This is a repository copy of *Zonal flow evolution and overstability in accretion discs*. 

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/116133/

Version: Published Version

**Article:**
Vanon, R. and Ogilvie, G. (2017) Zonal flow evolution and overstability in accretion discs. Monthly Notices of the Royal Astronomical Society, 466 (3). pp. 2590-2601. ISSN 0035-8711

https://doi.org/10.1093/mnras/stw3232

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Zonal flow evolution and overstability in accretion discs

R. Vanon* and G. I. Ogilvie

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

Accepted 2016 December 8. Received 2016 December 7; in original form 2016 October 28

ABSTRACT

This work presents a linear analytical calculation on the stability and evolution of a compressible, viscous self-gravitating (SG) Keplerian disc with both horizontal thermal diffusion and a constant cooling time-scale when an axisymmetric structure is present and freely evolving. The calculation makes use of the shearing sheet model and is carried out for a range of cooling times. Although the solutions to the inviscid problem with no cooling or diffusion are well known, it is non-trivial to predict the effect caused by the introduction of cooling and of small diffusivities; this work focuses on perturbations of intermediate wavelengths, therefore representing an extension to the classical stability analysis on thermal and viscous instabilities. For density wave modes, the analysis can be simplified by means of a regular perturbation analysis; considering both shear and thermal diffusivities, the system is found to be overstable for intermediate and long wavelengths for values of the Toomre parameter $Q \lesssim 2$; a non-SG instability is also detected for wavelengths $\gtrsim 18H$, where $H$ is the disc scale-height, as long as $\gamma \lesssim 1.305$. The regular perturbation analysis does not, however, hold for the entropy and potential vorticity slow modes as their ideal growth rates are degenerate. To understand their evolution, equations for the axisymmetric structure’s amplitudes in these two quantities are analytically derived and their instability regions obtained. The instability appears boosted by increasing the value of the adiabatic index and of the Prandtl number, while it is quenched by efficient cooling.

Key words: accretion, accretion discs – hydrodynamics – instabilities – turbulence.

1 INTRODUCTION

Accretion discs are subject to an assortment of instabilities; two of the most widely studied instances are the classical thermal and viscous instabilities (e.g. Pringle, Rees & Pacholczyk 1973; Lightman & Eardley 1974; Shakura & Sunyaev 1976; Livio & Shaviv 1977; Pringle 1977; Piran 1978; Pringle 1981). Their existence depends on assumptions about how the angular momentum transport and dissipation are modelled, which distinguishes them from more fundamental dynamical instabilities such as the magnetorotational instability (MRI), the gravitational instability and the vertical shear instability.

In a Keplerian disc of surface density $\Sigma$ and angular frequency $\Omega$, which is in thermal equilibrium, the heating and cooling rates $\mathcal{H}$ and $\mathcal{C}$ are equal and are given by

$$\mathcal{H} = \frac{9}{4} v \Sigma \Omega^2 \propto \alpha T_c \Sigma \Omega \tag{1}$$

$$\mathcal{C} = 2 \sigma T_{\text{eff}}^4 \propto \frac{\tau^4}{\tau} \tag{2}$$

where $v = \alpha c_{\text{visc}} H$ is the kinematic viscosity (with $c_{\text{visc}} \propto T_c^{1/2}$ and $H = c_{\text{visc}} / \Omega$ being the isothermal sound speed and the disc scale-height), $\sigma$ is the Stefan–Boltzmann constant, $\tau$ (here assumed $\gtrsim 1$) is the optical thickness and $T_c$ and $T_{\text{eff}}$ are the central and effective temperatures of the disc.

As both $\alpha$ and $\tau$ are potentially functions of $T_c$, the disc is thermally unstable to perturbations in $T_c$ if

$$\left. \frac{\partial \ln \mathcal{H}}{\partial \ln T_c} \right|_{\Sigma} > \left. \frac{\partial \ln \mathcal{C}}{\partial \ln T_c} \right|_{\Sigma}, \tag{3}$$

as it would lead to runaway heating (cooling) for an upward (downward) temperature perturbation. In the above criterion, the surface density $\Sigma$ is held constant as changes in temperature happen on a much shorter time-scale than changes in $\Sigma$ due to the thermal time-scale $\tau_{\text{th}}$ being given by

$$\tau_{\text{th}} \simeq \left( \frac{H}{R} \right)^2 \tau_{\text{visc}} \tag{4}$$

with $\tau_{\text{visc}}$ representing the viscous time-scale and $H/R \ll 1$ for a thin disc.

The $\alpha$ model of accretion discs (Shakura & Sunyaev 1973) predicts the disc to be thermally unstable in the inner regions (where the radiation pressure dominates), although it is uncertain whether

© 2016 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society
the thermal instability predicted by the α model takes place in real discs, with some observations seeming to have proven otherwise (eg. Gierliński & Done 2004; Done, Gierliński & Kubota 2007); a competing model exists (dubbed the β model) where the stress is proportional to the gas pressure, rather than the total pressure as in the α model. This produces a thermally stable disc (Sakimoto & Coroniti 1981; Stella & Rosner 1984; Merloni 2003). Moreover, the α model neglects other effects such as heating from MRI-induced turbulence (eg. Hirose, Kroll & Blaes 2009) and heat transport within the disc.

A disc is said to be viscously unstable if a perturbation δμ applied to the dynamic viscosity μ = νΣ grows. Substituting this perturbation into the equation of diffusive disc evolution

\[ \frac{\partial \Sigma}{\partial t} = \frac{3}{r} \frac{\partial}{\partial r} \left[ \frac{r^{1/2}}{\Sigma} \frac{\partial}{\partial r} \left( \nu \Sigma r^{1/2} \right) \right], \]  

(5)
gives

\[ \frac{\partial}{\partial t} (\delta \mu) = \frac{\partial \mu}{\partial \Sigma} \frac{3}{r} \frac{\partial}{\partial r} \left[ r^{1/2} \frac{\partial}{\partial r} (r^{1/2} \delta \mu) \right], \]  

(6)

with an instability being triggered if the diffusion coefficient is negative. This implies the viscous instability criterion to be (Lightman & Eardley 1974)

\[ \frac{\partial (\nu \Sigma)}{\partial \Sigma} < 0, \]  

(7)

with the derivative being taken at constant r, and under the assumption of both thermal balance and hydrostatic equilibrium.

The classical approach does, however, have limitations, the most notable of which being the consideration of long wavelength perturbations obeying \( H \ll \lambda_{\text{pert}} \ll R_0 \), only, which in turn allows the thermal and viscous instabilities to be distinct. A more general analysis can be conducted by considering perturbations of wavelength \( \lambda_{\text{pert}} \sim H \); in this case, the previously existing structure in the density also develops a significant perturbation in the azimuthal component of the velocity, therefore becoming a zonal flow, which modifies the shear rate from its Keplerian value. This more generic analysis can be used to study the stability of the slow modes and establish whether zonal flows grow or decay as a result of non-ideal effects such as viscous interactions, cooling and heating, as well as the coupling between the modes.

Zonal flows – axisymmetric shear flows consisting of parallel bands – represent an equilibrium solution to the equations governing the evolution of an accretion disc’s flow, including a geostrophic balance between the Coriolis force and the pressure gradient. This can, however, be unstable under certain conditions, in which case the flow can undergo a Kelvin–Helmholtz (or Rossby wave) instability (Vanon & Ogilvie 2016). Zonal flows have been observed to persist in certain conditions; one such example is 3D simulation of magnetohydrodynamic (MHD)-turbulent discs modelled using the shearing box approximation (Johansen, Youdin & Klahr 2009; Simon, Beckwith & Armitage 2012; Kunz & Lesur 2013; Bai & Stone 2014). In this scenario, zonal flows are seen to exhibit larger perturbations of wavelengths \( \lambda_{\text{pert}} \sim H \), rather than \( H \ll \lambda_{\text{pert}} \ll R_0 \) as in the classical works dealing with thermal and viscous instabilities; our work therefore represents an extension of the classical theory of said instabilities. The paper is arranged as follows: Section 2 serves as an introduction to the shearing sheet model, which is employed in this analysis, as well as the full non-linear, viscous equations governing the system described. Section 3 introduces the axisymmetric structure and the equations describing its temporal evolution; it also analyses the evolution and stability of both density waves (DWs) and slow modes. The work terminates in Section 4, where the conclusions drawn from the results are presented.

### 2 Model

The work presented in this paper is based on the local unstratified shearing sheet model, whose first use was by Goldreich & Lynden-Bell (1965) in the context of galactic discs. This consists of drawing a sheet of small dimensions compared to the disc size centred at a fiducial radius \( R_0 \) (i.e. \( L_r, L_z \ll R_0 \), where \( L_r \) and \( L_z \) are the radial and azimuthal dimensions of the chosen sheet). The frame of reference of the sheet, which is of a Cartesian nature, corotates with the disc at an angular frequency \( \Omega = \Omega_\text{e} \), with \( \mathbf{e} \) being the unit vector normal to the sheet; in the chosen frame of reference, the continuity and Navier–Stokes equations for a viscous, compressible fluid are given by

\[ \partial_t \Sigma + \nabla \cdot (\Sigma \mathbf{v}) = 0, \]  

(8)
\[ \partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + 2 \Omega \times \mathbf{v} = -\nabla \Phi - \nabla \Phi_{\text{d.m.}} - \frac{1}{\Sigma} \nabla P + \frac{1}{\Sigma} \nabla \cdot \mathbf{T}, \]  
\[ \text{where } \Sigma \text{ is the surface density of the disc, } \mathbf{v} \text{ is the velocity of the flow, } \Phi = -q \Omega^2 x^2 \text{ is the effective tidal potential (with } q = -\frac{\text{dln} \Omega}{\text{dln} r} \text{ representing the dimensionless shear rate), its value being } q = 3/2 \text{ for a Keplerian disc), } \Phi_{\text{d.m.}} \text{ is the disc potential evaluated at its mid-plane, } P \text{ is the 2D pressure and } T = 2 \mu_s \mathbf{S} + \mu_s (\nabla \cdot \mathbf{v}) \mathbf{I} \text{ is the viscous stress tensor, with } S = \frac{1}{2} \left[ (\nabla \mathbf{v} + (\nabla \mathbf{v})^T) - \frac{1}{2} \nabla (\mathbf{v} \cdot \mathbf{v}) \right] \text{ being the traceless shear tensor, } \mu_s \text{ the dynamic viscosity } (\mu_2 \text{ and } \mu_1 \text{ being the shear and bulk dynamic viscosities, respectively), } \nu \text{ the kinematic viscosity and } I \text{ the unit tensor.} \]

The quantity \( h = \ln \Sigma + \text{constant} \) is introduced, which turns the continuity equation into
\[ \partial_t h + \mathbf{v} \cdot \nabla h + \nabla \cdot \mathbf{v} = 0, \]  
while the disc potential can be readily evaluated at the disc’s mid-plane in Fourier space by means of Poisson’s equation \( \nabla^2 \Phi_d = 4 \pi G \Sigma \delta(z) \), its form being described by
\[ \Phi_{\text{d.m.}} = -\frac{2 \pi G \Sigma}{\sqrt{k_z^2 + k_r^2}}, \]  
where \( G \) is the gravitational constant and \( k_r \) and \( k_z \) are the radial and azimuthal components of the wave vector \( k \).

Another crucial equation in the setup described is that for the temporal evolution of the specific internal energy \( e \), which is given by
\[ \partial_t e + \mathbf{v} \cdot \nabla e = -\frac{P}{\Sigma} \nabla \cdot \mathbf{v} + 2 \nu_s S^2 + \nu_s (\nabla \cdot \mathbf{v})^2 + \frac{1}{\Sigma} \nabla \cdot (\nu_s \nabla e) \]
\[ - \frac{1}{\tau_c} (e - e_{\text{irr}}). \]  
where \( \nu_s \) and \( \nu_i \) are the bulk and shear kinematic viscosities, \( \nu_s \) the (horizontal) thermal diffusion, \( \tau_c \) the (constant) cooling time-scale and \( e_{\text{irr}} \) the equilibrium specific internal energy to which the disc would relax if it were not viscously heated. The Prandtl number is defined as
\[ \text{Pr} = \frac{\nu_s}{\nu_i}. \]  

The analysis conducted in this paper will also make use of two quantities, which are material invariants in ideal conditions (i.e. in the absence of diffusivities and cooling): potential vorticity \( \xi \) (PV) and the dimensionless specific entropy \( s \), whose forms are given by
\[ \xi = \frac{2 \Omega + (\nabla \times \mathbf{v})}{\Sigma}, \]
\[ s = \frac{1}{\gamma} \ln P - \ln \Sigma, \]  
where \( \gamma \) represents the adiabatic index. The pressure \( P \) is given in terms of the specific internal energy \( e \) by
\[ P = (\gamma - 1) \Sigma e. \]  
This allows us to evaluate the pressure gradient term in the momentum equation as
\[ \nabla P = (\gamma - 1) (\nabla e + e \nabla h). \]

The background state of the system is described by \( e = e_0 = e_i (\gamma (\gamma - 1)), \) where \( e_0 \) is the adiabatic sound speed; the introduction of an internal energy induced by external irradiation \( e_{\text{irr}} \) acts as a buffer in the thermal balance of the system. Whereas in its absence thermal balance can only be achieved with one combination of cooling time and shear viscosity, the assumption that \( e_{\text{irr}} \geq 0 \) allows us to explore multiple perturbations of the two parameters to gauge their effect on disc stability. The thermal balance of the background state is given by
\[ e_0 = e_{\text{irr}} \alpha_2 (\gamma - 1) q^2 \Omega \tau_c. \]  
It is possible to identify the quantity \( f_{\text{visc}} = \alpha_2 (\gamma - 1) q^2 \Omega \tau_c, \) which represents the fraction of viscously generated heat, with \( e_{\text{irr}} = 0 \) (ie. disc being entirely viscously heated) yielding the maximum value of \( f_{\text{visc}} \) = 1. Equation (18), under the assumption \( e_{\text{irr}} \geq 0 \), implies that \( \alpha_2 \tau_c \leq \frac{1}{q^2 \Omega (\gamma - 1)}. \)  
where \( \alpha_2 = v_i \left( \frac{2 \nu_s}{k_r^2} \right) \) is a dimensionless viscosity parameter, which defines our ranges of shear viscosity and cooling time-scale ranges for a specific dimensionless shear rate and adiabatic index.

The background state is then perturbed such that \( \mathbf{v} = \mathbf{v}_0 + \mathbf{v}' \) [with \( \mathbf{v}' = (u', v', 0)^T \)], etc. This yields the following set of linearized equations describing the temporal evolution of the disturbance:
\[ \partial_t h' = -\partial_t u', \]
\[ \partial_t u' - 2 \nu_s \Omega u' = -\partial_t \Phi_{\text{d.m.}} - (\gamma - 1) \left[ \partial_t e' + e_0 \partial_t h' \right] \]
\[ + \left( \nu_b + \frac{4}{3} \nu_s \right) \partial_t^2 u', \]
\[ \partial_t v' + (2 - q) \Omega u' = \nu_s \partial_t^2 v' - \nu_s q \Omega \partial_t h', \]
\[ \partial_t e' = -(\gamma - 1) e_0 \partial_t u' - 2 \nu_s q \Omega \partial_t v' + \nu_s \partial_t^2 e' \]
\[ - \frac{1}{\tau_c} (e' - e_{\text{irr}}), \]  
with the analysis being based on the assumptions of \( \tau_c = \text{const} \) and \( v_i = \text{const} \). It is worth noting that the assumption of constant diffusivities made can potentially affect the stability properties of the model described.

As further explored in Section 3, the solutions to the above equations – which are either DWs or non-oscillating structures in the entropy and PV – are deeply influenced by the viscosity and thermal diffusivity values, as well as the effectiveness of the imposed cooling. Depending on their combined effects, the solutions to the problem can be either damped, exponentially growing or overstable (i.e. growing oscillations).

3 Evolution

The system admits axisymmetric, sinusoidal standing-wave solutions of the form
\[ h'(x, t) = A_1(t) \cos(kx) \]
\[ u'(x, t) = A_2(t) \sin(kx) \]
\[ v'(x, t) = A_3(t) \sin(kx) \]
\[ e'(x, t) = e_0 A_4(t) \cos(kx), \]  
\[ (25) \]

MNRAS 466, 2590–2601 (2017)
where \(A_h, A_v, A_e, A_t\) represent the amplitudes in the respective quantities and \(k > 0\) is the wavenumber of the above structure.

It is possible to obtain a set of equations describing the temporal evolution of the axisymmetric structure by applying its form outlined above into the linearized equations describing the system (equations 21–24):

\[
\begin{align*}
\partial_t A_h &= -k A_v, \\
\partial_t A_u - 2\Omega A_v &= -2\pi G \Sigma_0 A_h + \frac{c_s^2 k (A_e + A_h)}{\gamma} \\
&\quad - \left( \frac{4}{\gamma} + \frac{2}{\gamma} \right) A_v, \\
\partial_t A_s + (2 - q)\Omega A_s &= -\gamma_e A_v + \gamma_s \frac{q\Omega}{k} A_h, \\
\partial_t A_e &= -(\gamma_0 - 1)k A_u - \frac{2\gamma_0}{k\Sigma_0} A_u - \gamma_e A_v,
\end{align*}
\]

where \(\gamma_s = v_sk^2, \gamma_e = vEk^2\) and \(\gamma_0 = v_0k^2 + 1/\tau_e\) are three damping coefficients.

If we assume that these equations have solutions of the form \(ae^{\lambda t}\), a quartic equation for the complex growth rate \(\lambda\) can be determined, and its solutions analysed. In the inviscid case with no cooling or diffusion, these will be

\[
\lambda = 0, 0, \pm i\omega_0,
\]

where the zero subscript indicates the ideal case considered, \(\omega_0^2 = k^2 - 2\pi G \Sigma_0 k + c_s^2 k^2\) is the square of the DW frequency and \(\omega_0^2 = 2(2 - q)\Omega^2\) is the epicyclic frequency squared. The two non-zero roots correspond to the DW modes, while the zero roots correspond to the PV and entropy slow modes, as indicated in Fig. 1 by the filled shapes. The DWs are stable for all \(k\) values if \(Q > 1\), where \(Q\) is the Toomre parameter – which represents the strength of SG within a disc, with \(Q < 1\) causing the disc to be gravitationally unstable – given by

\[
Q = \frac{c_s k}{\pi G \Sigma_0}.
\]

Introducing the damping coefficients (assumed to be small, i.e. \(\gamma_i \ll \Omega\)) back into the picture gives a non-zero real part to all the modes’ growth rates, as shown in Fig. 1 by the empty shapes. If the newly acquired real part is negative, the damping coefficients have a stabilizing effect on the modes, while if Re(\(\lambda\)) > 0 the modes exhibit exponential growth (entropy and PV modes) or viscous overstability (DW modes). Understanding how the introduction of the three diffusivities affects the values of the solutions is, however, non-trivial. It is expected that a regular perturbation analysis can be made for non-degenerate eigenvalues (i.e. for the DW modes with \(\lambda_0 = \pm i\omega_0\)), assuming the diffusivity values are small enough; in this case, the solutions to the full equations are

\[
\lambda = \lambda_0 + \sum_{i=1}^{3} \gamma_i \left( \frac{\partial \lambda}{\partial \gamma_i} \right) + O\left(\gamma_i^2\right),
\]

where \(\gamma_i\) can represent a bulk, shear or thermal damping coefficient, the latter also including effects due to cooling. In the degenerate case (i.e. entropy/PV modes with \(\lambda_0 = 0\)) it is, however, possible that a singular perturbation is necessary, meaning the solutions would not agree with the expression given by equation (32).

### 3.1 Density wave modes

The linearization assumption is found to hold for DW modes (i.e. the non-zero roots in the inviscid case), and the independent contributions to these modes from the damping coefficients are calculated using the eigenvalue problem; these are

\[
\begin{align*}
\frac{\partial \lambda}{\partial \gamma_b} &= -\frac{1}{2}, \\
\frac{\partial \lambda}{\partial \gamma_s} &= -\frac{k^2 c_s^2 (\gamma - 1)}{2\gamma_0\omega_0^2}, \\
\frac{\partial \lambda}{\partial \gamma_e} &= \left[(\gamma - 1)q^2 + 2(2 - \gamma)q - 2\right] \frac{\Omega^2}{\omega_0^2} - \frac{2}{3},
\end{align*}
\]

While it is clear to see that the contribution from the bulk viscosity is always negative, meaning it will always have a stabilizing effect on the DW modes, the situation is more intricate in the case of shear viscosity and thermal diffusion. Should the contribution from a specific diffusivity type happen to be positive, it would imply that diffusivity type would act towards causing the DW modes to be overstable. However, an overstability is only reached if the total contribution \(\sum \gamma_i \left( \frac{\partial \lambda}{\partial \gamma_i} \right)\) is positive.

While the thermal diffusion also has a stabilizing contribution when \(\omega_0^2 > 0\) (where the flow is dynamically stable), for the shear viscosity, the contribution is a more complicated expression that depends on \(\gamma\) and \(q\), as well as \(k\) and \(Q\). However, it should be noted that the expression enclosed within square brackets in \(\frac{\partial \lambda}{\partial \gamma_e}\) is positive for most realistic value combinations of \(q\) and \(\gamma\). The regions where overstability occurs when only shear viscosity is taken into account are shown in Fig. 2 for a range of values of the adiabatic index \(\gamma\), assuming \(q = 3/2\).

\(\gamma = 1\) (blue, dotted region) and 7/5 (orange, dashed) produce an overstable region in the \(k\_c/\Omega-Q\) plane that extends to arbitrarily high \(Q\) for sufficiently large wavelengths (\(\gg 9H\) and \(\gg 16H\), respectively, where \(H = c_{iso}/\Omega\) is the scaleheight of the disc), as well as a
In the non-SG limit, the coefficient \((1 - 1/Q^2)\) → 1, reducing the overstability condition to
\[
\gamma < \frac{16q - 3q^2 - 14}{3q(2 - q)},
\]
with the critical value being \(\gamma_{\text{crit}} \approx 1.444\) in the \(q = 3/2\) case.

The next step is to combine the contributions of different diffusivity types using equation (32) to find the regions of the \(kc_s/\Omega-Q\) plane where overstability would occur. We take the instance in which the bulk viscosity contribution is ignored; in this case, we find that the system would develop an overstability if
\[
\frac{\gamma_s}{\gamma} < \frac{2\gamma^2}{\Omega^2} \left[ (\gamma - 1)q^2 + 2(2 - \gamma)q - 2 - \frac{2}{3} \frac{\omega_0^2}{\Omega^2} \right],
\]
which, assuming \(q = 3/2\) and \(\gamma = 5/3\), simplifies to
\[
\frac{\gamma_s}{\gamma} < \frac{5}{6k^2c_s^2} \left[ 4k^2c_s^2 - 2c_s\kappa/\Omega + \Omega^2 \right],
\]
where the \(\omega_0^2/\Omega^2\) factor has been expanded to obtain a relationship as a function of \(k\). This highlights the stabilizing effect played by thermal diffusion and cooling, with an overstability developing only if the ratio \(\gamma_s/\gamma\) is below a critical value, which is dependent on the values of \(k\) and \(Q\) (as well as \(q\) and \(\gamma\)). The cooling in particular plays a dominant role in the long-wavelength limit as its contribution to \(\gamma_s\) is independent of \(k\), while both shear and thermal diffusivities produce damping coefficients that are proportional to \(k^2\). This hampers the triggering of overstability that, as seen in Fig. 2, prefers the small \(k\) limit, particularly for the non-SG case.

An analysis of the \(k \to 0\) limit, also taking into account the coupling between cooling time-scale and shear viscosity given by the thermal balance (equation 20), yields the following expression for the real part of the growth rate:
\[
\text{Re}(\lambda) = \gamma_s \frac{\Omega^2}{6kc_s^2} \left[ -28 + 4(8 - 3\gamma)q + 6(\gamma - 1)q^2 - 3(\gamma - 1)q^2 \frac{1}{f_{\text{visc}}} \right].
\]

From the expression above, it is possible to infer that a non-SG overstability is indeed possible as long as the adiabatic index obeys \(\gamma_s \lesssim 1.305\) (assuming \(q = 3/2\) still), with the threshold value \(\gamma_s \approx 1.305\) obtained when the disc is fully viscously heated \((f_{\text{visc}} = 1)\). This represents a stricter constraint than that obtained for shear viscosity only (equation 36), again underlining the stabilizing effect of \(\gamma_s\).

This is illustrated in Fig. 3, where the area obeying \(\omega_0^2 < 0\) has been ignored as any instability in that region would be of a dynamical nature. A range of cooling times satisfying thermal balance is explored, with the largest value chosen so that the flow is almost entirely heated by viscous dissipation. The Prandtl number is set to \(Pr = 5\) with \(\alpha_s = 0.05\), for both \(\gamma = 1.3\) and 1.4. It is possible to notice that as the cooling is made more efficient the overstable area shrinks, confirming its stabilizing role, particularly in the long-wavelength regime; indeed for \(\gamma = 1.4\), the system is found to be stable for all values of \(kc_s/\Omega\) and \(Q\) (for which \(\omega_0^2 > 0\) for the shortest cooling time explored \((\tau_c = 5\Omega^{-1})\). The \(\gamma = 1.3\) case, on the other hand, presents overstability for all cooling times explored, as predicted by equation (39); for non-SG or weak-SG conditions, overstability is also observed for \(\gamma = 1.3\) in the long-wavelength limit for the two longest cooling times analysed: \(\tau_c = 20\) (green, full lines) and \(29\Omega^{-1}\) (red, dot–dashed). Non-SG overstability, which requires wavelengths longer than \(\sim 18H\) for \(\gamma = 1.3\), is on the other
Figure 3. Overstability regions under the influence of shear viscosity, thermal diffusion and cooling for $q = 3/2$ and (a) $\gamma = 1.3$ and (b) $\gamma = 1.4$ (or $\gamma = 7/5$). The analysis is carried out for a Prandtl number $Pr = 5$ (with $\alpha_s = 0.05$) and various cooling time-scales permitted by thermal balance. The $\gamma = 1.3$ case retains a weak-SG/non-SG overstability at long wavelengths for the two longest cooling times: $\tau_c = 20$ (green, full lines) and $29/\Omega_1$ (red, dot–dashed). This is, however, not the case for the plot with $\gamma = 1.4$, which only shows overstability for $Q \lesssim 2$, as the value of $\gamma$ used in this case is larger than the predicted threshold value of $\gamma \approx 1.305$. In both plots, it is possible to see that cooling has a stabilizing effect on the system, with shorter time-scales progressively shrinking the overstability region. The hatched portion of the plot represents the region where the DW frequency $\omega_2 < 0$ and the system is therefore gravitationally unstable to axisymmetric disturbances.

hand suppressed for $\gamma = 1.4$, with overstable regions being contained to $Q \lesssim 2$. This is in agreement with the analytical prediction described above, which found that a weak-SG/non-SG overstability in the $k \to 0$ limit could only be achieved if the value of the adiabatic index was below the threshold value $\gamma \approx 1.305$.

A general form for the largest overstable value of $Q$ attainable over all $k \tau_c / \Omega_1$ in the absence of the bulk viscosity contribution can be derived analytically and is found to be

$$\frac{1}{Q_{\text{max}}^2} = \frac{[4\gamma + 3(\gamma - 1)Pr^{-1}]}{32(2 - q)\gamma^2} \left[ 28 - 4(8 - 3\gamma)q - 6(\gamma - 1)q^2 \right] + 3(\gamma - 1)^2 q^2 \frac{1}{f_{\text{visc}}} \quad (40)$$

Assuming the sum of the first three terms enclosed in square brackets is positive (as otherwise the system might be overstable for any $Q$ and there would therefore not be a critical $Q$ value), $Q_{\text{max}}$ is found to be an increasing function of $f_{\text{visc}}$ and $Pr$. A particular example of equation (40) is illustrated in Fig. 4; this shows the overstability growth rates, maximized over $k$, as a function of the adiabatic index $\gamma$ and Toomre parameter $Q$ for $q = 3/2$, $\tau_c = 15 \Omega_1^{-1}$, $Pr = 5$ and $\alpha = 0.05$. While $\gamma$ values up to $\gamma \approx 1.6$ are overstable at $Q \sim 1$ for the given cooling time-scale, the maximum $\gamma$ value needed for overstability gradually decreases to $\gamma \lesssim 1.25$ as the Toomre parameter reaches $Q \sim 5$. This is in agreement with the predicted maximum value of $\gamma$ that allows weak-SG/non-SG overstability ($\gamma \approx 1.305$), which is indicated by means of a dotted vertical line.
in the plot. Moreover, a dashed vertical line at \( \gamma \approx 1.6 \) represents the largest value of \( \gamma \) allowed by thermal balance.

The introduction of the bulk viscosity in the analysis further complicates the overstability analysis, with the full form of the overstability criterion being

\[
y_\ast \leq -\frac{6k^2c_1^2\gamma_1 + 5(4\alpha_0^2 - 3)\Omega_1^2}{15\alpha_0^2} \gamma, \quad (41)
\]

where the assumptions of \( \gamma = 5/3 \) and \( q = 3/2 \) have been made.

### 3.2 Slow modes

The analysis of the slow PV and entropy modes, having coinciding and degenerate solutions in the inviscid problem with no cooling or diffusion, requires a somewhat different approach from the regular perturbation method used for DW modes, as their solutions are found to depend non-linearly on \( \gamma_5 \) and \( \gamma_7 \); this is exemplified in Fig. 5 for \( \gamma = 5/3, \omega_c/\Omega = 2 \) and \( Q = 1.2 \). The real parts of both modes’ growth rates present non-linearities in their behaviour; interferences between the modes – where their growth rates form a complex conjugate pair – can also be observed for \( \gamma \lesssim 0.03 \). One of the two modes is also seen to be unstable in a sizeable part of the plot. Combinations of \( \gamma_5 \) and \( \gamma_7 \) values falling below the dashed line do not satisfy thermal balance (equation 20).

In order to gain a better understanding on the stability of these two modes, equations for the evolution of the structure in the specific entropy and PV (i.e. \( \partial_t A_s \) and \( \partial_t A_c \)) of the form

\[
\begin{align*}
\partial_t A_s &= c_1 A_s + c_2 A_c, \\
\partial_t A_c &= c_3 A_s + c_4 A_c,
\end{align*}
\]

were analytically derived from equations (27)–(29), where \( A_s \) and \( A_c \) are the dimensionless amplitudes of the axisymmetric structure in the respective quantities given by

\[
A_s = \frac{1}{\gamma} (A_v + A_b),
\]

\[
A_c = \frac{k A_b}{(2 - q)\Omega} - A_b,
\]

and \( c_1, c_2, c_3 \), and \( c_4 \) are coefficients that are independent of \( A_s, A_v, A_b \) and \( A_c \). The coefficients are found to be

\[
\begin{align*}
c_1 &= \frac{\gamma_1 (c_1^2k^2(\gamma - 1) - \gamma_1^2\alpha_0^2)}{\gamma_0^2}, \\
c_2 &= \frac{\kappa^2(\gamma - 1)\left(\frac{\alpha_0^2}{\gamma_0^2} + \frac{q\alpha_0^2}{\gamma_0^2}\left(k^2 - \alpha_0^2\right)\right)}{c_1^2\alpha_0^2}, \\
c_3 &= -4\gamma_1 c_1^2k^2(q - 1)\Omega^2, \\
c_4 &= -\gamma_0 \frac{(\alpha_0^2 + 4(q - 1)\Omega^2)}{\kappa^2\alpha_0^2}.
\end{align*}
\]

Assuming the solutions have an exponential form, a generic quadratic equation for the growth rate \( \lambda \) for the system described in equations (42)–(43) can be simply derived as

\[
\lambda^2 - (c_1 + c_2)\lambda + c_1 c_4 - c_2 c_3 = 0,
\]

with a generic solution being given by

\[
\lambda = \frac{(c_1 + c_4)}{2} \pm \sqrt{\frac{(c_1 - c_4)^2}{4} + c_2 c_3}.
\]

The regions of the \( \omega_c/\Omega - Q \) space where the system is unstable to slow modes can be found by either looking for areas where \( \text{Re}(\lambda) > 0 \) or by applying a relevant stability condition. This was found in the Routh–Hurwitz stability criteria, which represent necessary and sufficient stability conditions for a linear time-invariant system with a polynomial characteristic equation. The required stability condition in the case of a generic second-order polynomial of the form \( x^2 + a_1 x + a_0 = 0 \) is for all coefficients to satisfy \( a_1 > 0 \); in the particular instance of equation (50), this can be written as

\[
a_1 = -(c_1 + c_4) > 0,
\]

\[
a_0 = c_1 c_4 - c_2 c_3 > 0.
\]

Stability is achieved only if both of these conditions are satisfied.

If the coupling coefficients \( c_2 \) and \( c_3 \) are negligible compared to \( c_1 \) and \( c_4 \), entropy and PV evolve independently from each other,
with $c_1$ and $c_4$ representing the two quantities’ respective growth rates. Such is the case in both long-wavelength (i.e. $kc_1/\Omega \to 0$) and short-wavelength (i.e. $kc_1/\Omega \to \infty$) limits, the former being stable according to the classical approach. In these cases, the product of $c_1$ and $c_4$ – both coefficients being negative – dominates over the coupling product term $c_2c_3$; this means that both Routh–Hurwitz stability criteria are satisfied and the system is stable. The analysis presented in this paper focuses on the stability of the intermediate $kc_1/\Omega$ range, instead; this is somewhat more difficult to predict analytically as $c_2$ and $c_3$ are no longer negligible, meaning PV and entropy are coupled. This also implies that, should both $c_1$ and $c_4$ be negative under certain conditions, the system can nevertheless still be unstable by violating the $c_1c_4 - c_2c_3 > 0$ condition.

It is worth pointing out that the properties of the model used do affect the stability of the flow; should the $+q\Omega\gamma k^{-1}A_0$ term in equation (28) – which arises from the dynamic viscosities being linear functions of $\Sigma$ – be removed, the system would then be unstable to secular gravitational instability (Willering 1992; Gammie 1996). This occurs in the limit $kc_1/\Omega \to 0$ in systems that are marginally stable according to equation (7). However in the case analysed in this work, the system is stable to the onset of secular gravitational instability.

Fig. 6 illustrates the regions in the $kc_1^2/\Omega - \Omega$ plane where either (or both) of the stability conditions is not satisfied and the system is therefore unstable; the same instability regions are also obtained when looking for parts of the plane where $\text{Re}(\lambda) > 0$, therefore validating the instability criteria used. The analysis is carried out for a range of cooling times satisfying thermal balance (where again the largest value is such that $\epsilon_{\text{th}} \simeq 0$) with $q = 3/2$ and $\gamma = 5/3$ and for $Pr = 3$ (Fig. 6a) and $Pr = 1$ (Fig. 6b). The value of the shear viscosity is kept at $\alpha_s = 0.05$ throughout. Non-monotonic behaviour in the instability regions is observed thanks to a peak at $kc_1/\Omega \sim 2.5 - 3$ for $Pr = 3$, which is most prominent for $Pr = 3$ but is quenched as the Prandtl number decreases to unity. The overall region of instability also shrinks with decreasing $Pr$, highlighting the stabilizing effect of the thermal diffusion. A short cooling time seems to lightly boost instability at $kc_1/\Omega \sim 1$, but at the same time it appears to dampen the instability at $kc_1/\Omega \sim 2.5$–3. All instability regions seem to prefer intermediate $kc_1/\Omega$ values, ensuring the instability is again very relevant to the stability of zonal flows. The effect the value of the cooling time-scale has on the stability of the system appears to wane with decreasing Prandtl number, with the $Pr = 1$ case presenting a reduced difference between the $\tau_c = 3$ and $12\Omega^{-1}$ cases.

Fig. 7 represents a similar analysis to Fig. 6, but this time with the adiabatic index set to $\gamma = 2$. The increased value of $\gamma$ causes a boost in both peaks compared to the $\gamma = 5/3$ case, particularly the one located at $kc_1/\Omega \sim 2.5$–3. Once again, this latter peak is suppressed as the Prandtl number is decreased with the non-monotonic behaviour mostly suppressed for $Pr = 1$. Also, as seen in Fig. 6, the use of an effective cooling has the effect of

$^1$ Although the value $\gamma = 2$ bears questionable physical relevance, this has regularly been adopted in works of SG accretion discs since the seminal analysis by Gammie (2001). It is therefore useful in comparing our results to the relevant literature.
boosting the first peak (the one at \( k_c/\Omega \sim 1 \)), while quenching the second one.

The nature of the instability region is explored in Fig. 8 with \( \gamma = 2 \), \( \tau_c = 8.5 \, \Omega^{-1} \) and \( \Pr = 3 \) and 1 in Figs 8(a) and (b), respectively. The total unstable area is divided into the regions where each of the stability conditions given in equations (52a)-(52b) is violated. The first peak, located at \( k_c/\Omega \sim 1 \), is due to the \(- (c_1 + c_4) > 0 \) stability condition being violated and it therefore represents, as suggested by equation (51), an oscillatory instability. As \( c_1 < 0 \) (assuming \( \omega^2 > 0 \)), regardless of the values of \( Q \) or \( k_c/\Omega \), the unstable contribution must come from \( c_1 \), meaning that region is caused by an instability in the entropy; this is therefore a thermal instability. On the other hand, the second peak, found at \( k_c/\Omega \sim 2.5-3 \), is triggered by the second condition not being fulfilled (i.e. we therefore have \( c_1c_4 - c_2c_3 < 0 \)), implying the instability here has a non-oscillatory behaviour; this peak is therefore either due to the action of entropy or PV (orange, dashed region; \( c_1c_4 < 0 \)) or due to their coupling (green, dot–dashed region, \( -c_2c_3 < 0 \)), as seen in equations (42)-(43). The comparison between Figs 8(a) and (b) shows that decreasing the Prandtl number results in the quenching of the coupling’s destabilizing effect, with said coupling mostly driving the instability at \( k_c/\Omega \sim 2.5-3 \) for \( \Pr = 3 \) but it being largely suppressed in the \( \Pr = 1 \) case. A small boost of the entropy-driven instability is also observed upon decreasing \( \Pr \).

Fig. 9 shows the growth rates of the instability region, which have been maximized over \( k \), as a function of \( \alpha_s \) and \( Q \); these are obtained for \( \gamma = 2 \), \( \tau_c = 8.5 \, \Omega^{-1} \) and a fixed Prandtl number of \( \Pr = 3 \) (Fig. 9a) and = 1 (Fig. 9b). All values of \( \alpha_s \) used are allowed by thermal balance for the given cooling time-scale, with small \( \alpha_s \) values indicating the disc is predominantly heated by external irradiation, while the maximum explored value of \( \alpha_s = 0.05 \) means the disc is almost completely heated by viscous effects. The plot shows that while the value of the Prandtl number is of importance for the stability of the system, the value of \( \alpha_s \) – and therefore the source of internal energy – also affects the maximum value of \( Q \) at which the instability is observed. Indeed for \( \Pr = 3 \) the system is unstable up to \( Q \sim 2.7 \) for \( \alpha_s = 0.05 \) (viscously heated disc), but only up to \( Q \sim 1.4 \) when \( \alpha_s \lesssim 0.02 \) (external irradiation contributing at least as much as viscous effects). The \( \Pr = 1 \) case, on the other hand, presents little variation in \( Q \) over the diffusivity range, although a similar qualitative behaviour is observed.

The dependence of the \( k \)-maximized growth rates on \( Q \) and the cooling time \( \tau_c \) for \( \alpha_s = 0.05 \) and \( \Pr = 3 \) is instead explored in Fig. 10. This shows that while for most of the \( Q \) range shortening the cooling time has a stabilizing effect on the system, due to the peak at \( k_c/\Omega \sim 2.5-3 \) being quenched as seen in Figs 6 and 7, the situation is reversed for \( Q \lesssim 1.25 \). This is caused by the peak observed at \( k_c/\Omega \sim 1 \), which possesses a thermal nature as seen in Fig. 8, being instead boosted by efficient cooling.

**4 CONCLUSIONS**

We carried out an analytical calculation on the evolution of a viscous and compressible SG Keplerian disc having a constant cooling time-scale and horizontal thermal diffusion with an axisymmetric structure present in the analysed quantities. The analysis took into
Figure 8. Analysis showing which of the two stability criteria is violated in the instability regions obtained in Fig. 7 for $\gamma = 2$, $\tau_c = 8.5 \Omega^{-1}$ and (a) $Pr = 3$ and (b) $Pr = 1$. Since $c_4 < 0$ as long as $c_2^2 > 0$ the first peak (blue, full line) is caused by an instability in the entropy ($c_1$), meaning it has a thermal nature. The second peak is due to the second instability criterion being fulfilled, with it being split among its two components. In (a), this is predominantly driven by the coupling term between entropy and PV (i.e. $c_1 c_4 < 0$, green dot–dashed); in (b), the decreased Prandtl number $Pr$ quenches the coupling component almost completely with the $c_1 c_4 < 0$ (yellow, dashed) mostly causing the instability, meaning this is driven by either PV or entropy.

Figure 9. Instability growth rates maximized over $k$ as a function of the shear diffusivity $\alpha_s$ and Toomre parameter $Q$ for $q = 3/2$, $\gamma = 2$, $\tau_c = 8.5 \Omega^{-1}$ and a fixed Prandtl number of (a) $Pr = 3$ and (b) $Pr = 1$. Although the Prandtl number larger than unity remains a critical factor in boosting the instability, it is clear that the value of $\alpha_s$ is also of importance. For the smaller values of $\alpha_s$ plotted here, the disc is heated predominantly by external irradiation, while for the larger values of $\alpha_s$ it is mostly heated by viscous dissipation. For $Pr = 3$ no instability is seen above $Q \sim 1.4$ for $\alpha_s \lesssim 0.02$, although for larger values of $\alpha_s$ the instability spreads up to $Q \sim 2.7$; this points to the instability being boosted by a disc being viscously heated. The largest unstable $Q$ value is instead roughly constant in the $Pr = 1$ case.
account all solutions of the problem: both the DW modes and the PV and entropy slow modes.

While the solutions to the system are well known in the inviscid case with no cooling or thermal diffusion, the introduction of three types of diffusivity (bulk and shear viscosities and thermal diffusion) and cooling created a non-trivial problem in pinpointing whether they would have a stabilizing or destabilizing effect on the system. A simplification can be made for the DW modes, as their growth rates are found to be a linear function of each type of diffusivity used (regular perturbation method); this allowed us to individually derive the contribution from each diffusivity type to the final growth rate. These contributions can then be summed together to establish the actual growth rate of the modes. While the bulk and thermal diffusivities were found to always have a stabilizing effect, the situation was somewhat more complex for the shear viscosity. Ignoring the contribution made by the bulk viscosity, the system was found to be overstable for intermediate and long wavelengths for Toomre parameter values of \( Q \lesssim 2 \), although a weak-SG/non-SG overstability was also detected in the long-wavelength regime for inefficient cooling as long as the adiabatic index \( \gamma \lesssim 1.305 \). In the case of \( \gamma = 1.3 \), the system is overstable for non-SG conditions for wavelengths longer than roughly 18H. These results appear consistent with those by Latter & Ogilvie (2006) in the simplified 2D version of their calculation, although their work did not present any \( \gamma \) dependence due to the lack of thermal heating modulations in the azimuthal velocity equation. The \( k \)-maximized growth rates for overstability regions were plotted as a function of adiabatic index and \( Q \); while a sizeable range of \( \gamma \) values presented overstability for \( Q \sim 1 \), this gradually reduced as \( Q \) was increased. Only values obeying \( \gamma \lesssim 1.25 \) were found to be overstable in weak SG conditions for \( Q \sim 5 \), which is in agreement with the predicted threshold of \( \gamma \lesssim 1.305 \). Overstability criteria for shear and thermal diffusivities only and for all three diffusivity types were also derived, which highlight the stabilizing effect of thermal diffusivity in the weak-SG regime.

The situation was more complex for the entropy and PV slow modes as their degenerate solutions in the inviscid case with no cooling were found not to follow the regular perturbation method. In order to obtain their growth rates, equations for the evolution of the axisymmetric structure in these two quantities – which only depended on the structure’s amplitude in the entropy and PV themselves – were derived. The Routh–Hurwitz stability criteria, representing the conditions for which a linear time-invariant system with a polynomial characteristic equation is stable, were applied to the generic solution to these equations. The long- and short-wavelength limits, which are stable according to the classical stability analysis, were likewise found to be stable. Nevertheless, the flow was found to be unstable in the intermediate wavelength regime, in a clear extension to the classical approach. This instability was found to be aided by considering higher values of the adiabatic index and of the Prandtl number and by decreasing the values of the Toomre parameter, although it was also of importance whether the disc was heated by external irradiation or viscous effects. Efficient cooling, on the other hand, was found to have an overall stabilizing effect on the instability as long as \( Q \gtrsim 1.25 \). It is believed that this kind of instability – due to its tendency to operate at intermediate wavelengths – might result, in the appropriate conditions, in the formation of zonal flows; these might themselves be unstable, potentially giving rise to vortices in the flow. Further work is, however, required to obtain a more detailed link between the instability and the potential development of zonal flows.

**ACKNOWLEDGEMENTS**

We would like to thank the reviewer for providing a constructive set of comments. The research was conducted thanks to the funding received by the Science and Technology Facilities Council (STFC).

**REFERENCES**

Bai X.-N., Stone J., 2014, ApJ, 796, 31
Blumenthal G., Yang L., Lin D., 1984, ApJ, 287, 774
Borderies N., Goldreich P., Tremaine S., 1985, Icarus, 63, 406
Done C., Gierliński M., Kubota A., 2007, A&A, 15, 1
Fromang S., Nelson R., 2005, MNRAS, 364, L81
Gammie C., 1996, ApJ, 462, 725
Gammie C., 2001, ApJ, 553, 174
Gierliński M., Done C., 2004, MNRAS, 347, 885
Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 97
Hirose S., Krolik J., Blaes O., 2009, ApJ, 691, 16
Johansen A., Youdin A., Klahr H., 2009, ApJ, 697, 1269
Johnson B., Gammie C., 2005, ApJ, 635, 149
Kato S., 1978, MNRAS, 185, 629
Kato S., Fukue J., 1980, PASJ, 32, 337
Kato S., Honma F., Matsumoto R., 1988, MNRAS, 231, 37
Kato M., Nakamura K., Tandokoro K., Fujimoto M., Ida S., 2009, ApJ, 691, 1697
Klahr H., Lin D., 2001, ApJ, 554, 1095
Kunz M., Lesur G., 2013, MNRAS, 434, 2295
Latter H., Ogilvie G., 2006, MNRAS, 372, 1829
Lightman A., Eardley D., 1974, ApJ, 187, L1
Lithwick Y., 2007, ApJ, 670, 789
Lithwick Y., 2009, ApJ, 693, 85
Livio M., Shaviv G., 1977, A&A, 55, 95
Merloni A., 2003, MNRAS, 341, 1051
Papaloizou J., Lin D., 1988, ApJ, 331, 838
Papaloizou J., Stanley G., 1986, MNRAS, 220, 593
Piran T., 1978, ApJ, 221, 652
Pringle J., 1977, MNRAS, 177, 65

---

Figure 10. Instability growth rates, maximized over k, as a function of Q and \( \tau_c \) for \( q = 3/2 \), \( \gamma = 2 \), a Prandtl number of \( Pr = 3 \) and \( a_s = 0.05 \). For most of the Q range, making the cooling time-scale shorter has a stabilizing effect on the system, as the peak at \( k c_s/\Omega \sim 2.5–3 \) in Figs 6 and 7 is quenched; for \( Q \lesssim 1.25 \), the trend however reverses for very short cooling times, which is due to the peak at \( k c_s/\Omega \sim 1 \), having a thermal nature, being boosted.
This paper has been typeset from a \TeX/\LaTeX file prepared by the author.