The Astrophysical Journal, 983:89 (9pp), 2025 April 20 https://doi.org/10.3847/1538-4357/adc0ff

Gravitational Instability in Protoplanetary Disk with Cooling: 2D Global Analysis

Zehao Su and Xing Wei
IFAA, School of Physics and Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China; xingwei@bnu.edu.cn
Received 2025 February 22; revised 2025 March 13; accepted 2025 March 14; published 2025 April 10

Abstract

Self-gravity is important in protoplanetary disks for planet formation through gravitational instability (GI). We study the cooling effect on GI in a thin two-dimensional protoplanetary disk. By solving the linear perturbation equations in global geometry, we obtain all the normal modes. Faster cooling leads to faster growth rate of GI with lower azimuthal wavenumber m. According to the spatial structure of normal modes at different mass-averaged Toomre number $\overline{Q}$ and dimensionless cooling timescale β, we identify three modes: local, transitional, and global. The transitional modes are located in the outer disk, while the other two modes in the inner disk. At $\beta \approx 1$ (the resonance of dynamical timescale and thermal timescale) the growth rate changes sharply and the transitional modes dominate. The disk α due to GI is much higher in the transitional modes than in the other two. Our results imply that the transitional modes at $\overline{Q} \approx 1$ and $\beta \approx 1$ can plausibly interpret the substructures and planet/brown dwarf formation in the outer disk.

Unified Astronomy Thesaurus concepts: Protoplanetary disks (1300); Gravitational instability (668)

1. Introduction

Gravitational instabilities (GI) play a significant role in both the evolution of protoplanetary disks (PPDs) and the planet formation process. (e.g., A. G. W. Cameron 1978; A. P. Boss 1997; R. H. Durisen et al. 2007; A. C. Boley & R. H. Durisen 2010; S.-J. Paardekooper 2012; W. Xu & M. W. Kunz 2021). They likely manifest during the early stages of protostellar disk evolution when the disk mass occupies greater than 10% of the protostellar mass (F. C. Adams & D. N. C. Lin 1993; P. J. Armitage 2020). Disk mass estimates suggest that 50% of Class 0 and 25% of Class I disks exhibit GI instability (K. Kratter & G. Lodato 2016).

The observational evidence suggests that GI may play a substantial role in PPDs. Recent advancements in observational facilities and instruments, such as the Atacama Large Millimeter/submillimeter Array (ALMA), the Very Large Telescope/Spectro-Polarimetric High-contrast Exoplanet REsearch (VLT/SPHERE), the Very Large Telescope/CRyogenic high-resolution InfraRed Echelle Spectrograph (VLT/CRIRES), and the Gemini Planet Imager (GPI), have enabled spatially resolved observations of PPDs. Spiral structures have been detected in the submillimeter thermal emission reflected by the millimeter-sized dust around the disk midplane (e.g., L. M. Pérez et al. 2016; J. Huang et al. 2018; N. T. Kurtovic et al. 2018). These observations are biased tracers of gas-density profiles due to finite aerodynamic coupling between gas and dust. Moreover, optical and near-infrared (NIR) high-contrast, scattered-light images, arising from starlight scattered by μm-sized dust suspended in the disk surface, have also revealed spiral structures (e.g., A. Garufi et al. 2013; K. Wagner et al. 2015; H. Avenhaus et al. 2017; M. Benisty et al. 2017; T. Uyama et al. 2018). The origin of these spirals remains a debate, with GI offering a plausible explanation (L. M. Pérez et al. 2016; F. Meru et al. 2017; K. Tomida et al. 2017; C. Hall et al. 2018; J. Huang et al. 2018; J. Speedie et al. 2024). Additionally, the presence of substellar companions at large radii (with semimajor axes $a > 10$ au) implies the importance of GI (e.g., K. M. Kratter et al. 2010; D. Forgan & K. Rice 2013; D. Forgan et al. 2015).

The study of GI has a long history, dating back to James Jeans. Regarding disk GI, in recent years, J. Goodman & R. Narayan (1988) explored the fundamental modes in self-gravitating incompressible tori, demonstrating the existence of I mode (with corotation outside the tori) and J mode (with corotation at the density maximum). F. C. Adams et al. (1989) showed that one-armed density waves can be unstable when the disk mass is sufficiently high ($M_d/M_* = 1$). H. Noh et al. (1991) found nonaxisymmetric instability of high m's when the disk is not so massive ($M_d = 0.05 \sim 0.5 M_\odot$). K. Z. Hadley et al. (2014) studied the global nonaxisymmetric instabilities in a thick disk to produce a large parameter space of star–disk system. G. Lodato & W. K. M. Rice (2005) revealed that sufficiently massive disks can exhibit short-lived arms. The series of studies (G. Laughlin & M. Rozyczka 1996; R. Dong et al. 2015; C. Hall et al. 2019; E. Chen et al. 2021) demonstrated that GI-induced spirals can have more than two arms. In addition to classical GI, drag-mediated, two-fluid GI (A. Coradini et al. 1981; C. Longarini et al. 2023) and secular GI (W. R. Ward 2000; A. N. Youdin 2005; K. Shariff & J. N. Cuzzi 2011; S. Michikoshi et al. 2012; T. Takeuchi & S. Ida 2012; R. T. Tominaga et al. 2020; A. Pierens 2021) have been proposed for less-massive disks.

The dynamical consequences of GI are strongly influenced by disk thermodynamics. When a disk becomes gravitationally unstable, GI-induced spirals spread across the disk, forming shocks that heat the disk to increase thermal pressure against GI. However, radiative cooling can help remove heat from the disk, making it more unstable. The final state is determined by the balance between heating and cooling. When the cooling timescale is shorter than the dynamical timescale, fragmentation may occur, leading to the formation of gravitationally bound objects (e.g., A. P. Boss 2001; C. F. Gammie 2001; B. M. Johnson & C. F. Gammie 2003; W. K. M. Rice et al. 2005; H. Baehr et al. 2017). When the cooling timescale is larger than the dynamical timescale, the disk may settle into a quasi-steady, marginally unstable state (e.g., P. Goldreich & D. Lynden-Bell 1965; D. N. C. Lin & J. E. Pringle 1987;

C. F. Gammie 2001; G. Lodato & W. K. M. Rice 2004; A. C. Boley et al. 2006; S. Michael et al. 2012; T. Y. Steiman-Cameron et al. 2013), potentially sustaining long-lived substructures in the disk.

Motivated by the relationship of disk thermodynamics and GI, we perform a linear stability analysis to study the effects of cooling time for disk GI. In Section 2 we derive the linear perturbation equations and give the numerical methods. In Section 3 we show some representative modes in different GI regimes. In Section 4 we focus on the relationship of dominant modes and disk-cooling timescales. In Section 5 we discuss some implications and caveats, especially the disk α. In Section 6 and we summarize our results.

2. Methods

2.1. Linear Perturbation Equations

We consider an inviscid two-dimensional gas disk with self-gravity in cylindrical coordinates (r, ϕ). The equations of mass, momentum, internal energy, and self-gravity are given:

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r}\frac{\partial (r\Sigma u_r)}{\partial r} + \frac{1}{r}\frac{\partial (\Sigma u_\phi)}{\partial \phi} = 0, \tag{1}$$

$$\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + \frac{u_\phi}{r}\frac{\partial u_r}{\partial \phi} - \frac{u_\phi^2}{r} = -\frac{GM_*}{r^2} - \frac{1}{\Sigma}\frac{\partial P}{\partial r} - \frac{\partial \Psi}{\partial r}, \tag{2}$$

$$\frac{\partial u_\phi}{\partial t} + u_r\frac{\partial u_\phi}{\partial r} + \frac{u_\phi}{r}\frac{\partial u_\phi}{\partial \phi} + \frac{u_\phi u_r}{r} = -\frac{1}{r}\frac{\partial P}{\partial \phi} - \frac{1}{r}\frac{\partial \Psi}{\partial \phi}, \tag{3}$$

$$\frac{de}{dt} + P\frac{d}{dt}\left(\frac{1}{\Sigma}\right) = \left(\frac{\partial e}{\partial t}\right)_{\rm cool}, \tag{4}$$

$$\nabla^2 \Psi_d = 4\pi G\Sigma\delta(z), \tag{5}$$

where $\Sigma, P, e, u_r, u_\phi$ represent, respectively, surface density, pressure, internal energy, radial, and azimuthal velocities. Here Ψ_d denotes the self-gravity potential of the disk and $\delta(z)$ is the Dirac function. The thermodynamic effects are implemented on the right-hand side of Equation (4) using the same strategy as in R. Miranda & R. R. Rafikov (2020), with e relaxed to a prescribed unperturbed value $e_0(r) = c_{\rm s,adi}^2(r)/[\gamma(\gamma-1)]$ (γ: adiabatic index, $c_{\rm s,adi}$: adiabatic sound speed) on a cooling timescale $t_{\rm c}$

$$\left(\frac{\partial e}{\partial t}\right)_{\rm cool} = -\frac{e - e_0}{t_{\rm c}}. \tag{6}$$

Here, t_c is modeled with the standard β cooling approximation (C. F. Gammie 2001)

$$t_{\rm cool} = \beta\Omega_K^{-1} \tag{7}$$

where Ω_K is the Keplerian angular velocity.

We describe the physical quantities, including $\Sigma, P, e, u_r, u_\phi, \Psi_d$, in a general form, $X = X_0 + \delta X$, where the subscript X_0 denotes equilibrium and δX perturbation. In the following text, we drop the subscript "0" of the unperturbed variables for simplicity. The perturbations are assumed to be Fourier harmonics

$$\delta X(r, \phi, t) = \delta X(r)\exp[i(\omega_m t - m\phi)]. \tag{8}$$

Here, m represents the azimuthal wavenumber, and ω_m is the complex wave frequency that can be written as $\omega_m = m\omega_p - i\gamma_m$, where ω_p is the oscillation rate (or pattern speed) and γ_m the growth rate. Substituting Equation (8) into Equations (1)–(5) while keeping only the first-order terms yields the linear perturbation equations

$$-i\tilde{\omega}\delta\Sigma - \frac{1}{r}\frac{\partial}{\partial r}(r\Sigma\delta u_r) + \frac{im\Sigma}{r}\delta u_\phi = 0, \tag{9}$$

$$-i\tilde{\omega}\delta u_r + 2\Omega\delta u_\phi = \frac{1}{\Sigma}\frac{\partial}{\partial r}\delta P - \frac{1}{\Sigma^2}\frac{dP}{dr}\delta\Sigma + \frac{\partial}{\partial r}\delta\Psi_m, \tag{10}$$

$$-i\tilde{\omega}\delta u_\phi - \frac{\kappa^2}{2\Omega}\delta u_r = -\frac{im}{r}\left(\frac{\delta P}{\Sigma} + \delta\Psi_m\right), \tag{11}$$

$$-\left(\frac{1}{t_{\rm c}} + i\tilde{\omega}\right)\delta P + \left(\frac{1}{\gamma t_{\rm c}} + i\tilde{\omega}\right)c_{\rm s,adi}^2\delta\Sigma = \frac{\Sigma c_{\rm s,adi}^2}{L_S}\delta u_r \tag{12}$$

$$\delta\Psi_m(r) = -\int_{r_d}^{r_0} d\rho \int_{2\pi}^{0} \frac{G\cos(m\phi)\delta\Sigma(\rho)\rho}{\sqrt{\rho^2 + r^2 - 2\rho r\cos\phi}} d\phi. \tag{13}$$

Here, $\tilde{\omega} = \omega_m - m\Omega$ is the Doppler-shifted frequency of the perturbation where $\Omega(r)$ is the angular velocity of the unperturbed state. L_S is the length scale of entropy variation, defined as $1/L_S = (1/\gamma)(dS/dr)$, where S is the gas entropy and the adiabatic index γ is given to be $7/5$. The perturbed self-gravitational potential is rewritten in its integral form (F. H. Shu 1992) and the range $[r_0, r_d]$ will be discussed in the next subsection. The angular velocity of unperturbed state can be derived from the radial equilibrium

$$\Omega = \sqrt{\frac{1}{r}\left(\frac{GM_*}{r^2} + \frac{1}{\Sigma}\frac{dP}{dr} + \frac{d\Psi}{dr}\right)}, \tag{14}$$

with the epicyclic frequency

$$\kappa^2 = \frac{1}{r^3}\frac{d}{dr}[(r^2\Omega)^2]. \tag{15}$$

2.2. Disk Profiles

As is known, self-gravity becomes crucial, and GI can be triggered when M_d/M_* exceeds 0.1 (R. Dong et al. 2015; K. Kratter & G. Lodato 2016). To specifically investigate GI in the early stages of disks, we deliberately choose a relatively substantial disk–star mass ratio of $M_d/M_* = 0.4$. Usually the Toomre number $Q = \kappa c_s/\pi G\Sigma$ is used as the criterion of GI, i.e., a disk is susceptible to gravitational instability when $Q < 1$ (A. Toomre 1964). In our global analysis we employ the mass-averaged Toomre number $\overline{Q} = \int Q\Sigma dS/\int\Sigma dS$ rather than its local definition. We ensure $\overline{Q} \lesssim 2$ to trigger the nonaxisymmetric GI (E. Chen et al. 2021).

The exact profile of PPDs in their early stages remains elusive, despite advancements in observational facilities like ALMA and VLT. We acknowledge the challenge in constraining these profiles and for simplicity adopt straightforward and interpretable profiles for both surface density (Σ) and temperature (T).

We assume the surface density profile

$$\Sigma(r) = C_\Sigma f_{\rm tap}(r) r^{-p} f_{\rm trunc}(r) \tag{16}$$

where the constant C_Σ is determined with disk mass. To suppress the abrupt change of the self-gravitational potential at the inner and outer boundaries of the disk that will induce the edge modes, we apply tapering ($f_{\rm tap}$) and truncation ($f_{\rm trunc}$) functions.

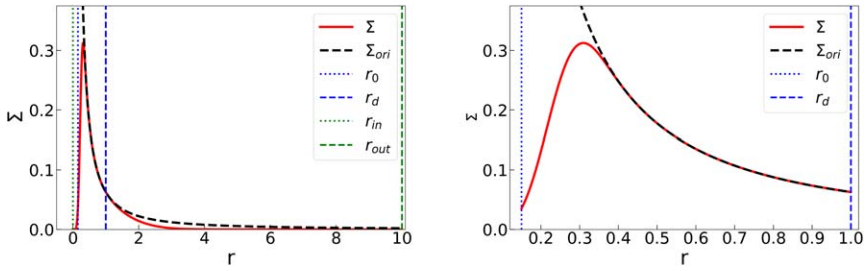

Figure 1. Surface density profiles within $[r_{\rm in}, r_{\rm out}]$ (left panel) and $[r_0, r_d]$ (right panel). The red solid line represents the density profile used in our calculation, while the black dashed line shows the profile $\Sigma_{\rm ori} \propto r^{-1.5}$. The green dotted–dashed lines indicate the positions of $r_{\rm in}$ and $r_{\rm out}$, respectively. The blue dotted–dashed lines denote the locations of r_0 and r_d, respectively.

The tapering function $f_{\rm tap}$ is a polynomial with its order higher than 10 defined to be $\in [0, 1]$ at the origin and the inner boundary r_0 (I. Backus & T. Quinn 2016). The truncation function is defined for $r > r_d$

$$f_{\rm trunc}(r) = e^{-(r-r_d)^2/L^2} \tag{17}$$

where r_d is the outer boundary and $L = 0.15 r_d$. Both the tapering and truncation functions are chosen to ensure that the surface density decreases rapidly when r extends beyond the calculation domain $r \in [r_0, r_d]$. They also smoothly approach 0 near $r = 0$, $r_{\rm out}$ (where $r_{\rm out}$ is the outer boundary used to calculate the equilibrium profile), while maintaining Σ and $d\Sigma/dr$ unchanged at $r = r_0$, r_d. The power-law index p for the surface density is set to 1.5. Figure 1 illustrates the density profile of our equilibrium state.

We assume a power-law temperature profile

$$T(r) = C_T r^{-q} \tag{18}$$

where the normalization factor C_T is determined by $\overline{Q}$. We choose $q = 0.6$, which is an intermediate value between $q = 3/7$ given by the balance of disk cooling and stellar irradiation and $q = 3/4$ given by the balance of disk cooling and accretion heating (K. Kratter & G. Lodato 2016).

2.3. Boundary Conditions

The perturbation equations consist of two first-order differential equations so that we need two boundary conditions. For the inner boundary, we use the reflecting boundary condition, i.e., the perturbed radial velocity vanishes when approaching the inner boundary of disk

$$\delta u_r = 0 \quad \text{at} \quad r = r_0. \tag{19}$$

For the outer boundary, we use the confining boundary condition, i.e., the Lagrangian perturbation of pressure vanishes at the outer boundary of disk

$$im\Omega\delta P - \frac{dP}{dr}\delta u_r = i\omega_m \delta P \quad \text{at} \quad r = r_d. \tag{20}$$

We refer these boundary condition as the "RC" boundary. We also test the other boundary conditions "RR," "CC," and "CR," and find that the boundary conditions cannot substantially change eigenfrequencies and eigenfunctions in the disk interior.

2.4. Numerical Method

We employ two different ranges in our analysis. The equilibrium is within $[r_{\rm in}, r_{\rm out}]$ to avoid self-gravitational divergence at the boundaries and obtain a consistent rotation profile, whereas the perturbation in $[r_0, r_d]$, with $r_{in} < r_0 < r_d < r_{\rm out}$. In other words, the disk range is $[r_0, r_d]$, and the regions between $[r_{\rm in}, r_0)$ and $(r_d, r_{\rm out}]$ serve solely as ghost ranges to obtain the equilibrium profile. The system is normalized with $G = M_* = \Omega_k|_{r=r_d} = 1$. We set $r_{\rm in} = 0$, $r_0 = 0.15$, $r_d = 1$, and $r_{\rm out} = 10$ in our calculation.

We translate the integro-differential Equations (9)–(13) to an eigenvalue problem, which has been widely used for searching normal modes in PPDs (F. C. Adams et al. 1989; H. Noh et al. 1991). We give a brief introduction here; the details are in the Appendix. The eigenvalue problem is

$$(\boldsymbol{M_1} + \boldsymbol{M_2})\boldsymbol{X} = \boldsymbol{i}\omega_{\boldsymbol{m}}\boldsymbol{X} \tag{21}$$

where $\boldsymbol{X}$ represents the eigenvector and $\boldsymbol{i}\omega_{\boldsymbol{m}}$ the eigenvalue. $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ are, respectively, the coefficient matrix for (9)–(13) without self-gravity terms and the coefficient matrix with only $\delta\Psi_m$ terms. We focus on the low $m \leqslant 5$, as GI is favored by low azimuthal wavenumbers and the multiple-arm spirals (>2) are not often observed (J. Bae et al. 2023). Only the modes with the highest growth rate are selected, as they dominate the other modes in the linear regime.

We use the logarithm-spaced grids within $[r_0, r_d]$ and take the resolution 800. To test the numerical convergence we double the resolution and the relative error of eigenfrequency is less than 2%.

3. Three Regimes

In this section we test with the mass-averaged Toomre numbers $\overline{Q} = 0.53$, 1.15, 1.65 across the entire range of β and show some representative results of GI modes. We focus on $m = 2$ modes because the two-armed spirals are commonly observed in both millimeter continuum (S. M. Andrews et al. 2018; J. Huang et al. 2018; N. T. Kurtovic et al. 2018) and near-infrared scattered-light observations (A. Garufi et al. 2013; M. Benisty et al. 2015, 2017; H. Canovas et al. 2018; T. Uyama et al. 2018). To characterize the modes, we compare the morphology of eigenfunctions, the oscillation rate (or pattern speed) ω_p, the growth rate γ_m, and the maximum location of $|\delta\Sigma|$ ($r_{|\delta\Sigma|_{\rm max}}$) under different $\overline{Q}$ and β. Additionally, we provide the locations of the inner and outer Lindblad resonances within which the spiral density wave propagates. The eigenvalues (oscillation frequencies and growth rates) are presented in Figures 2, 3 and 4, while the eigenfunctions (1D and 2D density perturbations) are shown in Figures 5, 6 and 7.

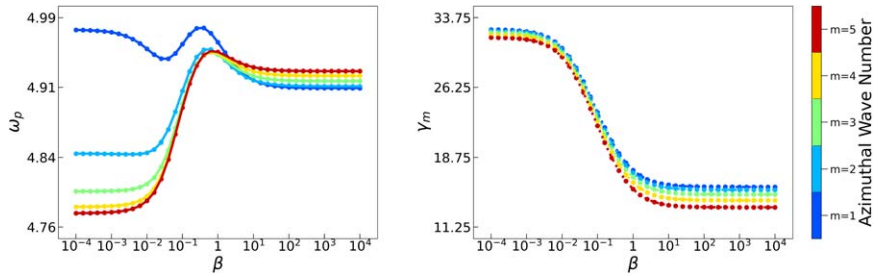

Figure 2. Oscillation frequency (left panel) and growth rate (right panel) vs. cooling timescale β for different azimuthal wavenumber m. Local regime at $\overline{Q} = 0.53$. All wave frequencies are normalized by $\Omega_k|_{r=r_d}$.

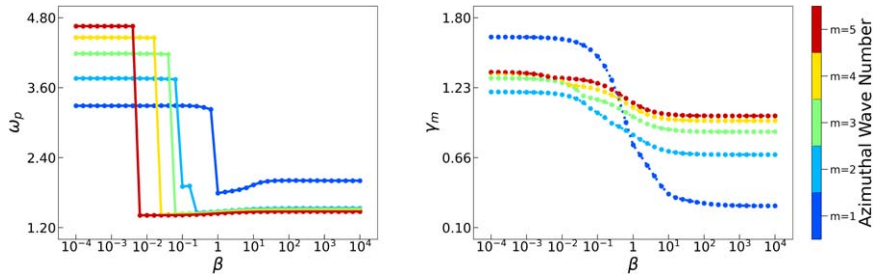

Figure 3. As in Figure 2, but transitional regime at $\overline{Q} = 1.15$.

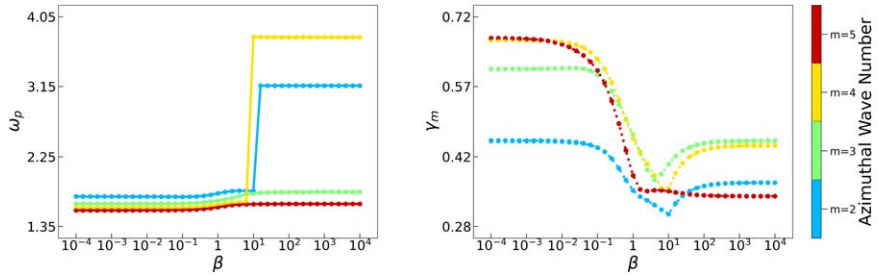

Figure 4. As in Figure 2, but global regime at $\overline{Q} = 1.65$.

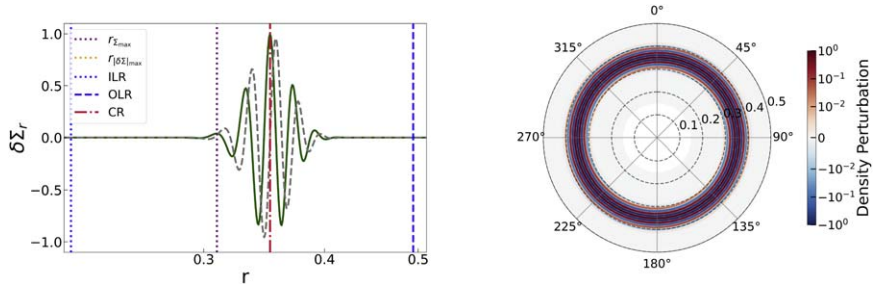

Figure 5. Morphology of the $m = 2$ mode. Local regime at $\overline{Q} = 0.53$ and $\beta = 1$. The left panel shows the real part of density perturbation by the green solid line, the imaginary part by the gray dashed line, Σ_{max} by the purple dashed line, and $r_{|\delta\Sigma|_{max}}$ by the yellow dashed line. The inner Lindblad resonance, outer Lindblad resonance, and corotation resonance are marked by blue-dotted, blue-dashed, and red-dashed lines, respectively. The right panel shows the (r, ϕ) distribution of density perturbation.

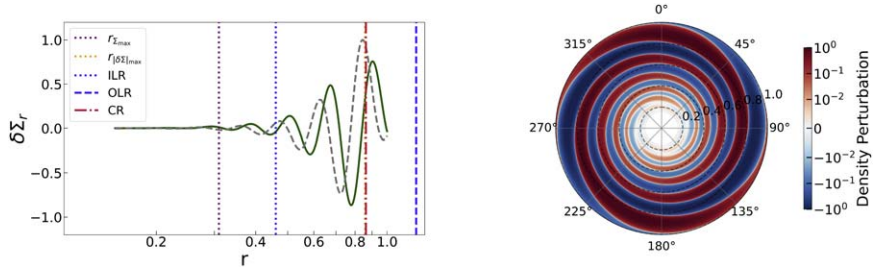

Figure 6. As in Figure 5, but transitional regime at $\overline{Q} = 1.15$ and $\beta = 1$.

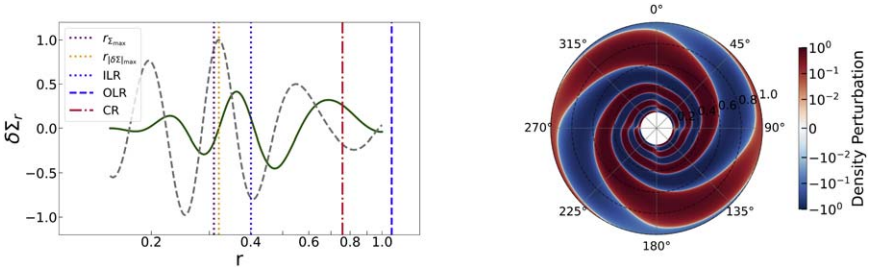

Figure 7. As in Figure 5, but global regime at $\overline{Q} = 1.65$ and $\beta = 10^{-2}$.

3.1. Local Regime

We study the low Toomre number regime with $\overline{Q} = 0.53$. The resulting eigenfunctions and eigenvalues are illustrated in Figures 2 and 5. In Figure 5, the left panels display the density perturbation as a function of radius, while the right panel shows the distribution of density perturbation on the disk plane. In the left panel, both real and imaginary parts of density perturbation are shown, and the phase angle between the two parts determines the spiral pattern depicted by the right panel.

1. *Oscillation rate*. In this regime, as shown in the left panel of Figure 2, all modes have a pattern speed $\omega_p \in [4, 5]$, indicating that their corotation radius (CR; vertical red line) are very close to the location of $\Sigma_{\max}$ where $\Omega = 5.8$ ($r_{\Sigma_{\max}} = 0.31$).

2. *Growth rate*. A notable feature of the growth rate γ_m in this regime is that the dominant modes always own the smallest m, regardless of cooling timescale, and they exhibit extremely large growth rates (>10) compared with the other regimes discussed below. Another significant feature, as shown in the right panel of Figure 2, is that γ_m decreases as β increases. This is because slow cooling (large β) leads to strong thermal pressure to oppose gravitational instability (J. H. Jeans 1902). Our result is consistent with C. F. Gammie (2001) that fast cooling triggers disk fragmentation. More interestingly, the growth rate changes sharply near $\beta \sim 1$ when orbital dynamical timescale is comparable to thermodynamic timescale, probably because of the resonance of orbital dynamics and thermodynamics.

3. *Mode morphology*. As shown in the left panel of Figure 5, the real and imaginary parts of density perturbation exhibit the similar distribution, and the peaks are located near the maximum of density perturbation $r_{|\delta\Sigma|_{\max}} = 0.36$ (vertical yellow line) as well as the corotation radius (vertical red line), and not far away from the maximum of equilibrium density $r_{\Sigma_{\max}} = 0.31$ (vertical purple line). The right panel of Figure 5 clearly show the concentration behavior of these local modes. Moreover, the strongest density perturbation has the radial wavenumber $k \gg 1/r$ given by the WKBJ dispersion relation (J. Binney & S. Tremaine 2008)

$$|k| = \frac{\pi G\Sigma}{c_s^2} \pm \frac{\pi G\Sigma}{c_s^2}\{1 - Q^2[1 - (\omega - m\Omega)^2/\kappa^2]\}^{1/2}. \tag{22}$$

3.2. Transitional Regime

$Q \sim 1$ is a threshold, below which the axisymmetric instability is triggered and slightly above which the non-axisymmetric instability is triggered (K. Kratter & G. Lodato 2016). We take $\overline{Q} = 1.15$ to investigate the GI behavior near this threshold, and the results are shown in Figures 3 and 6.

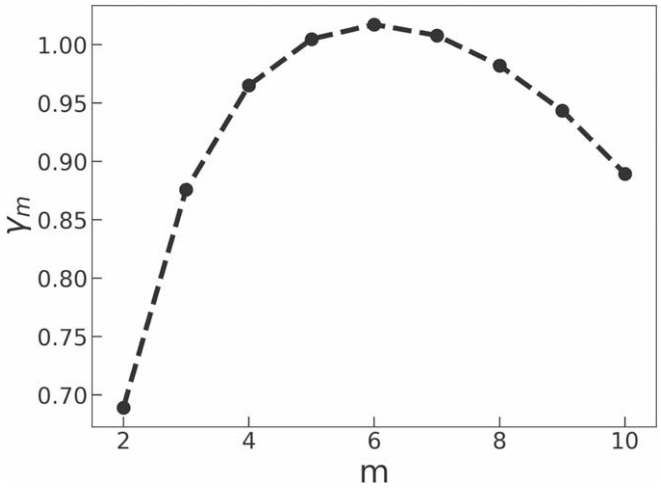

Figure 8. Growth rate γ_m for $m = 2$–10 modes at $\overline{Q} = 1.15$ and $\beta = 10^4$. The dominant mode is at $m = 6$.

1. *Oscillation rate*. When β gradually increases, the modes transit from the local regime to the transitional regime. A noticeable change occurs at $\beta \sim 1$(see the left panel of Figure 3). The GI modes of all m's enter the transitional regime when $\beta \gtrsim 1$. The typical ω_p in the transitional regime is around 1.4, which corresponds to the CR in the outer disk.

2. *Growth rate*. Similar to the local modes, the transitional modes with faster cooling are more unstable, albeit the growth rates are one or two orders lower than the local modes because of higher $\overline{Q}$. Unlike the local modes in which the lowest $m = 1$ always dominates, in the transitional modes the higher m dominates when $\beta > 1$. We also test even higher m at $\beta = 10^4$, shown in Figure 8. The growth rate peaks at $m = 6$ and then decreases as m increases. This indicates that the modes are indeed caused by gravitational instability which prefers long waves.

3. *Mode morphology*. Compared with the local modes, the maximum density perturbations are also located near $r_{|\delta\Sigma|_{\max}}$ and the CR in the outer disk but far away from $r_{\Sigma_{\max}}$ as shown in the left panel of Figure 6, which is different from the local modes, and the density perturbation is much more extended in the radial direction as shown in the right panel of Figure 6, which is also different from the high concentration of the local modes.

3.3. Global Regime

We study the high Toomre number regime $\overline{Q} = 1.65$ which reaches the boundary of GI. Only the $m = 2$, 3, 4, 5 modes are calculated and the $m = 1$ mode is stable. The results are shown in Figures 4 and 7.

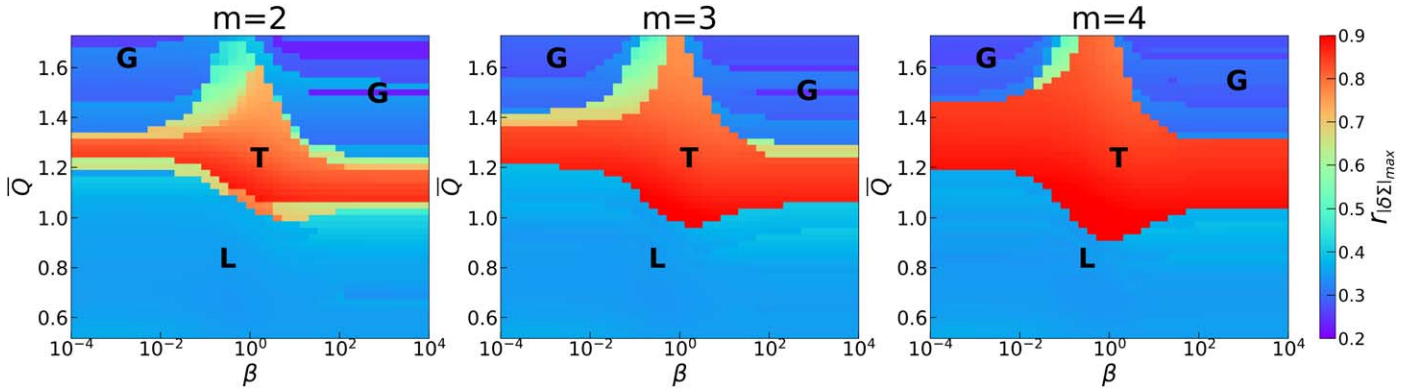

Figure 9. Parameter scan for $m = 2$, 3, 4 with $\overline{Q} \in [0.5, 1.7]$ and $\beta \in [10^{-4}, 10^{4}]$. The contour shows the location of $r_{|\delta\Sigma|_{\max}}$. The local regime marked by "L" dominates with low $\overline{Q}$, the transition regime marked by "T" occupies a large area for $\overline{Q} \gtrsim 0.9$, and the global regime is observed with high $\overline{Q}$ when β is away from unity.

1. *Oscillation rate*. The odd m's lead to $\omega_p \sim 1.5$ corresponding to CR ~ 0.76 in the outer disk. While the even m's lead to $\omega_p \sim 1.5$ when $\beta < 10$ but $\omega_p \sim 3$ when $\beta > 10$ corresponding to CR ~ 0.48 in the inner disk.
2. *Growth rate*. Similar to local and transitional modes, the growth rate γ_m of global modes changes at $\beta \sim 1$ but it is lower than 1 because of high $\overline{Q}$. The dominant modes nearly follow $m \sim M_*/M_d$, consistent with the numerical simulations R. Dong et al. (2015).
3. *Mode morphology*. As shown in Figure 7, a notable feature in this regime is the global pattern, i.e., the perturbations are distributed across the entire disk. This results in fewer peaks compared with the other two regimes (see left panel), and the spirals now have significantly larger pitch angles (wider phase difference between real and imaginary parts, see left panel), making them more loosely winded (see right panel). Another significant difference from the other two regimes is that the location of maximum of density perturbation $r_{|\delta\Sigma|_{\max}}$ departs away from the CR location (red and yellow lines, see left panel), whereas in the other two regimes the two locations are close (see left panels of Figures 5 and 6).

4. Parameter Scan

We have known that the growth rate and the mode transition depend not only on $\overline{Q}$ but also on β, therefore we scan the parameter space (β, $\overline{Q}$) for β from 10^{-4} to 10^{4} and $\overline{Q}$ from 0.5 to 1.7 with fixed disk–star mass ratio $M_d/M_* = 0.4$. We focus on the $m = 2$, 3, 4 modes, as the $m = 1$ mode may be stable and higher modes ($m \geqslant 5$) may not be GI but shear instability (recall that GI prefers long wave).

We choose $r_{|\delta\Sigma|_{\max}}$ to distinguish the transitional regime ($r_{|\delta\Sigma|_{\max}} > 0.5$) from the local and global regimes ($r_{|\delta\Sigma|_{\max}} \lesssim 0.5$). Figure 9 shows the contour of $r_{|\delta\Sigma|_{\max}}$ as a function of β and $\overline{Q}$. The bold characters "L, T, G" denote the local, transitional, and global regimes, respectively. For $\overline{Q} \lesssim 0.9$, the dominant modes in the disk are local regardless of disk thermodynamics β. As $\overline{Q}$ increases, at $\beta \sim 1$ GI enters the transitional regime, and moreover, slower cooling ($\beta > 1$) requires lower $\overline{Q}$ to enter the transitional regime than faster cooling ($\beta < 1$). When $\overline{Q}$ continues to increase, GI with both large and small β will eventually leave the transitional regime to enter the global regime. However, GI with $\beta \sim 1$ tends to retain in the transitional regime, and higher mode owns a wider transitional regime. Similar to growth rate, this behavior at $\beta \sim 1$ may also result from the resonance of orbital dynamics and thermodynamics.

5. Implication for Observations

Gravitational instability significantly influences the evolution of PPDs and the planet formation. First, ALMA has detected various substructures, e.g., spirals, rings, and gaps (S. M. Andrews et al. 2018), and GI is a potential mechanism for these large-scale substructures (C. C. Lin & F. H. Shu 1964; J. Binney & S. Tremaine 2008; F. Meru et al. 2017; J. Bae et al. 2023; J. Speedie et al. 2024). Second, although core accretion is considered to be the classical model to understand the giant planet formation even at large distance $\gtrsim$10 au (e.g., E. L. Nielsen et al. 2019; A. Vigan et al. 2021; T. Currie et al. 2023), GI is an alternative for the high-mass and long-period planet formation, especially to understand the metallicity noncorrelation of brown dwarfs, debris disks, and their host stars (e.g., S. Nayakshin 2017). Third, besides turbulent viscosity, GI can be responsible for the angular momentum transport (K. Rice & S. Nayakshin 2017).

The disk turbulent viscosity is related to accretion rate and angular momentum transport. Usually it is described by the disk α (N. I. Shakura & R. A. Sunyaev 1973)

$$\nu = \alpha H c_s \tag{23}$$

where ν is the turbulent viscosity, α is a dimensionless coefficient, H is the disk scale height and c_s is the local sound speed. The disk α is essentially the ratio of turbulent velocity of largest eddies to sound speed. The typical accretion rate $\sim 10^{-8}$ to $10^{-7} M_\odot$ in protoplanetary disks (e.g., P. J. Armitage 2020) corresponds to $\alpha \sim 10^{-3}$–10^{-2}. Turbulent viscosity arises from Reynolds stress which is given by

$$T^{\rm Rey}_{r\phi}(r) = \frac{1}{2\pi} \int \Sigma \delta u_r \delta u_\phi d\phi. \tag{24}$$

In addition to Reynolds stress, GI can also play a role in transporting angular momentum. Gravitational stress is given by (e.g., C. F. Gammie 2001; G. Lodato & W. K. M. Rice 2004; A. C. Boley et al. 2006; S. Michael et al. 2012; T. Y. Steiman-Cameron et al. 2023)

$$T^{\rm grav}_{r\phi}(r) = -\frac{1}{2\pi r^2} \int \delta\Sigma \frac{\partial \delta\Psi}{\partial \phi} dS \tag{25}$$

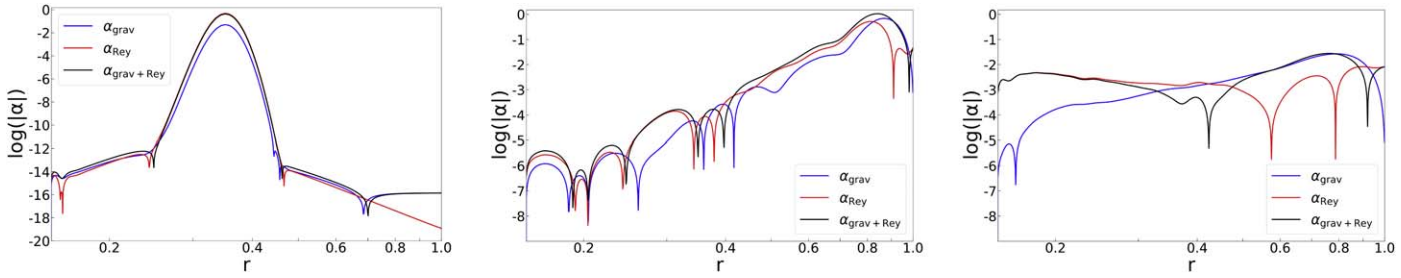

Figure 10. Radial distribution of $\alpha_{\rm Rey}$, $\alpha_{\rm grav}$, and $\alpha_{\rm grav+Rey}$ in the local regime at ($\overline{Q}=0.53$, $\beta=1$) (a), in the transitional regime at ($\overline{Q}=1.15$, $\beta=1$) (b), and in the global regime at $\overline{Q}=1.65$, $\beta=10^{-2}$ (c). The x-axis is plotted on a logarithmic scale. Red and blue solid lines represent $\alpha_{\rm Rey}$ and $\alpha_{\rm grav}$, respectively, while black solid lines show the sum.

where S is disk surface area. By Reynolds and gravitational stresses we can calculate the effective α

$$\alpha = \left| \frac{T_{r\phi}}{\Sigma c_s^2} \left(\frac{d\ln\Omega}{d\ln r} \right)^{-1} \right| \tag{26}$$

where the stress $T_{r\phi}$ can be either $T_{r\phi}^{\rm Rey}$ or $T_{r\phi}^{\rm grav}$. By the thermal balance between viscous heating and radiative cooling, we can derive $\alpha_{\rm grav} \approx 4/[9\gamma(\gamma-1)\beta]$ (P. J. Armitage 2020) such that $\alpha_{\rm grav} \approx 0.1$ at $\beta \approx 1$ and $\gamma \approx 2$. This argument is for a nonlinear self-sustained disk but our analysis is about linear stability (i.e., perturbation grows but not at the equilibrium balance) so that the disk α in our linear study is roughly proportional to disk mass (see Equation (26)).

Figure 10 shows the distribution of $\alpha_{\rm grav}$ and $\alpha_{\rm Rey}$ at different $\overline{Q}$ and β in the three regimes. The deep valleys are caused by the sign reversals of torque. In our 2D study, $\alpha_{\rm grav}$ and $\alpha_{\rm Rey}$ are comparable in most regions of disk, but it is reported that $\alpha_{\rm Rey}$ is less significant than $\alpha_{\rm grav}$ in 3D studies (e.g., S. Michael et al. 2012; T. Y. Steiman-Cameron et al. 2013; J. Bae et al. 2016; W. Béthune & H. Latter 2022). In the local and global regimes $\alpha_{\rm grav}$ at $r_{|\delta\Sigma|_{\rm max}}$ reaches $\sim 10^{-2}$, but in the transitional regime it reaches ~ 1. Although density perturbation becomes highest at $r_{|\delta\Sigma|_{\rm max}}$ ($|\delta\Sigma|/\Sigma \gtrsim 50\%$ in the local regime, 20%–40% in the transitional regime and 15% in the global regime), because of the other quantities in Equation (26) (e.g., $\delta\Psi$, r, etc.), $\alpha_{\rm grav}$ at $r_{|\delta\Sigma|_{\rm max}}$ in the transitional regime is much stronger than in the other two regimes.

Our linear studies in both local and global regimes show strong similarities with 3D simulations of disks governed by GI (e.g., A. P. Boss 2017). The rapid growth of instability in the local regime with low $\overline{Q}$ tends to create fragmentations and gravitationally bound objects in the inner disk, such as high-mass planets and brown dwarfs, regardless of cooling timescales. However, as GI saturates in the subsequent nonlinear phase and cooling is inefficient in the inner disk (e.g., <1 au), fragmentation can be suppressed. For disks in the global regime, the lower growth rate leads to long-lived nonaxisymmetric spirals. Notably, even for high $\overline{Q}$ ($\gtrsim 1.5$) and high β ($\gtrsim 10$), the effective α can still reach $\sim 10^{-3}$, consistent with results from marginally gravitationally unstable disk studies (A. P. Boss 2012), which suggest that GI could be the driving mechanism for generic accretion disk evolution.

The transitional regime is particularly intriguing, as robust instability develops in the outer disk, where $\beta \lesssim 1$ is typically satisfied (M.-K. Lin & A. N. Youdin 2015; T. Pfeil & H. Klahr 2019). With the efficient cooling, substructures, fragmentations, and substellar companions can be created by GI. As the transitional regime dominates at $\beta \approx 1$ (see Figure 9), we expect that spirals with pitch angles ranging from a few degrees to over 10° can be observed in the outer disk. Moreover, the large effective $\alpha_{\rm grav}$ in the transitional regime implies that GI is responsible for the angular momentum transport in the outer disk rather than viscosity.

6. Summary and Discussion

In this paper we explore the characteristics of GI in a protoplanetary disk at various $\overline{Q}$ and β, and our key results are summarized as follows. First, faster/slower cooling timescale (smaller/larger β) leads to faster/slower growth rate, which is consistent with numerical results. Moreover, growth rate changes sharply at $\beta \approx 1$. Second, different $\overline{Q}$ corresponds to different modes: low $\overline{Q}$ to local modes, intermediate $\overline{Q} \approx 1$ to transitional modes, and high $\overline{Q}$ to global modes. Density perturbation maximum $r_{|\delta\Sigma|_{\rm max}}$ of the local and global modes is located in the inner disk and close to equilibrium density maximum $r_{\Sigma_{\rm max}}$, but $r_{|\delta\Sigma|_{\rm max}}$ of the transitional modes in the outer disk and far away from $r_{\Sigma_{\rm max}}$. Moreover, the transitional modes lead to much stronger $\alpha_{\rm grav}$ than the other two modes, suggesting much more efficient transport of angular momentum. Third, the parameter scan reveals that the transitional modes dominate at $\beta \approx 1$. In summary, we infer that GI works in the outer disk at $\overline{Q} \approx 1$ and $\beta \approx 1$ to trigger the transitional modes for efficient transport of angular momentum, disk substructures (e.g., spirals and fragmentations), and planet/brown dwarf formation.

We assume a 2D thin disk. Although our calculations involve a relatively large aspect ratio, it should be noted that an early-age disk can be fairly thick even with an envelope, which will induce the additional mechanisms (K. Kratter & G. Lodato 2016). Moreover, our linear analysis cannot capture the nonlinear effects such as shocks, wave amplitude saturation, etc. In addition, some other factors such as accretion and magnetic fields can be taken into account.

Acknowledgments

We thank Hongping Deng and Cong Yu for the help on equilibrium, Qiang Hou for his initial assistance, Haiyang Wang for self-gravity, Yuru Xu for boundary conditions, and Cathie Clarke for disk α. X.W. is supported by the National Natural Science Foundation of China (NSFC) under grant No.12041301.

Appendix
Eigenvalue Problem

The basic linear perturbation Equations (9)–(13) are translated to an eigenvalue problem

$$im\Omega\delta\Sigma - \frac{1}{r}\frac{d(r\Sigma)}{dr}\delta u_r - \Sigma\frac{d}{dr}\delta u_r + \frac{im\Sigma}{r}\delta u_\phi = im\omega_p\delta\Sigma, \tag{A1}$$

$$\frac{1}{\Sigma^2}\frac{dP}{dr}\delta\Sigma - \frac{1}{\Sigma}\frac{d}{dr}\delta P + im\Omega\delta u_r + 2\Omega\delta u_\phi - \frac{d}{dr}\Phi_d = im\omega_p\delta u_r, \tag{A2}$$

$$\frac{im}{r\Sigma}\delta P - \frac{\kappa^2}{2\Omega}\delta u_r + im\Omega\delta u_\phi + \frac{im}{r}\Phi_d = im\omega_p\delta u_\phi, \tag{A3}$$

$$\begin{aligned}\frac{c_{s,adi}^2}{\gamma t_c}\delta\Sigma + \left(-\frac{1}{t_c} + im\Omega\right)\delta P - \left[\frac{\Sigma c_{s,adi}^2}{L_s} + \frac{c_{s,adi}^2}{r}\frac{d(r\Sigma)}{dr}\right]\delta u_r \\ -c_{s,adi}^2\Sigma\frac{d}{dr}\delta u_r + \frac{im\Sigma c_{s,adi}^2}{r}\delta u_\phi \\ = im\omega_p\delta P.\end{aligned} \tag{A4}$$

Furthermore, we write them as

$$\boldsymbol{M}\left[\delta\Sigma \quad \delta P \quad \delta u_r \quad \delta u_\phi\right]^T = i\omega_{\boldsymbol{m}}\left[\delta\Sigma \quad \delta P \quad \delta u_r \quad \delta u_\phi\right]^T, \tag{A5}$$

$$\boldsymbol{M} = \boldsymbol{M_1} + \boldsymbol{M_2} = \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} & \boldsymbol{M}_{13} & \boldsymbol{M}_{14} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} & \boldsymbol{M}_{23} & \boldsymbol{M}_{24} \\ \boldsymbol{M}_{31} & \boldsymbol{M}_{32} & \boldsymbol{M}_{33} & \boldsymbol{M}_{34} \\ \boldsymbol{M}_{41} & \boldsymbol{M}_{42} & \boldsymbol{M}_{43} & \boldsymbol{M}_{44} \end{bmatrix}. \tag{A6}$$

We divide the disk into N grids and the matrix $\boldsymbol{M}$ will be a $4N \times 4N$ coefficient matrix. The matrix $\boldsymbol{M_1}$ is sparse, containing coefficients for $\delta\Sigma, \delta P, \delta u_r, \delta u_\phi$ and their first-order derivatives. For the first-order differential operators, we employ a central finite difference scheme on logarithmically spaced radial grid points

$$\left.\frac{dX}{dr}\right|_i = \frac{1}{r_i}\sum_j D_{ij}^{(1)}X_j \tag{A7}$$

where D_{ij} represents the first-order differentiation matrices (e.g., F. C. Adams et al. 1989). The matrix $\boldsymbol{M_2}$ implements $\delta\Psi_m$ and $d\delta\Psi_m/dr$ terms. We calculate $\boldsymbol{M_2}$ with the method used in G. Laughlin & M. Rozyczka (1996) and W.-K. Lee et al. (2019) to avoid the singularity of self-gravity potential integration. Boundary conditions (19) and (20) are imposed on some rows of $\boldsymbol{M}$. It is noteworthy that we obtain $4N$ eigenvalues and eigenfunctions. However, we focus on the most unstable mode, which owns the largest growth rate.

ORCID iDs

Zehao Su https://orcid.org/0000-0001-5567-0309
Xing Wei https://orcid.org/0000-0002-3641-6732

References

Adams, F. C., & Lin, D. N. C. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson, AZ: Univ. Arizona Press), 721
Adams, F. C., Ruden, S. P., & Shu, F. H. 1989, ApJ, 347, 959
Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, ApJL, 869, L41
Armitage, P. J. 2020, Astrophysics of Planet Formation (2nd ed.; Cambridge: Cambridge Univ. Press)
Avenhaus, H., Quanz, S. P., Schmid, H. M., et al. 2017, AJ, 154, 33
Backus, I., & Quinn, T. 2016, MNRAS, 463, 2480
Bae, J., Isella, A., Zhu, Z., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Inutsuka et al. (San Francisco, CA: ASP), 423
Bae, J., Nelson, R. P., Hartmann, L., & Richard, S. 2016, ApJ, 829, 13
Baehr, H., Klahr, H., & Kratter, K. M. 2017, ApJ, 848, 40
Benisty, M., Juhasz, A., Boccaletti, A., et al. 2015, A&A, 578, L6
Benisty, M., Stolker, T., Pohl, A., et al. 2017, A&A, 597, A42
Béthune, W., & Latter, H. 2022, A&A, 663, A138
Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton, NJ: Princeton Univ. Press)
Boley, A. C., & Durisen, R. H. 2010, ApJ, 724, 618
Boley, A. C., Mejía, A. C., Durisen, R. H., et al. 2006, ApJ, 651, 517
Boss, A. P. 1997, Sci, 276, 1836
Boss, A. P. 2001, ApJ, 563, 367
Boss, A. P. 2012, AREPS, 40, 23
Boss, A. P. 2017, ApJ, 836, 53
Cameron, A. G. W. 1978, M&P, 18, 5
Canovas, H., Montesinos, B., Schreiber, M. R., et al. 2018, A&A, 610, A13
Chen, E., Yu, S.-Y., & Ho, L. C. 2021, ApJ, 906, 19
Coradini, A., Magni, G., & Federico, C. 1981, A&A, 98, 173
Currie, T., Biller, B., Lagrange, A., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Inutsuka et al. (San Francisco, CA: ASP), 799
Dong, R., Hall, C., Rice, K., & Chiang, E. 2015, ApJL, 812, L32
Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, K. Keil et al. (Tucson, AZ: Univ. Arizona Press), 607
Forgan, D., Parker, R. J., & Rice, K. 2015, MNRAS, 447, 836
Forgan, D., & Rice, K. 2013, MNRAS, 432, 3168
Gammie, C. F. 2001, ApJ, 553, 174
Garufi, A., Quanz, S. P., Avenhaus, H., et al. 2013, A&A, 560, A105
Goldreich, P., & Lynden-Bell, D. 1965, MNRAS, 130, 125
Goodman, J., & Narayan, R. 1988, MNRAS, 231, 97
Hadley, K. Z., Fernandez, P., Imamura, J. N., et al. 2014, Ap&SS, 353, 191
Hall, C., Dong, R., Rice, K., et al. 2019, ApJ, 871, 228
Hall, C., Rice, K., Dipierro, G., et al. 2018, MNRAS, 477, 1004
Huang, J., Andrews, S. M., Pérez, L. M., et al. 2018, ApJL, 869, L43
Jeans, J. H. 1902, RSPTA, 199, 1
Johnson, B. M., & Gammie, C. F. 2003, ApJ, 597, 131
Kratter, K., & Lodato, G. 2016, ARA&A, 54, 271
Kratter, K. M., Murray-Clay, R. A., & Youdin, A. N. 2010, ApJ, 710, 1375
Kurtovic, N. T., Pérez, L. M., Benisty, M., et al. 2018, ApJL, 869, L44
Laughlin, G., & Rozyczka, M. 1996, ApJ, 456, 279
Lee, W.-K., Dempsey, A. M., & Lithwick, Y. 2019, ApJ, 872, 184
Lin, C. C., & Shu, F. H. 1964, ApJ, 140, 646
Lin, D. N. C., & Pringle, J. E. 1987, MNRAS, 225, 607
Lin, M.-K., & Youdin, A. N. 2015, ApJ, 811, 17
Lodato, G., & Rice, W. K. M. 2004, MNRAS, 351, 630
Lodato, G., & Rice, W. K. M. 2005, MNRAS, 358, 1489
Longarini, C., Lodato, G., Bertin, G., & Armitage, P. J. 2023, MNRAS, 519, 2017
Meru, F., Juhász, A., Ilee, J. D., et al. 2017, ApJL, 839, L24
Michael, S., Steiman-Cameron, T. Y., Durisen, R. H., & Boley, A. C. 2012, ApJ, 746, 98
Michikoshi, S., Kokubo, E., & Inutsuka, S.-i. 2012, ApJ, 746, 35
Miranda, R., & Rafikov, R. R. 2020, ApJ, 892, 65
Nayakshin, S. 2017, PASA, 34, e002
Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, AJ, 158, 13
Noh, H., Vishniac, E. T., & Cochran, W. D. 1991, ApJ, 383, 372
Paardekooper, S.-J. 2012, MNRAS, 421, 3286
Pérez, L. M., Carpenter, J. M., Andrews, S. M., et al. 2016, Sci, 353, 1519
Pfeil, T., & Klahr, H. 2019, ApJ, 871, 150
Pierens, A. 2021, MNRAS, 504, 4522
Rice, K., & Nayakshin, S. 2017, MNRAS, 475, 921
Rice, W. K. M., Lodato, G., & Armitage, P. J. 2005, MNRAS, 364, L56
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shariff, K., & Cuzzi, J. N. 2011, ApJ, 738, 73
Shu, F. H. 1992, The Physics of Astrophysics. Volume II: Gas Dynamics. (Mill Valley, CA: Univ. Science books),
Speedie, J., Dong, R., Hall, C., et al. 2024, Natur, 633, 58
Steiman-Cameron, T. Y., Durisen, R. H., Boley, A. C., Michael, S., & McConnell, C. R. 2013, ApJ, 768, 192
Steiman-Cameron, T. Y., Durisen, R. H., Boley, A. C., et al. 2023, ApJ, 958, 139

Takeuchi, T., & Ida, S. 2012, ApJ, 749, 89
Tomida, K., Machida, M. N., Hosokawa, T., Sakurai, Y., & Lin, C. H. 2017, ApJL, 835, L11
Tominaga, R. T., Takahashi, S. Z., & Inutsuka, S.-i. 2020, ApJ, 900, 182
Toomre, A. 1964, ApJ, 139, 1217
Uyama, T., Hashimoto, J., Muto, T., et al. 2018, AJ, 156, 63
Vigan, A., Fontanive, C., Meyer, M., et al. 2021, A&A, 651, A72
Wagner, K., Apai, D., Kasper, M., & Robberto, M. 2015, ApJL, 813, L2
Ward, W. R. 2000, in Origin of the Earth and Moon, ed. R. M. Canup, K. Righter et al. (Tucson, AZ: Univ. Arizona Press), 75
Xu, W., & Kunz, M. W. 2021, MNRAS, 508, 2142
Youdin, A. N. 2005, arXiv:astro-ph/0508659