined through[7]

$$\left.\begin{aligned}\langle a,b\rangle_z &= \int \frac{\mathrm{d}z}{B_0\,\hat{\boldsymbol{b}}\cdot\nabla z} ab^*, \quad \langle a,b\rangle_{v,\nu} = \sum_\nu \frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2v\, ab^*,\\ \langle a,b\rangle_{z,v,\nu} &= \sum_\nu \int \frac{\mathrm{d}z}{B_0\hat{\boldsymbol{b}}\cdot\nabla z}\frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2v\, ab^*.\end{aligned}\right\} \tag{3.2}$$

We have introduced the set of adjoint variables λ_ν, ξ, ζ and σ, whose forms are to be determined.[8] We identify λ_ν as the adjoint variable to the distribution function, g_ν, whereas ξ, ζ and σ are adjoint to the field variables (referred to henceforth as adjoint fields). The quantities $\hat{G}_\nu$, $\hat{Q}$, $\hat{M}$ and $\hat{N}$ are our objective functions, and we remark that since $\hat{G}_\nu = \hat{Q} = \hat{M} = \hat{N} = 0$ for a consistent set of $\{\boldsymbol{p}_0, g_\nu(\boldsymbol{p}_0), \phi(\boldsymbol{p}_0), A_\parallel(\boldsymbol{p}_0), \delta B_\parallel(\boldsymbol{p}_0)\}$ we have $\mathcal{L}|_{\boldsymbol{p}_0} = 0$ for all choices of adjoint variables. For later convenience we decompose the functional operators into their components that act on g_ν, ϕ, $A_\parallel$ and $\delta B_\parallel$ separately

$$\begin{aligned}\hat{G}_\nu[\boldsymbol{p}; g_\nu, \phi, A_\parallel, \delta B_\parallel] = \hat{G}_{g,\nu}[\boldsymbol{p}; g_\nu] + \hat{G}_{\phi,\nu}[\boldsymbol{p}; \phi] + \hat{G}_{A_\parallel,\nu}[\boldsymbol{p}; A_\parallel] + \hat{G}_{B_\parallel,\nu}[\boldsymbol{p}; \delta B_\parallel]\\ - \hat{C}_\nu[\boldsymbol{p}; g_\nu],\end{aligned} \tag{3.3}$$

$$\hat{Q}[\boldsymbol{p}; g_\nu, \phi, \delta B_\parallel] = \left\langle \hat{Q}_{g,\nu}[\boldsymbol{p}; g_\nu], \mathbb{I}\right\rangle_{v,\nu} + \hat{Q}_\phi[\boldsymbol{p}; \phi] + \hat{Q}_{B_\parallel}[\boldsymbol{p}; \delta B_\parallel], \tag{3.4}$$

$$\hat{M}[\boldsymbol{p}; g_\nu, A_\parallel] = \left\langle \hat{M}_{g,\nu}[\boldsymbol{p}; g_\nu], \mathbb{I}\right\rangle_{v,\nu} + \hat{M}_{A_\parallel}[\boldsymbol{p}; A_\parallel], \tag{3.5}$$

$$\hat{N}[\boldsymbol{p}; g_\nu, \phi, \delta B_\parallel] = \left\langle \hat{N}_{g,\nu}[\boldsymbol{p}; g_\nu], \mathbb{I}\right\rangle_{v,\nu} + \hat{N}_\phi[\boldsymbol{p}; \phi] + \hat{N}_{B_\parallel}[\boldsymbol{p}; \delta B_\parallel], \tag{3.6}$$

with $\mathbb{I}$ simply equal to one. Explicit expressions for these operators are given in Appendix A.

We next consider taking the gradient of (3.1) with respect to the variables in our parameter space, $\boldsymbol{p}$. We now isolate all terms multiplying derivatives of $\{g_\nu, \phi, A_\parallel, \delta B_\parallel\}$ so that each of their coefficients can be set to zero. We are at liberty to do this because of the freedom that exists in choosing the adjoint variables introduced in (3.1). The gradient of (3.1) is expanded using (3.3)–(3.6)

$$\begin{aligned}\nabla_{\boldsymbol{p}}\mathcal{L} = \partial_{\boldsymbol{p}}\mathcal{L} &+ \left\langle \nabla_{\boldsymbol{p}} g_\nu, \hat{G}^\dagger_{g,\nu}[\boldsymbol{p}; \lambda_\nu] - \hat{C}^\dagger_\nu[\boldsymbol{p}; \lambda_\nu] + \hat{Q}^\dagger_{g,\nu}[\boldsymbol{p}; \xi] + \hat{M}^\dagger_{g,\nu}[\boldsymbol{p}; \zeta] + \hat{N}^\dagger_{g,\nu}[\boldsymbol{p}; \sigma]\right\rangle_{z,v,\nu}\\ &+ \left\langle \nabla_{\boldsymbol{p}}\phi, \hat{G}^\dagger_{\phi,\nu}[\boldsymbol{p}; \lambda_\nu] + \hat{Q}^\dagger_\phi[\boldsymbol{p}; \xi] + \hat{N}^\dagger_\phi[\boldsymbol{p}; \sigma]\right\rangle_z + \left\langle \nabla_{\boldsymbol{p}} A_\parallel, \hat{G}^\dagger_{A_\parallel,\nu}[\boldsymbol{p}; \lambda_\nu] + \hat{M}^\dagger_{A_\parallel}[\boldsymbol{p}; \zeta]\right\rangle_z\\ &+ \left\langle \nabla_{\boldsymbol{p}}\delta B_\parallel, \hat{G}^\dagger_{B_\parallel,\nu}[\boldsymbol{p}; \lambda_\nu] + \hat{Q}^\dagger_{B_\parallel}[\boldsymbol{p}; \xi] + \hat{N}^\dagger_{B_\parallel}[\boldsymbol{p}; \sigma]\right\rangle_z + \mathcal{B}.\end{aligned} \tag{3.7}$$

Here, the daggers that appear on the functional operators denote the adjoint of those operators with respect to the relevant inner product. The partial derivative, $\partial_{\boldsymbol{p}}$, is taken at fixed g_ν, ϕ, $A_\parallel$ and $\delta B_\parallel$. Note that it acts on the inner product itself in addition to the

[7]Note that these must be linear in their arguments, exhibit conjugate symmetry and satisfy the Cauchy–Schwartz inequality to be well defined. We also require that all arguments are square integrable in order for the optimisation to be finite.

[8]Note that the adjoint variables here are defined in Fourier space, so λ_ν, ξ, ζ and σ also contain suppressed $\boldsymbol{k}$ subscripts, and thus we are calculating each for a specific $\boldsymbol{k}$. The derivatives $\nabla_{\boldsymbol{p}}$ indicate how these Fourier components respond to external changes of the parameters $\boldsymbol{p}$ in our system.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

terms within it to account for the $\boldsymbol{p}$-dependence of the Jacobians[9] and functional operators. The term '$\mathcal{B}$' accounts for boundary terms that arise when we integrate by parts to invert operators such as $\langle \nabla_{\boldsymbol{p}} \partial_z g_\nu, \lambda_\nu \rangle_z$ onto the adjoint variables, $\langle \nabla_{\boldsymbol{p}} g_\nu, \partial_z \lambda_\nu \rangle_z$, and is given by

$$\begin{aligned}\mathcal{B} = \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}^2 v v_\parallel \lambda_\nu^* \Bigg[& \nabla_{\boldsymbol{p}} g_\nu + \frac{Z_\nu e}{T_\nu} \mathrm{J}_{0,\nu} F_{0,\nu} \nabla_{\boldsymbol{p}} \phi \\ & - 2 \frac{Z_\nu e}{T_\nu} v_\parallel \mathrm{J}_{0,\nu} F_{0,\nu} \nabla_{\boldsymbol{p}} A_\parallel + 4 \mu_\nu \frac{\mathrm{J}_{1,\nu}}{a_\nu} F_{0,\nu} \nabla_{\boldsymbol{p}} \delta B_\parallel \Bigg] \Bigg|_{z=-\infty}^{z=\infty} \\ & - \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}z \int \mathrm{d}\mu \, \mu_\nu \frac{\partial B_0}{\partial z} \lambda_\nu^* \nabla_{\boldsymbol{p}} g_\nu \Bigg|_{v_\parallel=-\infty}^{v_\parallel=\infty}. \end{aligned} \tag{3.8}$$

We can conveniently set these terms to zero by applying appropriate restrictions on our adjoint variables. These terms consequently define the boundary conditions we apply to our adjoint variables. The incoming boundary conditions along the magnetic field on g_ν are taken to be $g_\nu(z \to -\infty, v_\parallel > 0, \mu_\nu), g_\nu(z \to \infty, v_\parallel < 0, \mu_\nu) \to 0$, independently of $\boldsymbol{p}$, such that $d_{\boldsymbol{p}} g_\nu = 0$ at these limits. We impose the boundary condition on λ_ν^* to be $\lambda_\nu^*(z \to -\infty, v_\parallel < 0, \mu_\nu), \lambda_\nu^*(z \to \infty, v_\parallel > 0, \mu_\nu) \to 0$ in order to eliminate the boundary term arising from the z integration by parts, and hence remove the need to calculate $\nabla_{\boldsymbol{p}}\{g_\nu, \phi, A_\parallel, \delta B_\parallel\}$ at the boundaries in z. The boundary term arising from integration by parts in $v_\parallel$ is automatically satisfied as it is assumed that $g_\nu(z, v_\parallel \to \pm\infty, \mu_\nu) = 0$, $\forall\{z, \mu_\nu\}$ independently of $\boldsymbol{p}$. However, it is convenient to impose that $\lambda_\nu^*(z, v_\parallel \to \pm\infty, \mu_\nu) = 0$ such that λ_ν^* and g_ν satisfy similar boundary conditions, whilst also ensuring λ_ν is sensibly defined and normalisable.[10] The substitution $\lambda_\nu^{\leftrightarrow,*} = \lambda_\nu^*(z, -v_\parallel, \mu_\nu)$ is made such that the adjoint equations more closely resemble those in the original gyrokinetic system. This redefines the z-boundary condition on the adjoint variable $\lambda_\nu^{\leftrightarrow,*}(z \to -\infty, v_\parallel > 0, \mu_\nu), \lambda_\nu^{\leftrightarrow,*}(z \to \infty, v_\parallel < 0, \mu_\nu) \to 0$, which now mirrors those satisfied by g_ν.

Setting the remaining coefficients of $\nabla_{\boldsymbol{p}} g_\nu$, $\nabla_{\boldsymbol{p}}\phi$, $\nabla_{\boldsymbol{p}} A_\parallel$ and $\nabla_{\boldsymbol{p}} \delta B_\parallel$ in (3.7) equal to zero yields the constraint equations for the adjoint variables

$$\hat{G}_{g,\nu}^\dagger[\boldsymbol{p}; \lambda_\nu^{\leftrightarrow}] + \hat{Q}_{g,\nu}^\dagger[\boldsymbol{p}; \xi] + \hat{M}_{g,\nu}^\dagger[\boldsymbol{p}; \zeta] + \hat{N}_{g,\nu}^\dagger[\boldsymbol{p}; \sigma] - \hat{C}_\nu^\dagger[\boldsymbol{p}; \lambda_\nu^{\leftrightarrow}] = 0, \tag{3.9}$$

$$\langle \hat{G}_{\phi,\nu}^\dagger[\boldsymbol{p}; \lambda_\nu^{\leftrightarrow}] \rangle_{v,\nu} + \hat{Q}_\phi^\dagger[\boldsymbol{p}; \xi] + \hat{N}_\phi^\dagger[\boldsymbol{p}; \sigma] = 0, \tag{3.10}$$

$$\langle \hat{G}_{A_\parallel,\nu}^\dagger[\boldsymbol{p}; \lambda_\nu^{\leftrightarrow}] \rangle_{v,\nu} + \hat{M}_{A_\parallel}^\dagger[\boldsymbol{p}; \zeta] = 0, \tag{3.11}$$

$$\langle \hat{G}_{B_\parallel,\nu}^\dagger[\boldsymbol{p}; \lambda_\nu^{\leftrightarrow}] \rangle_{v,\nu} + \hat{Q}_{B_\parallel}^\dagger[\boldsymbol{p}; \xi] + \hat{N}_{B_\parallel}^\dagger[\boldsymbol{p}; \sigma] = 0. \tag{3.12}$$

The expressions for these adjoint operators are stated in Appendix B. The z-derivatives which appear in the $\hat{G}_\nu$ operators, under the velocity integrals, in (3.10)–(3.12) pose a potential difficulty; to calculate the adjoint fields, information is required for λ_ν^* at all z, for both positive and negative velocities. Given the boundary conditions on λ_ν and the propagation of information by advection this information is not readily available.

[9] It is noteworthy to point out that although the Jacobians present in the integrals (as well as the Lagrange multipliers themselves) are $\boldsymbol{p}$-dependent, we anticipate that when the distribution function and fields are evaluated at $\boldsymbol{p}_0$ their derivatives provide zero contribution as they multiply the functional operators, which are identically zero at $\boldsymbol{p}_0$. Hence, we would be justified in pulling the partial derivative through these inner products, and the final derivative of the growth rate obtained should be the same.

[10] An additional consequence is that this choice simplifies the implementation into an existing gyrokinetic code.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

This is akin to the problem faced when solving the gyrokinetic system; an incoming boundary condition is imposed on the distribution function at $z \to \pm\infty$ when $v_\parallel < 0$ and $v_\parallel > 0$, respectively. The outgoing boundary information is not known a priori but must instead be solved for. Anticipating the difficulties this will create for numerical solution, we take moments of (3.9) are taken to simplify the adjoint equations. This simplifies the adjoint equations, bringing them into a form more closely resembling the that of the gyrokinetic field equation. It should be emphasised that we retain a fully kinetic treatment in doing so.

A summary of this calculation can be found in Appendix C, and the result after algebraic manipulation is

$$\gamma^* \lambda_v^{\leftrightarrow} + v_\parallel \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial \lambda_v^{\leftrightarrow}}{\partial z} - \frac{\mu_v}{m_v} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial \lambda_v^{\leftrightarrow}}{\partial v_\parallel} - \mathrm{i}\omega_{d,v} \lambda_v^{\leftrightarrow} + Z_v e \mathrm{J}_{0,v} \xi - \frac{4\pi}{k_\perp^2} Z_v e \mathrm{J}_{0,v} \frac{v_\parallel}{c} \zeta + 8\pi \frac{\mathrm{J}_{1,v}}{a_v} \mu_v \sigma - \hat{C}_v[\lambda_v^{\leftrightarrow}] = 0, \tag{3.13}$$

$$\xi + \frac{1}{\eta} \sum_v \frac{2\pi B_0}{m_v} \int \mathrm{d}^2 v \left[\mathrm{i}\omega_{*,v} + \frac{Z_v e}{T_v} \gamma^* \right] \mathrm{J}_{0,v} F_{0,v} \lambda_v^{\leftrightarrow} = 0, \tag{3.14}$$

$$\zeta - \frac{1}{k_\perp^2} \sum_v \frac{2\pi B_0}{m_v} \int \mathrm{d}^2 v \frac{v_\parallel}{c} \left[\mathrm{i}\omega_{*,v} + \frac{Z_v e}{T_v} \gamma^* \right] \mathrm{J}_{0,v} F_{0,v} \lambda_v^{\leftrightarrow} = 0, \tag{3.15}$$

$$\sigma - \sum_v \frac{2\pi B_0}{m_v} \int \mathrm{d}^2 v \left(2 \frac{\mu_v}{Z_v e} \frac{\mathrm{J}_{1,v}}{a_v} \right) \left[\mathrm{i}\omega_{*,v} + \frac{Z_v e}{T_v} \gamma^* \right] F_{0,v} \lambda_v^{\leftrightarrow} = 0, \tag{3.16}$$

with $\eta = \sum_v (Z_v e)^2 n_v / T_v$. Noting that we can also rewrite $\hat{G}_v[\boldsymbol{p}; g_v, \phi, A_\parallel, \delta B_\parallel] = \gamma g_v + \hat{L}_v[\boldsymbol{p}; g_v, \phi, A_\parallel, \delta B_\parallel]$, and using $\nabla_p \mathcal{L}|_{\boldsymbol{p}_0} = 0$, we can rearrange equation (3.7) to obtain

$$\nabla_p \gamma \langle g_v, \lambda_v \rangle_{z,v,v} = - \left[\langle \partial_p \hat{L}_v, \lambda_v \rangle_{z,v,v} + \langle \partial_p \hat{Q}, \xi \rangle_z + \langle \partial_p \hat{M}, \zeta \rangle_z + \langle \partial_p \hat{N}, \sigma \rangle_z \right] \Big|_{\boldsymbol{p}_0}, \tag{3.17}$$

where the partial derivatives have been pulled inside the inner products as the contribution arising from the Jacobian derivatives is zero by virtue of $\hat{G}_v(\boldsymbol{p}_0) = \hat{Q}(\boldsymbol{p}_0) = \hat{M}(\boldsymbol{p}_0) = \hat{N}(\boldsymbol{p}_0) = 0$.

To solve for the derivative of the linear growth rate, the following procedure is taken: the gyrokinetic equations, (2.27)–(2.30), are solved to obtain g_v and γ, and then the equations (3.13)–(3.16) are used to solve for the adjoint variables. These quantities are then all fed into (3.17) to compute $\nabla_p \gamma$.

3.2. *Electrostatic, collisionless limit*

We consider now the electrostatic, collisionless limit, as we shall perform numerical tests in this regime. In the limit of small plasma β, meaning the ratio of plasma to magnetic pressure tends to zero, the magnetic field perturbations tend to zero. Hence, to extract the electrostatic, collisionless limit from these equations we set $A_\parallel = \delta B_\parallel = 0$, and $C_{v,v'} = 0$. The linear, collisionless, electrostatic gyrokinetic equation is

$$\hat{G}_v[\boldsymbol{p}; g_v, \phi] = \gamma g_v + v_\parallel \hat{\boldsymbol{b}} \cdot \nabla z \left[\frac{\partial g_v}{\partial z} + \frac{Z_v e}{T_v} \frac{\partial \mathrm{J}_{0,v} \phi}{\partial z} F_{0,v} \right] + \mathrm{i}\omega_{*,v} \mathrm{J}_{0,v} \phi F_{0,v} - \frac{\mu_v}{m_v} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial g_v}{\partial v_\parallel} + \mathrm{i}\omega_{d,v} \left[g_v + \frac{Z_v e}{T_v} \mathrm{J}_{0,v} \phi F_{0,v} \right], \tag{3.18}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

which is closed by the electrostatic limit of quasineutrality

$$\hat{Q}[\boldsymbol{p}; g_\nu, \phi] = \sum_\nu Z_\nu e \left[\frac{2\pi B_0}{m_\nu} \int \mathrm{d}^2 v \mathrm{J}_{0,\nu} g_\nu + \frac{Z_\nu n_\nu}{T_\nu} (\Gamma_{0,\nu} - 1)\phi \right], \tag{3.19}$$

with $\hat{G}_\nu$ and $\hat{Q}$ identically zero, and $\hat{M}$, $\hat{N}$ providing no contribution in the electrostatic limit. In the above we are once again considering the long time behaviour of a single wavenumber, and have suppressed the associated subscripts.

As in § 3.1 we decompose the functional operators $\hat{G}_\nu[\boldsymbol{p}; g_\nu, \phi]$ and $\hat{Q}[\boldsymbol{p}; g_\nu, \phi]$ into components that act on g_ν and ϕ separately, with all other operators in (3.3)–(3.6) set to zero. The derivation in § 3.1 is unchanged, with the exception that some terms may now be omitted. The resulting derivative of the growth rate in the electrostatic, collisionless regime is

$$\nabla_{\boldsymbol{p}} \gamma \langle g_\nu, \lambda_\nu \rangle_{z,v,\nu} = - \left[\langle \partial_{\boldsymbol{p}} \hat{L}_\nu, \lambda_\nu \rangle_{z,v,\nu} + \langle \partial_{\boldsymbol{p}} \hat{Q}, \xi \rangle_z \right] \Big|_{\boldsymbol{p}_0}, \tag{3.20}$$

where $\hat{L}_\nu = \hat{G}_\nu - \gamma g_\nu$ is given by equation (3.18), and the adjoint equations are

$$\gamma^* \lambda_\nu^\leftrightarrow + v_\parallel \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial \lambda_\nu^\leftrightarrow}{\partial z} - \frac{\mu_\nu}{m_s} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial v_\parallel} - \mathrm{i}\omega_{d,\nu} \lambda_\nu^\leftrightarrow + Z_\nu e \mathrm{J}_{0,\nu} \xi = 0, \tag{3.21}$$

$$\xi + \frac{1}{\eta} \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}^2 v \left[\mathrm{i}\omega_{*,\nu} + \frac{Z_\nu e}{T_\nu} \gamma^* \right] \mathrm{J}_{0,\nu} F_{0,\nu} \lambda_\nu^\leftrightarrow = 0, \tag{3.22}$$

with $\lambda_\nu^\leftrightarrow(v_\parallel) = \lambda_\nu(-v_\parallel)$, and $\eta = \sum_\nu (Z_\nu e)^2 n_\nu / T_\nu$ as before.

4. Normalisations and magnetic geometry

To evaluate (3.20), we need to solve for the set of variables $\{\gamma, g_\nu, \phi, \lambda_\nu^\leftrightarrow, \xi\}$, evaluated at the unperturbed geometric values, $\boldsymbol{p}_0$. We do this by implementing and combining the adjoint system within the local δf-gyrokinetic code `stella` (Barnes, Parra & Landreman 2019). In this section we write the gyrokinetic equations in normalised coordinates along with the corresponding normalised equations for the adjoint variables. We then introduce a specific choice for $\boldsymbol{p}$, the Miller parametrisation used to specify local magnetic equilibria in tokamaks (Miller *et al.* 1998), and detail how the adjoint method can be applied in this case.

4.1. *Normalisation*

4.1.1. *Gyrokinetic normalisations*

Here, we normalise the full electromagnetic gyrokinetic system and give the reduced electrostatic limit at the end of the section. We choose our reference quantities to be the same as in `stella` to aid in numerical implementation, and denote these with a subscript 'r'. A reference length scale that characterises the parallel lengths in the simulation is introduced as $L = a$; for the Miller formalism `stella` takes a to be the half diameter of the plasma volume (minor radius), at the height of the magnetic axis. The perpendicular reference length scale is taken as a reference gyroradius $\rho_r \doteq v_{\mathrm{th},r}/\Omega_r$, with $v_{\mathrm{th},r} = \sqrt{2T_r/m_r}$ the reference velocity, and $\Omega_r = eB_r/m_r c$, with T_r, n_r, B_r and m_r being user specified quantities. The normalised variables are denoted with a tilde, and are provided in table 1.

Multiplying the gyrokinetic equation, rewritten in normalised coordinates, by the factor $(a^2/\rho_r v_{\mathrm{th},r}) \exp(-v^2/v_{\mathrm{th},\nu}^2)/F_{0,\nu}$, we obtain the normalised low-flow, electromagnetic

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

Normalised Parameters

Parameter	Normalisation
$\tilde{t}$	$t\,a/v_{\mathrm{th},r}$
$\tilde{T}_\nu$	T_ν/T_r
$\tilde{B}$	B/B_r
$\tilde{m}_\nu$	m_ν/m_r
$\tilde{n}_\nu$	n_ν/n_r
$\tilde{\boldsymbol{\nabla}}$	$a\boldsymbol{\nabla}$
$\tilde{v}_\parallel$	$v_\parallel/v_{\mathrm{th},\nu}$
$\tilde{\mu}_\nu$	$\mu_\nu B_r/m_\nu v_{\mathrm{th},\nu}^2$
$\tilde{v}_{\mathrm{th},\nu}$	$v_{\mathrm{th},\nu}/v_{\mathrm{th},r}$
$\tilde{g}_{\boldsymbol{k},\nu}$	$g_{\boldsymbol{k},\nu}\left(\exp\left(-\tilde{v}_\nu^2\right)/F_{0,\nu}\right)(a/\rho_r)$
$\tilde{\phi}_{\boldsymbol{k}}$	$\phi_{\boldsymbol{k}}(e/T_r)(a/\rho_r)$
$\tilde{A}_{\parallel,\boldsymbol{k}}$	$(a/B_r\rho_r^2)A_{\parallel,\boldsymbol{k}}$
$\tilde{B}_{\parallel,\boldsymbol{k}}$	$(a/B_r\rho_r)\delta B_{\parallel,\boldsymbol{k}}$
$\tilde{a}_{\boldsymbol{k},\nu}$	$\tilde{k}_\perp\tilde{v}_\perp/\tilde{\Omega}_\nu$

Normalised Variables

$$\langle\tilde{\chi}\rangle_{\boldsymbol{k},\nu} = \mathrm{J}_{0,\boldsymbol{k},\nu}\tilde{\phi}_{\boldsymbol{k}} - 2\tilde{v}_{\mathrm{th},\nu}\tilde{v}_\parallel\mathrm{J}_{0,\boldsymbol{k},\nu}\tilde{A}_{\parallel,\boldsymbol{k}} + 4\tilde{\mu}_\nu(\tilde{T}_\nu/Z_\nu)(\mathrm{J}_{1,\boldsymbol{k},\nu}/\tilde{a}_{\boldsymbol{k},\nu})\delta\tilde{B}_{\parallel,\boldsymbol{k}}$$
$$\tilde{\omega}_{d,\boldsymbol{k},\nu} = (\tilde{T}_\nu\rho_r/Z_\nu\tilde{B})(\tilde{v}_\parallel^2\boldsymbol{v}_\kappa + \tilde{\mu}_\nu\boldsymbol{v}_{\nabla B})\cdot(k_x\tilde{\boldsymbol{\nabla}}x + k_y\tilde{\boldsymbol{\nabla}}y)$$
$$\tilde{\omega}_{*,\boldsymbol{k},\nu} = (k_y\rho_r/2)aB_r(\mathrm{d}y/\mathrm{d}\alpha)(\mathrm{d}\ln F_{0,\nu}/\mathrm{d}\psi)$$

TABLE 1. List of normalised parameters and variables.

gyrokinetic equation taken in the long time limit

$$\begin{aligned}\hat{G}_{\boldsymbol{k},\nu} = {}& \tilde{\gamma}\tilde{g}_{\boldsymbol{k},\nu} + \tilde{v}_{\mathrm{th},\nu}\tilde{v}_\parallel\,\hat{\boldsymbol{b}}\cdot\tilde{\boldsymbol{\nabla}}\tilde{z}\left[\frac{\partial\tilde{g}_{\boldsymbol{k},\nu}}{\partial\tilde{z}} + \frac{Z_\nu}{\tilde{T}_\nu}\frac{\partial\langle\tilde{\chi}\rangle_{\boldsymbol{k},\nu}}{\partial\tilde{z}}\mathrm{e}^{-\tilde{v}_\nu^2}\right]\\ & + \mathrm{i}\tilde{\omega}_{*,\boldsymbol{k},\nu}\mathrm{e}^{-\tilde{v}_\nu^2}\langle\tilde{\chi}\rangle_{\boldsymbol{k},\nu} + \mathrm{i}\tilde{\omega}_{d,\boldsymbol{k},\nu}\left[\tilde{g}_{\boldsymbol{k},\nu} + \frac{Z_\nu}{\tilde{T}_\nu}\langle\tilde{\chi}\rangle_{\boldsymbol{k},\nu}\mathrm{e}^{-\tilde{v}_\nu^2}\right]\\ & - \tilde{v}_{\mathrm{th},\nu}\tilde{\mu}_\nu\hat{\boldsymbol{b}}\cdot\tilde{\boldsymbol{\nabla}}\tilde{B}_0\frac{\partial\tilde{g}_{\boldsymbol{k},\nu}}{\partial\tilde{v}_\parallel} + 2\frac{Z_\nu}{\tilde{m}_\nu}\tilde{\mu}_\nu\hat{\boldsymbol{b}}\cdot\tilde{B}_0\mathrm{e}^{-\tilde{v}_\nu^2}\mathrm{J}_{0,\boldsymbol{k},\nu}\tilde{A}_{\parallel,\boldsymbol{k}} - \hat{C}_{\boldsymbol{k},\nu}[\tilde{g}_{\boldsymbol{k},\nu}].\end{aligned} \tag{4.1}$$

The corresponding transformed, normalised field equations are given by

$$\hat{Q}_{\boldsymbol{k}} = \sum_\nu Z_\nu\tilde{n}_\nu\left\{\frac{2\tilde{B}_0}{\sqrt{\pi}}\int\mathrm{d}^2\tilde{v}\mathrm{J}_{0,\boldsymbol{k},\nu}\tilde{g}_{\boldsymbol{k},\nu,0} + \frac{Z_\nu}{T_\nu}\left(\Gamma_{0,\boldsymbol{k},\nu} - 1\right)\tilde{\phi}_{\boldsymbol{k},0} + \frac{1}{\tilde{B}_0}\Gamma_{1,\boldsymbol{k},\nu}\delta\tilde{B}_{\parallel,\boldsymbol{k},0}\right\}, \tag{4.2}$$

$$\begin{aligned}\hat{M}_{\boldsymbol{k}} = {}& -\frac{\beta_r}{(k_\perp\rho_r)^2}\sum_\nu Z_\nu\tilde{n}_\nu\tilde{v}_{\mathrm{th},\nu}\frac{2\tilde{B}}{\sqrt{\pi}}\int\mathrm{d}^2\tilde{v}\tilde{v}_\parallel\mathrm{J}_{0,\boldsymbol{k},\nu}\tilde{g}_{\boldsymbol{k},\nu,0}\\ & + \left[1 + \frac{\beta_r}{(k_\perp\rho_r)^2}\sum_\nu\frac{Z_\nu\tilde{n}_\nu}{\tilde{m}_\nu}\Gamma_{0,\boldsymbol{k},\nu}\right]A_{\parallel,\boldsymbol{k},0},\end{aligned} \tag{4.3}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

$$\hat{N}_{\boldsymbol{k}} = 2\beta_r \sum_\nu \tilde{n}_\nu \tilde{T}_\nu \frac{2\tilde{B}_0}{\sqrt{\pi}} \int \mathrm{d}^2\tilde{v}\tilde{\mu}_\nu \frac{\mathrm{J}_{0,\boldsymbol{k},\nu}}{\tilde{a}_{\boldsymbol{k},\nu}} \tilde{g}_{\boldsymbol{k},\nu,0} + \left[\frac{\beta_r}{2\tilde{B}_0} \sum_\nu Z_\nu \tilde{n}_\nu \Gamma_{1,\boldsymbol{k},\nu}\right] \tilde{\phi}_{\boldsymbol{k},0}$$
$$+ \left[1 + \frac{\beta_r}{2\tilde{B}_0} \sum_\nu Z_\nu \tilde{n}_\nu \tilde{T}_\nu \Gamma_{2,\boldsymbol{k},\nu}\right] \delta\tilde{B}_{\parallel,\boldsymbol{k},0}, \tag{4.4}$$

with the reference plasma beta, $\beta_r = 8\pi n_r T_r / B_r^2$.

4.1.2. *Adjoint normalisation*

We choose the normalisation of the adjoint variables in such a way that our optimisation Lagrangian is dimensionless. In general, this is achieved by enforcing that the dimension of the adjoint variables in § 3 satisfy

$$\left.\begin{array}{ll} [\lambda_\nu] = [\hat{G}_\nu]^{-1} & [\xi] = [\hat{Q}]^{-1}, \\ [\zeta] = [\hat{M}]^{-1} & [\sigma] = [\hat{N}]^{-1}, \end{array}\right\} \tag{4.5}$$

with $[A]$ denoting the dimensionality of A. The normalised adjoint variables should then satisfy

$$\left.\begin{array}{ll} \tilde{\lambda}_\nu = \dfrac{\lambda_\nu}{[\lambda_\nu]} & \tilde{\xi} = \dfrac{\xi}{[\xi]}, \\ \tilde{\zeta} = \dfrac{\zeta}{[\zeta]} & \tilde{\sigma} = \dfrac{\sigma}{[\sigma]}. \end{array}\right\} \tag{4.6}$$

Analysis of (2.27), and (2.28)–(2.30) gives the normalisations consistent with those chosen for the gyrokinetic variables

$$\left.\begin{array}{ll} \tilde{\lambda}_\nu = \lambda_\nu \dfrac{F_{0,\nu}}{\mathrm{e}^{-\tilde{v}_\nu^2}} \dfrac{\rho_r v_{\mathrm{th},r}}{a^2} & \tilde{\xi} = \xi \dfrac{n_r e \rho_r}{a}, \\ \tilde{\zeta} = \zeta \dfrac{B_r \rho_r^2}{a} & \tilde{\sigma} = \sigma \dfrac{B_r \rho_r}{a}. \end{array}\right\} \tag{4.7}$$

We then multiply the adjoint equation (3.13), written in terms of normalised coordinates, by a factor of $(\rho_r/a)F_{0,\nu}/\mathrm{e}^{-\tilde{v}_\nu^2}$ to obtain the electromagnetic, collisional adjoint equations in normalised units

$$\frac{\partial\tilde{\lambda}_\nu^\leftrightarrow}{\partial\tilde{t}} + \tilde{\gamma}^*\tilde{\lambda}_\nu^\leftrightarrow + \tilde{v}_{\mathrm{th},\nu}\tilde{v}_\parallel \hat{\boldsymbol{b}}\cdot\tilde{\nabla}\tilde{z}\frac{\partial\tilde{\lambda}_\nu^\leftrightarrow}{\partial\tilde{z}} - \tilde{v}_{\mathrm{th},\nu}\tilde{\mu}_\nu \hat{\boldsymbol{b}}\cdot\tilde{\nabla}\tilde{z}\frac{\partial\tilde{B}_0}{\partial\tilde{z}}\frac{\partial\tilde{\lambda}_\nu^\leftrightarrow}{\partial\tilde{v}_\parallel} - \mathrm{i}\tilde{\omega}_{d,\nu}\tilde{\lambda}_\nu^\leftrightarrow$$
$$+ Z_\nu\tilde{n}_\nu \mathrm{J}_{0,\nu}\tilde{\xi} - \frac{\beta_r}{(k_\perp\rho_r)^2} Z_\nu\tilde{n}_\nu\tilde{v}_{\mathrm{th},\nu}\mathrm{J}_{0,\nu}\tilde{v}_\parallel\tilde{\zeta} + 2\beta_r\tilde{T}_\nu\tilde{\mu}_\nu\frac{\mathrm{J}_{1,\nu}}{\tilde{a}_\nu}\tilde{\sigma} - \hat{C}_\nu[\lambda_\nu] = 0, \tag{4.8}$$

$$\tilde{\xi} + \frac{1}{\tilde{\eta}}\sum_\nu \frac{2\tilde{B}}{\sqrt{\pi}} \int \mathrm{d}^2\tilde{v}\left[\mathrm{i}\tilde{\omega}_{*,\nu} + \frac{Z_\nu}{\tilde{T}_\nu}\tilde{\gamma}^*\right] \mathrm{J}_{0,\nu}\mathrm{e}^{-\tilde{v}_\nu^2}\tilde{\lambda}_\nu^\leftrightarrow = 0, \tag{4.9}$$

$$\tilde{\zeta} - \sum_\nu \frac{2\tilde{B}}{\sqrt{\pi}} \int \mathrm{d}^2\tilde{v}(2\tilde{v}_{\mathrm{th},\nu}\tilde{v}_\parallel)\left[\mathrm{i}\tilde{\omega}_{*,\nu} + \frac{Z_\nu}{\tilde{T}_\nu}\tilde{\gamma}^*\right] \mathrm{J}_{0,\nu}\mathrm{e}^{-\tilde{v}_\nu^2}\tilde{\lambda}_\nu^\leftrightarrow = 0, \tag{4.10}$$

$$\tilde{\sigma} - \sum_\nu \frac{2\tilde{B}}{\sqrt{\pi}} \int \mathrm{d}^2\tilde{v}\left(4\tilde{\mu}_\nu\frac{\tilde{T}_\nu}{Z_\nu}\frac{\mathrm{J}_{1,\nu}}{\tilde{a}_\nu}\right)\left[\mathrm{i}\tilde{\omega}_{*,\nu} + \frac{Z_\nu}{\tilde{T}_\nu}\tilde{\gamma}^*\right] \mathrm{e}^{-\tilde{v}_\nu^2}\tilde{\lambda}_\nu^\leftrightarrow = 0, \tag{4.11}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

with $\tilde{\lambda}_\nu^{\leftrightarrow} = \tilde{\lambda}_\nu(\tilde{z}, -\tilde{v}_\parallel, \tilde{\mu}_\nu)$, and $\tilde{\eta} = \sum_\nu Z_\nu^2 \tilde{n}_\nu / \tilde{T}_\nu$. This change of variables has been made such that the adjoint equations more closely resemble the gyrokinetic equations, with the boundary conditions on $\tilde{\lambda}_\nu^{\leftrightarrow}$ mimicking those on $\tilde{g}_{\boldsymbol{k},\nu,0}$. This is convenient as the numerical machinery from existing gyrokinetic codes can be re-used. An artificial time dependence has also been introduced in (4.8) in order to make computation easier, and we solve for the steady-state solution for $\tilde{\lambda}_\nu$, when $\partial_{\tilde{t}} \tilde{\lambda}_\nu = 0$.[11]

As before, we can use $\hat{G}_\nu[\boldsymbol{p}; \tilde{g}_\nu, \tilde{\phi}, \tilde{A}_\parallel, \delta\tilde{B}_\parallel] = \tilde{\gamma}\tilde{g}_\nu + \hat{L}_\nu[\boldsymbol{p}; \tilde{g}_\nu, \tilde{\phi}, \tilde{A}_\parallel, \delta\tilde{B}_\parallel]$, with the added constraint of $\nabla_{\boldsymbol{p}} \mathcal{L}|_{\boldsymbol{p}_0} = 0$ to rearrange the above and obtain

$$\nabla_{\boldsymbol{p}} \tilde{\gamma} \langle \tilde{g}_\nu, \tilde{\lambda}_\nu \rangle_{\tilde{z},\tilde{v},\nu} = -\left[\langle \partial_{\boldsymbol{p}} \hat{L}_\nu, \tilde{\lambda}_\nu \rangle_{\tilde{z},\tilde{v},\nu} + \langle \partial_{\boldsymbol{p}} \hat{Q}, \tilde{\xi} \rangle_{\tilde{z}} + \langle \partial_{\boldsymbol{p}} \hat{M}, \tilde{\zeta} \rangle_{\tilde{z}} + \langle \partial_{\boldsymbol{p}} \hat{N}, \tilde{\sigma} \rangle_{\tilde{z}} \right]\Big|_{\boldsymbol{p}_0}, \quad (4.12)$$

with closure equations provided by (4.8)–(4.11).

4.1.3. *Electrostatic collisionless limit*

Finally, in the electrostatic, collisionless regime, the system of equations to solve in the normalised `stella` coordinates is given by

$$\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}} + \tilde{\gamma}^* \tilde{\lambda}_\nu^{\leftrightarrow} + \tilde{v}_{\mathrm{th},\nu} \tilde{v}_\parallel \hat{\boldsymbol{b}} \cdot \tilde{\nabla} \tilde{z} \frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{z}} - \tilde{v}_{\mathrm{th},\nu} \tilde{\mu}_\nu \hat{\boldsymbol{b}} \cdot \tilde{\nabla} \tilde{z} \frac{\partial \tilde{B}_0}{\partial \tilde{z}} \frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial v_\parallel} - \mathrm{i}\tilde{\omega}_{d,\nu} \tilde{\lambda}_\nu^{\leftrightarrow}$$
$$+ Z_\nu \tilde{n}_\nu \mathrm{J}_{0,\nu} \mathrm{e}^{-\tilde{v}_\nu^2} \xi = 0, \quad (4.13)$$

$$\xi + \frac{1}{\tilde{\eta}} \sum_\nu \frac{2\tilde{B}}{\sqrt{\pi}} \int \mathrm{d}^2 \hat{v} \left[\mathrm{i}\tilde{\omega}_{*,\nu} + \frac{Z_\nu}{\tilde{T}_\nu} \tilde{\gamma}^* \right] \mathrm{J}_{0,\nu} \mathrm{e}^{-\tilde{v}_\nu^2} \tilde{\lambda}_\nu^{\leftrightarrow} = 0, \quad (4.14)$$

and

$$\nabla_{\boldsymbol{p}} \tilde{\gamma} \langle \tilde{g}_\nu, \tilde{\lambda}_\nu \rangle_{\tilde{z},\tilde{v},\nu} = -\left[\langle \partial_{\boldsymbol{p}} \hat{L}_\nu, \tilde{\lambda}_\nu \rangle_{\tilde{z},\tilde{v},\nu} + \langle \partial_{\boldsymbol{p}} \hat{Q}, \tilde{\xi} \rangle_{\tilde{z}} \right]\Big|_{\boldsymbol{p}_0}. \quad (4.15)$$

4.2. *Magnetic geometry*

The coefficients in equations (4.1)–(4.4), and thus the associated linear growth rates, are implicitly dependent on the magnetic geometry. For the remainder of the paper we will take $\boldsymbol{p}$ to be an appropriate set of parameters that specifies the local magnetic geometry. In particular, we will use the Miller formalism (Miller *et al.* 1998) to parameterise the magnetic field on the flux surface of interest. The Miller approach ensures that the Grad–Shafranov (Shafranov 1966) equation is always satisfied locally by using a set of independent parameters to define a single flux surface in an axisymmetric device. The model equations describing the shape of the flux surface with flux label r are

$$R(r, \theta) = R_0(r) + r \cos\{\theta + \sin(\theta)\delta(r)\}, \quad (4.16)$$

$$Z(r, \theta) = r\kappa(r) \sin(\theta). \quad (4.17)$$

Here, $R(r, \theta)$ and $Z(r, \theta)$ define the major radial and vertical locations for a given poloidal location, θ, which is related to the cylindrical angle, and δ and κ indicate the triangularity and elongation of the surface, respectively. Specifying the full set of Miller parameters provides all of the information required to compute the geometric coefficients in the gyrokinetic-adjoint system consistent with the local magnetohydrodynamic

[11] Here, the time derivative serves as an iterative scheme to converge to the correct steady-state solution for $\tilde{\lambda}_\nu^{\leftrightarrow}$.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

(MHD) equilibrium (though consistency with a global MHD equilibrium is not guaranteed).

The user specified input parameters used in the local version of `stella` are $\{r_{\psi_0}, R_{\psi_0}, \Delta_{\psi_0}, q_{\psi_0}, \hat{s}_{\psi_0}, \kappa_{\psi_0}, \kappa'_{\psi_0}, I_{\psi_0}, \delta_{\psi_0}, \delta'_{\psi_0}, \beta'_{\psi_0}\}$, corresponding to markers for the minor and major radii, horizontal Shafranov shift ($\Delta_{\psi_0} = R'_{\psi_0}$), safety factor, magnetic shear ($\hat{s} \doteq (r/q)q'$), elongation and its radial derivative, the external axial discharge current (which is used as a proxy for the reference magnetic field), triangularity and its radial derivative and the radial pressure derivative, respectively, with each being specified at the flux surface of interest. Further information about how the Miller geometry is treated in `stella` is given in Appendix E. We henceforth set $\boldsymbol{p} := \{r_{\psi_0}, R_{\psi_0}, \Delta_{\psi_0}, q_{\psi_0}, \hat{s}_{\psi_0}, \kappa_{\psi_0}, \kappa'_{\psi_0}, I_{\psi_0}, \delta_{\psi_0}, \delta'_{\psi_0}, \beta'_{\psi_0}\}$.

5. Numerical implementation

The aim is to find the magnetic geometry that minimises the linear growth rate for the ITG instability, and maximises the linear critical temperature gradient across the device. This requires three distinct stages: first, the computation of $\nabla_{\boldsymbol{p}}\gamma$ at a fixed ion temperature gradient, T'_i; second, its subsequent use in an optimisation algorithm to find the $\boldsymbol{p}$ that minimises γ for this given T'_i; third, iteration of this procedure with variable T'_i to find the maximum temperature gradient for which $\gamma \leq 0$ for the range of $\boldsymbol{p}$ considered.

5.1. *Initial simulation*

Solving, at an initial set of $\boldsymbol{p}_0$, for $\tilde{\gamma}$, $\tilde{g}_v$ and $\tilde{\phi}$ is achieved by allowing `stella` to run for a sufficiently long time, such that the solution is dominated by a single normal mode. To determine the time at which this is satisfied, we employ a convergence test: the growth rate is calculated at each time step and if the value of this is constant in time (within a specified tolerance) then the system is deemed to be converged. There are two components of the convergence test. The first is to check that the growth rates calculated at adjacent time steps are within a given tolerance of each other. The second is to perform a windowed average to check that the growth rate remains consistent over a defined number of time steps. Two windowed averages are done; one over N_t time steps and one over $\mathbb{Z}(N_t/2)$ time steps. When these two window-averaged growth rates agree within a set tolerance, $\tilde{g}_v$ and $\tilde{\phi}$ are taken to have converged. The corresponding growth rate is then calculated from the windowed average.[12]

5.2. *Adjoint simulation*

As previously introduced, an artificial time dependence is added to the adjoint equations to facilitate computation. The solution is found in the steady-state limit, in which the time derivative appearing in the adjoint equations goes to zero.

The resulting adjoint equations are treated in a similar way to the treatment of the usual gyrokinetic system of equations in `stella`. The main aim is to ensure that the parallel streaming term may be treated separately from the rest of the dynamics through operator splitting. This is done by discretising in time, and splitting the time derivative into a series

[12] In cases where there exist two modes of similar growth rates, it may take a long run time to isolate the dominant mode. In these instances of ambiguity, the adjoint method as demonstrated here, may fail to provide the correct gradient of the linear growth rate in the desired parameter space at that point. However, it should be noted that as soon as you move away from this region of ambiguity in the growth rate, the adjoint method can continue to be applicable, and will provide the correct gradient in the parameter space. Hence, even if you take a step in the wrong direction within the space, the adjoint method should correct itself to then step in the direction of stability.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

of three steps

$$\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}} = \left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_1 + \left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_2 + \left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_3, \tag{5.1}$$

where

$$\left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_1 = -\tilde{\gamma}^* \tilde{\lambda}_\nu^{\leftrightarrow} + \mathrm{i}\tilde{\omega}_{d,\nu} \tilde{\lambda}_\nu^{\leftrightarrow} - Z_\nu \tilde{n}_\nu \tilde{\xi}, \tag{5.2a}$$

$$\left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_2 = \tilde{v}_{\mathrm{th},\nu}\, \tilde{\mu}_\nu\, \hat{\boldsymbol{b}} \cdot \tilde{\boldsymbol{\nabla}} \tilde{z} \frac{\partial \tilde{B}_0}{\partial \tilde{z}} \frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{v}_\parallel}, \tag{5.2b}$$

$$\left(\frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{t}}\right)_3 = -\tilde{v}_{\mathrm{th},\nu} \tilde{v}_\parallel\, \hat{\boldsymbol{b}} \cdot \tilde{\boldsymbol{\nabla}} \tilde{z} \frac{\partial \tilde{\lambda}_\nu^{\leftrightarrow}}{\partial \tilde{z}}. \tag{5.2c}$$

Analogous to `stella`, the terms in (5.2*a*) are treated explicitly using a strong-stability-preserving, third-order Runge–Kutta method (Gottlieb, Shu & Tadmor 2001). In combination with the operator splitting employed within `stella` the overall algorithm is second-order accurate in time step size, $\Delta\tilde{t}$.

The parallel streaming and mirror terms, given by (5.2*c*) and (5.2*b*), respectively, are treated separately, due to the presence of the prefactor $\tilde{v}_{\mathrm{th},\nu}$, which increases the relative amplitude of these terms when considering electron dynamics. As a result these terms have the potential to exert a stringent Courant–Friedrichs–Lewy condition on the simulation, and require a small time step to be taken in order to retain accuracy. Thus these terms are treated implicitly in time to relax this constraint.

The mirror term, (5.2*b*), is a simple advection equation of $\tilde{\lambda}_\nu^{\leftrightarrow}$ in $\tilde{v}_\parallel$, which is treated using a semi-Lagrange method, akin to the algorithm used to advect the distribution function in $\tilde{v}_\parallel$ within `stella`. The mirror coefficients are independent of both time and $\tilde{v}_\parallel$ and hence the exact characteristics of this equation are known. The interpolation in $\tilde{v}_\parallel$ is forth order accurate in $\tilde{v}_\parallel$ grid spacing, $\Delta\tilde{v}_\parallel$(Barnes *et al.* 2019).

The streaming term, (5.2*c*), is also an advection equation in $\tilde{z}$, which is treated using the Thompson algorithm for tri-diagonal solve. When fully centred in time the discretisation reduces to the Crank–Nicolson method, which is second-order accurate in time, and $\tilde{z}$ grid spacings, $\Delta\tilde{t}$ and $\Delta\tilde{z}$, respectively.

We are seeking a steady-state solution to the adjoint equations. The same convergence test is performed on $\tilde{\lambda}_\nu^{\leftrightarrow}$ as that performed on the distribution function in order to check that the complex growth rate has converged to zero within a given tolerance. When this is satisfied, the resulting $\tilde{\lambda}_\nu^{\leftrightarrow}$ is used to solve for $\tilde{\lambda}_\nu$. This is then stored for use in the remainder of the calculation.

5.3. *Optimisation loop*

Once the gradients $\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma$ are obtained we use them inside an optimisation loop to find the $\boldsymbol{p}$ that minimises γ. We employ the Levenberg–Marquardt (LM) algorithm, (Transtrum & Sethna 2012), to find the local minimum. The method adopts a steepest descent behaviour when the location in parameter space is considered to be far from the minimum, and progresses towards Gauss–Newton behaviour as one approaches the minimum. This is achieved by introducing a damping factor, Γ, which is updated with each iteration. The

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

algorithm iteratively solves the following:

$$[\boldsymbol{H} + \Gamma \mathrm{diag}(\boldsymbol{H})]\,\mathrm{d}\boldsymbol{p} = -\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma, \tag{5.3}$$

with $\boldsymbol{H} = \boldsymbol{\nabla}^2_{\boldsymbol{p}}\gamma \approx (\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma)^{\dagger}\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma$ the Hessian matrix. When Γ is large we have $\boldsymbol{p}_{\mathrm{new}} \approx \boldsymbol{p}_{\mathrm{old}} - \boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma$, with $\boldsymbol{\alpha} \doteq [\Gamma \mathrm{diag}(\boldsymbol{H})]^{-1}$, which mimics the gradient-descent algorithm. However, when Γ is small, (5.3) reduces to $\boldsymbol{p}_{\mathrm{new}} \approx \boldsymbol{p}_{\mathrm{old}} - \boldsymbol{H}^{\dagger}\cdot\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma$, which matches with the Gauss–Newton algorithm.

The LM formalism is derived using the Taylor expansion, and as such a trust region is included within the optimisation loop to ensure that the updated value of $\boldsymbol{p}$ is close enough to the previous, such that the Taylor approximation is valid within the limits for which the algorithm is applied. The trust region for $\boldsymbol{p}$ is defined via

$$\rho = \frac{0.5\,\mathrm{d}\boldsymbol{p}^{\dagger}\cdot\boldsymbol{H}\cdot\mathrm{d}\boldsymbol{p}}{\mathrm{d}\boldsymbol{p}\cdot\boldsymbol{\nabla}_{\boldsymbol{p}}\gamma} < \bar{\epsilon}, \tag{5.4}$$

where $\bar{\epsilon}$ is a chosen tolerance. If $\rho > \bar{\epsilon}$ for a given $\mathrm{d}\boldsymbol{p}$ the algorithm rejects the output p_{new} and increases the weight Γ in an attempt to improve the accuracy of the approximation. This helps ensure that the updated value of $\boldsymbol{p}$ is a reasonable one.

It is worth noting that the LM algorithm is designed to find local minima, so there is no guarantee that the minimum obtained is the global minimum of the system.

We also emphasise that the gradient-based optimisation algorithm is independent of the adjoint method that has been developed for gyrokinetic microstability. The optimisation loop may be itself optimised to efficiently search for regions of stability given a gradient input. Different algorithms, and indeed different parameter choices within each algorithm, will yield different efficiencies in finding stable solutions. An illustrative example of this is given later in figure 1, where the step size for the optimisation loop is varied to yield two distinct paths through the parameter space using the adjoint gradient. However, this is not the focus of this paper and we will not labour on enhancing the gradient-based optimisation loop.

6. Numerical results

6.1. *Initial benchmark*

The first numerical check we perform is to ensure that the values of $\mathrm{d}_{\boldsymbol{p}}\gamma_0$ obtained from our adjoint method agree with those obtained using a finite-difference approach. Following this, we perform a more extensive benchmark by conducting a parameter scan in the growth rate using `stella` for different values of the Miller parameters. We then choose an initial set of parameters and perform the adjoint-optimisation scheme described above, and overlay the results of this with the parameter scan to see how these compare, whilst checking the growth rates at each point considered by the adjoint-optimisation scheme against those obtained using a finite-difference approach. As a proof of principle we chose to vary two parameters: triangularity, δ, and elongation, κ, whilst holding the other Miller variables fixed. Given that the Miller parametrisation is local to a given flux surface, this variation is not necessarily consistent with a global solution to the Grad–Shafranov equation. The choice to vary these two parameters in isolation is driven by two considerations; first, it is only intended as a proof-of-principle check of the adjoint approach so simplicity is desirable, and second, previous research has shown that maximal shaping, with large elongation and triangularity, minimises the linear ITG instability, whereas parameters such as κ' and δ' have an order $a/R_0 \ll 1$ is the inverse aspect ratio, with a and R_0 the scales associated with the minor and major radii, respectively (Highcock

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

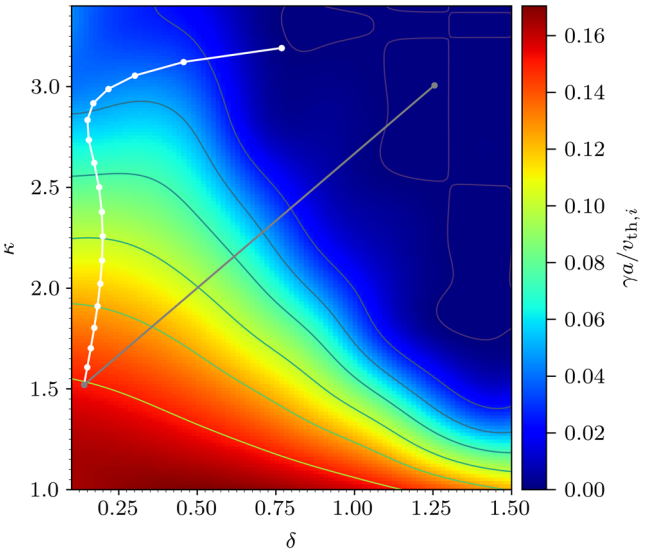


FIGURE 1. Two-dimensional parameter scan over elongation and triangularity, with the colour map indicating the amplitude of the linear growth rate. Here, $k_y = 0.68$, $k_x = 0.0$, $m_e/m_i = 2.7 \times 10^{-4}$, $T_i = T_e = 1$, $n_i = n_e = 1$, $a/L_{T_i} = a/L_{T_e} = 2.42$, $a/L_{n_i} = a/L_{n_e} = 0.81$, with a the minor radius of the last closed flux surface. The path taken by the optimisation algorithm is indicated in white, with the initial point $\kappa = 1.5$, and $\delta = 0.14$. A second path, drawn in grey, is shown indicating the adjoint optimisation with a different step size for the optimisation loop.

Miller Parameter	Initial Value
ρ	0.5
R_0	2.94
R_{geo}	2.94
Δ	−0.11
κ	1.52
κ'	0.10
q	2.02
$\hat{s}$	0.34
δ	0.14
δ'	0.29
β'	0.069

Equilibrium Miller parameter values used in the initial benchmark simulations.

TABLE 2. List of Miller Parameters.

et al. 2018). Table 2 lists the values of input equilibrium variables in the Miller geometry. These have been chosen to coincide with values used in Beeke *et al.* (2020) in order to verify the qualitative behaviour found.

Given these initial values of δ and κ, we determine the linear growth rates for a grid of perpendicular wavenumbers within $k_x < 2.0$, $k_y < 2.0$, which reveals that the most unstable mode is found at $\{k_y, k_x\} = \{0.68, 0.0\}$ for mass ratio $m_e/m_i = 2.7 \times 10^{-4}$,

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

normalised species temperatures and densities of $T_i = T_e = 1, n_i = n_e = 1$ and normalised species temperature and density gradients of $a/L_{T_i} = a/L_{T_e} = 2.42, a/L_{n_i} = a/L_{n_e} = 0.81$.

In figure 1 we show a scan in the linear growth in elongation and triangularity obtained by running `stella` with the Miller parameters specified in table 2, and the values of κ and δ adjusted accordingly. The contour colour indicates the magnitude of the growth rate, and the plot extends over a range of values that has been set by reasonable physical constraints on devices.

Figure 1 shows that increasing the elongation and triangularity of our flux surface reduces the linear growth rate, and that there exists a region of stability when the shaping is maximal, in agreement with previous work. The path taken using the adjoint method is indicated in white. At the chosen starting point, located in the region of instability, the gradient $\mathrm{d}_p\gamma$ is calculated and the value of $\boldsymbol{p} = \{\delta, \kappa\}$ is updated using the previously mentioned LM method. The final point is found to be locally stable as the growth rate here is negative. The algorithm then checks a nearby point to determine if a small region of stability exists and, once this has been verified, outputs this as the final $\boldsymbol{p}$ value. In this particular case of a two-dimensional parameter scan, a finite-difference approach could utilise the same optimisation loop, requiring only three simulations, plus one additional simulation performed at the next iteration point to verify the growth rate. The adjoint method must simulate the gyrokinetic equation once, and also solve the adjoint equations – which are of similar computational cost as the gyrokinetic equation – in order to achieve the same result. In this particular low-dimensionality case, the advantage of the adjoint approach is comparatively small over using a finite-difference approach. However, the computation time required for the traditional method scales linearly with the number of parameters, so that for an N-dimensional parameter space, $N + 1$ simulations are necessary at each iteration point, whilst the adjoint method is effectively independent of the number of parameters, and only the steady-state solutions to the gyrokinetic and adjoint equations are required. It can therefore be readily applied to a high-dimensional parameter space without any significant increase in computational cost.

A second path is plotted on figure 1 in grey. This path is taken using the same adjoint technique, but increasing the step size within the optimisation loop. When the step size is small, as with the white path, the LM algorithm more closely resembles a gradient-descent method, however, when the step size is larger, as with the grey path, the LM algorithm resembles Newton's method for gradient optimisation. The figure illustrates how the adjoint algorithm can be combined with an optimiser to quickly and efficiently converge to a stable region of parameter space.

6.2. *Increasing the critical temperature gradient*

We have seen that the adjoint method is a powerful technique for computing stable points in a large parameter space; figure 1 shows that it can efficiently be used inside an optimisation loop to find a minimum of the linear growth rate for a given temperature gradient. Once the adjoint-optimisation loop locates a region in the parameter space with negative or zero growth rate the local temperature gradient (or other plasma profile variable of interest) may be increased and the process repeated. Conversely, if a minimum positive (unstable) growth rate is found, the temperature gradient can be reduced to seek out the optimal shape that maximises the critical temperature gradient at which linear instability occurs to seek out microstability.

Figure 2 demonstrates an iterative use of the adjoint optimisation. Here, we have iterated the temperature gradient, and inside each temperature-gradient iteration the adjoint method is used to find a stable geometric configuration. However, this principle could be extended further by continuing to increase the temperature gradient until no

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

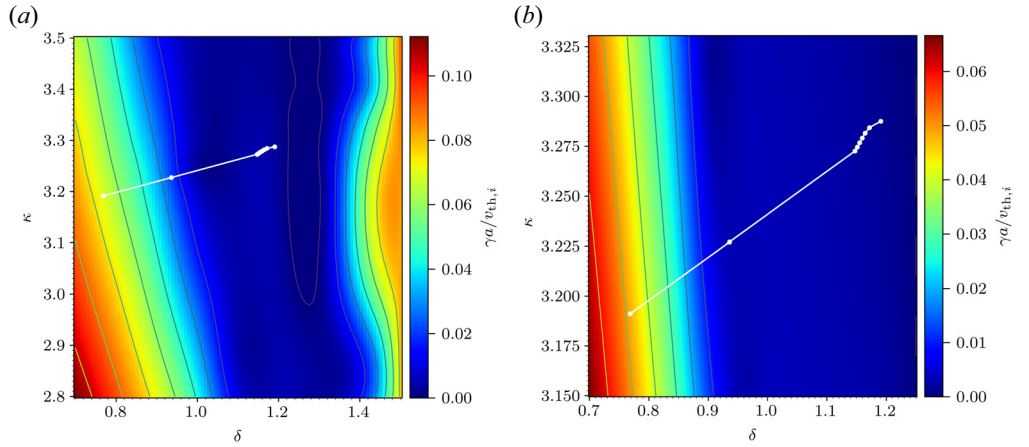


FIGURE 2. Plots showing a parameter scan in elongation and triangularity, with a temperature gradient of $a/L_{T_i} = 3.80$, increased from the $a/L_{T_i} = 2.42$ value used in figure 1. The geometry of the initial point, located in the unstable region, is provided by final point in figure 1 and is now unstable due to the increased temperature gradient. Here, $k_y = 0.68$, $k_x = 0.0$, $m_e/m_i = 2.7 \times 10^{-4}$, $T_i = T_e = 1$, $n_i = n_e = 1$, $a/L_{T_i} = a/L_{T_e} = 3.80$, $a/L_{n_i} = a/L_{n_e} = 0.81$. Note that the colour scales used in the figures above are different than that used in figure 1. The right-hand side plot is a zoomed in figure of the left.

region of stability is available, indicating a limiting temperature gradient that can be achieved through geometric considerations alone. Figure 2 is a demonstration that the adjoint method may be used to increase the temperature gradient, whilst retaining stability using geometry. Though we have opted to consider only two parameters in the above as a demonstration and for clarity have focused on ITG, it is possible to employ the adjoint method to optimise over a large number of geometric parameters simultaneously. Such an exploration would be expensive using traditional finite-difference methods.

6.3. *Negative triangularity*

To form a final example here, we note there has been previous evidence that negative triangularity can offer improvements for microstability, Pochelon *et al.* (1999), so we repeated the benchmark scan using negative triangularity. All values of equilibrium Miller parameters are the same as in table 2, except we now set the initial value of triangularity to $\delta = -0.14$.

The path taken using the adjoint-optimisation loop is again shown in figure 3 by the white path, starting in the dark red region at $\{\delta, \kappa\} = \{-0.14, 1.52\}$.

This highlights a key feature of the gradient-based optimisation method: the solution is not unique, and the output can depend on the starting region within parameter space. See figure 4 for an illustrative example of this.

Finally, we consider increasing the temperature gradient for the negative triangularity case, and we repeat the procedure to find a stable region of parameter space. We again take the input parameters $\{\delta, \kappa\} = \{-0.9965, 2.4488\}$ to be the outputs from the previously optimised case at a temperature gradient of $a/L_{T_i} = 2.42$. Then we use the adjoint algorithm coupled with the gradient optimiser to look for a stable region of parameter space at an increased temperature gradient of $a/L_{T_i} = 3.8$. The results of this are shown in figure 5.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

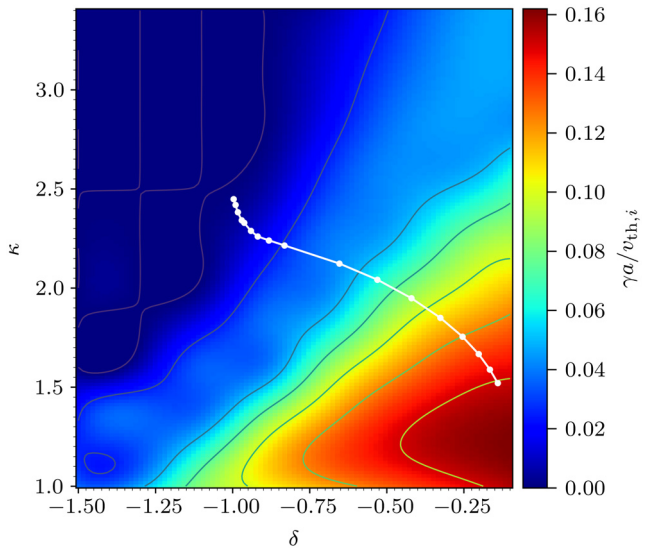


FIGURE 3. Growth rate contours for a parameter scan with negative triangularity for $k_y = 0.68$, $k_x = 0.0$ and equilibrium parameters $m_e/m_i = 2.7 \times 10^{-4}$, $T_i = T_e = 1$, $n_i = n_e = 1$, $a/L_{T_i} = a/L_{T_e} = 3.80$, $a/L_{n_i} = a/L_{n_e} = 0.81$. The white line indicates the path taken by the optimisation algorithm. The initial values of $\{\delta, \kappa\}$ are taken to be $\{-0.14, 1.52\}$.

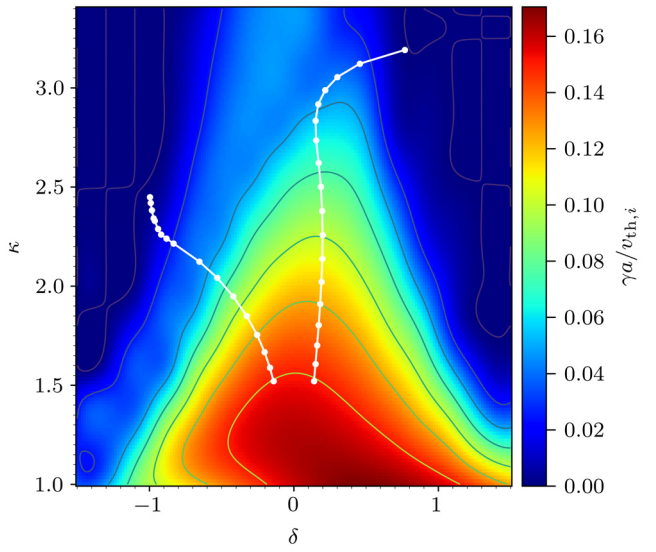


FIGURE 4. Growth rate contours for a parameter scan with both positive and negative triangularity for $k_y = 0.68$, $k_x = 0.0$ and equilibrium parameters $m_e/m_i = 2.7 \times 10^{-4}$, $T_i = T_e = 1$, $n_i = n_e = 1$, $a/L_{T_i} = a/L_{T_e} = 3.80$, $a/L_{n_i} = a/L_{n_e} = 0.81$. The white line indicates the two paths taken by the optimisation algorithm starting in different regions in parameter space. The initial values of $\{\delta, \kappa\}$ are taken to be $\{0.14, 1.52\}$ and $\{-0.14, 1.52\}$.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

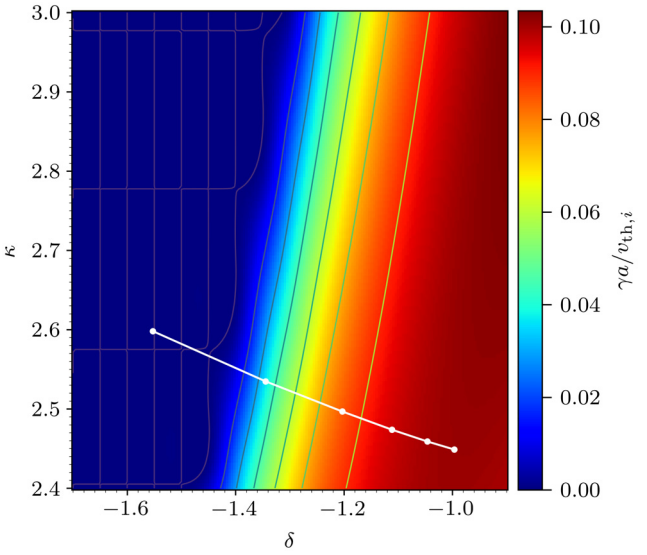


FIGURE 5. Growth rate contours for a parameter scan with negative triangularity at a temperature gradient of $a/L_{T_i} = 3.80$, increased from $a/L_{T_i} = 2.42$. The geometry of the initial point is taken as the final point in figure 3, and is now unstable due to the increased temperature gradient. The scan is performed at the same parameter values as those in figure 3 – $k_y = 0.68$, $k_x = 0.0$ $m_e/m_i = 2.7 \times 10^{-4}$, $T_i = T_e = 1$, $n_i = n_e = 1$, $a/L_{T_i} = a/L_{T_e} = 3.80$, $a/L_{n_i} = a/L_{n_e} = 0.81$. The white line indicates the path taken by the optimisation algorithm. The initial values of $\{\delta, \kappa\}$ are taken to be $\{-0.9965, 2.4488\}$.

7. Numerical efficiency improvement

Recall that for a conventional finite-difference approach method for a parameter vector, $\boldsymbol{p}$, of size $\mathcal{N}$ we require $\mathcal{N} + 1$ simulations to be carried out until convergence at each of the points considered (these are shown by the white dots in the figures above) in order to compute one gradient in our parameter space. However, when using the adjoint method the same gradient may be computed for the cost of roughly two simulations – one gyrokinetic, and one adjoint – which is of similar computational cost as a gyrokinetic simulation. This approach is essentially independent of the number of parameters used, with the only additional cost incurred being that associated with calculating the partial derivatives that appear. However, these are extremely inexpensive for computers and can be calculated using one processor.

To illustrate this example we will show the progressive computational improvement of using the adjoint method, as compared with finite differences, and show how this scales favourably with increasing $\mathcal{N}$. In order for the comparisons to be fair, we shall run all simulations to a standard time of $100a/v_{\mathrm{th},i}$.

7.1. *Numerical demonstration of improved efficiency*

For the case where $\mathcal{N} = 2$ we are optimising with respect to two parameters, $\{\delta, \kappa\}$. We find that to calculate a gradient at each point in our parameter space requires a total of ~ 4.752 CPU (central processing unit) hours on 4 nodes, each with 48 cores on the supercomputer Marconi. The same calculation as performed using the adjoint method requires a total of $\sim$3.168 CPU hours. Though this is only a modest improvement,

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

we shall show that as $\mathcal{N}$ increases the advantage of using the adjoint method, over a finite-difference scheme, becomes increasingly apparent.

Increasing the number of parameters to $\mathcal{N} = 4$, optimising over the miller parameters $\{\Delta, \kappa, q, \delta\}$, we find that the finite-difference approach requires ∼7.920 CPU hours to compute a gradient at each point in parameter space. However, when computing the same gradient using the adjoint method the CPU time, to the precision of the CPU clock, is ∼3.168 CPU hours.

If we increase the number of parameters further to $\mathcal{N} = 7$, optimising over the miller parameters $\{R_0, R_{\text{geo}}, \Delta, \kappa, q, \hat{s}, \delta\}$ we find that using the finite-difference approach requires $\sim$ 12.672 CPU hours to compute each gradient at a point in parameter space. When using the adjoint method to compute the same gradient, to the precision of the CPU clock, is still $\sim$ 3.168 CPU hours.

Hence, we conclude that the numerical speed up is significant with increasing $\mathcal{N}$. This allows for the potential of including multiple harmonics within our shaping optimisation for very little additional computing cost.

For the example including seven parameters, we iterate the process of stepping through parameter space to find a point of stability. We take our set of initial parameters to be those given in table 2, and we simulate using a temperature and density gradient of $a/L_{T_i} = a/L_{T_e} = 2.42$ and $a/L_{n_i} = a/L_{n_e} = 0.81$, respectively. We perturb the magnetic geometry by varying $\{R_0, R_{\text{geo}}, \Delta, \kappa, q, \hat{s}, \delta\}$. Following iterations of the coupled adjoint-LM system we find a configuration that is stable when $\{R_0, I, \Delta, \kappa, q, \hat{s}, \delta\} = \{2.979, 2.846, 0.562, 1.656, 2.085, 0.167, 0.225\}$. A cross-section of the initial, unstable configuration and the final, stable configuration, with their corresponding neighbouring flux surfaces, are shown in figure 6.

8. Conclusion and discussion

We have derived an adjoint method tailored for local, linear gyrokinetics and demonstrated its numerical integration within the δf-gyrokinetic code `stella`. As a proof-of-principle case we have demonstrated the effectiveness of our adjoint method, as applied to a gyrokinetic system, by finding the geometric configuration of the magnetic field that is stable to microinstabilities. The illustrative example given has focused on increasing the temperature gradient, whilst preserving microstability; ITG instabilities are often prevalent in fusion devices due to the existence of large temperature gradients, and thus it is conceivable that geometric considerations could help mitigate their growth and improving overall efficiency.

As that the computational cost of the adjoint method remains independent of the number of parameters, its applicability to high-dimensional parameter spaces is readily apparent. The advantages become more pronounced with an expanding number of parameters, as the adjoint method outperforms traditional techniques for calculating gradients, where the computation cost scales with parameter count. This becomes especially beneficial when examining devices like stellarators, which have a large number of geometric parameters that can influence the microinstability of the confined plasma.

It is important to stress that although we have demonstrated a specific example focused on increasing the ITG, this approach can be readily extended to optimise the density gradient or other plasma properties, by adapting the overarching LM optimisation loop, without necessitating alterations to the adjoint calculation itself. Such adaptability enables the application of `stella` and other local, δf-gyrokinetic codes to explore the impact of shaping on various types of microinstabilities and assess how geometry can be instrumental in mitigating their growth.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

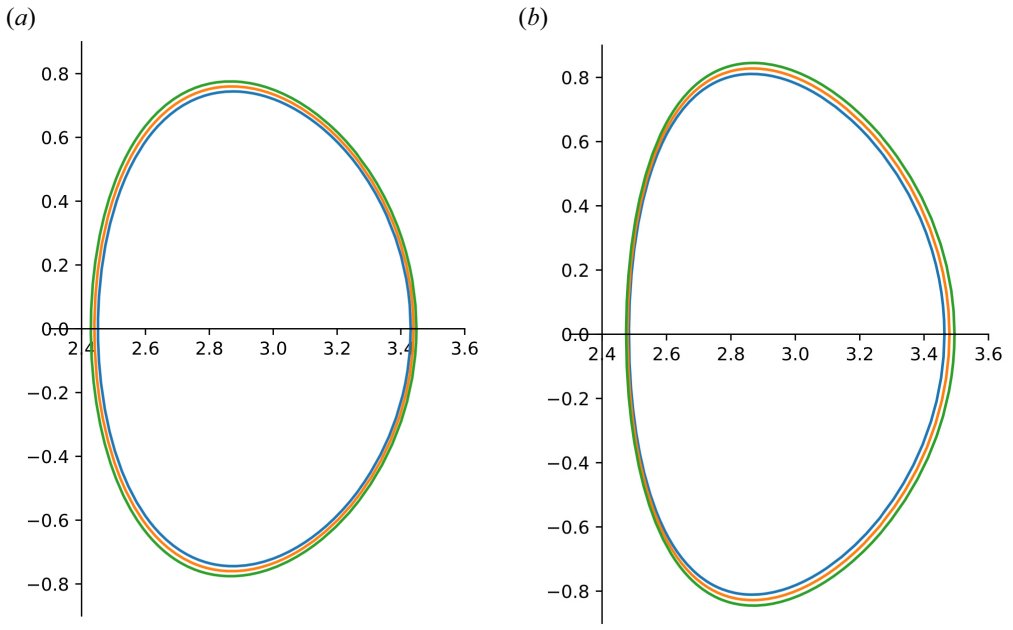


FIGURE 6. Plots of the flux surfaces in the poloidal cross-section. The orange surface is the flux surface at $\rho = 0.5$, and the blue and green surfaces are the two adjacent flux surfaces. The image (*a*) is the initial unstable configuration, before optimising. The image (*b*) is the stable, optimised configuration found using the adjoint-LM system.

A crucial point to note is that, while the numerical examples presented above have focused on the electrostatic, collisionless regime for optimisation with respect to the Miller geometry, equations (3.13)–(3.17) maintain generality. They can be applied to an electromagnetic system, including collisions, and can be optimised using any appropriate set $\{p_i\}$. It should also be stressed that the adjoint method is agnostic to the underlying mechanism driving the linear instability, and provides a gradient in the desired parameter space independently of the drive. As a result the method presented here is applicable to a wide range of microinstabilities.

It should also be emphasised that the adjoint method and the gradient-based optimisation presented are independent of one another. We have considered the simplified case of optimising the linear growth rate of a single wavenumber, $\boldsymbol{k}$. For more practical purposes one may wish to apply the adjoint method to a range of wavenumbers, and design the external optimisation routine to take an appropriately weighted average of these as the most unstable wavenumber may change as one moves within the parameter space.

Future extensions of this work may include: implementing the electromagnetic equations derived into gyrokinetic codes, applying this work to three-dimensional stellarator geometries, as well as potentially adapting the analytical framework to include global effects (radial or poloidal), whereby different k_x or k_y modes are coupled, and considering the time-dependent adjoint case to include effects such as $\boldsymbol{E} \times \boldsymbol{B}$ flow shear.

Acknowledgements

We are indebted to T. Adkins, W. Dorland, R. Ewart, M. Hardman and F. Parra for their helpful discussions on various topics throughout this project.

Editor P. Helander thanks the referees for their advice in evaluating this article.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

Funding

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 – EUROfusion) and from the EPSRC (grant number EP/W006839/1). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. Part of the simulations presented were performed on the CINECA Marconi supercomputer within the framework of the FUA37_STELTURB project.

Declaration of interests

The authors report no conflict of interest.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

Appendix A. Decomposition of operators

Here, we give the definitions of the operators introduced in § 3 in (3.3)–(3.4)

$$
\begin{aligned}
\hat{G}_{g,\nu}[\boldsymbol{p};\, g_\nu] &= \gamma g_\nu + v_\parallel \hat{\boldsymbol{b}}\cdot\nabla z \frac{\partial g_\nu}{\partial z} - \frac{\mu}{m_\nu}\hat{\boldsymbol{b}}\cdot\nabla z \frac{\partial B_0}{\partial z}\frac{\partial g_\nu}{\partial v_\parallel} + \mathrm{i}\omega_{d,\nu} g_\nu,\\
\hat{G}_{\phi,\nu}[\boldsymbol{p};\, \phi] &= v_\parallel \hat{\boldsymbol{b}}\cdot\nabla z \frac{Z_\nu e}{T_\nu} F_{0,\nu}\frac{\partial \mathrm{J}_{0,\nu}\phi}{\partial z} + \mathrm{i}\frac{Z_\nu e}{T_\nu}\omega_{d,\nu}\mathrm{J}_{0,\nu}F_{0,\nu}\phi + \mathrm{i}\omega_{*,\nu}\mathrm{J}_{0,\nu}F_{0,\nu}\phi,\\
\hat{G}_{A_\parallel,\nu}[\boldsymbol{p};\, A_\parallel] &= -\frac{v_\parallel^2}{c}\hat{\boldsymbol{b}}\cdot\nabla z \frac{Z_\nu e}{T_\nu}F_{0,\nu}\frac{\partial \mathrm{J}_{0,\nu}A_\parallel}{\partial z} - \mathrm{i}\frac{Z_\nu e}{T_\nu}\omega_{d,\nu}\mathrm{J}_{0,\nu}F_{0,\nu}\frac{v_\parallel}{c}A_\parallel - \mathrm{i}\omega_{*,\nu}\mathrm{J}_{0,\nu}F_{0,\nu}\frac{v_\parallel}{c}A_\parallel\\
&\quad + \frac{Z_\nu e}{T_\nu}\frac{\mu_\nu}{m_\nu c}\hat{\boldsymbol{b}}\cdot\nabla B_0 F_{0,\nu}\mathrm{J}_{0,\nu}A_\parallel,\\
\hat{G}_{B_\parallel,\nu}[\boldsymbol{p};\, \delta B_\parallel] &= 2v_\parallel\hat{\boldsymbol{b}}\cdot\nabla z F_{0,\nu}\frac{\partial}{\partial z}\left(\frac{\mu_\nu}{T_\nu}\frac{\mathrm{J}_{1,\nu}}{a_\nu}\delta B_\parallel\right) + 2\mathrm{i}\omega_{d,\nu}\frac{\mu_\nu}{T_\nu}\frac{\mathrm{J}_{1,\nu}}{a_\nu}F_{0,\nu}\delta B_\parallel\\
&\quad + 2\mathrm{i}\omega_{*,\nu}\frac{\mu_\nu}{Z_\nu e}\frac{\mathrm{J}_{1,\nu}}{a_\nu}F_{0,\nu}\delta B_\parallel,\\
\hat{Q}_{g,\nu}[\boldsymbol{p};\, g_\nu] &= Z_\nu e \mathrm{J}_{0,\nu} g_\nu,\\
\hat{Q}_{\phi}[\boldsymbol{p};\, \phi] &= \sum_\nu \frac{(Z_\nu e)^2 n_\nu}{T_\nu}(\Gamma_{0,\nu} - 1)\phi,\\
\hat{Q}_{B_\parallel}[\boldsymbol{p};\, \delta B_\parallel] &= \sum_\nu Z_\nu e n_\nu \frac{\Gamma_{1,\nu}}{B_0}\delta B_\parallel,\\
\hat{M}_{g,\nu}[\boldsymbol{p};\, g_\nu] &= -\frac{4\pi}{k_\perp^2}\frac{v_\parallel}{c}Z_\nu e \mathrm{J}_{0,\nu} g_\nu,\\
\hat{M}_{A_\parallel}[\boldsymbol{p};\, A_\parallel] &= \left[1 + \frac{4\pi}{k_\perp^2 c^2}\sum_\nu \frac{(Z_\nu e)^2 n_\nu}{m_\nu}\Gamma_{0,\nu}\right]A_\parallel,\\
\hat{N}_{g,\nu}[\boldsymbol{p};\, g_\nu] &= 8\pi\frac{\mathrm{J}_{1,\nu}}{a_\nu}\mu_\nu g_\nu,\\
\hat{N}_{\phi}[\boldsymbol{p};\, \phi] &= \left[4\pi\sum_\nu \frac{Z_\nu e n_\nu}{B_0}\Gamma_{1,\nu}\right]\phi,\\
\hat{N}_{B_\parallel}[\boldsymbol{p};\, \delta B_\parallel] &= \left[1 + 16\pi\sum_\nu \frac{n_\nu T_\nu}{B_0^2}\Gamma_{2,\nu}\right]\delta B_\parallel,
\end{aligned}
\tag{A1}
$$

and $\hat{C}_{\nu,\nu'}$ is an appropriate collision operator.

Appendix B. Adjoints of operators

The adjoint operators appearing in equations (3.9)–(3.12) are obtained by performing integration by parts wherever a derivative acts on the distribution function or a field variable. After performing the change of variables $\tilde{v}_\parallel \to -\tilde{v}_\parallel$ the adjoint operators take

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

the form

$$\left.\begin{aligned}
\hat{G}^{\dagger}_{g,\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] &= \gamma^{*}\lambda^{\leftrightarrow}_{\nu} + v_{\parallel}\hat{\boldsymbol{b}}\cdot\nabla z\frac{\partial\lambda^{\leftrightarrow}_{\nu}}{\partial z} - \frac{\mu_{\nu}}{m_{\nu}}\hat{\boldsymbol{b}}\cdot\nabla z\frac{\partial B_0}{\partial z}\frac{\partial\lambda^{\leftrightarrow}_{\nu}}{\partial v_{\parallel}} - \mathrm{i}\omega_{d,\nu}\lambda^{\leftrightarrow}_{\nu},\\
\hat{G}^{\dagger}_{\phi,\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] &= \frac{Z_{\nu}e}{T_{\nu}}\mathrm{J}_{0,\nu}F_{0,\nu}\hat{S}_{\nu},\\
\hat{G}^{\dagger}_{A_{\parallel},\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] &= \frac{Z_{\nu}e}{T_{\nu}}\mathrm{J}_{0,\nu}F_{0,\nu}\frac{v_{\parallel}}{c}\hat{S}_{\nu} + \frac{Z_{\nu}e}{T_{\nu}}\frac{\mu_{\nu}}{m_{\nu}c}\hat{\boldsymbol{b}}\cdot\nabla B_0F_{0,\nu}\mathrm{J}_{0,\nu}\lambda^{\leftrightarrow}_{\nu},\\
\hat{G}^{\dagger}_{B_{\parallel},\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] &= 2\frac{\mu_{\nu}}{T_{\nu}}\frac{\mathrm{J}_{1,\nu}}{\tilde{a}_{\nu}}F_{0,\nu}\hat{S}_{\nu},\\
\hat{Q}^{\dagger}_{g,\nu}[\boldsymbol{p};\xi] &= Z_{\nu}e\mathrm{J}_{0,\nu}F_{0,\nu}\xi,\\
\hat{Q}^{\dagger}_{\phi}[\boldsymbol{p};\xi] &= \sum_{\nu}\frac{(Z_{\nu}e)^2n_{\nu}}{T_{\nu}}\left(\Gamma_{0,\nu}-1\right)\xi,\\
\hat{Q}^{\dagger}_{B_{\parallel}}[\boldsymbol{p};\xi] &= 4\pi\sum_{\nu}\frac{Z_{\nu}en_{\nu}}{B_0}\Gamma_{1,\nu}\xi,\\
\hat{M}^{\dagger}_{g,\nu}[\boldsymbol{p};\zeta] &= -\frac{4\pi}{k_{\perp}^2}\frac{v_{\parallel}}{c}Z_{\nu}e\mathrm{J}_{0,\nu}\zeta,\\
\hat{M}^{\dagger}_{A_{\parallel}}[\boldsymbol{p};\zeta] &= \left[1+\frac{4\pi}{k_{\perp}^2c^2}\sum_{\nu}\frac{(Z_{\nu}e)^2n_{\nu}}{m_{\nu}}\Gamma_{0,\nu}\right]\zeta,\\
\hat{N}^{\dagger}_{g,\nu}[\boldsymbol{p};\sigma] &= 8\pi\frac{\mathrm{J}_{1,\nu}}{a_{\nu}}\mu_{\nu}\sigma,\\
\hat{N}^{\dagger}_{\phi}[\boldsymbol{p};\sigma] &= \left[4\pi\sum_{\nu}\frac{Z_{\nu}en_{\nu}}{B_0}\Gamma_{1,\nu}\right]\sigma,\\
\hat{N}^{\dagger}_{B_{\parallel}}[\boldsymbol{p};\sigma] &= \left[1+16\pi\sum_{\nu}\frac{n_{\nu}T_{\nu}}{B_0^2}\Gamma_{2,\nu}\right]\sigma,\\
\hat{C}^{\dagger}_{\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] &= \hat{C}_{\nu}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}],
\end{aligned}\right\} \tag{B1}$$

where we have defined

$$\hat{S}[\boldsymbol{p};\lambda^{\leftrightarrow}_{\nu}] = v_{\parallel}\hat{\boldsymbol{b}}\cdot\nabla z\frac{\partial\lambda^{\leftrightarrow}_{\nu}}{\partial z} - \mathrm{i}\omega_{d,\nu}\lambda^{\leftrightarrow}_{\nu} - \mathrm{i}\omega_{*,\nu}\lambda^{\leftrightarrow}_{\nu}. \tag{B2}$$

Appendix C. Simplifying adjoint equations

Consider taking the following moments of (3.9):

$$\left\langle\frac{Z_{\nu}e}{T_{\nu}}\mathrm{J}_{0,\nu}F_{0,\nu},\cdot\right\rangle_{v,\nu},\quad \left\langle\frac{Z_{\nu}e}{T_{\nu}}\mathrm{J}_{0,\nu}F_{0,\nu}\frac{v_{\parallel}}{c},\cdot\right\rangle_{v,\nu},\quad \left\langle 2\frac{\mathrm{J}_{1,\nu}}{a_{\nu}}F_{0,\nu}\frac{\mu_{\nu}}{T_{\nu}},\cdot\right\rangle_{v,\nu}, \tag{C1a–c}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

giving

$$0 = \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}v_\parallel \int \mathrm{d}\mu_\nu \alpha_\nu(z, v_\parallel, \mu_\nu) F_{0,\nu} \left\{ \gamma^* \lambda_\nu^\leftrightarrow + v_\parallel \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial \lambda_\nu^\leftrightarrow}{\partial z} \right.$$
$$\left. - \frac{\mu_\nu}{m_\nu} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial v_\parallel} - \mathrm{i}\omega_{d,\nu} \lambda_\nu^\leftrightarrow + Z_\nu e \mathrm{J}_{0,\nu} \xi - \frac{4\pi}{k_\perp^2} Z_\nu e \mathrm{J}_{0,\nu} \frac{v_\parallel}{c} \zeta + 8\pi \frac{\mathrm{J}_{1,\nu}}{a_\nu} \mu_\nu \sigma \right\}, \tag{C2}$$

where α_ν can take the forms

$$\alpha_\nu = \begin{cases} \dfrac{Z_\nu e}{T_\nu} \mathrm{J}_{0,\nu}, \\ \dfrac{Z_\nu e}{T_\nu} \mathrm{J}_{0,\nu} \dfrac{v_\parallel}{c}, \\ 2 \dfrac{\mathrm{J}_{1,\nu}}{a_\nu} \dfrac{\mu_\nu}{T_\nu}. \end{cases} \tag{C3}$$

We now identify different terms in (C2) for each α_ν which are odd in $v_\parallel$ so evaluate to zero when integrated over the domain $\{-\infty, \infty\}$

$$0 = \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}^2 v \frac{Z_\nu e}{T_\nu} F_{0,\nu} \left\{ \gamma^* \mathrm{J}_{0,\nu} \lambda_\nu^\leftrightarrow + v_\parallel \hat{\boldsymbol{b}} \cdot \nabla z \mathrm{J}_{0,\nu} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial z} - \mathrm{i}\omega_{d,\nu} \mathrm{J}_{0,\nu} \lambda_\nu^\leftrightarrow \right.$$
$$\left. - \frac{\mu_\nu}{m_\nu} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial v_\parallel} + Z_\nu e \mathrm{J}_{0,\nu}^2 \xi + \underbrace{\frac{4\pi}{k_\perp^2} Z_\nu e \mathrm{J}_{0,\nu}^2 \frac{v_\parallel}{c} \zeta}_{\text{odd in } v_\parallel} + 8\pi \frac{\mathrm{J}_{1,\nu} \mathrm{J}_{0,\nu}}{a_\nu} \mu_\nu \sigma \right\}, \tag{C4}$$

$$0 = \sum_\nu \frac{2\pi B_0}{m_\nu} \int \mathrm{d}^2 v \frac{Z_\nu e}{T_\nu} F_{0,\nu} \left\{ \gamma^* \mathrm{J}_{0,\nu} \frac{v_\parallel}{c} \lambda_\nu^\leftrightarrow + \frac{v_\parallel^2}{c} \hat{\boldsymbol{b}} \cdot \nabla z \mathrm{J}_{0,\nu} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial z} \right.$$
$$- \frac{\mu_\nu}{m_\nu} \frac{v_\parallel}{c} \hat{\boldsymbol{b}} \cdot \nabla z \frac{\partial B_0}{\partial z} \frac{\partial \lambda_\nu^\leftrightarrow}{\partial v_\parallel} - \mathrm{i}\omega_{d,\nu} \frac{v_\parallel}{c} \mathrm{J}_{0,\nu} \lambda_\nu^\leftrightarrow$$
$$\left. + \underbrace{Z_\nu e \mathrm{J}_{0,\nu}^2 \frac{v_\parallel}{c} \xi}_{\text{odd in } v_\parallel} + \frac{4\pi}{k_\perp^2} Z_\nu e \mathrm{J}_{0,\nu}^2 \left(\frac{v_\parallel}{c} \right)^2 \zeta + \underbrace{8\pi \frac{\mathrm{J}_{1,\nu} \mathrm{J}_{0,\nu}}{a_\nu} \frac{v_\parallel}{c} \mu_\nu \sigma}_{\text{odd in } v_\parallel} \right\}, \tag{C5}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

$$0 = 2\sum_{\nu}\frac{2\pi B_0}{m_\nu}\int \mathrm{d}^2 v F_{0,\nu}\left\{\frac{\mathrm{J}_{1,\nu}}{a_\nu}\frac{\mu_\nu}{T_\nu}\gamma^*\lambda_\nu^{\leftrightarrow} + v_\parallel\frac{\mathrm{J}_{1,\nu}}{a_\nu}\frac{\mu_\nu}{T_\nu}\hat{\boldsymbol{b}}\cdot\nabla z\frac{\partial\lambda_\nu^{\leftrightarrow}}{\partial z}\right.$$
$$-\frac{\mu_\nu}{m_\nu}\hat{\boldsymbol{b}}\cdot\nabla z\frac{\partial B_0}{\partial z}\frac{\mathrm{J}_{1,\nu}}{a_\nu}\frac{\mu_\nu}{T_\nu}\frac{\partial\lambda_\nu^{\leftrightarrow}}{\partial v_\parallel} - \mathrm{i}\omega_{d,\nu}\frac{\mathrm{J}_{1,\nu}}{a_\nu}\frac{\mu_\nu}{T_\nu}\lambda_\nu^{\leftrightarrow} + \frac{Z_\nu e\mu_\nu}{T_\nu}\frac{\mathrm{J}_{0,\nu}\mathrm{J}_{1,\nu}}{a_\nu}\xi$$
$$\left.+\underbrace{\frac{4\pi}{k_\perp^2}Z_\nu e\frac{\mathrm{J}_{0,\nu}\mathrm{J}_{1,\nu}}{a_\nu}\frac{v_\parallel}{c}\frac{\mu_\nu}{T_\nu}\zeta}_{\text{odd in } v_\parallel} + 8\pi\left(\frac{\mathrm{J}_{1,\nu}}{a_\nu}\right)^2\frac{\mu_\nu^2}{T_\nu}\sigma\right\}. \tag{C6}$$

We can then perform integration by parts on the remaining $v_\parallel$ derivative, using velocity independence of the fields. Using equations (3.10)–(3.12) it is then possible to simplify the above equations to produce the results (3.14)–(3.16).

Appendix D. Integral weights

The code `stella` computes its $\tilde{\mu}_\nu$ grid points according to Gauss–Laguerre quadrature such that the $\tilde{\mu}_\nu$-integration of a variable, f, may be approximated numerically as

$$B\int_0^\infty \mathrm{d}\mu_\nu f(\mu_\nu) \approx \sum_i^{N_{\mu_\nu}} w_i \mathrm{e}^{\hat{\mu}_\nu}\frac{B}{B_0}f(\hat{\mu}_\nu), \tag{D1}$$

where N_{μ_ν} is the number of grid points in μ_ν, and the definition $\hat{\mu}_\nu = \mu_\nu B_0$ is made, with B_0 relating to the minimum of the normalised magnetic field. This B_0 acts as a scaling factor that acts on the upper limit of the μ_ν grid such that the grid may be defined solely based on the number of grid points whilst still covering the necessary domain as $\hat{\mu}_\nu$ is an independent coordinate that is independent of $\boldsymbol{p}$. Previously, the importance of the role of the Jacobians was mentioned with a stress that certain variables are being considered as fixed, dummy variables, whilst others are dependent on the geometric inputs, $\boldsymbol{p}$. We take the $\hat{\mu}_\nu$ variable to be the dummy variable that is kept fixed in the perpendicular velocity integral such that the weights $w_i = B/B_0$ are the only factors to be perturbed in the integrals and, in a similar manner, the fixed variable, $\hat{\mu}_\nu$, is weighted by a varying factor, B_0, when it appears in the equations. It should be noted that since the $\hat{\mu}_\nu$ grid is calculated independently of the geometry, it is the same for each perturbed value of $\boldsymbol{p}$, and as such the $\hat{\mu}_\nu$ grids will always align even when multiplying terms that are evaluated at different $\boldsymbol{p}$ values, such as when coefficients are perturbed but are multiplied by $\tilde{g}_\nu(\boldsymbol{p}_0)$.

Appendix E. Geometry implementation

The code `stella` has an input option to use the Miller parametrisation of a flux surface; it takes a set of input variables to describe the local geometry of a specified flux surface along with the two adjacent flux surfaces on either side. We take (4.16) in the form

$$R(r,\theta) = R_0(r) + r\cos[\theta + \sin(\theta)\delta(r)], \tag{E1}$$

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

with the triangularity redefined as $\delta(r) := \arcsin[\bar{\delta}(r)]$. We now consider Taylor expanding in r about $r = r_{\psi_0}$

$$R_0(r) = R_0(r_{\psi_0}) + \left.\frac{\mathrm{d}R_0}{\mathrm{d}r}\right|_{r_{\psi_0}} (r - r_{\psi_0}) + \cdots \approx R_{\psi_0} + \Delta_{\psi_0}\,\mathrm{d}r + O(\mathrm{d}r^2), \tag{E2}$$

with $\mathrm{d}r = r - r_{\psi_0}$, $R_{\psi_0} = R_0(r_{\psi_0})$, and $\Delta_{\psi_0} = \mathrm{d}R_0/\mathrm{d}r|_{r_{\psi_0}}$. Similarly

$$\delta(r) = \delta(r_{\psi_0}) + \left.\frac{\mathrm{d}\delta}{\mathrm{d}r}\right|_{r_{\psi_0}} (r - r_{\psi_0}) + \cdots \approx \delta_{\psi_0} + \delta'_{\psi_0}\,\mathrm{d}r + O(\mathrm{d}r^2), \tag{E3}$$

with $\delta_{\psi_0} = \delta(r_{\psi_0})$, and $\delta'_{\psi_0} = \mathrm{d}\delta/\mathrm{d}r|_{r_{\psi_0}}$. Combining (E2) and (E3) gives

$$\begin{aligned} R(r,\theta) &\approx R_{\psi_0} + r_{\psi_0}\cos[\theta + \sin(\theta)\delta_{\psi_0}] \\ &+ \left\{\Delta_{\psi_0} + \cos[\theta + \sin(\theta)\delta_{\psi_0}] - r_{\psi_0}\sin[\theta + \sin(\theta)\delta_{\psi_0}]\sin(\theta)\delta'_{\psi_0}\right\}\mathrm{d}r, \end{aligned} \tag{E4}$$

such that the above definition holds on any given flux surface, ψ_0. Equivalently, (4.17) can also be expanded about $r = r_{\psi_0}$ by first expanding the elongation

$$\kappa(r) = \kappa(r_{\psi_0}) + \left.\frac{\mathrm{d}\kappa}{\mathrm{d}r}\right|_{r_{\psi_0}} (r - r_{\psi_0}) + \cdots \approx \kappa_{\psi_0} + \kappa'_{\psi_0}\,\mathrm{d}r + O(\mathrm{d}r^2), \tag{E5}$$

with $\kappa_{\psi_0} = \kappa(r_{\psi_0})$, and $\kappa'_{\psi_0} = \mathrm{d}\kappa/\mathrm{d}r|_{r_{\psi_0}}$ to then write

$$Z(r,\theta) \approx r_{\psi_0}\kappa_{\psi_0}\sin(\theta) + \left[\kappa_{\psi_0} + r_{\psi_0}\kappa'_{\psi_0}\right]\sin(\theta)\,\mathrm{d}r. \tag{E6}$$

These functions are used to describe the geometry of the flux surface of interest and the two adjacent flux surfaces by setting $\mathrm{d}r = \{0, \pm\Lambda\}$, with $\Lambda \ll 1$ a constant, in order to evaluate their radial derivatives. These quantities are then used to compute the Jacobian, magnetic field strength and configuration along with other functions defined on the flux surface.

REFERENCES

Abel, I.G., Plunk, G.G., Wang, E., Barnes, M., Cowley, S.C., Dorland, W. & Schekochihin, A.A. 2013 Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. *Rep. Prog. Phys.* **76** (11), 116201.

Antonsen, T.M. Jr. & Lane, B. 1980 Kinetic equations for low frequency instabilities in inhomogeneous plasmas. *Phys. Fluids* **23** (6), 1205–1214.

Barnes, M., Parra, F.I. & Landreman, M. 2019 `stella`: an operator-split, implicit-explicit δf-gyrokinetic code for general magnetic field configurations. *J. Comput. Phys.* **391**, 365–380.

Beeke, O., Barnes, M., Romanelli, M., Nakata, M. & Yoshida, M. 2020 Impact of plasma shaping on tokamak microstability. Nuclear Fusion 61 (6), 066020.

Beer, M.A., Cowley, S.C. & Hammett, G.W. 1995 Field-aligned coordinates for nonlinear simulations of tokamak turbulence. *Phys. Plasmas* **2** (7), 2687–2700.

Catto, P.J. 1978 Linearized gyro-kinetics. *Plasma Phys.* **20** (7), 719.

Cowley, S.C., Kulsrud, R.M. & Sudan, R. 1991 Considerations of ion-temperature-gradient-driven turbulence. *Phys. Fluids* B **3** (10), 2767–2782.

Dimits, A.M., Bateman, G., Beer, M.A., Cohen, B.I., Dorland, W., Hammett, G.W., Kim, C., Kinsey, J.E., Kotschenreuther, M., Kritz, A.H., *et al.* 2000 Comparisons and physics basis of tokamak transport models and turbulence simulations. *Phys. Plasmas* **7** (3), 969–983.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press

DOYLE, E.J., HOULBERG, W.A. & KAMADA, Y. 2007 Chapter 2: plasma confinement and transport [Progress in the ITER Physics Basis (PIPB)]. *Nucl. Fusion* **47**, S20–S36.

DRAKE, J.F. & LEE, Y.C. 1977 Kinetic theory of tearing instabilities. *Phys. Fluids* **20** (8), 1341–1353.

GERALDINI, A., LANDREMAN, M. & PAUL, E. 2021 An adjoint method for determining the sensitivity of island size to magnetic field variations. *J. Plasma Phys*. **87** (3), 905870302.

GOTTLIEB, S., SHU, C.-W. & TADMOR, E. 2001 Strong stability-preserving high-order time discretization methods. *SIAM Rev*. **43** (1), 89–112.

HIGHCOCK, E.G., MANDELL, N.R., BARNES, M. & DORLAND, W. 2018 Optimisation of confinement in a fusion reactor using a nonlinear turbulence model. *J. Plasma Phys*. **84** (2), 905840208.

MANTICA, P., ANGIONI, C., BAIOCCHI, B., BARUZZO, M., BEURSKENS, M.N.A., BIZARRO, J.P.S., BUDNY, R.V., BURATTI, P., CASATI, A., CHALLIS, C., *et al.* 2011 Ion heat transport studies in JET. *Plasma Phys. Control. Fusion* **53** (12), 124033.

MILLER, R.L., CHU, M.S., GREENE, J.M., LIN-LIU, Y.R. & WALTZ, R.E. 1998 Noncircular, finite aspect ratio, local equilibrium model. *Phys. Plasmas* **5** (4), 973–978.

NIES, R., PAUL, E.J., HUDSON, S.R. & BHATTACHARJEE, A. 2022 Adjoint methods for quasi-symmetry of vacuum fields on a surface. *J. Plasma Phys*. **88** (1), 905880106.

PAUL, E.J., LANDREMAN, M., BADER, A. & DORLAND, W. 2018 An adjoint method for gradient-based optimization of stellarator coil shapes. *Nucl. Fusion* **58** (7), 076015.

PIRONNEAU, O. 1974 On optimum design in fluid mechanics. *J. Fluid Mech*. **64**, 97–110.

POCHELON, A., GOODMAN, T.P., HENDERSON, M., ANGIONI, C., BEHN, R., CODA, S., HOFMANN, F., HOGGE, J.P., KIRNEVA, N., MARTYNOV, A.A., *et al.* 1999 Energy confinement and MHD activity in shaped TCV plasmas with localized electron cyclotron heating. *Nucl. Fusion* **39** (11Y), 1807–1818.

ROGERS, B.N., DORLAND, W. & KOTSCHENREUTHER, M. 2000 Generation and stability of zonal flows in ion-temperature-gradient mode turbulence. *Phys. Rev. Lett*. **85**, 5336–5339.

RUDAKOV, L.I. & SAGDEEV, R.Z. 1961 On the instability of a nonuniform rarefied plasma in a strong magnetic field. *Sov. Phys. Dokl*. **6**, 415.

SHAFRANOV, V.D. 1966 Plasma equilibrium in a magnetic field. *Rev. Plasma Phys*. **2**, 103.

TANG, W.M., CONNOR, J.W. & HASTIE, R.J. 1980 Kinetic-ballooning-mode theory in general geometry. *Nucl. Fusion* **20** (11), 1439.

TRANSTRUM, M.K. & SETHNA, J.P. 2012 Improvements to the Levenberg–Marquardt algorithm for nonlinear least-squares minimization. arXiv:1201.5885.

https://doi.org/10.1017/S0022377824000709 Published online by Cambridge University Press