Physics – Uspekhi **58** (11) 1031 – 1058 (2015)

REVIEWS OF TOPICAL PROBLEMS PACS numbers: **02.30. – f**, 02.60.Pn, 47.32.C –, 97.10.Gz, 97.82.Jw, 98.62.Mw

Transient dynamics of perturbations in astrophysical disks

D N Razdoburdin, V V Zhuravlev

DOI: 10.3367/UFNe.0185.201511a.1129

Contents

Abstract. We review some aspects of a major unsolved problem in understanding astrophysical (in particular, accretion) disks: whether the disk interiors can be effectively viscous in spite of the absence of magnetorotational instability. A rotational homogeneous inviscid flow with a Keplerian angular velocity profile is spectrally stable, making the transient growth of perturbations a candidate mechanism for energy transfer from regular motion to perturbations. Transient perturbations differ qualitatively from perturbation modes and can grow substantially in shear flows due to the nonnormality of their dynamical evolution operator. Because the eigenvectors of this operator, also known as perturbation modes, are not pairwise orthogonal, they can mutually interfere, resulting in the transient growth of their linear combinations. Physically, a growing transient perturbation is a leading spiral whose branches are shrunk as a result of the differential rotation of the flow. We discuss in detail the transient growth of vortex shearing harmonics in the spatially local limit, as well as methods for identifying the optimal (fastest growth) perturbations. Special attention is given to obtaining such solutions variationally by integrating the respective direct and adjoint equations forward and backward in time. The presentation is intended for experts new to the subject.

Keywords: hydrodynamics, turbulence, accretion disks

D N Razdoburdin Lomonosov Moscow State University,
Faculty of Physics,
Leninskie gory, 119991 Moscow, Russian Federation;
Lomonosov Moscow State University,
Sternberg State Astronomical Institute,
Universitetskii prosp. 13, 119991 Moscow, Russian Federation
V V Zhuravlev Lomonosov Moscow State University,
Sternberg State Astronomical Institute,
Universitetskii prosp. 13, 119991 Moscow, Russian Federation
E-mail: zhuravlev@sai.msu.ru

Received 18 May 2015, revised 1 September 2015
Uspekhi Fizicheskikh Nauk **185** (11) 1129 – 1161 (2015)
DOI: 10.3367/UFNr.0185.201511a.1129
Translated by K A Postnov; edited by A M Semikhatov

1. Introduction: modal and nonmodal analysis of perturbations

A salient feature of disk accretion is that it is impossible without a dissipation mechanism of the differential rotation energy of matter. It is the internal friction in the disk, i.e., irreversible interaction of its adjacent rings, that leads to the transformation of the gravitational energy of the accreting matter into heat and electromagnetic radiation, which simultaneously allows the matter to flow toward the center and the angular momentum to flow outwards to the disk periphery.

Direct dissipation is already possible due to the microscopic viscosity of a gas (plasma); however, in astrophysical conditions, it turns out to be absolutely insufficient for explaining the observed properties of the disks. Essentially, the disks are too large for the characteristic accretion time t_v to be explained by the microscopic viscosity. For example, in protoplanetary disks with the typical size $L \sim 10$ a.u., where the kinematic viscosity is estimated to be $\nu_m \sim 10^7$ cm^2 s^{-1}, the accretion time is $t_v = L^2/\nu \sim 10^{13}$ years (see Section 3.3.2 in [1]). Apparently, t_v is several orders of magnitude longer than the age of the Universe. At the same time, observations of gas–dust disks around young stars suggest that their lifetime is as short as only several million years (see, e.g., review [2]). A similar conclusion is obtained for hot accretion disks, in particular, those around black holes in close binary systems. In this case, for much smaller scales $L \sim 10^{10}$ cm and somewhat smaller viscosity of hydrogen plasma $\nu_m \sim 10^5$ cm^2 s^{-1}, we obtain $t_v \sim 3 \times 10^7$ years, which, for example, by many orders of magnitude exceeds the duration of X-ray Nova outbursts caused by nonstationary disk accretion (see review [3]).

At the same time, it is known from statistical hydromechanics (see a discussion of the Reynolds equations in [4], vol. 1, ch. 3) that the presence of significant correlating fluctuations of velocity components in a flow is equivalent to the presence of a high effective viscosity that exceeds the microscopic viscosity, because the mixing scale of matter in

the flow is much larger than the mean free path of individual particles. In turn, the high effective viscosity enhances the angular momentum transfer toward the disk periphery, thus decreasing t_v to the observed values. The perturbations under discussion can be regular in general: for example, the accretion can be due to tidal waves generated in the disk by the secondary companion of a binary system (see [5]). But it is more natural to assume that these perturbations are generated by *turbulence* in the fluid. The turbulence, on the one hand, takes energy from the rotational motion of matter on large scales, and on the other hand, via the interaction of perturbation components with different wave numbers, cascades this energy to small scales, where its direct dissipation into heat occurs due to microscopic viscosity.

It is important to recognize that energy transfer from regular flow to perturbations must be mediated by some linear mechanism that follows from the dynamics of small perturbations described by linearized hydrodynamic equations. This can be rigorously proved for vortex fluid motion using the Navier–Stokes equations (see [6], Section 1.4, as well as [7]). Therefore, the first natural step in the theoretical study of turbulence generation in some (stationary) flow is to search for exponentially growing linear perturbations on a steady-state background. Such perturbations are usually referred to as *modes*, and the corresponding analysis is called the *modal* or spectral analysis of perturbations, because it is used to determine eigenvalues of the corresponding dynamical operator of the problem, the (complex) mode frequencies. Turbulence arising from growing modes is called supercritical. In astrophysical flows with Keplerian angular frequencies, the spectral (magneto-rotational) instability and the corresponding supercritical (MHD) turbulence have been found in analytic and numerical calculations [8–10] (see also reviews [11, 12]) for disks with a frozen seed magnetic field. But the magneto-rotational instability does not operate in cold low-ionized disks. Protoplanetary disks, accretion disks in quiescent states of cataclysmic variables, and the outer parts of accretion disks in active galactic nuclei provide examples. Hence, it would be very important to show that differential rotation alone is capable of exciting turbulence in Keplerian disks. This property of Keplerian flows is universal, unlike the presence of a seed magnetic field together with a sufficiently high degree of ionization of matter, or the existence of flow inhomogeneities due to a vorticity jump (see, e.g., [13]), or the appearance of radial velocity gradients (see [14]) or of vertical and/or horizontal gradients of some thermodynamic values (see, e.g., [15, 16]).

But the generation of turbulence in a *homogeneous* Keplerian flow without a magnetic field remains questionable so far. The main difficulty here is that such a flow is spectrally stable: the specific angular momentum for Keplerian rotation increases with the radial distance from the center; therefore, according to the Rayleigh criterion ([17] and [18], vol. 6, paragraph 27), the growth of axially symmetric modes is impossible; in turn, nonaxisymmetric modes cannot grow because the necessary Rayleigh condition for the existence of an extremum of vorticity in a background flow [19, 20] is not satisfied. Nevertheless (as follows from laboratory experiments and numerical simulations), turbulence arises in spectrally stable flows as well. In this case, it is called subcritical. A planar Couette flow provides the simplest and most prominent example (see classic monographs [21, 22]).

In the theory of hydrodynamic stability, the transition of some flow (with nonzero microscopic viscosity) to a turbulent state is usually characterized by a set of critical Reynolds numbers Re (see Section 1.3.2 in [6]). The smallest of them is the number $\mathrm{Re_E}$ such that at $\mathrm{Re} < \mathrm{Re_E}$ there are no initial perturbations, irrespective of their amplitudes, whose energy would grow at the initial instant $t = 0$. $\mathrm{Re_E}$ can be derived from the Reynolds–Orr energy equation (see Section 1.4 in [6]). For the Couette flow, $\mathrm{Re_E} \approx 20$. For $\mathrm{Re} > \mathrm{Re_E}$, perturbations initially growing at $t = 0$ arise, but as long as $\mathrm{Re} < \mathrm{Re_G}$, again, there are no initial perturbations with any amplitude that would not decay as $t \to \infty$. This is the definition of the second critical number $\mathrm{Re_G} > \mathrm{Re_E}$. Finally, at higher values $\mathrm{Re} > \mathrm{Re_G}$, perturbations that can sustain their amplitude at all times appear, and starting from some $\mathrm{Re_T} > \mathrm{Re_G}$, the transition to a turbulent state is experimentally observed. For the Couette flow, $\mathrm{Re_T} \approx 360$. The largest of the critical Reynolds numbers is $\mathrm{Re_L} > \mathrm{Re_T}$, starting from which growing modes arise, i.e., the flow becomes spectrally unstable. For the Couette flow, as well as for the Keplerian flow of interest here, $\mathrm{Re_L} = \infty$. However, the Keplerian flow is different in that the value of $\mathrm{Re_G}$ is currently unknown and $\mathrm{Re_T}$ has not been measured, either theoretically or experimentally.

On the one hand, the general opinion emerged that for Keplerian flows, $\mathrm{Re_G} = \mathrm{Re_T} \to \infty$. It is based on the indirect argument that (locally) the action of the tidal and Coriolis forces on a perturbation, which are absent in the Couette flow, strongly stabilizes the shear flow (see Fig. 9 in [11], where the results from [23] are shown). This conclusion is supported by local numerical simulations [24, 25] and a series of laboratory experiments [26–28], in which the stability of a quasi-Keplerian flow was observed up to $\mathrm{Re} = 2 \times 10^6$. Here, we assume the quasi-Keplerian flow to be the so-called anti-cyclonic flow (see, e.g., the definition in [29]), where the specific angular momentum increases and the angular velocity, in contrast, decreases toward the periphery.[1]

On the other hand, in a cyclonic flow, subcritical turbulence is observed at finite, albeit large, values of $\mathrm{Re_T}$; see [30, 31] on experiments with a spectrally stable Taylor–Couette flow, as well as their analysis in the astrophysical context in Zel'dovich's paper [32] and later in [33]. In addition, the negative results obtained in the numerical experiments mentioned above can be explained by insufficient numerical resolution, as discussed in [34]. In a subsequent paper [29], the dynamics of perturbations in cyclonic and anti-cyclonic flows were compared numerically. It was concluded that the required numerical resolution in the second case is much higher than in the first case, and current computational power is insufficient to discover turbulence in a Keplerian flow; it is also impossible to argue that the stabilizing action of the Coriolis force in this case rules out the existence of a finite value $\mathrm{Re_T} < \infty$. Last, another laboratory experiment presented in [35, 36] shows the appearance of subcritical turbulence and angular momentum transfer to the periphery of a quasi-Keplerian flow. The contradictory results claimed by different experimental groups show the complexity of the experiment due to the inevitable secondary flows induced by experimental tools. Presently, the influence of axial boundaries on laboratory flow is being discussed (see [37, 38]).

[1] In a cyclonic flow, both these quantities increase with the distance from the center.

Anyway, it can be stated that of all types of homogeneous rotating flows, quasi-Keplerian (anti-cyclonic) flows are the most stable with respect to finite-amplitude perturbations. Nevertheless, the smallness of microscopic viscosity in the astrophysical conditions mentioned above simultaneously means that huge Reynolds numbers should exist in the disks: for example, for the protoplanetary disk discussed above, taking the thickness $H \approx 0.05L = 0.5$ a.u. as the natural limiting scale of the problem, which corresponds to the sound velocity in the disk at this radius $c_s \sim 0.5$ km s^{-1}, yields $\mathrm{Re} \approx 10^{10}$. In other astrophysical disks, Re can be even higher. Apparently, considering all negative results, there are still several orders of magnitude for the possibility of turbulence in astrophysical Keplerian flows: $10^6 < \mathrm{Re_T} < 10^{10}$.

Searches for the critical value of $\mathrm{Re_T}$ for Keplerian flows continue, and in this paper we discuss in detail the necessary condition for turbulence and/or the transfer of extra angular momentum to the disk periphery — the transition of energy from the regular flow to perturbations in such a flow. As mentioned above, this transition must be mediated by a linear mechanism. Because the Keplerian flow is spectrally stable, only (small) perturbations different from modes can provide such a mechanism. The existence of such transiently growing nonmodal perturbations in shear flows was already suggested by Kelvin [39] and Orr [40, 41]. In astrophysics, this problem was studied in stellar dynamics (see [42, 43]). However, in the context of hydrodynamic stability, the rigorous treatment of such perturbations and methods for determining them were elaborated only in the 1990s and were called the *nonmodal* perturbation analysis. To stress the inapplicability of the traditional modal analysis here, the corresponding concept of the transition to subcritical turbulence due to transient perturbation growth was called the *bypass* transition. The nonmodal analysis of perturbations was formulated in [44–47] (see also reviews [48, 49] and book [6]). These papers showed that the nonmodal growth is mathematically due to the nonorthogonality of perturbation modes. If modes with a physically motivated norm are not orthogonal to each other, their linear combinations can grow in norm even if each separate mode decays, as in a spectrally stable flow (see Fig. 5 in Section 3.1). In turn, the modes are nonorthogonal due to the *nonnormality* of the linear dynamical operator governing the perturbation evolution (see the introductory information about the operators in the same section). A nonnormal operator does not commute with its adjoint operator, which is due to a nonzero velocity shear in the regular flow (see the concluding part of Section 3.4 below for more details). Here, the higher Re is, the higher the degree of nonorthogonality of modes to each other and, correspondingly, the greater the probability of transient growth. Papers mentioned above argue that the maximum possible transient growth of perturbations by a fixed time, called the *optimal growth*, is determined by the norm of a dynamical operator; this norm can be obtained by calculating singular vectors of the operator (see Section 3.1 for the details). Finally, the operator norm is closely related to the notion of the operator pseudospectrum (see [48] and book [6]).

Later, this method was applied to astrophysical flows in [50–52], where different models were used to search for optimal perturbations demonstrating the optimal growth. In particular, it was shown that for the Keplerian velocity profile, the growth can be substantial only starting from $\mathrm{Re} \sim 10^6$, while in a similar setup for an isomomentum profile and the Couette flow, the growth starts already at $\mathrm{Re} \sim 10^3$ (see the discussion in [52]). Here, papers [53, 54] should also be mentioned, which discuss the transient dynamics in the spectrally stable Taylor–Couette flow and include both cyclonic and anti-cyclonic regimes. In [53], a correlation was discovered between the experimentally obtained stability boundary in a laminar flow (see [55]) and the optimal growth value; it was found in [54] that for the same Re number, the transient growth is minimal in the quasi-Keplerian regime. Using the correlation from [53], the authors of [54] estimated $\mathrm{Re_T} \sim 10^5$ for the quasi-Keplerian regime. As in the numerical experiments [23–25] mentioned above, the effective Re caused by numerical viscosity was barely above $\sim 10^4 - 10^5$; it is not surprising that the Keplerian profile was stable against perturbations in these studies.

In addition, presently, there are many astrophysical studies of the transient growth of local perturbations by the Lagrangian method, where the transformation to a reference frame comoving with the shear is done, and separate shear harmonics are considered (see Section 2.2). It was found that in the local space limit, there always exist vortex shear harmonics with transient growth that at the instant of swing (see Section 2.2) can emit different wave shear harmonics (depending on the account for the compressibility or some inhomogeneities in the flow), which also demonstrate nonmodal growth [56–68].

Finally, in [69, 70], the nonlinear transient dynamics of three-dimensional perturbations was investigated with the global structure of the flow taken into account in the model of a geometrically thin disk with α-viscosity. As in [50], these papers discussed the possibility of exciting nonmodal perturbations by weak turbulence already present in the disk, which gives rise to low effective viscosity parameterized by the α parameter. In Section 2.3, we also consider the influence of the effective viscosity on the transient growth of vortices with different scales relative to the disk thickness. Hence, the transient growth of perturbations can be discussed not only in the context of the bypass transition of a laminar flow to turbulence but also as a mechanism to enhance the angular momentum transfer in a disk with pre-existing weak turbulence producing low viscosity. In the last case, this turbulence can be mathematically treated as an external stochastic perturbation in a shear flow, which transits to a quasi-stationary state with a significant increase in the amplitude of perturbations due to the nonnormality of the linear operator governing their dynamics (see [50]).

The purpose of this paper is to consider the transient growth phenomenon in detail using the simplest example of two-dimensional adiabatic perturbations in a homogeneous rotating shear flow with a quasi-Keplerian angular velocity profile. In Section 2, we present an analysis of shear vortex harmonics that are responsible for the transient growth in a spatially local treatment of the problem, and discuss the mechanism of perturbation growth using them as an example. Sections 3 and 4 are mainly devoted to methods for studying nonmodal perturbation growth and to searches for optimal perturbations with maximum growth. Two methods for obtaining the optimal growth curve are presented: a matrix one and a variational one. The variational method is less in use, especially in astrophysical studies (see [71]); however, it is essentially more universal than the matrix one. For example, using this method, we here calculate one optimal transient perturbation in a geometrically thin quasi-

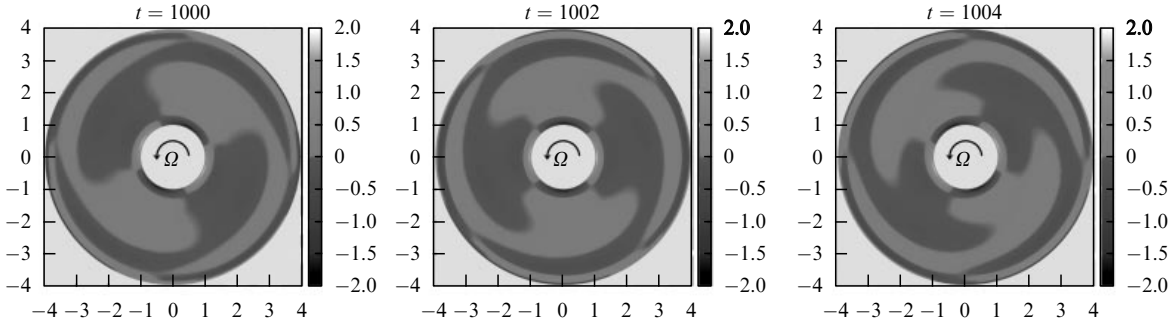

Figure 1. Contours of the most unstable perturbation mode with the azimuthal wave number $m = 2$ in the model of a quasi-Keplerian thin disk described in Section 3.2. Parameters of the calculation: the characteristic disk aspect ratio $\delta = 0.3$, the inner and outer boundaries are at $r_1 = 1$ and $r_2 = 4$, the polytropic index of matter is $n = 3/2$. The mode increment and phase velocity are $\Im[\omega] \approx 0.001$ and $\Re[\omega] \approx 0.26$. Shown is the time (in units of the inverse Keplerian frequency at the inner disk edge) after the conventional moment when the mode has the unit amplitude. The arrow shows the rotational direction of matter in the disk. The calculation method is described in Section 4.2.

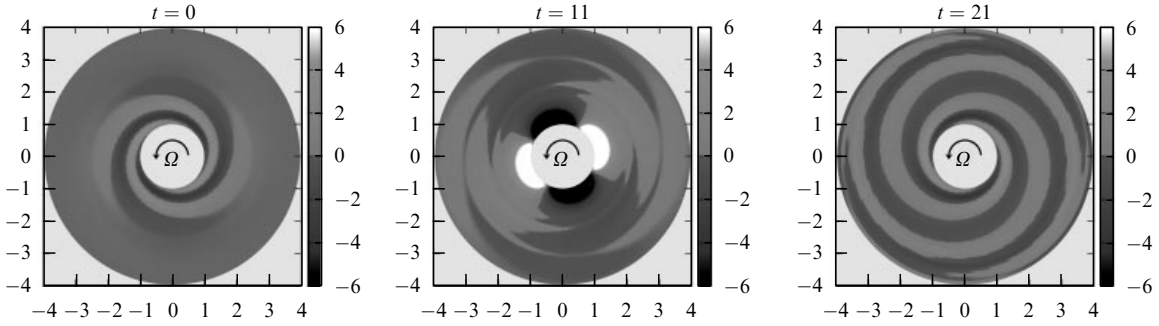

Figure 2. Contours of the $m = 2$ perturbation demonstrating a maximum possible transient growth of acoustic energy at the time $t_{\rm opt} = 10$ measured from the beginning of the perturbation evolution in units of the inverse Keplerian frequency at the inner disk edge. The initial perturbation conventionally has a unit amplitude; the model of the flow is the same as in Fig. 1. The calculation method is described in Section 4.2.

Keplerian flow with free boundaries (Fig. 2) and the most unstable perturbation mode (Fig. 1), which we discuss in detail in the concluding part of Section 4.2. A comparison of Figs 2 and 1 shows that these two types of perturbations are indeed qualitatively different: the transient spiral is wound up by the flow and its amplitude increases, while the modal spiral rotates as a solid body and demonstrates a monotonic but very weak growth because of a low instability increment. Here, the phase velocity of the modal spiral is such that its corotation radius, at which the energy is transferred from the regular flow, lies inside the flow.

2. Analytic solution for two-dimensional vortices

2.1 Adiabatic perturbations in rotational shear flow

We first consider the dynamics of small adiabatic perturbations in a perfect fluid with an isentropic equation of state. Perturbations are described using the Euler approach, i.e., in terms of variations of physical quantities such as the density ρ, the velocity $\mathbf{v}$, and the pressure p at a given point in space at a given time in the perturbed flow relative to the unperturbed background.[2] We assume for simplicity that there are no entropy gradients in the fluid. In the right-hand side of the Euler equations, it is then convenient to pass from the pressure gradient to the enthalpy gradient. Indeed, at constant entropy, the enthalpy differential per unit mass is $\mathrm{d}h = \mathrm{d}p/\rho$ (see [73]), and this is valid in both the background and perturbed flows. For Euler perturbations, we thus obtain $\delta(\nabla p/\rho) = \nabla\delta h$. Using this relation, we write equations for $\delta\rho$, δh, and $\delta\mathbf{v}$ (see also [18], paragraph 26) in the form

$$\frac{\partial\delta\mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\delta\mathbf{v} + (\delta\mathbf{v}\nabla)\mathbf{v} = -\nabla\delta h\,, \tag{1}$$

$$\frac{\partial\delta\rho}{\partial t} + \nabla(\rho\delta\mathbf{v}) + \nabla(\delta\rho\mathbf{v}) = 0\,, \tag{2}$$

where we assume that $\mathbf{v}$ and ρ are the velocity and density in the unperturbed background flow, which itself can evolve in time. Equations (1) and (2) are linear because the perturbations are small and all quadratic terms are omitted.

2.2.1 The model and basic equations. To write the corresponding equations for scalar quantities, we specify the model we wish to consider to illustrate the transient dynamics. First of all, we assume that the background flow is *stationary* and is purely rotational; this condition is satisfied well in astrophysical disks. This implies that the flow is axially symmetric, and it is convenient to use a cylindric coordinate system (r, φ, z) in which the velocity has only the azimuthal nonzero component, $\mathbf{v} = (0, v_\varphi, 0)$. Below, we also use the angular

[2] See [72] concerning applications of hydrodynamics to astrophysical problems, in particular, the application of the theory of hydrodynamic perturbations.

velocity of the flow $\Omega = v_\varphi/r$. It is important to note that the isentropicity of the fluid (which is a particular case of barotropicity) immediately implies that v_φ and Ω depend only on the radial coordinate (see [74], paragraph 4.3). At the same time, the density in Eqns (1) and (2) is a function of both r and z: $\rho = \rho(r, z)$. The case of a geometrically thin disk, where $H(r)/r \ll 1$ and H is the disk half-thickness, is the most common. This assumption is useful in finding how the density ρ changes with the height above the equatorial disk plane. We use the hydrostatic equilibrium condition in the background flow,

$$\frac{\partial h}{\partial z} = -\Omega^2(r) z\,, \tag{3}$$

where the right-hand side is the vertical acceleration of gravity due to the central gravitating body around which the disk rotates. This acceleration is written here ignoring quadratic corrections in the small parameter z/r. Integrating (3) with the condition $h(z = H) = 0$ yields the vertical enthalpy distribution

$$h = \frac{1}{2}(\Omega H)^2\left(1 - \frac{z^2}{H^2}\right). \tag{4}$$

Next, due to the constant entropy assumption, $p \propto \rho^\gamma$, where $\gamma = 1 + 1/n$ is the adiabatic index of matter written in terms of the polytropic index n. This means that the square of the sound velocity in the background flow is $a^2 = \gamma p/\rho$, and the density is mainly dependent on z as follows:

$$a^2 \propto \left(1 - \frac{z^2}{H^2}\right), \qquad \rho \propto \left(1 - \frac{z^2}{H^2}\right)^n. \tag{5}$$

Finally, for simplicity, we consider only perturbations in which $\delta\mathbf{v}$ *is independent of* z. Generally, this very strong assumption needs a justification. In particular, it is relevant to ask: if we take initial perturbations with such a property, is it conserved in further evolution, and if not, how rapidly is this assumption violated? The answer depends on the vertical disk structure. For example, it was shown in [75] that in the particular case of an isothermal vertical density distribution ($n \to \infty$), small perturbations with a velocity field homogeneous in z are exact solutions of Eqns (1) and (2). In the more general case with a finite n, this is no longer the case; however, for example, three-dimensional simulations of barotropic toroidal flows indicate that the most unstable perturbations there weakly depend on z (see [76]). This can be related to the fact that when the angular velocity is independent of z, the Reynolds stresses, responsible for the energy transfer from the main flow to perturbations, do not depend on the vertical component of the velocity perturbation [77, 78]. Finally, a three-dimensional study of the transient dynamics of vortices in a Keplerian flow [51] also shows that the most rapidly growing perturbations in a vertically nonstratified medium are almost independent of z (see also [54]). Now, considering the vertical, radial, and azimuthal projections of (1), we see that our assumption implies the independence of δh from z, and therefore the right-hand side of the vertical projection of (1) vanishes. Then, if we additionally assume that initial vertical velocity perturbations are absent, $\delta v_z = 0$, they do not appear later. Therefore, in the perturbed flow, as well as in the background flow, vertical hydrostatic equilibrium occurs. It can be shown that the assumption of vertical hydrostatic equilibrium in the perturbed flow is equivalent to the assumption that the velocity perturbation field is homogeneous in z (one assumption always follows from the other). On the other hand, if the fluid is not isentropic and there is a radial entropy gradient in the disk, the simplifying assumptions made above are not sufficient to set δv_z to zero.

Thus, we come to the conclusion that we can deal with a flat velocity perturbation field, i.e., $\delta\mathbf{v} = \{\delta v_r, \delta v_\varphi, 0\}$, with δv_r and δv_φ, like δh, being dependent on the radial coordinate only. However, it is important to emphasize that this is not the case with $\delta\rho$ that enters continuity equation (2). Here, it is convenient to use the relation between the pressure and density variations in an isentropic fluid, $\mathrm{d}p = a^2\,\mathrm{d}\rho$, which is a consequence of the barotropic equation of state. Due to the universal character of this relation, small Eulerian perturbations are related in the same way, i.e., $\delta p = a^2\delta\rho$, where a is the speed of sound in the background flow. Consequently,

$$\delta\rho = \frac{\rho}{a^2}\,\delta h\,, \tag{6}$$

and this expression can be substituted in (2), after which only background quantities in (2) depend on the radial coordinate. When integrating Eqn (2) in its new form over z, we should keep in mind that

$$\begin{aligned} &\int_{-H}^{H} \frac{\rho}{a^2}\,\mathrm{d}z = \sqrt{\pi}\,\frac{\Gamma(n)}{\Gamma(n+1/2)}\,\frac{\rho}{a^2}\bigg|_{z=0}\,,\\ &\int_{-H}^{H} \rho\,\mathrm{d}z \equiv \Sigma = \sqrt{\pi}\,\frac{\Gamma(n+1)}{\Gamma(n+3/2)}\,\rho\bigg|_{z=0}\,, \end{aligned} \tag{7}$$

where we use relation (5) and introduce the surface density Σ.

Using the fundamental property of the gamma-function $\Gamma(z+1) = z\Gamma(z)$ in (2), we can explicitly write the system of equations (1), (2) for azimuthal complex Fourier harmonics $\delta v_r, \delta v_\varphi, \delta h \propto \exp(\mathrm{i}m\varphi)$:

$$\frac{\partial\delta v_r}{\partial t} = -\mathrm{i}m\Omega\delta v_r + 2\Omega\delta v_\varphi - \frac{\partial\delta h}{\partial r}\,, \tag{8}$$

$$\frac{\partial\delta v_\varphi}{\partial t} = -\frac{\kappa^2}{2\Omega}\,\delta v_r - \mathrm{i}m\Omega\delta v_\varphi - \frac{\mathrm{i}m}{r}\,\delta h\,, \tag{9}$$

$$\frac{\partial\delta h}{\partial t} = -\frac{a_*^2}{r\Sigma}\,\frac{\partial}{\partial r}(r\Sigma\delta v_r) - \frac{\mathrm{i}ma_*^2}{r}\,\delta v_\varphi - \mathrm{i}m\Omega\delta h\,, \tag{10}$$

where $a_*^2 \equiv na_{\mathrm{eq}}^2/(n+1/2)$ and a_{eq} is the background speed of sound in the equatorial disk plane. In addition, $\kappa^2 = (2\Omega/r)\,\mathrm{d}(\Omega r^2)/\mathrm{d}r$ is the square of the epicyclic frequency, i.e., the frequency of free oscillations of the fluid in the (r, φ) plane, as can easily be verified by writing (8) and (9) for $\delta h = 0$ and substituting the solution $\delta v_r, \delta v_\varphi \propto \exp(-\mathrm{i}\omega t)$ there. We note that reducing the three-dimensional problem to an effectively two-dimensional one in a thin disk can clearly be performed by simply replacing the volume density with the surface density and the polytropic index with $n + 1/2$ in the original equations that are not integrated over z, as was first shown in [79].

2.1.2 Types of perturbations. System of equations (8)–(10) describes the dynamics of two types of perturbations inside a disk that are possible in the two-dimensional formulation of the problem: vortices and density waves.[3] The separation between them for transient perturbations is described below in the local problem setup, which allows giving a simpler

[3] Density waves are also frequently referred to as inertial acoustic waves.

physical interpretation of the behavior of perturbations in a differentially rotating flow. In addition, when there are free radial boundaries in the background flow (for example, in a disk of a finite radial extension such that Σ vanishes and the shear acquires a super-Keplerian angular velocity gradient at some inner and outer radii), surface gravity waves arise near the boundaries (see [80–82]). This occurs because the presence of any significant radial pressure gradient in the flow is equivalent to a nonzero gravitational acceleration, which gives rise to waves similar to ocean waves running over free surfaces (or radial density jumps).

2.1.3. Perturbation modes. These types of perturbations were studied in detail in the 1980s by the spectral method, where system of equations (8)–(10) was solved for particular temporal Fourier harmonics $\propto \exp(-\mathrm{i}\omega t)$, called *modes* (see reviews [83, 84]). In this analysis, the local dispersion relation gives only real values of ω in all astrophysically important cases where $\Omega(r)$ is such that the specific angular momentum Ωr^2 increases with the radius outwards. This implies a local stability of the disks and prohibits exponential growth of small-scale perturbations, which also agrees with the well-known Rayleigh criterion in the particular case of axially symmetric perturbations (see paragraph 27 in [18]). Unlike this case, the global setup of the problem for axially nonsymmetric modes, when the system of differential equations for the radial coordinate with the corresponding boundary conditions at the inner disk radius and at infinity (or at the outer disk boundary) is to be solved, yields a discrete set of ω, possibly including complex frequencies (see, e.g., [81, 82, 85–91]). The nonzero real part of the frequency corresponds to the angular velocity of *solid-body* rotation of a given mode in the flow. Generally, the solid-body azimuthal motion of a constant-phase contour of perturbations with the same azimuthal velocity $\Re[\omega]/m$ at all r is the main feature of modes that distinguishes them from other perturbations. Here, $\Re$ denotes the real part of a frequency ω. A nonzero imaginary part of the frequency, $\Im[\omega]$, means that the canonical energy and angular momentum (see [92]) are exchanged between this mode and either the background flow [87, 93–95] or the mode with (canonical) energy with the opposite sign [82, 96, 97]. In the literature, the first mechanism is also referred to as the Landau mechanism, and the second one as mode coupling. The energy exchange is resonant in both cases, i.e., always occurs in the so-called critical layer at the radius where $\omega = m\Re[\Omega]$, which is called the corotation radius. We refer the reader to [98] for a detailed discussion of the physics of these resonant mechanisms of mode growth (decay). Nevertheless, in flows with almost Keplerian rotation, both mode coupling and their interaction with the background occur extremely slowly, and the corresponding increments, even for a large disk aspect ratio $H/r \sim 0.1$, is only one hundred thousandth of the characteristic Keplerian frequency [99, 100]. This result led to the conclusion that at least in the simplest barotropic disks, the modes cannot underlie any hydrodynamic activity and, in particular, cannot induce turbulence or another variant of enhanced angular momentum transfer to the flow periphery.

2.1.4. Measuring the perturbations. To conclude this section, we discuss the problem of measuring the perturbations. In this paper, we are interested in how strongly some perturbations can grow in a given time interval. To describe this quantitatively, it is necessary to introduce a norm of perturbations that would characterize the amplitude of δv_r, δv_φ, and δh at a given time. It must be a real and positive definite quantity. The total acoustic energy of the perturbation in the disk

$$E = \pi \int \Sigma \left(|\delta v_r|^2 + |\delta v_\varphi|^2 + \frac{|\delta h|^2}{a_*^2} \right) r \, \mathrm{d}r \,, \tag{11}$$

where we integrate over the azimuthal coordinate, is the most natural choice.

After taking the derivative of (11) with respect to time and using (8)–(10), we obtain (also see expression (8) in [97])

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -2\pi \int \frac{\mathrm{d}\Omega}{\mathrm{d}r} \, r\Sigma \, \Re[\delta v_r \delta v_\varphi^*] \, r \, \mathrm{d}r - 2\pi r \Sigma \, \Re[\delta v_r \delta h^*] \Big|_{r_1, r_2} \,, \tag{12}$$

where the symbol $*$ denotes complex conjugation and r_1 and r_2 are the respective inner and outer boundaries of the flow. Here, r_2 can be at infinity. Because $\Sigma \to 0$ at the flow boundaries, the second term in the right-hand side of (12) vanishes, and we see that E can change exactly in a differentially rotating body. Without rotation or for solid-body rotation, E remains constant in time. It is important to note that the increase/decrease in E would imply that the average of the flow amplitudes δv_r, δv_φ, and δh also increases/decreases, because (11) contains squared moduli of these quantities taken with the same signs. For modes, Eqn (12) implies that

$$\frac{\mathrm{d}E}{\mathrm{d}t} \propto \exp\left(2\Im[\omega]t\right) , \tag{13}$$

i.e., small increments obtained for quasi-Keplerian flows allow us to conclude that the total acoustic energy of modes there is $E \simeq \text{const}$ on dynamic ($\sim \Omega^{-1}$) and sound ($\sim (\Omega H/r)^{-1}$) time scales.

Our task is now to understand how E can change over the same time intervals for arbitrary perturbations. To summarize, introducing the perturbation vector $\mathbf{q}(t)$ as the collection of functions $\{\delta v_r(r), \delta v_\varphi(r), \delta h(r)\}$ taken at some time t, the norm of the perturbation can be chosen as

$$\left|\left|\mathbf{q}(t)\right|\right|^2 = E(t) \,. \tag{14}$$

2.2 Local approximation: transition to shear harmonics

The easiest solution of the problem formulated above can be obtained in the local space approximation, where the characteristic scale of perturbations λ is assumed to be a small fraction of some fiducial radial coordinate r_0 around which the dynamics of perturbations are studied, $\lambda \ll r_0$. We introduce the new radial variable $x \equiv r - r_0 \ll r_0$ and the new azimuthal variable $y \equiv r_0(\varphi - \Omega_0 t) \ll r_0$, where $\Omega_0 \equiv \Omega(r_0)$ is the angular velocity of rotation of the new coordinate system. In Eqns (8)–(10), only leading terms in small x are retained. In practice, this means that only the dependence linear in x should be taken into account in the angular velocity profile,

$$\Omega = \left.\frac{\mathrm{d}\Omega}{\mathrm{d}x}\right|_{r_0} x = -q\Omega_0 \, \frac{x}{r_0} \ll \Omega_0 \,, \tag{15}$$

where $q \equiv -(r/\Omega)(\mathrm{d}\Omega/\mathrm{d}r)|_{r=r_0}$ and $\Omega(x=0) = 0$ because we are working in the frame rotating with the angular velocity Ω_0. The corresponding linear background velocity is $v_y^{\mathrm{loc}} = r_0\Omega = -q\Omega_0 x$.

Next, in the right-hand side of Eqns (8)–(10), we keep only terms through the order $\sim x/\lambda$ and drop the terms $\sim x/r_0$ and smaller terms. For clarity, we also write the coefficient before δv_r in the term in (9) that includes κ^2:

$$-\frac{\kappa^2}{2\Omega} = -2\Omega - r\,\frac{\mathrm{d}\Omega}{\mathrm{d}r} = 2q\Omega_0\,\frac{x}{r_0} + (r_0 + x)\,\frac{q\Omega_0}{r_0}$$
$$= 3q\Omega_0\,\frac{x}{r_0} + q\Omega_0\,.$$

It suffices to take only the term $q\Omega_0$ into account. Next, bearing in mind that the new reference frame is not inertial, it is necessary to add the perturbed Coriolis force components $2\Omega_0\delta v_\varphi$ to the right-hand side of (8) and $-2\Omega_0\delta v_r$ to the right-hand side of (9).

After substituting $\mathrm{i}m \to \partial/\partial\varphi$ in system (8)–(10), i.e., after returning to an arbitrary dependence of the Eulerian perturbations on φ and letting u_x, u_y, and W denote the local analogs of perturbations of the velocity components, we arrive at the equations

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x\,\frac{\partial}{\partial y}\right)u_x - 2\Omega_0 u_y = -\frac{\partial W}{\partial x}\,, \tag{16}$$

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x\,\frac{\partial}{\partial y}\right)u_y + (2-q)\Omega_0 u_x = -\frac{\partial W}{\partial y}\,, \tag{17}$$

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x\,\frac{\partial}{\partial y}\right)W + a_*^2\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y}\right) = 0\,. \tag{18}$$

System of equations (16)–(18) was first derived in [42] (see also [102]), where it is described for different background flow models.[4]

2.2.1 Transition to shear harmonics. A very convenient property of system of equations (16)–(18) is that by a change of variables corresponding to the transition to the co-moving shear reference frame, it is possible to make it homogeneous in both x and y, which, in turn, allows separating arbitrary perturbation into individual spatial Fourier harmonics (SFHs) with certain wave numbers k_x and k_y. We introduce new *dimensionless* variables $x' = \Omega_0 x/a_*$, $y' = \Omega_0(y + q\Omega_0 x t)/a_*$, and $t' = \Omega_0 t$.[5] Such a substitution corresponds to changing the partial derivatives as

$$\frac{a_*}{\Omega_0}\,\frac{\partial}{\partial x} = \frac{\partial}{\partial x'} + qt'\,\frac{\partial}{\partial y'}\,, \qquad \frac{a_*}{\Omega_0}\,\frac{\partial}{\partial y} = \frac{\partial}{\partial y'}\,,$$
$$\Omega_0^{-1}\,\frac{\partial}{\partial t} = \frac{\partial}{\partial t'} + qx'\,\frac{\partial}{\partial y'}\,. \tag{19}$$

Using (19), we obtain a system of equations in which all coefficients depend only on t'. In this system, we now substitute the SFHs written in the form

$$f = \hat{f}(k_x, k_y, t')\exp\left(\mathrm{i}k_x x' + \mathrm{i}k_y y'\right), \tag{20}$$

where f is any unknown variable, $\hat{f}$ is its Fourier amplitude, and k_x and k_y are dimensionless wave numbers along axes x' and y', expressed in units Ω_0/a_*. Changing back to the variables x and y in particular solutions (20) reveals that they represent perturbations periodic in space whose phase forms a plane front with the orientation depending on time for $k_y \neq 0$. The dimensionless wave number along x has the form $\tilde{k}_x(t) \equiv k_x + qk_y t$ and changes with time: the wave vector turns during advection by the shear flow, which was first noted by Kelvin [39] and Orr [40], and that is why the SFHs are often called *shear harmonics*. We note from the very beginning that for $\tilde{k}_x < 0$, the wave vector is directed to the interior of the disk, and on the global scale for Fourier harmonics with a wave number m, this corresponds to the so-called *leading spirals*, whose arms are turned to the disk rotation direction. Conversely, the case $\tilde{k}_x > 0$ corresponds to the *trailing spirals*, whose arms are turned oppositely to the disk rotation. If $k_x < 0$ at the initial instant, the arms of the initially leading spiral are deformed and shortened by the flow, and then the so-called *swing* moment t_s occurs when the wave vector of the SFHs is strictly azimuthal and $\tilde{k}_x(t_\mathrm{s}) = 0$, after which the spiral becomes trailing, and its arms start stretching due to their deformation by the flow (see Fig. 2). This process is well known in the dynamics of stellar galactic disks (see paragraph 6.3.2 in [103]).

Thus, for the SFHs, we arrive at the system of ordinary differential equations

$$\frac{\mathrm{d}\hat{u}_x}{\mathrm{d}t} = 2\hat{u}_y - \mathrm{i}\,\tilde{k}_x(t)\,\hat{W}\,, \tag{21}$$

$$\frac{\mathrm{d}\hat{u}_y}{\mathrm{d}t} = -(2-q)\hat{u}_x - \mathrm{i}\,k_y\hat{W}\,, \tag{22}$$

$$\frac{\mathrm{d}\hat{W}}{\mathrm{d}t} = -\mathrm{i}\left(\tilde{k}_x(t)\hat{u}_x + k_y\hat{u}_y\right), \tag{23}$$

where $\hat{u}_x$ and $\hat{u}_y$ are expressed in units a_*, and $\hat{W}$ in units a_*^2. Here and below, we omit the prime for the time variable notation.

2.2.2 Potential vorticity. Equations (21)–(23) have an important property: the quantity

$$I = \tilde{k}_x(t)\hat{u}_y - k_y\hat{u}_x + \mathrm{i}(2-q)\,\hat{W} \tag{24}$$

is an invariant of motion, as can be easily verified by the direct calculation of $\mathrm{d}I/\mathrm{d}t$.

It turns out that I (up to the factor i) is an SFH of the Eulerian perturbation of a potential vorticity. The potential vorticity $\boldsymbol{\zeta}$, which is by definition the vorticity itself divided by density, $\boldsymbol{\zeta} \equiv \boldsymbol{\omega}/\rho$ (see [104]), is conserved in all fluid elements in planar barotropic flows. Therefore, for its Eulerian perturbation, we have

$$\delta\left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}t}\right) = \frac{\mathrm{d}\delta\boldsymbol{\zeta}}{\mathrm{d}t} + (\delta\mathbf{v}\nabla)\,\boldsymbol{\zeta}_0 = 0\,, \tag{25}$$

where $\boldsymbol{\zeta}_0$ is the potential vorticity of the background flow. Because the velocity fields are planar in both background and perturbed flows, the vorticity has only one nonzero component, the z component, which we consider a scalar in what follows.

Next, by definition (in a nonrotating cylindrical coordinate system), the potential vorticity in the background flow is

$$\zeta_0 = \frac{1}{r\Sigma}\,\frac{\mathrm{d}}{\mathrm{d}r}(\Omega r^2) = \frac{\kappa^2}{2\Omega\Sigma} = \frac{(2-q)\Omega}{\Sigma} \tag{26}$$

and must be constant in the local space approximation in use, because the velocity shear is then constant [cf. (15)]. There-

[4] Even earlier, in the context of lunar dynamics, the local approach was used by Hill to study the motion of matter [101].

[5] Due to the vertical hydrostatic equilibrium in the disk, this means that we express length in units of the disk half-thickness $H = a_*/\Omega_0$.

fore, the second term in the last equality in (25) vanishes, and we see that $\delta\zeta$ is indeed conserved. Clearly, the first two terms in (24) arise due to a perturbation of the vorticity itself, which is equal to the curl of the velocity perturbation, and the third term emerges due to a nonzero density perturbation represented by the dimensionless quantity $\hat{W}$ [the coefficient $2-q$ here arises due to multiplication by a constant background vorticity; cf. (26)].

2.2.3 Inhomogeneous wave equations. Density waves and vortices. We now differentiate Eqn (22) with respect to t and use the relations following from the other two equations, (21) and (23), as well as the definition (24), to obtain the new equation

$$\frac{\mathrm{d}^2\hat{u}_y}{\mathrm{d}t^2}+K(t)\hat{u}_y=\tilde{k}_x(t)I, \tag{27}$$

where $K(t)\equiv\tilde{k}_x^2(t)+k_y^2+2(2-q)$. It follows that (27) is a *decoupled wave* equation for the azimuthal velocity component perturbation $\hat{u}_y$, with the *inhomogeneous part* $\sim I$ [62].

In a similar way, from (21) and (23), we derive two equations of the same type:

$$\frac{\mathrm{d}^2\hat{u}_x}{\mathrm{d}t^2}+K(t)\hat{u}_x+2\mathrm{i}qk_y\hat{W}=-k_yI, \tag{28}$$

$$\frac{\mathrm{d}^2\hat{W}}{\mathrm{d}t^2}+K(t)\hat{W}+2\mathrm{i}qk_y\hat{u}_x=-2\mathrm{i}I, \tag{29}$$

which can be decoupled by changing the variables as $\hat{u}_\pm=(\hat{u}_x\pm\hat{W})/2$ [64].

We consider Eqn (27), for example, in more detail. Its general solution is the sum of the general solution of the corresponding homogeneous equation and a particular solution of the inhomogeneous equation. We first consider both these solutions in the solid-body rotation limit, i.e., without the shear, $q=0$. Then all coefficients in (27) become constant and

• the homogeneous equation has particular fundamental solutions $\hat{u}_y^{\mathrm{dw}}\propto\exp(\pm\mathrm{i}\omega t)$ with the frequency $\omega=\sqrt{K}$, corresponding to density waves propagating in opposite directions,

• a particular solution with the nonzero right-hand side can be taken as the constant $\hat{u}_y^{\mathrm{v}}=(k_x/K)\,I$. In other words, u_y^{v} corresponds to the zero frequency $\omega=0$ and represents a static perturbation. This perturbation, as we can see, has a nonzero vorticity and corresponds to a vortex (it can be shown that the divergence of the velocity perturbation for this solution vanishes by taking a similar solution for $\hat{u}_x$ from equation (28), $\hat{u}_x^{\mathrm{v}}$, and verifying that $k_x\hat{u}_x^{\mathrm{v}}+k_y\hat{u}_y^{\mathrm{v}}=0$).

2.2.4 Amplification of density waves. With a nonzero shear taken into account, the density wave frequency becomes a function of time. For example, for leading/trailing spirals, this frequency gradually decreases/increases with the simultaneous wavelength increase/decrease, which in the absence of viscosity leads to a monotonic decrease/increase in the energy and amplitude of the density waves. Such a growth of the density wave amplitude was studied in [58, 105]. The reason can be understood from the fact that due to the axial symmetry of the background flow, the canonical angular momentum of the wave, J_{c}, must be conserved (see [92]). Hence, we find that in accordance with Eqn (522) in [92], the canonical energy $E_{\mathrm{c}}\sim\omega J_{\mathrm{c}}$, *linearly* increases after from some sufficiently late time, because $\omega=\sqrt{K}$ (see above). The conservation of J_{c} for the local perturbation considered here is discussed in paragraph 3.2 in [64]. Unlike J_{c}, the canonical energy itself is no longer conserved in this case because the time-dependent frequency makes the problem inhomogeneous in time. This growth of (or decrease in) the energy, despite the wave frequency ω involved here, is already essentially nonmodal because ω is a function of time, which, in turn, is connected exactly to the deformation of SFHs by the shear flow.

In this paper, however, we are more interested in the 'classic' variant of nonmodal growth, which is called 'transient' in the literature. In the simplest model considered here, it is represented by the vortex solution, which for $q\neq0$ becomes dynamical and, in contrast to waves, is aperiodic.

2.2.5 Vortex existence criterion. Before discussing the behavior of the vortex solution in detail, we analyze the justification of the separation of perturbations into waves and vortices made above in the presence of a shear. Indeed, immediately after $\tilde{k}_x$ becomes variable, the solution $\hat{u}_y^{\mathrm{v}}$ no longer satisfies Eqn (27) exactly because a nonzero second derivative of $\hat{u}_y^{\mathrm{v}}$ appears. Moreover, as $\tilde{k}_x\to0$, Eqn (27) becomes homogeneous in the limit, and its solution describes density waves only. The region in which $\tilde{k}_x\to0$ corresponds to the swing of the SFHs, and we thus see that the vortex solution there becomes ill-defined: the vortex must share wave properties. This means that we can no longer neglect the second time derivative in Eqn (27) for slowly evolving solutions. In other words, $\hat{u}_y^{\mathrm{v}}$ cannot be considered a solution of Eqn (27), even approximately. We discuss the criterion of the possibility of the decoupling of density waves and vortices in a shear flow in more detail.

For this, we use the fact that the vortex dynamics are possible only in subsonic flows (see [18], the end of paragraph 10). In the considered case of an infinite flow, this means that the difference in the fluid velocity on the characteristic scale of the problem must be smaller than the sound velocity. The characteristic spatial scale is determined by the instant spatial period of the SFHs in the radial direction, $\lambda_x\sim H|\tilde{k}_x|^{-1}$. Because infinitesimal perturbations are considered here, its is sufficient to apply the condition of the vortex dynamics for the background flow, and then the velocity difference is given simply by the change in the flow azimuthal velocity; for a flow with constant shear, we thus obtain

$$\frac{\lambda_x q\Omega_0}{a_*}=\frac{q}{|\tilde{k}_x|}\ll1. \tag{30}$$

Hence, the spatial radial period of the vortex harmonics must be smaller than the disk thickness. It is important to note that condition (30) does not directly contain the azimuthal wave number k_y, and hence perturbations can be vortex-like even if their azimuthal spatial scale exceeds the disk thickness. In this connection, it is very important to consider the case of the initially leading spirals, i.e., SFHs with $k_x<0$. For such spirals, the swing occurs at

$$t_{\mathrm{s}}=-\frac{k_x}{qk_y}>0, \tag{31}$$

i.e., when $\tilde{k}_x=0$. Clearly, if the initial spiral is vortex-like, and therefore $k_x\gg1$, and its evolution is initially described by

an approximate solution $\hat{u}_y^{\rm v}$, then the vortex approximation is invalid in some time interval around $t_{\rm s}$, and the full equation (27) must be integrated. We call this time interval 'the swing interval' and obtain the condition under which its duration is much shorter than the characteristic time of evolution of the SFHs determined by the time of the spiral unwinding, $t_{\rm s}$ (see [71]).

The instant at which the vortex approximation breaks down can be estimated from the limit case of the equality in condition (30):

$$t_{\rm s_1,\,s_2} = t_{\rm s}\left(1 \pm \frac{q}{k_x}\right), \tag{32}$$

whence we see that the swing interval is much shorter than the evolution time of the entire vortex spiral, $t_{\rm s_2} - t_{\rm s_1} \ll t_{\rm s}$, if

$$|k_x| \gg 2q\,, \tag{33}$$

which does not contain k_y. Condition (33) implies that in order to study the vortex dynamics, we can use the solution $\hat{u}_y^{\rm v}$ whenever the spiral is sufficiently strongly wound at the initial instant irrespective of the value of k_y, i.e., in both the truly short-wave limit $k_y \gg 1$ and the long-wave limit $k_y \ll 1$. In the latter case, the vortices are referred to as 'large-scale'.

Here, we exclude the case $k_y \sim 1$ because, as was shown numerically in [58, 62] and analytically studied in the WKB approximation in [64], during the swing, the vortex then additionally generates a pair of density waves corresponding to trailing spirals and propagating inside and outside the disk. This process is asymmetric because density wave generation is only possible by vortices, and not vice versa. In [64], analytic expressions for the amplitude and phase of the generated wave were obtained. It was shown that its amplitude is proportional, first, to the vortex vorticity I and, second, to the combination $\epsilon^{-1/2}\exp(-4\pi/\epsilon)$ (see formula (53) in [64]). Here,

$$\epsilon = \frac{qk_y}{k_y^2 + \kappa^2/\Omega_0^2} \tag{34}$$

is the small WKB parameter, where, we recall, $\kappa^2/\Omega_0^2 = 2(2-q)$. Expression (34) implies that the excitation of density waves is exponentially suppressed in both short-wave and long-wave limits and is significant only for $k_y \sim 1$ (we clarify that we do not consider the extreme cases where $q \ll 1$ and therefore $\epsilon \ll 1$ even for $k_y \sim 1$, or $q \to 2$ and hence $\epsilon \gtrsim 1$ even for $k_y \ll 1$).

Thus, the vortex solution of Eqn (27) exists when condition (33) holds together with the requirement $k_y \ll 1$ or $k_y \gg 1$, which excludes density wave generation with nonzero vorticity during the swing of a vortex SFH. At the same time, these restrictions provide a criterion for the decoupling of waves and vortices in the perturbed flow. Indeed, under such constraints, density waves with zero vorticity propagate in the flow independently of vortices and represent the high-frequency branch of solutions of Eqn (27) with a zero right-hand side. Similarly, for example, sound and wind exist independently in Earth's atmosphere.

2.2.6 Vortex solution. Below, we only consider the evolution of a vortex SFH in a shear flow. To conclude Section 2.2, we also obtain vortex solutions for $\hat{u}_x$ and $\hat{W}$. This can be done most easily by neglecting second time derivatives of $\hat{u}_x$ and $\hat{W}$ in Eqns (28) and (29), as has been done with Eqn (27) to obtain $u_y^{\rm v}$. Thus, for all three quantities, we have

$$\hat{u}_x^{\rm v} = -\frac{K+4q}{K^2+4q^2k_y^2}\,k_y I\,, \tag{35}$$

$$\hat{u}_y^{\rm v} = \frac{\tilde{k}_x}{K}\,I\,, \tag{36}$$

$$\hat{W}^{\rm v} = 2{\rm i}\,\frac{qk_y^2 - K}{K^2+4q^2k_y^2}\,I\,. \tag{37}$$

It is important to note that the existence of an aperiodic vortex solution in form (35)–(37) is possible because of the main simplifying assumption about the local constant velocity shear, which provides the existence of the time invariant I. This enables us to reduce the system of three homogeneous first-order equations (21)–(23) to one inhomogeneous second-order equation (27) (other dynamical variables can be obtained from the known solution $\hat{u}_y(t)$), which gives two independent wave solutions (the general solution of the corresponding homogeneous equation) and one aperiodic vortex solution [particular solution (27)]. However, with the velocity shear gradient in the flow taken into account, the invariant I disappears, and the reduction of system of equations (21)–(23) becomes impossible, and from this system we need to obtain three independent solutions directly, two of which, as before, correspond to the density waves, and the third describes the vortex wave called the Rossby wave (see the discussion in paragraph 4 in [62]).[6]

2.3 Vortex amplification factor

To measure the growth of local perturbations, the average density of their acoustic energy can be taken as a local analog of norm (11):

$$E = \frac{1}{2\bar{S}}\int_S \left(\left(\Re[u_x]\right)^2 + \left(\Re[u_y]\right)^2 + \frac{\left(\Re[W]\right)^2}{a_*^2}\right){\rm d}x\,{\rm d}y\,, \tag{38}$$

where $\bar{S}$ is the area of the integration region S.

After substituting dimensionless SFHs (20) in (38) and integrating over their spatial period, we obtain the local variant of norm (14):

$$||\mathbf{q}||^2 = \frac{1}{2}\left(|\hat{u}_x|^2 + |\hat{u}_y|^2 + |\hat{W}|^2\right). \tag{39}$$

Using the vortex solution for SFHs (35)–(37), we obtain the norm in the form

$$||\mathbf{q}||^2 = \left(\frac{\tilde{k}_x^2}{K^2} + \frac{4+k_y^2}{K^2+4q^2k_y^2}\right) I^2\,. \tag{40}$$

In what follows, the main quantity characterizing the perturbation dynamics is the growth factor

$$g(t) \equiv \frac{\left|\left|\mathbf{q}(t)\right|\right|^2}{\left|\left|\mathbf{q}(0)\right|\right|^2}\,, \tag{41}$$

which is the perturbation norm divided by its initial value.

[6] See book [106], paragraph 43, for a discussion of Rossby waves arising due to the gradient of velocity shear (the gradient of vorticity) in an incompressible rotating flow.

Short-wave perturbations. For $k_y \gg 1$, we can in any case omit the 4 in the numerator of the second term in (40), the term $4q^2k_y^2$ in the denominator of the second term, and the term $2(2-q) = \kappa^2/\Omega_0^2$ in K. Then

$$g \approx \frac{k_x^2 + k_y^2}{\tilde{k}_x^2 + k_y^2}\,, \tag{42}$$

which is the result obtained in [56] (also see formula 4 in [61]). Expression (42) shows that the SFH initially taken as a leading spiral with $k_x < 0$ increases in amplitude until the instant (31), and at the swing moment, when $\tilde{k}_x = 0$, reaches a maximum in the norm, and then decays. The energy transfer from the background flow to perturbations is described in detail in terms of fluid particles in [107] (see Fig. 2 therein). It is based on the mechanism of entrainment of fluid particles by the main flow as they move into the region with a different shear velocity (see book [6], paragraph 2.3.3 for more details). Here, the interaction of particles with each other is important, which eventually results in the growth of their additional velocity relative to the main flow.

2.3.1 On the transient growth mechanism. Here, we provide an additional consideration clarifying the transient growth mechanism. As mentioned in the Introduction and discussed in Section 2.2, a differentially rotating flow shortens the length of the leading spiral arms of a transiently growing vortex until the swing instant (see Fig. 2). Due to the barotropicity of the perturbed flow, the velocity circulation along a fluid contour coinciding with the boundary of spiral arms must be constant. Consequently, the contour shortening must lead to a compensating increase in gas velocity along the boundary of the spiral. We consider this heuristic argument more rigorously in the local space limit (see the layout in Fig. 3). We calculate the velocity circulation for the simplest fluid contour. Without perturbations, this is naturally a parallelogram with one pair of sides (call them the base of the parallelogram) running along the background stream lines, i.e., parallel to the y axis and symmetric on both sides of the level $x = 0$. The condition that these sides move synchronously with the fluid automatically implies that the entire contour is comoving with the background flow, since the velocity in the flow is linear in x. We now pass to the reference frame comoving with the shear, in which Eqns (21)–(23) were written: in this frame, the background velocity and the velocity circulation along the given contour are zero. Next, with small perturbations taken into account, the velocity circulation must change, strictly speaking, for two reasons: first, a velocity perturbation $\mathbf{u}$ arises in the shear reference frame and, second, even the contour taken at the time $t = 0$ as a parallelogram starts being deformed due to additional shifts caused by perturbations. But in the second case, only the addition due to the corresponding change in the background velocity circulation is important for small perturbations considered here. But this addition is absent because the background velocity is zero at all points in the shear reference frame. Hence, all we need to do is to calculate the circulation $\mathbf{u}$ along a contour comoving with the background flow. At the instant $t = 0$, we take it such that the parallelogram sides coincide with the SFH front lines separated by the phase π (see Fig. 3, where the initial front direction is indicated by the wave vector $\mathbf{k}_0$). In the shear frame, by definition, an SFH has constant spatial phase front lines, and hence it is clear that they continue to coincide with the contour sides at times $t > 0$. We next note that we are considering the case $k_y \gg 1$; therefore, $\hat{W}^{\nu} \to 0$, and from (23) we derive the orthogonality condition $\mathbf{u} \perp \mathbf{k}$. Consequently, the velocity perturbation is directed along the parallelogram sides and agrees with the chosen orientation (the direction of going around the parallelogram). As regards the parallelogram bases, their contributions to the circulation cancel, because the projection of the velocity $\mathbf{u}$ does not change along them, but the direction of going around is opposite. Hence, the perturbed flow circulation in the comoving shear frame for the left contour in Fig. 3 is

$$\mathcal{C}\big|_{t=0} = 2\Delta y \left(1 + \frac{k_y^2}{k_x^2}\right)^{1/2} |\mathbf{u}|_{t=0}\,.$$

For the right contour in Fig. 3 taken at the spiral swing instant, we similarly find

$$\mathcal{C}\big|_{t=t_s} = 2\Delta x\, |\mathbf{u}|_{t=t_s}\,.$$

By equating these two expressions, we see that for a vortex SFH with $k_y \gg 1$, the circulation conservation law yields

$$g(t_s) = \frac{\left|\mathbf{u}(t_s)\right|^2}{\left|\mathbf{u}(0)\right|^2} = \frac{k_x^2 + k_y^2}{k_y^2}\,. \tag{43}$$

This coincides with the result following from (42) for the spiral swing time.

Therefore, we are quite certain that the transient growth of a vortex is in fact due to its perimeter shortening (its 'size') by the background shear flow with a constant velocity circulation, $\mathcal{C} = \text{const}$, along its perimeter. It is important to note that $\mathcal{C}$, as well as the corresponding vorticity flux, is a measure of the vortex rotation. Therefore, it is appropriate to compare it with a contracting body whose angular momentum is conserved, because in that case the body angular velocity increases inversely with the moment of inertia, $\omega_{\text{rot}} \propto I_{\text{rot}}^{-1}$, and the rotation energy $E_{\text{rot}} = 1/2 I_{\text{rot}}\omega_{\text{rot}}^2 \propto I_{\text{rot}}^{-1}$ increases with time. In our case, the background flow shortens the vortex size and imparts kinetic energy to the vortex.

Finally, we also note that because the differential rotation is purely shear, i.e., occurs with zero divergence of the

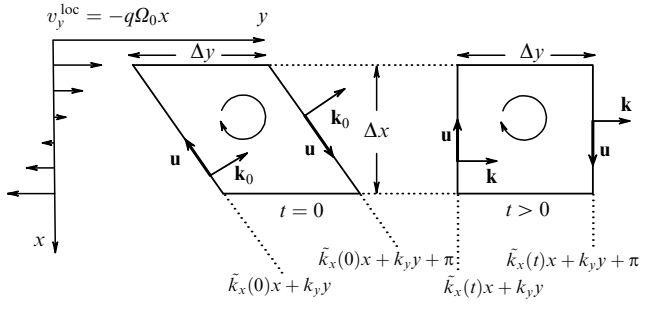

Figure 3. Illustration of the physical reasons for the transient growth of two-dimensional vortices in the local space limit (see Section 2.2). The case of a short-wave ($k_y \gg 1$) vortex SFH with $k_x < 0$. Shown is a liquid contour comoving with the background flow at two instants: the initial time $t = 0$ and the time of the SFH swing when $\tilde{k}_x = 0$. At $t = 0$, the contour has the form of a parallelogram with one pair of sides along the y axis symmetric with respect to $x = 0$ and the other pair along two SFH fronts, with the phase difference π between them. $\mathbf{u}$ is the velocity perturbation vector, $\mathbf{k}_0$ and $\mathbf{k}$ show the SFH wave vector at different instants. Δx and Δy are the parallelogram height and base.

background flow, the area subtended by the contour considered above must remain constant. Indeed, the area of the parallelogram is the product of its base (which is constant because the flow is homogeneous in y) and its height (which is constant because there is no radial background velocity). Therefore, due to the constant $\mathcal{C}$ and, hence, the vorticity perturbation flux through the contour, the vorticity perturbation itself is constant. The same conclusion was obtained in Section 2.2 in the discussion of invariant (24).

2.3.2 Estimation of the optimal growth. Knowing the physical mechanism of the transient vortex growth, we return to expression (42) for the growth factor in the short-wave case. Clearly, the growth factor of an individual SFH is a function of three arguments, $g = g(k_x, k_y, t)$. However, it is possible to consider a more general characteristic of the transient dynamics, which is called the *optimal growth* G of perturbations. By definition,

$$G \equiv \max_{\forall k_x} \{g\} . \tag{44}$$

Formula (44) gives the maximum possible amplification among all vortices with a given k_y that can occur in a time interval t. We note that below we use an analogue of (44) for the global space problem described by system of equations (8)–(10) [see formula (90)]. There, G is determined for perturbations with a fixed azimuthal wave number m.

There are rigorous mathematical algorithms to search for the optimal growth, which we discuss in the next section. Here, for analytic estimates in the local space limit, it suffices to recognize that because the growth factor $g(k_x, k_y, t)$ of a separate SFH has a maximum at $\tilde{k}_x = 0$, it is reasonable to assume that G can be estimated as

$$G \approx g(k_x = -k_y q t) ; \tag{45}$$

in other words, we assume that, of all the SFHs with a given k_y, the harmonics that show swing at an instant t reach maximum growth by this instant.

Using definition (45), from (42) we obtain the simple expression

$$G_1 \approx (qt)^2 , \tag{46}$$

which can also be found in [61] [see formula (5) therein]. We note that corrections to G_1 due to a nonzero vertical projection of the wave vector and a finite value of k_y were also obtained in that paper. As we see, in a sufficiently long time, it is possible to reach arbitrarily large amplitude growth of small-scale vortices $k_y \gg 1$. This growth, however, is power-law and not exponential, as would be expected in a modal instability of the flow.

Long-wave perturbations. We now turn to the other limit case where $k_y \ll 1$ and the azimuthal space period of an SFH is much larger than the disk thickness (see [71]). In the second term in (40), we then omit k_y^2 in the numerator and $4q^2k_y^2$ in the denominator, and also assume that $K = \tilde{k}_x^2 + \kappa^2/\Omega_0^2$. Here, by virtue of condition (33), we see that $||\mathbf{q}(0)||^2 \approx k_x^{-2}$.

For the SFH growth factor, we then obtain

$$g \approx k_x^2 \, \frac{\tilde{k}_x^2 + 4}{(\tilde{k}_x^2 + \kappa^2/\Omega_0^2)^2} . \tag{47}$$

This quantity increases as $\tilde{k}_x$ decreases with time, i.e., similarly to short-wave vortices, the transient growth occurs for $k_x < 0$. We note that the maximum g attained during the spiral swing is now proportional to the square of k_x itself, and not to the square of the ratio k_x/k_y, as in the case of short-wave vortices [cf. (42)]. In addition, another important difference is that g now depends on the epicyclic frequency as κ^{-4}. Such a strong dependence can be important in disks with a higher-than-Keplerian angular velocity gradient: in thin disks, this can occur in the inner regions of relativistic disks, where $\kappa \to 0$ in approaching their inner boundary.

With definition (45), we obtain the corresponding optimal growth factor from (47):

$$G_2 \approx \frac{4\Omega_0^4}{\kappa^4} \, k_y^2 (qt)^2 . \tag{48}$$

We note that both (46) and (48) are valid only for sufficiently large times, because we used the condition $k_x = -qk_y t$ to obtain this expression, but the condition $k_x \gg 1$ must hold at the same time, as is required by (33). Formula (48) shows that for rotation profiles weakly differing form the Keplerian one, when $\kappa \sim \Omega_0$, for equal time intervals $G_2 \ll G_1$, because the azimuthal wave number explicitly entering the optimal growth factor is small, $k_y \ll 1$.[7] Therefore, in the local space limit considered here, small-scale vortices take energy from the flow more efficiently than large-scale ones. However, it is interesting to see which of them can display the highest growth over the entire time interval. In an inviscid flow, $G_{1,2} \to \infty$, mostly due to small-scale SFHs, as we just noted. Nevertheless, a shear flow can have noticeable effective viscosity, for example, due to some weak turbulence. Then the dependence $G(t)$ turns out to have the global maximum $G_{\max}$ corresponding to the maximum possible nonmodal growth of perturbations irrespective of the time intervals we have considered so far. Physically, the decrease in $G(t)$ after some long time occurs because more tightly wound spirals have larger swing times t_s. This, in turn, means a smaller radial scale of perturbations and hence a smaller dissipation time of perturbations due to viscosity. Ultimately, the leading transient spirals start decaying faster than they grow due to the unwinding by the flow. It is the values $G_{\max}$ in the cases $k_y \gg 1$ and $k_y \ll 1$ that we compare below.

2.3.3 Taking the viscosity into account. The effect of viscosity on the maximum possible transient growth of vortices can be estimated as follows. For sufficiently long time intervals $qt \gg 1$, we have $k_x \gg k_y$ for any of the two limits of k_y we are considering. Therefore, in a shear-free flow, the spiral would decay in the characteristic viscous time $\Delta t_\nu \sim \lambda_x^2/\nu$, where ν is the kinematic viscosity coefficient. Using the standard viscosity parameterization by the Shakura–Sunyaev α parameter, $\nu = \alpha a_* H$, we find that $\Delta t_\nu \sim (\Omega_0^{-1}\alpha k_x^2)^{-1}$ rapidly decreases with increasing $|k_x|$. At the same time, the larger $|k_x|$ is, the longer the transient growth time of the spiral, $\Delta t_{tg} \sim |k_x/(qk_y)|$. Simultaneously with the shear arising in the flow, the spiral starts unwinding, and therefore the viscous dissipation is delayed. Hence, the equality of these characteristic times, $\Delta t_{tg} = \Delta t_\nu$, gives the

[7] In Section 4.2 below, we calculate G in the global problem (see Fig. 11), which implies that as $m \to 1$, the difference in the transient growth rate between vortices with an azimuthal wavelength shorter or longer than the disk thickness is significantly smaller.

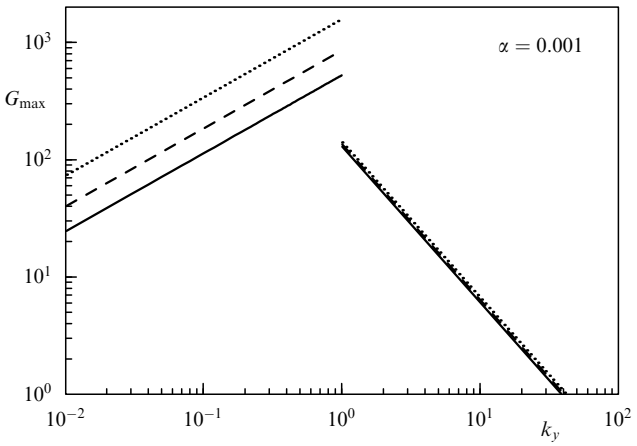

Figure 4. Estimate of the maximum possible transient growth of acoustic energy in a disk with the efficient viscosity $\alpha = 0.001$. The respective solid, dashed, and dotted lines correspond to $q = 1.5$, 1.6, and 1.7. The three curves to the right and left are obtained using respective formulas (50) and (51).

lower bound for the duration of the transient growth of vortices in a viscous flow. Using it, we obtain

$$\max(\Delta t_{\rm tg}) \gtrsim \alpha^{-1/3}(qk_y)^{-2/3}\,. \tag{49}$$

It can be verified that expression (49) reproduces the estimate in [61] [see formula (81) therein].

The upper bound for the optimal growth time (49), $G_{\max} \equiv G(\max(\Delta t_{\rm tg}))$, is given by its inviscid value taken for G_1 or G_2. We then work out that for $k_y \gg 1$,

$$(G_{\max})_1 \approx \alpha^{-2/3} q^{2/3} k_y^{-4/3} \tag{50}$$

(see also formula (83) in [61]). At the same time, for $k_y \ll 1$, we have

$$(G_{\max})_2 \approx \frac{4\Omega_0^4}{\kappa^4}\,\alpha^{-2/3} q^{2/3} k_y^{2/3}\,. \tag{51}$$

This result is shown in Fig. 4 for some small α and several shears q: Keplerian and super-Keplerian. We see that even for the Keplerian shear, when $\kappa = \Omega_0$, for k_y different from 1, $(G_{\max})_2 \gtrsim (G_{\max})_1$. This occurs because the large-scale vortices are much less dissipative, which more than compensates for their low growth rate compared to low-scale vortices. We also note that despite $(G_{\max})_2$ decreasing with decreasing k_y, this occurs at a lower rate than when $(G_{\max})_1$ decreases with increasing k_y. As a result, the integral transient growth of large-scale vortices at all k_y increases in comparison with small-scale ones. An even more significant advantage of large-scale vortices appears for super-Keplerian shears, when $q > 3/2$, due to $(G_{\max})_2 \propto \kappa^{-4}$ [see the comment after formula (47)]. Clearly, a deviation from $q = 3/2$ by several percent would increase the transient growth rate of perturbations several-fold.

As discussed in [71], estimate (51) is in reasonable agreement with exact calculations of the optimal growth rate in thin disks in the global space limit for low azimuthal wave numbers m. Thus, large-scale vortices are also able to provide additional transportation of the angular momentum to the periphery of a disk with weak turbulence already present.

In Section 3, we provide a rigorous mathematical justification of algorithms to search for the most rapidly growing perturbations in shear flows. Such perturbations are called optimal, and the corresponding amplification, as we already mentioned, is referred to as the optimal growth G. The solutions presented in the Introduction and shown in Figs 1 and 2 were obtained using one of these algorithms. We also provide another example of the calculation of G by solving the general system of equations (8)–(10) in a geometrically thin disk (see Fig. 11 below). When discussing mathematical aspects of the nonmodal dynamics of perturbations in shear flows, already in the introductory part to the next section, we see that the transient growth phenomenon can be treated as a consequence of the *nonorthogonality* of perturbation modes, which is evident, in particular, from a consideration of simple analogs presented in Figs 5 and 6.

3. Search for optimal perturbations

3.1 Definition and properties of singular vectors

General solutions of the initial value problem of the small perturbation evolution described by general equations (1) and (2) can be conveniently studied using abstract concepts of the functional space of the so-called state vectors of the system, as well as the notion of linear operators acting on these vectors. In Section 2.1, in application to system of equations (8)–(10), we introduced a particular case of the state vector as a set of azimuthal Fourier harmonics $\mathbf{q}(t) \equiv \{\delta v_r(r), \delta v_\varphi(r), \delta h(r)\}$ of Eulerian perturbations taken at some fixed instant t. In this section, we assume the initial general case where $\mathbf{q}(t) \equiv \{\delta\mathbf{v}(\mathbf{r}), \delta h(\mathbf{r}), \delta\rho(\mathbf{r})\}$. We consider some properties of a dynamical operator $\mathbf{Z}$ acting in the Banach space of vectors $\mathbf{q}$ and corresponding to system (1), (2). This operator transforms the initial perturbation vector $\mathbf{q}(0)$ into a subsequent vector $\mathbf{q}(t)$, and hence in the operator form, the system of equations can be written as

$$\mathbf{q}(t) = \mathbf{Z}\mathbf{q}(0)\,. \tag{52}$$

All functions entering $\mathbf{q}(t)$ are assumed to be infinitely smooth and to have a uniformly bounded derivative in their domain. The last condition follows from physical considerations: in realistic gas flows, there cannot be perturbations with arbitrarily small wavelengths. In addition, due to the linearity of the problem, all vectors $\mathbf{q}(0)$ are assumed to have the unit norm at the initial instant.

In this section, we show that the general assumptions given above imply important properties of the operator $\mathbf{Z}$. For example, we show that the norm of all initial vectors $\mathbf{q}(0)$ can grow by the time t only by less than some factor. We present two methods for calculating this perturbation growth limit. In addition, we show that in the space of initial conditions, there is an orthonormal basis that can be found by solving the eigenvalue problem for some operator *different* from $\mathbf{Z}$.

3.1.1 Continuity of the dynamical operator. Continuity is the first important property of the operator $\mathbf{Z}$. To see this, we write the operator in the integral form:

$$\mathbf{Z}\mathbf{q}(0) = \mathbf{q}(t) = \mathbf{q}(0) + \int_0^t \left(\mathbf{M}_1(s)\mathbf{q}(s) + \mathbf{M}_2(s)\,\frac{\partial\mathbf{q}(s)}{\partial\mathbf{r}}\right) \mathrm{d}s\,. \tag{53}$$

Here, we introduce the matrices $\mathbf{M}_1(s)$ and $\mathbf{M}_2(s)$ composed of coefficients in the dynamical equations before the corresponding spatial derivatives of $\mathbf{q}$. The explicit form of these matrices can be obtained from Eqns (1) and (2). The number of rows and the number of columns in the matrices are respectively equal to the number of quantities forming the state vector $\mathbf{q}$ and the number of spatial variables. Because the quantities describing the background flow are bounded and continuous, all elements of the matrices $\mathbf{M}_1(s)$ and $\mathbf{M}_2(s)$ are also bounded and continuous in the domain of the operator $\mathbf{Z}$. In addition, we note that if viscous forces are included, one more term appears in Eqn (53) corresponding to the second time derivative of $\mathbf{q}$. This case can be treated analogously.

We see that operator (53) is a superposition of continuous maps (see [108], Ch. 1) and, hence, is itself a continuous map (see [109], Ch. 2). This implies that it is bounded on bounded sets (see [109], Ch. 4).

Thus, the map defined by (53) is continuous and bounded, which implies that vectors $\mathbf{q}$ are *uniformly bounded* at any time t. We now use this property.

3.1.2 Essentially continuous dynamical operator. The next property of the operator $\mathbf{Z}$ is its essential continuity. We recall the definition of this notion.

Definition 1 (essentially continuous operator). *An operator* $\mathbf{Z}$ *mapping a Banach space* $\mathbf{E}$ *into itself is called essentially continuous if it takes any bounded set to a relatively compact set* ([109], Ch. 4).

To prove the essential continuity of operator (53), it suffices to prove the relative compactness of its range, because the boundedness of its domain was postulated by assuming that all $\mathbf{q}(0)$ have unit norm. We use the Arzelá–Ascoli theorem. According to this theorem, a sequence of continuous functions defined on a closed and bounded interval is relatively compact if and only if this sequence is uniformly bounded and equicontinuous ([109], Ch. 4).

The uniform boundedness was shown above, and for a sequence of differentiable functions to be equicontinous, it is sufficient that their derivatives be uniformly bounded ([110], Ch. 2), which was initially postulated. Thus, we see that the range of the operator $\mathbf{Z}$ is relatively compact, and hence the operator is essentially continuous.

Now, if we introduce an inner product [in physical problems, as a rule, it is introduced such that the norm of a vector coincides with the perturbation energy, as was done, for example, in Eqn (14)], it is possible to define the adjoint operator $\mathbf{Z}^\dagger$ using the Lagrange identity for arbitrary vectors $\mathbf{f}$ and $\mathbf{g}$ (see, e.g., [108], Ch. 1, for more details on the adjoint operators):

$$(\mathbf{Zf}, \mathbf{g}) = (\mathbf{f}, \mathbf{Z}^\dagger \mathbf{g}) \,. \tag{54}$$

Here, if the operator $\mathbf{Z}$ is essentially continuous, so are the adjoint operator $\mathbf{Z}^\dagger$ and the self-adjoint composite operators $\mathbf{ZZ}^\dagger$ and $\mathbf{Z}^\dagger\mathbf{Z}$ ([109], Ch. 4).

3.1.3 Linear operators: from the particular to the general. There can be different linear operators, depending on their properties. We list those that we need below, from the more particular to the more general ones. We start from positive definite operators, for which the scalar product $(\mathbf{Zq}, \mathbf{q}) > 0$ for any vector $\mathbf{q}$. By definition, eigenvalues of a positive definite operator are positive. Indeed, by multiplying both sides of the equation $\mathbf{Zq} = \lambda\mathbf{q}$ by $\mathbf{q}$, we see that its left-hand side is positive, and the right-hand side is the product of the eigenvalue and a positive quantity; hence, the eigenvalue is positive.

Self-adjoint (Hermitian) operators, which coincide with their adjoint operators, $\mathbf{Z} = \mathbf{Z}^\dagger$ ([111], paragraph 14.4), are most frequently used in various physical problems. The eigenvalues of a self-adjoined operator are real ([111], paragraph 14.8).

Self-adjoined operators are a particular case of normal operators. An operator $\mathbf{Z}$ is called normal if it commutes with its adjoint operator: $\mathbf{ZZ}^\dagger = \mathbf{Z}^\dagger\mathbf{Z}$ ([111], paragraph 14.4). All eigenvalues of a normal operator are complex conjugate of its adjoint operator's eigenvalues. Eigenfunctions of the operators $\mathbf{Z}$ and $\mathbf{Z}^\dagger$ coincide. Additionally, the eigenvectors of a normal operator corresponding to different eigenvalues are orthogonal ([111], paragraph 14.8). Therefore, to calculate the *operator norm* of these operators, it is sufficient to find their eigenvalues. We recall that the norm of an operator $\mathbf{Z}$ mapping a Banach space H into itself is the number $||\mathbf{Z}|| = \sup_{\mathbf{x}\in H} (||\mathbf{Zx}||/||\mathbf{x}||)$ ([108], Ch. 1). The norm of the master operator is very useful, because it allows calculating the limit of the vector norm growth under the action of this operator.

For a normal operator, this problem is solved quite easily. To illustrate this, following [49], we consider an important particular case where the operator $\mathbf{Z}$ can be represented as an operator exponential: $\mathbf{Z} = \exp(\mathbf{A}t)$ (see Section 3.3.1 for more details). The operator $\mathbf{A}$ is time-independent, and its eigenvalues are traditionally denoted as $\{-\mathrm{i}\omega_1, -\mathrm{i}\omega_2, \ldots, -\mathrm{i}\omega_N\}$; here, ω can take both real and complex values. In this case, the eigenvalues of $\mathbf{Z}$ are $\{\exp(-\mathrm{i}\omega_1 t), \exp(-\mathrm{i}\omega_2 t), \ldots, \exp(-\mathrm{i}\omega_N t)\}$. Now, we use the definition of the eigenvectors and eigenvalues of an operator by writing it in the matrix form

$$\mathbf{ZX} = \mathbf{XP} \,, \tag{55}$$

where $\mathbf{P}$ is a diagonal matrix with the eigenvalues of $\mathbf{Z}$ and the columns of $\mathbf{X}$ are the eigenvectors of $\mathbf{Z}$ placed in the order of their eigenvalues encountered in $\mathbf{P}$.

From (55), we find the decomposition $\mathbf{Z} = \mathbf{XPX}^{-1}$. Next, we use the submultiplicativity of the operator norm ([111], paragraph 14.2): $||\mathbf{Z}|| \leqslant ||\mathbf{X}||\,||\mathbf{P}||\,||\mathbf{X}^{-1}||$. For orthonormal eigenvectors, the matrix $\mathbf{X}$ is unitary, $\mathbf{XX}^\dagger = \mathbf{I}$, and therefore its norm is unit, $||\mathbf{X}|| = 1$, and $||\mathbf{Z}|| \leqslant ||\mathbf{P}|| = \exp(\omega_{\max} t)$, where $\omega_{\max} = \max_{j\leqslant N} (\Im[\omega_j])$.

Finally, the most general operators are those that do not commute with their adjoint: $\mathbf{ZZ}^\dagger \neq \mathbf{Z}^\dagger\mathbf{Z}$. Eigenvalues of these operators can be both purely real and complex, and eigenvectors are *nonorthogonal* to each other. The nonorthogonality of the eigenvectors complicates the calculation of the operator norm, because the matrix $\mathbf{X}$ introduced above is no longer unitary. For this reason, the energy of a combination of modes is not equal to the sum of the energies of each mode, i.e., the Parceval rule is not valid, and nonzero cross terms appear. In other words, due to the interference in time between nonorthogonal modes, perturbations described by such an operator can increase even if there are no growing modes. This energy growth of perturbations, which is mathematically related to the nonnormality of the dynamical operator, was called the transient growth of perturbations. In the context of stability of hydrodynamical flows, nonnormal operators and examples were discussed in [112], as well as in Sections 3 and 4 in [6].

3.1.4 Simple geometrical example of the nonorthogonality of eigenvectors. A simple geometrical example can illustrate the transient growth mechanism. On the plane (x, y), we introduce two vectors symbolizing two perturbation modes. We write them in the form of two complex numbers $\mathbf{f}_1 = f_0 \exp(-\mathrm{i}\omega_1 t)$ and $\mathbf{f}_2 = f_0 \exp(-\mathrm{i}\omega_2 t + \mathrm{i}\psi)$, where the numbers $\omega_{1,2}$ can also be complex. In this form, the analogy between $\mathbf{f}_{1,2}$ and perturbation modes is the clearest. The real and imaginary parts of each $\mathbf{f}_{1,2}$ yield the x- and y-vector components. Clearly, $\Re[\omega_{1,2}]$ corresponds to the angular velocity with which both vectors rotate in the plane, and $\Im[\omega]_{1,2}$ corresponds to the rate of change in their lengths. Below, we assume that the imaginary parts of $\omega_{1,2}$ are negative, which corresponds to the length shortening of $\mathbf{f}_{1,2}$. We recall that in the case of modes, real parts give angular velocities of the solid-body rotation of the spiral pattern in the flow (see Fig. 1), and imaginary parts give their decay rate, in analogy with a spectrally stable flow. In addition, we assume that at the instant $t = 0$, the vectors have the same length f_0 and the angle between them is ψ.

We now take the vector $\mathbf{q} = \mathbf{f}_1 + \mathbf{f}_2$ and calculate a quantity similar to (41) that gives the rate of change with time of the length of $\mathbf{q}$ squared:

$$g = \frac{1}{2(1+\cos\psi)}\Big[\exp\big(2\Im[\omega_1]t\big) + \exp\big(2\Im[\omega_2]t\big) + 2\exp\big(\Im[\omega_1+\omega_2]t\big)\cos\big(\Re[\omega_1-\omega_2]t+\psi\big)\Big]. \quad (56)$$

This shows that for the angles ψ close to π, the denominator in (56) is small, and any insignificant increase in the numerator leads to a large increase in g. We consider two particular examples. In the first case, we assume that $\Re[\omega_{1,2}] = 0$, and in the second case, that $\Im[\omega_{1,2}] = 0$. For simplicity, we set $\cos\psi \approx -1 + \epsilon$, where $\epsilon \ll 1$.

Then, with $\Re[\omega_{1,2}] = 0$, we see that if we additionally assume a large difference in decrements, $|\Im[\omega_1]| \gg |\Im[\omega_2]|$, after some large time we have

$$g \approx \frac{\exp\big(2\Im[\omega_2]t\big)}{2\epsilon}, \quad (57)$$

which corresponds to $g \gg 1$ on time intervals such that $|\Im[\omega_1]t| \gg 1$ but simultaneously $|\Im[\omega_2]t| \ll 1$. This means that despite the decrease in the length of each individual vector, in the case of strong nonorthogonality (which is characterized by a significant difference between ϵ and 1), their sum experiences a transient growth up to values $\sim \epsilon^{-1}$ (Fig. 5) and only at later times g decreases again at a rate determined by the most slowly decreasing vector. A similar effect takes place for transient perturbations, which can be represented as a sum of decaying modes with zero phase velocity.

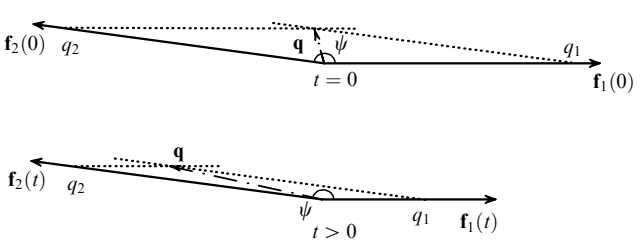

Figure 5. Increase in the sum of two nonorthogonal vectors, $\mathbf{q} = \mathbf{f}_1 + \mathbf{f}_2$, upon shortening their lengths but maintaining the angle between them. It is assumed that $q_1 = q_2 = 1$.

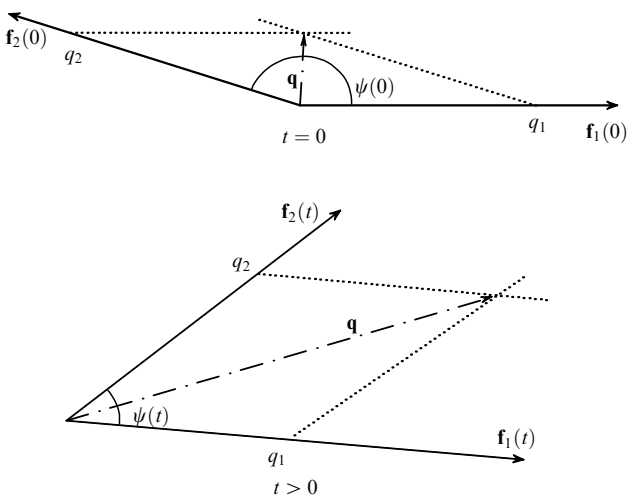

Figure 6. Increase in the sum of two nonorthogonal vectors, $\mathbf{q} = \mathbf{f}_1 + \mathbf{f}_2$, upon maintaining their lengths but changing the angle between them. It is assumed that $q_1 = q_2 = 1$.

In the opposite case $\Im[\omega_{1,2}] = 0$, the following approximate formula can be derived from (56):

$$g \approx \frac{1 - \cos\big(\Re[\omega_1 - \omega_2]t\big)}{\epsilon}, \quad (58)$$

which is valid when the value of the cosine in the numerator is not too close to unity. In contrast to the example with the sum of nonorthogonal vectors with decreasing length (when the length $\mathbf{q}$ first increases to a maximum and then monotonically decreases to zero as $t \to \infty$), it follows that the length of the sum of rotating vectors experiences oscillating growth by returning many times to ever increasing values $\sim \epsilon^{-1}$ in equal time intervals $\sim |\Re\omega_1 - \Re\omega_2|^{-1}$, as is evident from the illustration in Fig. 6. In contrast to the first case, it would be inappropriate to refer to this second possible variant of the mode superposition growth as 'transient growth', as we did, for example, when analyzing local SFHs in Section 2.3. Therefore, it is more appropriate to call it 'nonmodal growth'. One example of such nonmodal growth of a superposition of neutral modes with nonzero phase velocities was studied in [113] and is discussed in Section 3.2.

3.1.5 Singular vectors. We have just demonstrated how the nonorthogonality of the modes leads to transient growth of perturbations. In many physical and astrophysical problems, the evolution of linear perturbations is determined just by nonnormal operators with nonorthogonal eigenvectors. Here, the nonnormality of $\mathbf{Z}$ is provided by a shear in the background flow. We can justify this by deriving the system of *adjoint* dynamical equations corresponding to the action of the adjoint operator $\mathbf{Z}^\dagger$ (see Section 3.4.1).

It follows that knowing only the eigenvalues of a nonnormal operator is insufficient in order to fully describe the possible (transient) growth of perturbations in the system. In addition, at least, the eigenvectors or, more precisely, the pairwise scalar products ('angles') between the eigenvectors in the chosen norm of perturbations must be known. One more potential complication of the problem with a nonnormal dynamical operator is that it is no longer possible to

guarantee the completeness of the set of its eigenvectors and, hence, to guarantee the validity of the solution of the problem when using the eigenvectors as a basis for decomposing an arbitrary perturbation.

For all these reasons, to compute the maximal transient growth rate of perturbations, we use the technique of *singular* values and vectors in what follows. As is to be shown, singular vectors form a complete orthonormal system, which allows using them as a basis to describe the evolution of perturbations. Moreover, singular values, unlike eigenvalues, allow calculating the perturbation energy growth by any given time even for nonnormal operators.

Definition 2 (singular values and vectors). *A nonnegative real number σ is called a singular number of a linear operator* $\mathbf{Z}$ *if there are unit-length vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *such that*

$$\begin{aligned} \mathbf{Z}\mathbf{v} &= \sigma\mathbf{u}\,, \\ \mathbf{Z}^{\dagger}\mathbf{u} &= \sigma\mathbf{v}\,. \end{aligned} \tag{59}$$

The vectors $\mathbf{u}$ *and* $\mathbf{v}$ *are called the respective left and right singular vectors corresponding to the singular value σ.*

We note that the singular values and vectors are related to the eigenvalues and eigenvectors of the composite self-adjoint operators $\mathbf{Z}\mathbf{Z}^{\dagger}$ and $\mathbf{Z}^{\dagger}\mathbf{Z}$. To see this, we act with $\mathbf{Z}^{\dagger}$ on the vector $\mathbf{Z}\mathbf{v}$ and with $\mathbf{Z}$ on the vector $\mathbf{Z}^{\dagger}\mathbf{u}$ and then use the definition:

$$\mathbf{Z}^{\dagger}(\mathbf{Z}\mathbf{v}) = \mathbf{Z}^{\dagger}(\sigma\mathbf{u}) = \sigma\mathbf{Z}^{\dagger}\mathbf{u} = \sigma^2\mathbf{v}\,, \tag{60}$$

$$\mathbf{Z}(\mathbf{Z}^{\dagger}\mathbf{u}) = \mathbf{Z}(\sigma\mathbf{v}) = \sigma\mathbf{Z}\mathbf{v} = \sigma^2\mathbf{u}\,. \tag{61}$$

Hence, vectors $\mathbf{v}$ and $\mathbf{u}$ are eigenvectors of the respective operators $\mathbf{Z}^{\dagger}\mathbf{Z}$ and $\mathbf{Z}\mathbf{Z}^{\dagger}$. The singular value squares are eigenvalues of the composite operators.

The operators $\mathbf{Z}\mathbf{Z}^{\dagger}$ and $\mathbf{Z}^{\dagger}\mathbf{Z}$ are positive definite because for any vector $\mathbf{f}$, the inequalities $(\mathbf{f}, \mathbf{Z}\mathbf{Z}^{\dagger}\mathbf{f}) = (\mathbf{Z}^{\dagger}\mathbf{f}, \mathbf{Z}^{\dagger}\mathbf{f}) > 0$ and $(\mathbf{f}, \mathbf{Z}^{\dagger}\mathbf{Z}\mathbf{f}) = (\mathbf{Z}\mathbf{f}, \mathbf{Z}\mathbf{f}) > 0$ hold. Because all eigenvalues of a positive definite operator are positive, the singular values are real.

Because the operators $\mathbf{Z}\mathbf{Z}^{\dagger}$ and $\mathbf{Z}^{\dagger}\mathbf{Z}$ are self-adjoint and essentially continuous, their limit spectrum consists of one point, zero ([108], Ch. 4). Next, because the limit spectrum of an operator is the union of all points of the continuous spectrum, limit points of the discrete spectrum, and infinite-multiplicity eigenvalues, the essential continuity of the composite operators implies that for any small $\epsilon > 0$, the set of eigenvalues greater than ϵ is discrete.

Therefore, the set of singular values is bounded from above due to the boundedness of operator (53), is discrete, and has the limit point $\sigma = 0$. The singular vectors are usually numbered in the order of their decrease [114], and therefore the perturbation growth by a time t is limited by the first singular value by that time, and the first right singular vector is the perturbation exhibiting this growth.

It hence follows that to calculate the maximum possible perturbation growth rate, it suffices to calculate the first singular value, called the optimal growth in the literature, and the right singular vector corresponding to this value is then the sought (optimal) perturbation demonstrating the maximum possible growth rate. Below, we present two methods for calculating singular values and the corresponding singular vectors.

Another important consequence of the essential continuity of the dynamical operator $\mathbf{Z}$ is the validity of the Hilbert–Schmidt theorem for the operators $\mathbf{Z}\mathbf{Z}^{\dagger}$ and $\mathbf{Z}^{\dagger}\mathbf{Z}$. The theorem states that for any self-adjoint linear operator, there is an orthonormal sequence $\{\boldsymbol{\varphi}_n\}$ of eigenvectors corresponding to eigenvalues $\{\lambda_n\}$ such that each element $\boldsymbol{\xi}$ can be uniquely written in the form

$$\boldsymbol{\xi} = \sum c_k\boldsymbol{\varphi}_k + \boldsymbol{\xi}'\,,$$

where the vector $\boldsymbol{\xi}'$ satisfies the condition $\mathbf{U}\boldsymbol{\xi}' = 0$; here,

$$\mathbf{U}\boldsymbol{\xi} = \sum \lambda_k c_k\boldsymbol{\varphi}_k\,, \qquad \lim_{n\to\infty}\lambda_n = 0\,.$$

It hence follows that the set of singular functions is orthogonal and complete, as a sequence of eigenvectors of a self-adjoint operator, and can be used as a basis for decomposing any perturbation.

3.2 Matrix method for optimal solutions

The first method to calculate singular vectors is conventionally referred to as the *matrix* method. It is based on the singular value decomposition of the matrix of the dynamical operator. As a rule, the set of eigenvectors is used as the basis for calculating the operator matrix.

We note that there is another possibility, which was used, for example, in [50], when the space is covered by a grid of points and each perturbation corresponds to a column of numbers corresponding to the values of the perturbation at these points. A dynamical operator corresponds to a matrix obtained by the difference approximation of derivatives in the dynamical equations. The singular value decomposition of this matrix allows calculating the singular vectors at the grid points. The large size of the operator matrix is a shortcoming of this approach, which requires a long time to calculate the singular value decomposition; an advantage is that it is not necessary to calculate the operator eigenvectors. In this section, we describe the matrix method in the eigenvector basis.

The problem is to find a linear combination of the dynamical operator modes whose norm exhibits the largest growth by a given time. We assume that the sequence of eigenvectors $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3 \ldots \mathbf{f}_N\}$ and the corresponding eigenvalues $\{\exp(-\mathrm{i}\omega_1 t), \exp(-\mathrm{i}\omega_2 t), \exp(-\mathrm{i}\omega_3 t), \ldots, \exp(-\mathrm{i}\omega_N t)\}$ of the operator $\mathbf{Z}$ are known. In the space of linear combinations of the eigenvectors, an arbitrary perturbation vector can be represented in the form (see paragraph 4.3.2 and Section 4.4. in [6] for more details)

$$\mathbf{q} = \sum_{j=1}^{N} \kappa^j \hat{f}_j\,, \tag{62}$$

where the numbers $\{\kappa^1, \kappa^2, \kappa^3, \ldots, \kappa^N\}$ are coordinates of the vector $\mathbf{q}$ in the eigenvector basis. We note that the time dependence of $\mathbf{q}$ is contained in its coordinates.

The scalar product of two vectors $\mathbf{q}$ and $\mathbf{g}$ in this representation can be calculated from the known coordinates using the metric matrix $\mathbf{M}$:

$$(\mathbf{q}, \mathbf{g}) = (\mathbf{q}^{\dagger})^i M_{ij}\,\mathbf{g}^j\,, \tag{63}$$

where the elements of the metric matrix are equal to the scalar products of eigenvectors,

$$M_{ij} = (\mathbf{f}_i, \mathbf{f}_j)\,. \tag{64}$$

The matrix $\mathbf{M}$ is positive definite because the norm of a nonzero vector is always positive.

Now, the problem of calculating the maximum possible perturbation growth reduces to finding the κ^j at which the growth of the perturbation norm determined in accordance with (62) is maximal by a given time moment.

The representation of an operator $\mathbf{Z}$ in the eigenvector basis can be easily calculated by letting this operator act on a basis element:

$$\mathbf{Z}\mathbf{f}_j = \mathbf{f}_j(\tau) = \exp(-\mathrm{i}\omega_j\tau)\,\mathbf{f}_j\,. \tag{65}$$

Therefore, in the set of its basis eigenvectors, an operator can be represented by a diagonal matrix $\mathbf{P}$ with complex exponentials on the main diagonal: $\mathbf{P} = \mathrm{diag}\{\exp(-\mathrm{i}\omega_1\tau), \exp(-\mathrm{i}\omega_2\tau), \exp(-\mathrm{i}\omega_3\tau), \ldots, \exp(-\mathrm{i}\omega_N\tau)\}$. We next use the first equality in Definition 2, $\mathbf{Z}\mathbf{v} = \sigma\mathbf{u}$, and rewrite it in matrix form:

$$\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{-1}\,. \tag{66}$$

The matrix $\boldsymbol{\Sigma}$ is diagonal, with the singular values on its diagonal, $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_N\}$. Columns of the matrices $\mathbf{U}$ and $\mathbf{V}$ respectively represent right and left singular vectors.

We now write the scalar product of two arbitrary singular vectors $\mathbf{q}$ and $\mathbf{g}$ as

$$(\mathbf{q}, \mathbf{g}) = (\mathbf{q}^\dagger)^i M_{ij}\,\mathbf{g}^j = \left((\mathbf{F}\mathbf{q})^\dagger\right)^i (\mathbf{F}\mathbf{g})^j\,, \tag{67}$$

where the matrix $\mathbf{F}$ is the *Cholesky decomposition* of the metric matrix, $\mathbf{M} = \mathbf{F}^{\mathrm{T}}\mathbf{F}$. Because $\mathbf{M}$ is positive define, its Cholesky decomposition always exists and is unique.

Systems of singular vectors are orthonormalized; therefore, the following relations for the matrices $\mathbf{V}$ and $\mathbf{U}$ hold:

$$\mathbf{V}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{V} = \mathbf{I}\,, \tag{68}$$

$$\mathbf{U}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{U} = \mathbf{I}\,, \tag{69}$$

where $\mathbf{I}$ is the identity matrix. The inverse matrices to $\mathbf{V}$ and $\mathbf{U}$ are expressed in terms of their Hermitian conjugates as

$$\mathbf{V}^{-1} = \mathbf{V}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F}\,, \tag{70}$$

$$\mathbf{U}^{-1} = \mathbf{U}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F}\,. \tag{71}$$

Using these relations in (66) yields

$$\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{F}^{-1}\mathbf{F}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\mathbf{F}^{\mathrm{T}}\mathbf{F}\,. \tag{72}$$

We rewrite this in the form

$$\mathbf{F}\mathbf{P}\mathbf{F}^{-1} = (\mathbf{F}\mathbf{U})\,\boldsymbol{\Sigma}\,(\mathbf{F}\mathbf{V})^\dagger \equiv \tilde{\mathbf{U}}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^\dagger\,. \tag{73}$$

It is clear that the right-hand side of this equality coincides with the so-called *singular value decomposition* of the matrix $\mathbf{F}\mathbf{P}\mathbf{F}^{-1}$. We remind the reader that the singular value decomposition is a factorization of a matrix in the form $\tilde{\mathbf{U}}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^\dagger$, where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are orthogonal matrices and $\boldsymbol{\Sigma}$ is a diagonal matrix with positive numbers on the main diagonal [115]. This factorization exists for any real matrix and is unique. It can be easily ascertained that the matrices $\tilde{\mathbf{U}}$, $\tilde{\mathbf{V}}$, and $\boldsymbol{\Sigma}$ satisfy the conditions for singular value decomposition, and therefore, to calculate singular values and vectors, it suffices to perform this decomposition for the matrix $\mathbf{F}\mathbf{P}\mathbf{F}^{-1}$. The singular value decomposition procedure is a standard tool in many linear algebra software packages.

The original matrices $\mathbf{U}$ and $\mathbf{V}$ are calculated using $\mathbf{F}^{-1}$: $\mathbf{U} = \mathbf{F}^{-1}\tilde{\mathbf{U}}$ and $\mathbf{V} = \mathbf{F}^{-1}\tilde{\mathbf{V}}$. The maximum of the numbers on the diagonal of $\boldsymbol{\Sigma}$ is the first singular value by the time t, and the corresponding column of the matrix $\mathbf{V}$ is the first singular vector in the eigenvector basis.

3.2.1 Illustration of the matrix method. The matrix method to search for optimal perturbations has been used in many studies on the stability of laboratory flows (see, e.g., [45, 47, 53, 54, 116, 117]) and in astrophysical papers [51, 52, 118]. Here, we illustrate it with a simple semi-analytic study [113], where the eigenvector basis[8] was calculated in the WKB approximation in a geometrically thin and barotropic quasi-Keplerian torus with free boundaries. For simplicity, only the modes whose corotation radius lies outside the outer boundary of the torus were considered. (See Section 2.1 for a discussion of the energy exchange mechanism between the modes and with the background flow at the corotation radius in the context of the spectral problem corresponding to Eqns (8)–(10).) When the corotation radius is outside the flow, the energy of the modes is conserved. This means that they do not show exponential growth or decay, i.e., their frequencies ω are real [see expression (13)]. These are referred to as *neutral* modes. Nevertheless, due to their mutual nonorthogonality, in other words, due to the nonorthogonality of eigenvectors of the dynamical operator of the system of perturbations, we expect a nonmodal growth of their linear combinations (see the analogy in Fig. 6 and the comment on it in the text).

The modes we wish to obtain below physically correspond to inertial–acoustic waves that form a solid-body rotating pattern on the disk, i.e., that have the azimuthal projection of the wave vector constant in time and space. As we see from the WKB analysis in what follows, their characteristic radial wavelength is close to the disk thickness H. Their characteristic azimuthal scale λ_φ can be both larger and smaller than H, depending on the azimuthal wave number m entering system of equations (8)–(10). Results concerning the optimal perturbation growth are presented below in the case $\lambda_\varphi \gg H$ (see Fig. 7). We see in what follows that the optimal perturbation is then not a spiral unwinding by the flow, which we discussed in the context of the transient growth of vortices (see Fig. 2), but a wave packet initially located at the outer boundary of the torus and further propagating toward its inner boundary. At the instant of reflection from the inner boundary, its total acoustic energy reaches a maximum and then decreases when the packet goes back to the flow periphery. After the reflection from the outer boundary, the process repeats. Thus, the nonmodal growth in this case is not transient but oscillating, as must be the case according to the analogy shown in Fig. 6.

3.2.2 Background flow. We consider a toroidal flow of a finite radial extension as a model background flow. The azimuthal velocity component corresponds to the power-law angular velocity radial profile

$$\Omega = \Omega_0\left(\frac{r}{r_0}\right)^{-q}, \tag{74}$$

[8] Here, the eigenvectors of $\mathbf{Z}$ multiplied by the eigenvalues, i.e., by the time dependence $\exp(-\mathrm{i}\omega t)$, are referred to as perturbation modes.

where r_0 is the distance to the gravity center in the equatorial plane of the torus at which the rotation occurs with the Keplerian frequency Ω_0, $2 > q > 3/2$. We assume that the matter moves in the external Newtonian gravitational potential produced by a central point-like mass:

$$\Phi = -\frac{\Omega_0^2 r_0^3}{(r^2+z^2)^{1/2}}\,.$$

As is to become clear below, the parameter q then characterizes the torus thickness, which tends to zero as the angular velocity profile approaches the Keplerian one. As in Section 2.1, we here use the polytropic equation of state and write the force balance for the flow using the enthalpy h,

$$\begin{aligned}\frac{\partial h}{\partial r} &= \Omega^2 r - \frac{\partial \Phi}{\partial r}\,,\\ \frac{\partial h}{\partial z} &= -\frac{\partial \Phi}{\partial z}\,,\end{aligned} \tag{75}$$

where the two equations correspond to projections of the Euler equation on the radial and vertical directions. The joint integration of (75) yields

$$h(r,z) = \frac{\Omega_0^2 r_0^3}{(r^2+z^2)^{1/2}} + \frac{\Omega_0^2 r_0^{2q}}{2(1-q)}\, r^{2(1-q)} + C\,,$$

where the integration constant C is determined from the condition that $h(r_1,0)=0$ at the inner boundary of the torus $r_1 < r_0$.

In dimensionless coordinates $\hat{x} \equiv r/r_0$ and $\hat{y} \equiv z/r_0$, we then obtain

$$h = (\Omega_0 r_0)^2 \left[(\hat{x}^2+\hat{y}^2)^{-1/2} - \hat{x}_1^{-1} + \frac{1}{2(q-1)}\left(\hat{x}_1^{-2(q-1)} - \hat{x}^{-2(q-1)}\right)\right], \tag{76}$$

where $\hat{x}_1 \equiv r_1/r_0$. Enthalpy distribution (76) also determines the outer radial boundary of the torus $\hat{x}_2 > 1$, where $h(\hat{x}_2,0)=0$. The quantity $\hat{x}_\mathrm{d} = \hat{x}_2 - \hat{x}_1$ is called the radial extension of the flow.

It is now easy to pass to the case of a quasi-Keplerian, geometrically thin torus of interest here: $q = 3/2 + \epsilon^2/2$, $\epsilon \ll 1$. Using this assumption, the enthalpy profile can be simplified to

$$\frac{h}{\Omega_0^2 r_0^2} = \frac{\hat{H}^2}{2\hat{x}^3}\left[1 - \left(\frac{\hat{y}}{\hat{H}}\right)^2\right], \tag{77}$$

where $\hat{H}(x)$ is the dimensionless thickness of the torus expressed in units of r_0:

$$\hat{H} = \delta\,\hat{x}\left[\frac{\hat{x}_1(1+\ln\hat{x}) - \hat{x}(1+\ln\hat{x}_1)}{\hat{x}_1 - 1 - \ln\hat{x}_1}\right]^{1/2}. \tag{78}$$

Here, we introduce the small parameter

$$\delta \equiv \hat{H}(\hat{x}=1) = 2^{1/2}\epsilon\left(1 - \frac{1+\ln\hat{x}_1}{\hat{x}_1}\right)^{1/2} \ll 1\,,$$

which is clearly interpreted as the characteristic aspect ratio of a disk-like torus with $\delta \ll \hat{x}_\mathrm{d}$. It can be verified that expression (77) coincides with (4).

Equations (77) and (78) completely determine the quasi-Keplerian background flow that we use to illustrate the matrix method of calculation of nonmodal growth of the perturbation mode superposition. In the next section, we solve the spectral problem for such a flow, i.e., find the perturbation mode profiles.

3.2.3 Modes. Modes are nonstationary perturbations with an exponential time dependence $\propto \exp(-\mathrm{i}\omega t)$. They are also solutions of operator equation (52) for the evolution of a linear perturbation in the flow. This means that the modes are state vectors that we obtain by acting with the operator $\mathbf{Z}$ on its eigenvectors $\mathbf{f}_i$:

$$\mathbf{f}_i(t) = \mathbf{Z}\mathbf{f}_i = \exp(-\mathrm{i}\omega t)\,\mathbf{f}_i\,.$$

We repeat that the numbers $\exp(-\mathrm{i}\omega t)$ are eigenvalues of $\mathbf{Z}$, which we have to find along with its eigenvectors.

In practice, we do not use the equation exactly in form (52), but instead derive an equivalent ordinary differential equation of the second order in the radial coordinate for a Eulerian enthalpy perturbation. As everywhere in this paper, we assume that the hydrostatic equilibrium holds, i.e., $\delta v_z = 0$. Because we are dealing with a torus thin in z, $\delta \ll 1$, our perturbations taken initially in the form of azimuthal Fourier harmonics $\propto \exp(\mathrm{i}m\varphi)$ satisfy system of equations (8)–(10), which contains the background variables integrated over z (see Section 2.1). The transition to the mode analysis means the substitution $\partial/\partial t \to \mathrm{i}\omega$, after which we find from (8) and (9) that the complex Fourier harmonics of the Eulerian velocity perturbations, which are denoted here as $\mathbf{v}_r$ and $\mathbf{v}_\varphi$, are expressed in terms of the Fourier harmonics of the enthalpy perturbation, denoted here as $\mathbf{W}$, as

$$\mathbf{v}_r = \frac{\mathrm{i}}{D}\left(\bar{\omega}\,\frac{\mathrm{d}\mathbf{W}}{\mathrm{d}\hat{x}} - \frac{2m\Omega\mathbf{W}}{\hat{x}}\right), \tag{79}$$

$$\mathbf{v}_\varphi = \frac{1}{D}\left(\frac{\kappa^2}{2\Omega}\,\frac{\mathrm{d}\mathbf{W}}{\mathrm{d}\hat{x}} - \frac{m\bar{\omega}\mathbf{W}}{\hat{x}}\right), \tag{80}$$

where $D \equiv \kappa^2 - \bar{\omega}^2$, $\kappa^2 = (2\Omega/x)\,\mathrm{d}(\Omega\hat{x}^2)/\mathrm{d}\hat{x}$ is, as usual, the epicyclic frequency squared, and $\bar{\omega} \equiv \omega - m\Omega$ is the shifted frequency. In the rest of this section, we assume that all frequencies are taken in units of the frequency Ω_0 and time in units Ω_0^{-1}.

Substituting (79) and (80) in continuity equation (10), we obtain the equation for $\mathbf{W}$

$$\frac{D}{\hat{x}\Sigma}\,\frac{\mathrm{d}}{\mathrm{d}\hat{x}}\left(\frac{\hat{x}\Sigma}{D}\,\frac{\mathrm{d}\mathbf{W}}{\mathrm{d}\hat{x}}\right) - \left[\frac{2m}{\bar{\omega}}\,\frac{D}{\hat{x}\Sigma}\,\frac{\mathrm{d}}{\mathrm{d}\hat{x}}\left(\frac{\Omega\Sigma}{D}\right) + \left(n+\frac{1}{2}\right)\frac{D}{h_*} + \frac{m^2}{\hat{x}^2}\right]\mathbf{W} = 0\,, \tag{81}$$

where

$$\Sigma(r) = \int_{-H}^{H} \rho\,\mathrm{d}z \propto \hat{H}\left(\frac{\hat{H}^2}{\hat{x}^3}\right)^n, \qquad h_* = \frac{\hat{H}^2}{2\hat{x}^3}\,. \tag{82}$$

Here, h_* is the dimensionless background enthalpy in the equatorial disk plane [cf. (77)]. To reproduce the surface density dependence of Σ on r given above, it is enough to recall that $\Sigma \sim H\rho|_{z=0}$, and $\rho \sim h^n$ for the polytorpic equation of state. Equation (81), as well as its more general analog for three-dimensional perturbation modes, is frequently used in the literature. Their derivation and analysis can be found, e.g., in [77, 88, 90, 91, 119].

As we already noted, solving Eqn (81) is complicated by resonances: the corotation one, where $\bar{\omega} = 0$, and the Lindblad one, where $D = 0$. These points are singular for Eqn (81). However, to illustrate the matrix method of optimization, we restrict ourselves to calculating only some of the modes with resonances outside the outer boundary of the flow $\hat{x}_2$. The condition that the inner Lindblad resonance lies at $\hat{x} > \hat{x}_2$ implies that

$$\omega < (m-1)\,\Omega(\hat{x}_2)\,, \tag{83}$$

where we set $\kappa \approx \Omega$ in the condition $D = 0$ because of the almost Keplerian angular velocity profile in a thin torus. We also recall that ω is a real quantity. We note that for $m = 1$, the inner Lindblad resonance is at $\hat{x} = 0$, and hence there are no modes with $m = 1$ satisfying condition (83). For this reason, we consider only modes with $m > 1$. Hence, under the restrictions made, the term $\propto D/h_* \sim \delta^{-2}$ is large everywhere in the flow, and therefore the solution of the equation can be sought in the WKB approximation.

A WKB solution of Eqn (81) can be written as

$$\mathbf{W} = \mathbf{C}_0 S_1 \cos(S_0 + \varphi_0)\,, \tag{84}$$

where $S_0 \sim \delta^{-1}$ and $S_1 \sim \delta^0$. Substituting (84) in (81) yields its expansion in δ. By collecting terms with like powers of δ, namely, δ^{-2} and δ^{-1}, we find the explicit form of the functions S_0 and S_1:

$$S_0 = \int_{\hat{x}_1}^{\hat{x}} \left[\left(n+\frac{1}{2}\right)\frac{-D}{h_*} - \frac{m^2}{\hat{x}^2}\right]^{1/2} \mathrm{d}\hat{x}\,,$$

$$S_1 = \left(\frac{-D}{\hat{x}\Sigma}\right)^{1/2} \left[\left(n+\frac{1}{2}\right)\frac{-D}{h_*} - \frac{m^2}{\hat{x}^2}\right]^{-1/4}.$$

The phase φ_0 is fixed by the boundary conditions.

WKB solution (84) is irregular at the boundary points $\hat{x}_1$ and $\hat{x}_2$, at which $h_* \to 0$. It is possible to find a solution that is regular at the boundaries by the WKB method (see [120]), but we here use the traditional matching of (84) to an approximate regular solution of the original equation (81) near $\hat{x}_1$ and $\hat{x}_2$. This matching should yield a discrete set of eigenfrequencies ω, as well as the value of φ_0.

To find the regular solution near $\hat{x}_1$ and $\hat{x}_2$, we pass to the new radial coordinate $\tilde{x} \equiv |\hat{x} - \hat{x}_{1,2}|$ and expand Eqn (81) in the leading order in the variable $\tilde{x} \ll 1$. Technically, this means that all variables in (81) that are nonzero at $\hat{x}_{1,2}$ are set to their values exactly at the points $\hat{x}_{1,2}$. The disk half-thickness, vanishing at the boundaries, is approximated as $\hat{H} = \hat{H}_{1,2}\tilde{x}^{1/2}$, where the constant $\hat{H}_{1,2}$ is

$$\hat{H}_{1,2} = \delta\hat{x}_{1,2}\left|\frac{\ln\hat{x}_{1,2}}{1+\ln\hat{x}_{1,2}-\hat{x}_{1,2}}\right|^{1/2}.$$

We obtain the following near-boundary equation:

$$\tilde{x}\,\frac{\mathrm{d}^2\mathbf{W}}{\mathrm{d}\tilde{x}^2} + (n+1/2)\,\frac{\mathrm{d}\mathbf{W}}{\mathrm{d}\tilde{x}} + E_{1,2}\mathbf{W} = 0\,, \tag{85}$$

where

$$E_{1,2} = \frac{(2n+1)(-D_{1,2})\,\hat{x}_{1,2}^3}{H_{1,2}^2}\,,$$

and $D_{1,2}$ are the values of D at the points $\hat{x}_{1,2}$.

The solution of (85) that is regular at $\tilde{x} = 0$ has the form

$$\mathbf{W} = \mathbf{C}_{1,2}\,\tilde{x}^{-(2n-1)/4}\,J_{n-1/2}(\tilde{z})\,, \tag{86}$$

where $\tilde{z} = 2E_{1,2}^{1/2}\,\tilde{x}^{1/2}$.

We note that Eqn (85) at $\tilde{x} \to 0$ is equivalent to the boundary condition for the enthalpy perturbation at the free boundary of the flow, which states that the Lagrangian enthalpy perturbation vanishes at the boundary points $\hat{x}_{1,2}$, $\Delta h|_{x_{1,2}} = 0$ (see, e.g., [81]).

Because the denominator of $\tilde{z}$ contains the small δ, we have $\tilde{z} \gg 1$ at some distance from the boundary points under the condition $\tilde{x} \ll 1$. In this region, $\mathbf{W}$ has asymptotic form (86) for a large argument:

$$\mathbf{W} \approx \mathbf{C}_{1,2}\,\tilde{x}^{-n/2}(4\pi^2 E_{1,2})^{-1/4}\cos\left(2E_{1,2}^{1/2}\,\tilde{x}^{1/2} - \frac{n\pi}{2}\right). \tag{87}$$

Matching (87) to the WKB decomposition of the solution near $\hat{x}_1$ and $\hat{x}_2$ yields the zeroth phase $\varphi_0 = -n\pi/2$ in Eqn (84) and the dispersion equation

$$\int_{\hat{x}_1}^{\hat{x}_2}\left((2n+1)\,\frac{-D\hat{x}^3}{\hat{H}^2} - \frac{m^2}{\hat{x}^2}\right)^{1/2}\mathrm{d}\hat{x} = \pi(n+p)\,, \tag{88}$$

where p is an integer. Solving (88) for different p yields a discrete set of ω entering D. This is the sequence of eigenfrequencies of neutral modes we are interested in.

The mode profiles themselves are given by Eqns (84) and (86) with the relations between the corresponding constants taken into account:

$$\frac{\mathbf{C}_0}{\mathbf{C}_1} = \left(\frac{\hat{H}_1^{2n+1}}{2\pi\hat{x}_1^{3n-1}(-D_1)}\right)^{1/2},$$

$$\frac{\mathbf{C}_2}{\mathbf{C}_1} = (-1)^p\left[\left(\frac{\hat{x}_2}{\hat{x}_1}\right)^{3n-1}\frac{D_2}{D_1}\left(\frac{\hat{H}_1}{\hat{H}_2}\right)^{2n+1}\right]^{1/2}. \tag{89}$$

After obtaining the profile $\mathbf{W}(\hat{x})$ for a given ω_i, the corresponding complex Fourier harmonics of the Eulerian velocity perturbations $\mathbf{v}_r(\hat{x})$ and $\mathbf{v}_\varphi(\hat{x})$ can be calculated from (79) and (80). Thus, we find the complete eigenvector $\mathbf{f}_i \equiv \{\mathbf{v}_r, \mathbf{v}_\varphi, \mathbf{W}\}$ of the operator $\mathbf{Z}$ corresponding to its eigenvalue $\exp(-\mathrm{i}\omega_i t)$.

3.2.4 Optimal growth. Knowing the eigenvectors of the dynamical operator allows calculating the optimal growth, i.e., finding a linear combination of these vectors that demonstrates the maximum increase in the norm by a given time. The optimal growth by a time t has the form

$$G(t) = \max_{\mathbf{q}(0)}\frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^2}\,. \tag{90}$$

This is a generalization of (44) to the spatially global case.

The scalar product of two vectors from the linear span of N eigenvectors of the operator $\mathbf{Z}$ is introduced such that the square of the corresponding norm is coincident with the acoustic energy of perturbation (14):

$$(\mathbf{f},\mathbf{g}) = \pi\int_{r_1}^{r_2}\Sigma\left[(\delta v_r)_f(\delta v_r)_g^* + (\delta v_\varphi)_f(\delta v_\varphi)_g^* + \left(n+\frac{1}{2}\right)\frac{(\delta h)_f(\delta h)_g^*}{h_*}\right] r\,\mathrm{d}r\,, \tag{91}$$

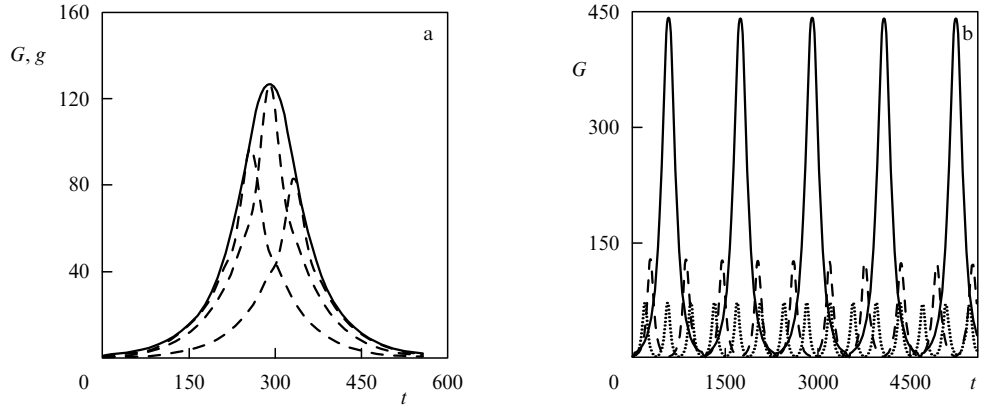

Figure 7. (a) The optimal growth curve $G(t)$ (solid curve) as a linear combination of slow modes in a thin disk for $\delta = 0.002$. The dashed lines show the growth factor of the total acoustic energy $g(t)$ of individual optimal perturbations as a function of time. These perturbations are optimal for time instants $t = 250, 290, 390$ expressed in units of the characteristic Keplerian period $2\pi\Omega_0^{-1}$. (b) The curves $G(t)$. The solid, dashed, and dotted lines correspond to $\delta = 0.001, 0.002, 0.003$, respectively. The linear combination shown has the dimensionality $N = 20$; the parameters are $x_d = 1.0$, $m = 25$, and $n = 3/2$. (Figure from [113].)

where the indices f or g indicate the relation of some physical variable to the vector $\mathbf{f}$ or $\mathbf{g}$. We recall that δv_r, δv_φ, and δh here denote the azimuthal Fourier harmonics of the corresponding Eulerian perturbations of the velocity components and the enthalpy.

We now apply the procedure for calculating the optimal combination of eigenvectors described above. With the eigenvectors in analytic form, the matrix $\mathbf{M}$ can be obtained by simple numerical integration of the combination of elementary functions using scalar product formula (91):

$$M_{ij} = (\mathbf{f}_i, \mathbf{f}_j)\,. \tag{92}$$

Next, we perform the Cholesky decomposition $\mathbf{M} = \mathbf{F}^{\mathrm{T}}\mathbf{F}$ and then the singular value decomposition of the matrix $\mathbf{FPF}^{-1}$. Both these procedures are standard in numerical methods of matrix algebra.

As an example in Fig. 7 a and 7 b, we show the dependence of the maximum possible energy growth $G(t)$ among all superpositions of 20 neutral modes by the time $t_s \sim (\delta\Omega_0)^{-1}$ on a time scale of the order of the sound time t, and $t_s \sim (\delta\Omega_0)^{-1}$. Figure 7a also shows the energy growth of the optimal mode combinations $g(t)$. Clearly, the curves $g(t)$ touch the common optimal growth curve $G(t)$, as must be the case, each at its own optimization time. The optimal growth itself in this model has a quasi-periodic form, reaching maxima at times $\sim t_s$, and the thinner the torus, the higher g the mode superposition can reach.

3.2.5 Angular momentum flux. In Section 3.2.4, we have shown that some combinations of modes can demonstrate a significant growth in the acoustic energy. We consider in more detail what the optimal perturbation is. The perturbation amplitude growth suggests that the main flow transfers energy to perturbations. The first term in the right-hand side of (12) is responsible for this, and the integrand there is sometimes referred to as the Reynolds force power (see [77]), denoted here by F_R. It turns out that F_R is simply related to the density of the specific angular momentum flux F related to perturbations: $F_R = -(\mathrm{d}\Omega/\mathrm{d}\hat{x})\,F$ (see Sections 2.3 and 4 in [97]). Clearly, for Keplerian rotation, F_R and F have the same sign: when the perturbation energy increases, $F > 0$, an angular momentum flux to the torus periphery occurs, and vice versa.

We use the expression

$$F = \hat{x}\Sigma\langle\delta v_r \delta v_\varphi\rangle \tag{93}$$

to calculate the evolution of the profile of F for the optimal mode superposition presented by the curve $g(t)$ for $t = 290$ in Fig. 7 a. Figure 8 shows how the radial distribution F changes in the interval $(\hat{x}_1, \hat{x}_2)$. We see that F is first localized in the radial direction, and its localization region moves with time: during the perturbation amplitude growth phase, it shifts toward the inner torus boundary, and during the amplitude decay phase, it shifts back to the outer boundary. Therefore, a nonmodal growing perturbation is here represented by a wave packet containing a sequence of neutral modes (each of which, as we recall, solid-body rotates with an angular velocity that is smaller than the angular velocity of the flow). Initially, this wave packet is localized near the outer boundary of the torus and moves toward the inner boundary. This causes an angular momentum outflow to the disk periphery, because $F > 0$, and its acoustic energy increases. At the instant of reflection from the inner boundary, the sign of F and the direction of motion of the wave packet reverse, which later leads to a decrease in its acoustic energy, and the angular momentum inflows back to the inner parts of the torus. Because there is no viscous dissipation and the background flow is stationary, it is evident that if we continue to track the evolution of the optimal mode superposition, then this scenario must recur: the wave packet, after reflecting from the outer boundary, goes back toward the inner boundary. We also note that the form of $G(t)$ obtained suggests that during the evolution of this particular type of perturbation, there are epochs (time intervals measured from the conventional start of the perturbation evolution) during which no combination of modes can be amplified. These epochs correspond to minima on the $G(t)$ curve (see Fig. 7b). This is because only wave packets localized near the outer disk boundary can exhibit significant growth. At the same time, the velocity of their radial motion is determined by the speed of sound in the flow, and hence the time intervals 'favorable' to nonmodal growth are always $\sim \hat{x}_d/\delta$.

If, on the plane (r, φ), we plot the constant-phase lines of perturbations corresponding to the wave packet discussed, it turns out that at the growth stage they correspond to a trailing

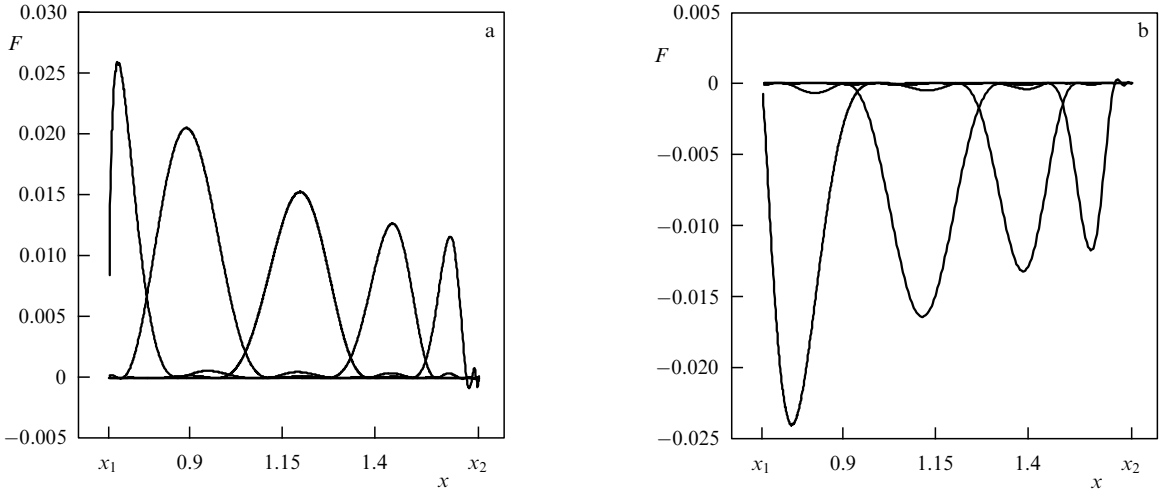

Figure 8. Radial profiles of the azimuthally averaged angular momentum flux density F of the optimal perturbation $g(t)$ shown in Fig. 7a for the optimization time $t = 290$. (a) Profiles of F at time instants $t = 50$, 100, 150, 200, 240 before the $g(t)$ maximum. Each profile has one large maximum, shifting from the outer disk boundary x_2 to the inner disk boundary x_1 as t increases. (b) Profiles of F at time instants $t = 290$, 350, 400, 450 after the maximum of $g(t)$. Each profile, similarly, has one pronounced minimum, whose position is shifted from the inner boundary of the disk at x_1 to the outer boundary at x_2 as t increases further. The linear combination shown has the dimensionality $N = 20$, the parameters are $\delta = 0.002$, $x_d = 1.0$, $m = 25$, and $n = 3/2$. (Figure paper [113].)

spiral. It has the maximum opening at the initial time, but while propagating toward the inner boundary, it winds up stronger and stronger. After reflection from the inner boundary, it transforms into a tightly wound leading spiral, and in the process of motion back toward the outer boundary, the degree of winding decreases. This behavior of the optimal perturbation is similar to the process of enhancement/weakening by a shear flow that we discussed in Section 2.2 in the context of a spatially local problem.

3.3 Alternative: the variational approach

Singular vectors can be found differently by the variational method. This method is a generalization of the method of power iterations for matrix eigenvalues and eigenvectors with finite dimensions (see, e.g., [115]). The variational method requires less computational power than the matrix method [121] and, importantly, can be applied for nonstationary background flows, as well as used to solve nonlinear problems of the transient dynamics of finite-amplitude perturbations. Unlike the matrix method, it does not require a discrete representation of the dynamical operator, for example, the expansion of perturbations in proper eigenvectors, whose calculation in a shear flow encounters a known difficulty when bypassing the corotation and Lindblad resonances (see [122]).

For the linear dynamics, the variational method turns out to be equivalent to solving the simpler problem of finding the maximum eigenvalue of the operator $\mathbf{Z}^\dagger\mathbf{Z}$ (see Section 3.1 and, for example, [123]), and we therefore start with solving exactly this problem, and postpone the derivation of the variational method directly from the variational principle until the generalization to the nonlinear case.

3.3.1 Linear autonomous operators. In Section 3.1, after introducing the notion of singular values, we discussed that the first singular value is simultaneously the maximum eigenvalue of the composite operator $\mathbf{Z}^\dagger\mathbf{Z}$, and the first right singular vector is the corresponding eigenvector of this operator. Here, we try to understand what the action of $\mathbf{Z}^\dagger\mathbf{Z}$ on the initial state vector $\mathbf{q}(0)$ is equivalent to. The action of the first (right) part of the composite operator is known from its definition (52): this is the integration of equations of the perturbation dynamics, for example, system (8)–(10), which we symbolically rewrite here as

$$\frac{\partial \mathbf{q}}{\partial t} = \mathbf{A}\mathbf{q}\,, \tag{94}$$

until the time t with the initial condition $\mathbf{q}(0)$. We note that due to the linearity of the problem, the operator $\mathbf{A}$ in (94) does not depend on $\mathbf{q}$.

The subsequent action of the operator $\mathbf{Z}^\dagger$ on $\mathbf{q}(t)$ is easy to understand if the operator $\mathbf{A}$ is autonomous, i.e., time-independent [112]. The solution of Eqn (94) can then be written in operator form: $\mathbf{q}(t) = \exp(\mathbf{A}t)\,\mathbf{q}(0)$, which means that $\mathbf{A}$ and $\mathbf{Z}$ are related as

$$\mathbf{Z} = \exp(\mathbf{A}t)\,. \tag{95}$$

The right-hand side of (95) is called the operator exponential and is to be understood as the infinite series $\mathbf{I} + \mathbf{A}t + (\mathbf{A}t)^2/2 + \ldots$.

The operator adjoint to $\mathbf{Z}$ can also be written in terms of the operator exponential $\mathbf{Z}^\dagger = \exp(\mathbf{A}^\dagger t)$, where $\mathbf{A}^\dagger$ is the operator adjoint to $\mathbf{A}$ defined by the Lagrange relation $(\mathbf{A}\mathbf{q}, \tilde{\mathbf{q}}) = (\mathbf{q}, \mathbf{A}^\dagger\tilde{\mathbf{q}})$, and $\mathbf{q}$ and $\tilde{\mathbf{q}}$ are arbitrary vectors. This expression for $\mathbf{Z}^\dagger$ follows by taking the adjoint of the infinite operator series given above. We now consider the scalar product

$$\left(\frac{\partial \mathbf{q}}{\partial t}, \tilde{\mathbf{q}}\right) = (\mathbf{A}\mathbf{q}, \tilde{\mathbf{q}}) = (\mathbf{q}, \mathbf{A}^\dagger\tilde{\mathbf{q}})\,. \tag{96}$$

On the other hand,

$$\left(\frac{\partial \mathbf{q}}{\partial t}, \tilde{\mathbf{q}}\right) = \frac{\partial}{\partial t}(\mathbf{q}, \tilde{\mathbf{q}}) - \left(\mathbf{q}, \frac{\partial \tilde{\mathbf{q}}}{\partial t}\right) = \frac{\partial}{\partial t}\left(\exp(\mathbf{A}t)\,\mathbf{q}(0), \tilde{\mathbf{q}}\right)$$

$$- \left(\mathbf{q}, \frac{\partial \tilde{\mathbf{q}}}{\partial t}\right) = \left(\mathbf{q}(0), \frac{\partial}{\partial t}\left(\exp(\mathbf{A}^\dagger t)\,\tilde{\mathbf{q}}\right)\right) - \left(\mathbf{q}, \frac{\partial \tilde{\mathbf{q}}}{\partial t}\right). \tag{97}$$

Combining (96) and (97) yields the identity

$$\left(\mathbf{q}(0), \frac{\partial}{\partial t}\left(\exp\left(\mathbf{A}^{\dagger}t\right)\tilde{\mathbf{q}}\right)\right) - \left(\mathbf{q}, \frac{\partial\tilde{\mathbf{q}}}{\partial t}\right) = \left(\mathbf{q}, \mathbf{A}^{\dagger}\tilde{\mathbf{q}}\right). \quad (98)$$

It is easy to see that if $\tilde{\mathbf{q}}$ and $\partial\tilde{\mathbf{q}}/\partial t$ are related as

$$\frac{\partial\tilde{\mathbf{q}}}{\partial t} = -\mathbf{A}^{\dagger}\tilde{\mathbf{q}}, \quad (99)$$

then $\tilde{\mathbf{q}}(t) = \exp\left(-\mathbf{A}^{\dagger}t\right)\tilde{\mathbf{q}}(0)$ and identity (98) is satisfied for any $\mathbf{q}$.

Thus, the action of the operator $\mathbf{Z}^{\dagger} = \exp\left(\mathbf{A}^{\dagger}t\right)$ is equivalent to the integration of Eqn (99) backward in time from the instant t with the initial condition $\mathbf{q}(t)$ to the instant $t = 0$. Equation (99) is called the adjoint equation.

We note additionally that although the operator $\mathbf{Z}$ can be represented as $\mathbf{Z} = \exp\left(\mathbf{A}t\right)$ and $\mathbf{Z}^{\dagger}$ as $\mathbf{Z}^{\dagger} = \exp\left(\mathbf{A}^{\dagger}t\right)$, the composite operator cannot be represented as $\mathbf{Z}^{\dagger}\mathbf{Z} = \exp\left(\left(\mathbf{A}^{\dagger} + \mathbf{A}\right)t\right)$. To see this, it is sufficient use the series expansion of the operator exponential.

The action of the composite operator $\mathbf{Z}^{\dagger}\mathbf{Z}$ on the initial vector $\mathbf{q}(0)$ is therefore equivalent to the integration forward in time of the original equation (94) with the initial condition $\mathbf{q}(0)$ up to the instant t, and then to the integration of adjoint equation (99) backward in time to the initial instant with the initial condition in the form of the vector $\mathbf{q}(t)$ we obtained by integrating (94).

Any vector for which the action of the composite operator $\mathbf{Z}^{\dagger}\mathbf{Z}$ is equivalent to multiplication by a constant is a right singular vector of the dynamical operator, and the constant is the square of the corresponding singular value: $\mathbf{Z}^{\dagger}\mathbf{Z}\mathbf{v} = \sigma^2\mathbf{v}$. However, we need only the first, i.e., the largest, right singular vector. To calculate it, we consider an iteration procedure of applying the composite operator $\mathbf{Z}^{\dagger}\mathbf{Z}$ with the subsequent normalization of the result to unity. To show the convergence of iterations to the first singular vector, we consider the decomposition of an arbitrary state vector in terms of singular vectors $\mathbf{q}(0) = \sum_{k=1}^{\infty} q_k\mathbf{v}_k(0)$ and act on it by the iteration operator, $\mathbf{Z}^{\dagger}\mathbf{Z}\mathbf{q}(0) = \sum_{k=1}^{\infty} \sigma_k^2 q_k\mathbf{v}_k(0)$.

It is easy to see that the iteration operator increases the weight of each singular vector in proportion to the square of the singular value. Therefore, the limit $(\mathbf{Z}^{\dagger}\mathbf{Z})^{p\to\infty}\mathbf{q}(0)$, where p is a natural number, for an arbitrary initial state vector $\mathbf{q}(0)$ is equal to the first right singular vector, because it corresponds to the maximum singular value. The rate of convergence depends on the difference between the singular vectors.

We note that to converge exactly to the first singular vector, the initial approximation must not be orthogonal to it, because if the weight of the first singular vector in the decomposition of $\mathbf{q}(0)$ is zero, $q_1 = 0$, then the action of the iteration operator does not increase this weight: $\sigma_1^2 q_1 = 0$. The iteration scheme in this case converges to the singular vector with the maximal singular value of all vectors that have a nonzero weight in the initial approximate decomposition.

Thus, to find the first right singular vector, it is necessary to apply an iteration procedure that includes the integration of the original equation (94) forward in time and of the adjoint equation (99) backward in time, with the subsequent normalization to unity after each iteration.

3.3.2 Linear nonautonomous operators. In the case of a time-dependent operator $\mathbf{Z}^{\dagger}$ (so-called nonautonomous operator; see [124]), the action of $\mathbf{A}$ also corresponds to the integration of Eqn (99) backward in time, which can be verified as follows.

For a nonautonomous operator $\mathbf{A}$, the action of the operator $\mathbf{Z}$ can be factored into the product of infinitesimal operators:

$$\mathbf{Z}(\tau) = \lim_{n\to\infty}\prod_{j=1}^{n}\exp\left(\mathbf{A}(t_j)\right)\delta t, \quad (100)$$

where $\delta t = \tau/n$, $(j-1)\,\delta t < t_j < j\,\delta t$ (see [124]). Taking the adjoint of the product of operators yields

$$\mathbf{Z}^{\dagger}(\tau) = \lim_{n\to\infty}\prod_{j=n}^{1}\exp\left(\mathbf{A}^{\dagger}(t_j)\right)\delta t. \quad (101)$$

Clearly, at each time interval δt, the integration is performed backward in time, and the intervals themselves are arranged in the order of decreasing j; therefore, the action of $\mathbf{Z}^{\dagger}$ is again equivalent to the integration of (99) backward in time.

Thus, as in the case of autonomous operators, the application of $\mathbf{Z}^{\dagger}\mathbf{Z}$ is equivalent to the consecutive integration of (94) forward in time and of (99) backward in time. Accordingly, the iteration procedure to seek the first singular vector presented above is applicable to nonautonomous operators as well.

3.3.3 Calculation of the next singular vectors. Singular vectors form an orthogonal set of functions and can be used as a basis for the decomposition of any linear perturbation. Hence, it can be useful to calculate not only the first but also the consecutive singular vectors; below, we briefly describe their calculation by the variational method.

In order that the iterations described above converge not to the first singular vector but to the Nth vector, it is sufficient that the domain of the iteration operator be the complement of the subset of linear combinations of the preceding $N-1$ vectors, or, equivalently, that the initial approximation be orthogonal to the singular vectors already calculated, i.e., the condition $(\mathbf{q}(0), \mathbf{v}_j(0)) = 0$ be satisfied for $j < N$. In this case, the action of the iteration operator is orthogonal to the calculated singular vectors:

$$\begin{aligned}\left(\mathbf{Z}^{\dagger}\mathbf{Z}\mathbf{q}(0), \mathbf{v}_j(0)\right) &= \left(\mathbf{Z}^{\dagger}\mathbf{Z}\sum_{k=N}^{\infty} q^k\mathbf{v}_k(0), \mathbf{v}_j(0)\right)\\ &= \left(\sum_{k=N}^{\infty}\sigma_k^2 q^k\mathbf{v}_k(0), \mathbf{v}_j(0)\right) = 0. \end{aligned} \quad (102)$$

Therefore, if we decompose some vector with respect to singular vectors as

$$\mathbf{q}(0) = \sum_{k=1}^{\infty} q^k\mathbf{v}_k(0), \quad (103)$$

then changing the initial condition in the iteration procedure by $\mathbf{q}(0) - \sum_{k=1}^{N-1} q^k\mathbf{v}_k(0)$ provides the convergence of direct iterations to the Nth singular vector. Thus, from known $N-1$ singular vectors, it is always possible to calculate the next one.

3.3.4 Generalization to the nonlinear case. In the case of nonlinear dynamics, the justification of iteration calculations of optimal growth presented in the two preceding sections becomes invalid; however, in a somewhat generalized form, it

can be obtained from the variational principle, as we show below.

The problem is formulated as a search for the initial condition demonstrating the maximum growth of the norm by a given time. In other words, it is required to find a vector $\mathbf{q}(0)$ such that the functional

$$\mathcal{G}(\tau) = \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^2} \tag{104}$$

reaches a maximum if the vector $\mathbf{q}$ satisfies the dynamic equations written in operator form (94). For this, a technique similar to the well-known method of Lagrange multipliers for finding the conditional extremum of a function is used.

The Lagrangian needed to find the conditional extremum in this case includes two terms: the functional whose maximum is being sought and the so-called 'penalty' term, which is nonzero only when $\mathbf{q}$ stops satisfying dynamic equations (94) (see also [125, 126] and review [49]):

$$\mathcal{L}(\mathbf{q}, \tilde{\mathbf{q}}) = \mathcal{G}(\mathbf{q}) - \int_0^t \left(\tilde{\mathbf{q}}, \dot{\mathbf{q}} - \mathbf{A}(\mathbf{q})\,\mathbf{q}\right) \mathrm{d}\tau\,. \tag{105}$$

The penalty term is written in (105) as the scalar product of the Lagrange multipliers entering $\tilde{\mathbf{q}}$, additionally integrated over time, and Eqn (94). In contrast to the well-known problem of finding the conditional extremum of a function, the Lagrangian in this case is a functional defined for all possible values of $\mathbf{q}$, and the Lagrange multipliers themselves are functions.

The extremum of (104) is reached when variations of the Lagrangian with respect to $\mathbf{Q}$ and $\tilde{\mathbf{q}}$ vanish simultaneously. These variations are defined as (see book [127])

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}}\,\delta\mathbf{q} = \lim_{\epsilon \to 0} \frac{\mathcal{L}(\mathbf{q} + \epsilon\delta\mathbf{q}, \tilde{\mathbf{q}}) - \mathcal{L}(\mathbf{q}, \tilde{\mathbf{q}})}{\epsilon}\,, \tag{106}$$

$$\frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{q}}}\,\delta\tilde{\mathbf{q}} = \lim_{\epsilon \to 0} \frac{\mathcal{L}(\mathbf{q}, \tilde{\mathbf{q}} + \epsilon\delta\tilde{\mathbf{q}}) - \mathcal{L}(\mathbf{q}, \tilde{\mathbf{q}})}{\epsilon}\,, \tag{107}$$

where $\delta\mathbf{q}$ and $\delta\tilde{\mathbf{q}}$ are arbitrary functions at any time.

The variation with respect to the indefinite multipliers is clearly given by

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{q}}}\,\delta\tilde{\mathbf{q}} &= -\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^t \left(\epsilon\delta\tilde{\mathbf{q}}, \dot{\mathbf{q}} - \mathbf{A}(\mathbf{q})\,\mathbf{q}\right) \mathrm{d}\tau \\ &= -\int_0^t \left(\delta\tilde{\mathbf{q}}, \dot{\mathbf{q}} - \mathbf{A}(\mathbf{q})\,\mathbf{q}\right) \mathrm{d}\tau\,.\end{aligned} \tag{108}$$

Equating (108) to zero, by virtue of the arbitrariness of $\delta\tilde{\mathbf{q}}$, we obtain Eqn (94). To compute variations with respect to the state vectors, we use the Lagrange identity $(\tilde{\mathbf{q}}, \mathbf{A}\mathbf{q}) = (\mathbf{A}^\dagger\tilde{\mathbf{q}}, \mathbf{q})$ (see, e.g., [128] for more details about adjoint operators in nonlinear problems) and integrate the penalty term by parts, after which the Lagrangian becomes

$$\mathcal{L}(\mathbf{q}, \tilde{\mathbf{q}}) = \mathcal{G}(\mathbf{q}) - (\tilde{\mathbf{q}}, \mathbf{q})\Big|_0^t + \int_0^t \left(\dot{\tilde{\mathbf{q}}} + \mathbf{A}^\dagger(\tilde{\mathbf{q}})\,\tilde{\mathbf{q}}, \mathbf{q}\right) \mathrm{d}\tau\,. \tag{109}$$

Using the smallness of ϵ and the real-valuedness of the scalar product,[9] we calculate the variation with respect to state vectors:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{q}}\,\delta\mathbf{q} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \Bigg[&\frac{||\mathbf{q}(t) + \epsilon\delta\mathbf{q}(t)||^2}{||\mathbf{q}(0) + \epsilon\delta\mathbf{q}(0)||^2} - \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^2} - \left(\tilde{\mathbf{q}}(t), \epsilon\delta\mathbf{q}(t)\right) \\ &+ \left(\tilde{\mathbf{q}}(0), \epsilon\delta\mathbf{q}(0)\right) + \int_0^t \left(\dot{\tilde{\mathbf{q}}} + \mathbf{A}^\dagger(\tilde{\mathbf{q}})\,\tilde{\mathbf{q}}, \epsilon\delta\mathbf{q}\right) \mathrm{d}\tau\Bigg]\,.\end{aligned} \tag{110}$$

Here, the first term can be rewritten in the form

$$\begin{aligned}&\lim_{\epsilon \to 0} \frac{1}{\epsilon} \frac{||\mathbf{q}(t) + \epsilon\delta\mathbf{q}(t)||^2}{||\mathbf{q}(0) + \epsilon\delta\mathbf{q}(0)||^2} \\ &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \frac{||\mathbf{q}(t)||^2 + \epsilon\left(\delta\mathbf{q}(t), \mathbf{q}(t)\right) + \epsilon\left(\mathbf{q}(t), \delta\mathbf{q}(t)\right)}{||\mathbf{q}(0)||^2 + \epsilon\left(\delta\mathbf{q}(0), \mathbf{q}(0)\right) + \epsilon\left(\mathbf{q}(0), \delta\mathbf{q}(0)\right)}\,.\end{aligned} \tag{111}$$

Because the scalar product is real-valued, we have $(\delta\mathbf{q}(t), \mathbf{q}(t)) = (\mathbf{q}(t), \delta\mathbf{q}(t))$, and therefore the transformation can be continued:

$$\begin{aligned}&\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\frac{||\mathbf{q}(t) + \epsilon\delta\mathbf{q}(t)||^2}{||\mathbf{q}(0) + \epsilon\delta\mathbf{q}(0)||^2} - \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^2}\right] \\ &\quad= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\frac{||\mathbf{q}(t)||^2 + 2\epsilon\left(\delta\mathbf{q}(t), \mathbf{q}(t)\right)}{||\mathbf{q}(0)||^2 + 2\epsilon\left(\delta\mathbf{q}(0), \mathbf{q}(0)\right)} - \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^2}\right] \\ &\quad= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\frac{2\epsilon\left(\delta\mathbf{q}(t), \mathbf{q}(t)\right)}{||\mathbf{q}(0)||^2} - \frac{2\epsilon\left(\delta\mathbf{q}(0), \mathbf{q}(0)\right) ||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^4}\right] \\ &\quad= \frac{2\left(\delta\mathbf{q}(t), \mathbf{q}(t)\right)}{||\mathbf{q}(0)||^2} - 2\left(\delta\mathbf{q}(0), \mathbf{q}(0)\right) \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^4}\,,\end{aligned} \tag{112}$$

which ultimately gives the variation

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{q}}\,\delta\mathbf{q} = &\frac{2\left(\delta\mathbf{q}(t), \mathbf{q}(t)\right)}{||\mathbf{q}(0)||^2} - 2\left(\delta\mathbf{q}(0), \mathbf{q}(0)\right) \frac{||\mathbf{q}(t)||^2}{||\mathbf{q}(0)||^4} \\ &- \left(\tilde{\mathbf{q}}(t), \delta\mathbf{q}(t)\right) + \left(\tilde{\mathbf{q}}(0), \delta\mathbf{q}(0)\right) + \int_0^t \left(\dot{\tilde{\mathbf{q}}} + \mathbf{A}^\dagger(\tilde{\mathbf{q}})\,\tilde{\mathbf{q}}, \delta\mathbf{q}\right) \mathrm{d}\tau\,.\end{aligned} \tag{113}$$

The variation of $\delta\mathbf{q}$ is independent at different time instants, and therefore equating (113) to zero at one time yields an equation for indefinite multipliers (99), which provides the vanishing of the Lagrangian variation in the interval $0 < \tau < t$, and additional relations between $\mathbf{q}$ and $\tilde{\mathbf{q}}$, which are needed for the vanishing of the Lagrangian variations at the instants $\tau = 0$ and $\tau = t$:

$$\tilde{\mathbf{q}}(t) = \frac{2}{||\mathbf{q}(0)||^2}\,\mathbf{q}(t)\,, \tag{114}$$

$$\mathbf{q}(0) = \frac{||\mathbf{q}(0)||^4}{2||\mathbf{q}(t)||^2}\,\tilde{\mathbf{q}}(0)\,. \tag{115}$$

For the vectors $\mathbf{q}$ and $\tilde{\mathbf{q}}$ satisfying Eqns (94) and (99) and constraints (114) and (115), the Lagrangian variations vanish, and hence functional (104) reaches an extremum at these vectors.

As in linear systems, the joint solution of equations can be sought by the power iteration method schematically shown in Fig. 9. This issue is further discussed in [129–131].

We note once again that for linear perturbations, the optimization of functional (104) reduces to seeking the maximal eigenvalue of the composite operator $\mathbf{Z}^\dagger\mathbf{Z}$.

[9] The real-valuedness of the scalar product is additionally required only in this section to obtain constraints (114) and (115) in a simple form; see below.

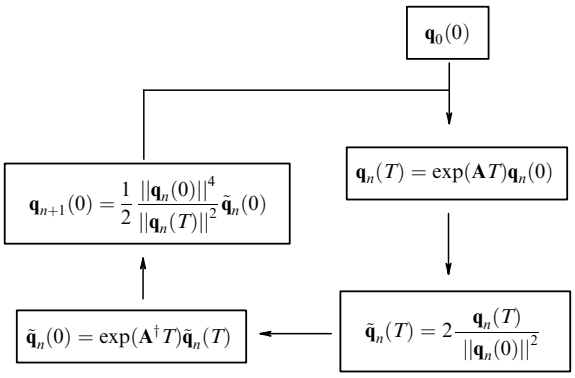

Figure 9. Flow chart of the iteration loop to seek the optimal perturbation for a time instant T satisfying the general system (94) (see review [49]).

3.4 Adjoint equations

3.4.1 Derivation of the adjoint equations. To obtain the explicit form of the equations adjoint to (8)–(10), we use the norm identical to the total acoustic energy of perturbations (14). The scalar product corresponding to this norm coincides with formula (91), which we already used. Using it, we represent $(\tilde{\mathbf{q}}, \mathbf{A}\mathbf{q})$ as

$$(\tilde{\mathbf{q}}, \mathbf{A}\mathbf{q}) = \pi \int_{r_{\rm in}}^{r_{\rm out}} \Sigma \left[\delta\tilde{v}_r \left(\mathrm{i}m\Omega\, \delta v_r^* + 2\Omega\, \delta v_\varphi^* - \frac{\partial\, \delta h^*}{\partial r} \right) + \delta\tilde{v}_\varphi \left(-\frac{\kappa^2}{2\Omega}\, \delta v_r^* + \mathrm{i}m\Omega\, \delta v_\varphi^* + \frac{\mathrm{i}m\, \delta h^*}{r} \right) + \frac{\delta\tilde{h}}{a_*^2} \left(-\frac{a_*^2}{\Sigma r} \frac{\partial}{\partial r} (r\Sigma \delta v_r^*) + \frac{\mathrm{i}m a_*^2}{r}\, \delta v_\varphi^* + \mathrm{i}m\Omega\, \delta h^* \right) \right] r\, \mathrm{d}r\,. \tag{116}$$

Using the Lagrange identity $(\tilde{\mathbf{q}}, \mathbf{A}\mathbf{q}) = (\mathbf{A}^\dagger\tilde{\mathbf{q}}, \mathbf{q})$ and Eqn (99) in the left-hand side of this expression, we represent the scalar product in accordance with formula (91). The right-hand side can be rearranged to show the dependence on the components of $\delta\mathbf{q}$ in factored form. Here, the spatial derivatives are calculated using integration by parts. We obtain

$$\pi \int_{r_{\rm in}}^{r_{\rm out}} \Sigma r\, \mathrm{d}r \left[-\delta v_r^* \frac{\partial\tilde{v}_r}{\partial t} - \delta v_\varphi^* \frac{\partial\tilde{v}_\varphi}{\partial t} - \delta h^* \frac{\partial\tilde{h}}{\partial t} \right] = \pi \int_{r_{\rm in}}^{r_{\rm out}} \Sigma r\, \mathrm{d}r \left[\delta v_r^* \left(\mathrm{i}m\Omega\, \delta\tilde{v}_r - \frac{\kappa^2}{2\Omega}\, \delta\tilde{v}_\varphi + \frac{\partial\delta\tilde{h}}{\partial r} \right) + \delta v_\varphi^* \left(2\Omega\, \delta\tilde{v}_r + \mathrm{i}m\Omega\, \delta\tilde{v}_\varphi + \frac{\mathrm{i}m}{r}\, \delta\tilde{h} \right) + \delta h^* \left(\frac{1}{r\Sigma} \frac{\partial}{\partial r} (r\Sigma\, \delta\tilde{v}_r) + \frac{\mathrm{i}m}{r}\, \delta\tilde{v}_\varphi + \frac{\mathrm{i}m\Omega}{a_*^2}\, \delta\tilde{h} \right) \right] - \pi r\Sigma \delta\tilde{h} \delta v_r^* \Big|_{r_{\rm in}}^{r_{\rm out}} - \pi r\Sigma \delta\tilde{v}_r\, \delta h^* \Big|_{r_{\rm in}}^{r_{\rm out}}. \tag{117}$$

The substitutions in the right-hand side of (117) vanish because $\Sigma \to 0$ at the boundaries.

Components of the variation $\delta\mathbf{q}$ are arbitrary and independent, and therefore (117) is transformed into three independent equalities, each corresponding to one of the components of $\delta\mathbf{q}$. These equalities result in a system of adjoint equations:

$$\frac{\partial\delta\tilde{v}_r}{\partial t} = -\mathrm{i}m\Omega\, \delta\tilde{v}_r + \frac{\kappa^2}{2\Omega}\, \delta\tilde{v}_\varphi - \frac{\partial\delta\tilde{h}}{\partial r}\,, \tag{118}$$

$$\frac{\partial\delta\tilde{v}_\varphi}{\partial t} = -2\Omega\, \delta\tilde{v}_r - \mathrm{i}m\Omega\, \delta\tilde{v}_\varphi - \frac{\mathrm{i}m}{r}\, \delta\tilde{h}\,, \tag{119}$$

$$\frac{\partial\delta\tilde{h}}{\partial t} = -\frac{a_*^2}{r\Sigma} \frac{\partial}{\partial r} (r\Sigma \delta\tilde{v}_r) - \frac{\mathrm{i}m a_*^2}{r}\, \delta\tilde{v}_\varphi - \mathrm{i}m\Omega\, \delta\tilde{h}\,. \tag{120}$$

Passing to the local space limit in (118)–(120) [as we did in Section 2.2 to obtain system (16)–(18) from Eqns (8)–(10)], we obtain the explicit form of the equations adjoint to (16)–(18):

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x \frac{\partial}{\partial y} \right) \tilde{u}_x - (2-q)\, \Omega_0 \tilde{u}_y = -\frac{\partial\tilde{W}}{\partial x}\,, \tag{121}$$

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x \frac{\partial}{\partial y} \right) \tilde{u}_y + 2\Omega_0 \tilde{u}_x = -\frac{\partial\tilde{W}}{\partial y}\,, \tag{122}$$

$$\left(\frac{\partial}{\partial t} - q\Omega_0 x \frac{\partial}{\partial y} \right) \tilde{W} + a_*^2 \left(\frac{\partial\tilde{u}_x}{\partial x} + \frac{\partial\tilde{u}_y}{\partial y} \right) = 0\,, \tag{123}$$

where tildes above u_x, u_y, and W mean that these quantities compose an adjoint state vector.

Finally, passing to the comoving reference frame in (121)–(123) yields the adjoint equations for separate SFHs:

$$\frac{\mathrm{d}\hat{\tilde{u}}_x}{\mathrm{d}t} = (2-q)\, \hat{\tilde{u}}_y - \mathrm{i}\, \tilde{k}_x(t) \hat{\tilde{W}}\,, \tag{124}$$

$$\frac{\mathrm{d}\hat{\tilde{u}}_y}{\mathrm{d}t} = -2\hat{\tilde{u}}_x - \mathrm{i}\, k_y \hat{\tilde{W}}\,, \tag{125}$$

$$\frac{\mathrm{d}\hat{\tilde{W}}}{\mathrm{d}t} = -\mathrm{i} \left(\tilde{k}_x(t)\, \hat{\tilde{u}}_x + k_y \hat{\tilde{u}}_y \right). \tag{126}$$

By applying the power iteration method jointly to (8)–(10) and (118)–(120) for global azimuthal Fourier harmonics of two-dimensional perturbations or to systems (21)–(22) and (124)–(126) for local SFHs, we automatically arrive at the optimal initial profiles of the enthalpy and velocity component perturbations that maximize the total acoustic energy growth in a given time interval. For Keplerian flows, this problem was solved in [71].

3.4.2 Nonnormality condition for Z. Here, we show that the nonnormality of the dynamical operator determined by system of equations (8)–(10) is a direct consequence of the angular velocity gradient in the flow. We already discussed this in Section 3.1, where we introduced the notion of singular vectors. We can now prove this rigorously in a very general case, because the explicit form of the operator $\mathbf{A}^\dagger$ determined by system (118)–(120) is known. First, we calculate the commutator of $\mathbf{A}$ and $\mathbf{A}^\dagger$:

$$[\mathbf{A}, \mathbf{A}^\dagger] = \begin{pmatrix} \frac{16\Omega^4 - \kappa^4}{4\Omega^2} & 0 & \frac{\mathrm{i}m}{2r\Omega} (4\Omega^2 - \kappa^2) \\ 0 & \frac{\kappa^4 - 16\Omega^4}{4\Omega^2} & \frac{4\Omega^2 - \kappa^2}{2\Omega} \frac{\partial}{\partial r} \\ \frac{\mathrm{i}m a_*^2}{2r\Omega} (\kappa^2 - 4\Omega^2) & \frac{a_*^2}{r\Sigma} \frac{\partial}{\partial r} \left(\frac{r\Sigma}{2\Omega} (\kappa^2 - 4\Omega^2) \right) + \frac{a_*^2}{2\Omega} (\kappa^2 - 4\Omega^2) \frac{\partial}{\partial r} & 0 \end{pmatrix}. \tag{127}$$

It is easy to see that $[\mathbf{A}, \mathbf{A}^\dagger]$ vanishes for $\kappa = 2\Omega$, which corresponds to solid-body rotation. In this case, the commutator $[\mathbf{Z}, \mathbf{Z}^\dagger] = [\exp(\mathbf{A}t), \exp(\mathbf{A}^\dagger t)]$ can easily be calculated, because for commuting operators, the product of their operator exponentials is equal to the exponential of their sum, which can easily be verified by writing the operator exponentials as the corresponding infinite series:

$$\begin{aligned}\left[\exp(\mathbf{A}t), \exp(\mathbf{A}^\dagger t)\right] &= \exp(\mathbf{A}t)\exp(\mathbf{A}^\dagger t)\\ &- \exp(\mathbf{A}^\dagger t)\exp(\mathbf{A}t) = \exp\left((\mathbf{A}+\mathbf{A}^\dagger)\,t\right)\\ &- \exp\left((\mathbf{A}^\dagger+\mathbf{A})\,t\right) = 0\,. \end{aligned} \tag{128}$$

Thus, the operator $\mathbf{Z}$ becomes normal for solid-body rotation.

The converse statement is also valid: if $\mathbf{Z}$ is normal for any time instant t, the rotation is solid-body. To see this, we use the Campbell–Baker–Hausdorff formula ([132], Ch. 25) to represent the composite operators $\mathbf{Z}\mathbf{Z}^\dagger$ and $\mathbf{Z}^\dagger\mathbf{Z}$:

$$\begin{aligned}\exp(\mathbf{A}t)\exp(\mathbf{A}^\dagger t) &= \exp\Bigg((\mathbf{A}+\mathbf{A}^\dagger)\,t + \frac{t^2}{2}[\mathbf{A},\mathbf{A}^\dagger]\\ &+ \frac{t^3}{12}\left[\mathbf{A},[\mathbf{A},\mathbf{A}^\dagger]\right] - \frac{t^3}{12}\left[\mathbf{A}^\dagger,[\mathbf{A},\mathbf{A}^\dagger]\right] + \ldots\Bigg)\,, \end{aligned} \tag{129}$$

$$\begin{aligned}\exp(\mathbf{A}^\dagger t)\exp(\mathbf{A}t) &= \exp\Bigg((\mathbf{A}^\dagger+\mathbf{A})\,t + \frac{t^2}{2}[\mathbf{A}^\dagger,\mathbf{A}]\\ &+ \frac{t^3}{12}\left[\mathbf{A}^\dagger,[\mathbf{A}^\dagger,\mathbf{A}]\right] - \frac{t^3}{12}\left[\mathbf{A},[\mathbf{A}^\dagger,\mathbf{A}]\right] + \ldots\Bigg)\\ &= \exp\Bigg((\mathbf{A}^\dagger+\mathbf{A})\,t - \frac{t^2}{2}[\mathbf{A},\mathbf{A}^\dagger] + \frac{t^3}{12}\left[\mathbf{A},[\mathbf{A},\mathbf{A}^\dagger]\right]\\ &- \frac{t^3}{12}\left[\mathbf{A}^\dagger,[\mathbf{A},\mathbf{A}^\dagger]\right] + \ldots\Bigg)\,. \end{aligned} \tag{130}$$

The equality $\left[\exp(\mathbf{A}t), \exp(\mathbf{A}^\dagger t)\right] = 0$ is satisfied for any t; therefore, the terms with the same powers of t must be independently equal to zero, which is possible only if $[\mathbf{A}, \mathbf{A}^\dagger] = 0$. The last equality is valid only for solid-body rotation.

This implies that solid-body rotation is necessary and sufficient for the dynamical operator $\mathbf{Z}$ of the system of equations (8)–(10) to be normal. Hence, any deviation from solid-body rotation, i.e., the appearance of an angular velocity gradient in astrophysical disks, makes the dynamical operator nonnormal and perturbation modes nonorthogonal to each other.

4. Optimal perturbations in Keplerian disks

In the concluding section of this paper, we briefly discuss the examples of using the variational method for seeking optimal perturbations in astrophysical disks. These last are understood as geometrically thin disks with an almost Keplerian azimuthal velocity profile in a background flow. In numerical calculations, we consider a radially infinite disk with only the inner (free) boundary and a thin quasi-Keplerian torus with inner and outer radial boundaries. The latter configuration was used in the analysis of superpositions of neutral modes in Section 3.2 in studying the superpositions of neutral modes to illustrate the matrix method of optimization. However, for methodological purposes, we start with the simplest analytically tractable problem of the transient growth of local short-wave perturbations with $k_y \gg 1$, which we discussed in detail in Section 2.3.

4.1 Local approximation

We apply the power iteration method to system (21)–(23), (124)–(126) in the limit $k_y \gg 1$ corresponding to an incompressible fluid. In this limit, system (21)–(23) can be reduced to one equation for $\hat{u}_x$:

$$\frac{\mathrm{d}\hat{u}_x}{\mathrm{d}t} + 2qk_y\,\frac{\tilde{k}_x}{k_y^2+\tilde{k}_x^2}\,\hat{u}_x = 0\,, \tag{131}$$

which has the analytic solution

$$\hat{u}_x(t) = \hat{u}_x(0)\,\frac{k_x^2+k_y^2}{\tilde{k}_x^2+k_y^2}\,, \tag{132}$$

which, of course, repeats (35) for $k_y \gg 1$.

At the same time, adjoint equations (124)–(126) in the limit of a quasi-incompressible fluid suggest that the quantity $\hat{u}_x$ conjugate to $\hat{\tilde{u}}_x$ is time-conserved:[10]

$$\frac{\mathrm{d}\hat{\tilde{u}}_x}{\mathrm{d}t} = 0\,. \tag{133}$$

Obviously, after p iterations of an arbitrary initial profile $\hat{u}_x^{\mathrm{in}}(k_x, k_y, t=0)$, we find that it is multiplied by the factor

$$\left[\frac{k_x^2+k_y^2}{(\tilde{k}_x(t))^2+k_y^2}\right]^p\,. \tag{134}$$

With the solution renormalized at each iteration, factor (134) at $p \to \infty$ suppresses all SFHs composing $\hat{u}_x^{\mathrm{in}}(k_x, k_y, t=0)$ except the optimal SFH corresponding to the maximum of (134) as a function of k_x. For a fixed time interval t, this k_x is

$$k_x = \frac{1}{2}\,k_y\left(-qt - \left((qt)^2+4\right)^{1/2}\right). \tag{135}$$

Substituting (135) in the SFH growth factor (42) yields the sought optimal growth G, which for the local problem is defined as (44):

$$G(t) = \frac{(qt)^2 + qt\left[(qt)^2+4\right]^{1/2} + 4}{(qt)^2 - qt\left[(qt)^2+4\right]^{1/2} + 4}\,. \tag{136}$$

Expression (136) gives the first singular value to which the iteration cycle for short-wave local vortices converges. Clearly, for large time intervals $qt \gg 1$, it gives $G \approx (qt)^2$, which reproduces the approximate estimate of G in formula (46).

We also note that the exact result in (136) could be obtained in this simple example directly from the expression for growth factor (42) by calculating the maximum of g as a function of k_x at a fixed t.

For an arbitrary k_y, the optimal growth can be obtained by numerical forward–backward integration of the full system of direct and adjoint equations, which are ordinary differential equations for the SFHs.

4.2 Global problem

When the azimuthal scale of perturbations is comparable to the horizontal disk scale, it is necessary to solve the system of partial differential equations (8)–(10) and (118)–(120)

[10] It can be verified that the quantity I, which was conserved for direct equations (21)–(23), becomes time-dependent in adjoint equations (124)–(126) (see the appendix in [71]).

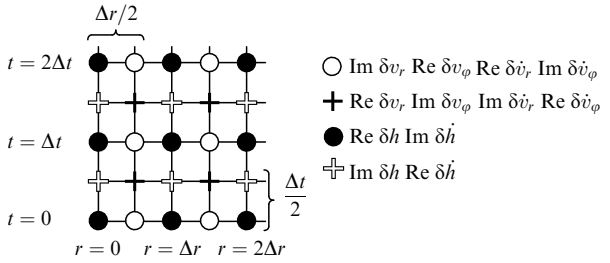

Figure 10. Illustration of the numerical setup for the integration of the system of equations (8)–(10) and (118)–(120).

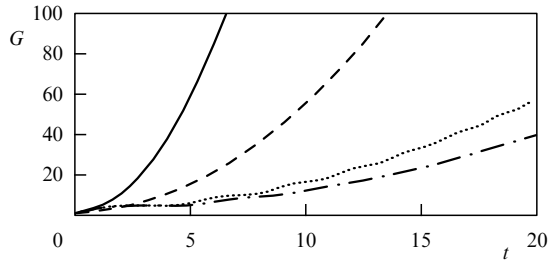

Figure 11. Comparison of the optimal growth for small-scale and large-scale vortices (see Section 2.3) in the global and local space limits. The respective solid and dotted curves are calculated for the local SFHs using formula (44) by the iteration cycle for Eqns (21)–(23) and (124)–(126) for $k_y = 12.5$ and $k_y = 0.125$. Harmonics with $m = 5$ are taken as global perturbations. The optical perturbations are calculated using formula (90) by the iteration cycle for Eqns (8)–(10) and (118)–(120) with the polytropic index $n = 3/2$. Using the relation $(m/r)\, H \sim k_y$, for the respective similar large-scale and small-scales vortices, a disk with $\delta = 0.05$ (dashed-dotted line) and a formally thick disk with $\delta = 5$ (dashed line) were considered. In both cases, the time is in units of $\Omega(r_1)^{-1}$.

numerically, which was done in [71] using a second-order explicit difference procedure (leap-frog) (see, e.g., [76]). In this difference procedure, each equation is separated into real and imaginary parts, and four grids are introduced on the plane (r, t). Unknown variables are calculated at the nodes of these grids using the corresponding differences (Fig. 10). The nodes are shifted with respect to each other by half the time step Δt and/or by half the radial step Δr. This allows using the central approximation to calculate derivatives with respect to r and t, which provides an accuracy of the order of $(\Delta r)^2$ and $(\Delta t)^2$. The time step is determined by the radial step using the Courant condition that follows from a local dispersion relation and can be obtained from the equations being integrated.

4.2.1 Comparison of the transient growth of vortices in global and local space limits. As a background flow, we consider an unbounded Keplerian disk that has only an inner boundary at $r = r_1$. To see how the axial symmetry of the disk and, mainly, the exactly Keplerian angular velocity law $\Omega = \Omega(r_1)(r/r_1)^{-3/2}$ affect the transient growth, we assume for simplicity that all other quantities in the equations for perturbations are constant:

$$\Sigma = \text{const}, \quad a_{\text{eq}} = \frac{\delta}{\sqrt{2n}}\,(\Omega r)\big|_{r_1}\,. \tag{137}$$

As shown in [71], taking a more realistic distribution of Σ and a_{eq} (for example, as in standard accretion disks) does not change the qualitative conclusions presented below. The results of local and global calculations of optimal perturbations by the variational method are shown in Fig. 11.

Here, we compare the transient growth of vortices with azimuthal scales both smaller and larger than the disk thickness. The main qualitative conclusion is that the growth rate of low-scale vortices ($\lambda_\varphi < H$) decreases much faster on scales of the order of the radial distance ($m \sim 1$) than that of large-scale vortices. It can be verified that in the limit case of global perturbations with $m = 1$, the values of G for low-scale and large-scale vortices differ by a factor of 1.5–2 only for the given parameters on time intervals up to $t \approx 20$. On the other hand, for local vortices, the values of G for $\lambda_\varphi < H$ and $\lambda_\varphi > H$ differ by several orders of magnitude. Thus, this calculation suggests that global large-scale vortices in thin Keplerian disks can also exhibit a tenfold increase in very short time intervals of the order of several Keplerian periods at the inner disk boundary. In turn, this may imply the importance of the transient growth of perturbations for the angular momentum transfer on scales much larger than the disk thickness.

4.2.2 Transient spirals and modes in a quasi-Keplerian torus. To conclude, we return to the disk model considered above to illustrate the matrix method for obtaining optimal perturbations (see Section 3.2). As is well known (see, e.g., [81, 82, 96], as well as [100]), this flow demonstrates a weak spectral instability, because there are exponentially growing inertial–acoustic modes. Their increments, as we already mentioned in Section 2.1, rapidly decrease with decreasing the relative geometrical thickness of the torus, i.e., when Ω approaches the Keplerian profile. Perturbations can then grow only due to the transient mechanism of shortening of the leading spirals by the shear flow (see the discussion in Section 2.3), which occurs on short time scales of the order of several Keplerian periods in the flow. However, in the intermediate case, where the pressure gradient in the torus is sufficiently high, both nonmodal and modal perturbation growth can occur simultaneously but at different time scales. The exponential growth of modes always dominates over the transient growth starting from some large time intervals. Interestingly, this essentially means that by calculating the first singular value of the dynamical operator by the variational method, i.e., the optimal growth for a given time interval t, the curve $G(t)$ should become exponential starting from some time, corresponding to the most unstable mode growth. Also, the iteration cycle, which always converges to the optimal initial perturbation vector $\mathbf{q}(t = 0)$, must now yield not a leading spiral but a mode. If at a time $t > 0$ the spiral starts being deformed by the shear flow and is enhanced due to perimeter shortening (see the discussion in Section 2.1), the mode solid-body rotates with the angular velocity coinciding with Ω at the corotation radius inside the flow, because its amplitude increases exactly due to the resonance energy exchange with the flow at this radius. Thus, the method of optimization of perturbations can be applied both to studying the transient perturbation growth and to finding the profiles and increments of the most unstable modes in arbitrary complex shear flows, i.e., to solving the spectral problem as well.

An example of the calculation of a transient spiral and an unstable mode in one toroidal flow by the joint solution of systems (8)–(10) and (118)–(120) by the variational method was presented in Figs 1 and 2 in the Introduction. We see here

that already at $\delta = 0.3$, the maximal increment is very low, and it takes $\sim 10^3$ Keplerian periods for the amplitude of the most unstable mode to at least double. On the other hand, the transient growing spiral increases by a factor of 6 already after several rotational periods at the inner disk boundary.

5. Conclusion

This paper is devoted to the transient dynamics of perturbations, which is of special interest in the theory of astrophysical disks, in particular, accretion disks. If there are no conditions for the magneto-rotational instability in a homogeneous inviscid Keplerian flow, there are no exponentially growing perturbations. Nevertheless, observations suggest that even in this case, an angular flux outwards should somehow occur in the disk. This means, at the very least, that there should be some mechanism of energy transfer from the regular rotational motion to hydrodynamical perturbations. In spectrally stable flows, this can be the transient perturbation growth. Here, we discussed it in terms of the simplest two-dimensional vortices and came to the conclusion that the reason for their growth is the shortening of the length of leading spirals by the differential rotation of the flow (see Figs 2 and 3). Physically, the energy growth in vortices is due to their angular momentum conservation, which in the local limit is expressed by the conservation of their potential vorticity and the existence of the invariant I (see Section 2.2). Here, we considered both low-scale ($k_y \gg 1$) and large-scale ($k_y \ll 1$) vortices and compared their optimal growth, with the nonzero effective viscosity in the disk taken into account (see Fig. 4). Importantly, the transient growth of large-scale vortices increases greatly for a super-Keplerian rotation, which can be significant in relativistic disks with $q > 3/2$. Special attention in this paper was given to the mathematical aspects of nonmodal analysis and to methods of searching for optimal perturbations. We have discussed in detail that the transient growth is a consequence of the nonnormality of the governing dynamical operator of the problem and the nonorthogonality of its eigenvectors — perturbation modes (see Figs 5 and 6). Therefore, the growth of arbitrary perturbations can be properly studied by calculating the singular vectors, not eigenvectors, of this operator. We have considered two methods, a matrix one and a variational one, and applied them to a specific problem (see the corresponding results in Figs 7 and 11). The matrix method requires a discrete representation of the dynamical operator, for example, in the basis of its eigenvectors. The variational method reduces to iterative integration of the system of direct and adjoint equations forward and backward in time. We have emphasized that the variational method is more universal and can be applied to studying nonmodal dynamics of perturbations in nonstationary flows, as well as to nonlinear problems.

As was discussed, transient growth of perturbations is used in the concept of bypass transition to turbulence in laminar flows. It can also be important as a mechanism of additional angular momentum removal and the accretion rate enhancement in weakly turbulized disks. We note that turbulence due to the bypass mechanism is fundamentally different from 'classical' turbulence, in which the energy transfer from the background flow is mediated by modes exponentially growing on large spatial scales, and nonlinear interactions redistribute this energy between modes with other wave vectors k (the so-called direct or inverse cascade). This means that an energy flux $\epsilon_T(k)$ arises in the phase space, which (in the case of direct cascade) brings the kinetic energy of perturbations to small scales where viscous dissipation occurs. In this scenario, the mode direction distribution in the phase space $\mathbf{k}$ is of minor importance, and ϵ_T can be nonzero only along the direction of change of the modulus of $\mathbf{k}$. An entirely different situation must occur when the transient perturbation growth is responsible for the energy redistribution from the background flow. This linear mechanism manifests itself as leading spirals in the disk, i.e., spatial Fourier harmonics corresponding to only such values of $\mathbf{k}$ that $k_x/k_y < 0$. In a spectrally stable flow, where there is no energy supply to the leading spirals, initial perturbations inevitably decay, because the leading spirals turn into trailing ones. Hence, the turbulent state is here possible only due to a positive nonlinear response, which can be the appearance of a nonzero ϵ_T in the direction of the positional change of $\mathbf{k}$, i.e., in the phase space angles, when the trailing spirals give part of their energy back to the leading ones, a part sufficient to sustain the transient growth. Simultaneously, another part of the energy in the trailing spirals dissipates to heat due to their transition to higher k. Here, the heat dissipation can be due not to the direct cascade but to a purely linear winding of the trailing spiral by the flow, i.e., due to the increase in the ratio $k_x/k_y > 0$ with time at $k_y = \text{const}$. As we see, the transverse cascade is an essential part of the alternative picture of turbulence in a shear flow, which is the angular redistribution of spatial Fourier harmonics of perturbations (see, e.g., the appendix in [59]). Maintaining the transient growth of small perturbations by the transverse cascade was studied in detail in [113] for a two-dimensional Couette flow. Such dramatic changes in the concept of the possible structure of turbulent flows should affect both analytic estimates of the turbulent viscosity coefficient (see, e.g., [134]) and numerical simulations of turbulence in astrophysical disks (see, e.g., [135, 136], where mostly spectral properties of the turbulence averaged over directions of $\mathbf{k}$ were studied). We note that we deliberately cited numerical simulations in disks with a magnetic field, in which the modal growth of perturbations due to magneto-rotational instability occurs. The point is that recent studies [137, 138] show that even in Keplerian flows, where the magneto-rotational instability operates, the optimal transiently growing perturbations are dominant over exponentially growing modes on short time-scales. As in an unmagnetized flow, these transient perturbations are locally represented by shear harmonics. Therefore, the nonmodal dynamics of perturbations can also be essential in taking energy from the background flow in MHD turbulent accretion disks. This can also be suggested by recent paper [139], which, like [133], studied the transverse cascade of shear harmonics in a spectrally stable planar magnetized flow numerically and demonstrated that two-dimensional turbulence arises due to a positive feedback with linear transient growth of shear harmonics. The planar Poiseuille flow provides another example of shear flow in which the bypass transition to turbulence turns out to be preferable to the 'classical' mechanism despite the presence of growing modes. Here, we can cite papers [140, 141], which numerically studied some scenarios of turbulence generation from regular initial small perturbations of different types, not the developed turbulence (as is usually done in most papers on MHD turbulence in Keplerian flows) (see also Ch. 9 in [6]). It turns out that the popular early scenario of the flow transition into the turbulent state due to the secondary instability of

saturated modes requires much more time and/or significantly higher initial perturbation amplitudes in the form of so-called streaks growing due to the transient mechanism. We also note that as follows from Fig. 1 in [140], the time of turbulence development from regular initial perturbations strongly depends on their amplitudes. This is not surprising, however, because shear spatial Fourier harmonics of a smaller amplitude require more time to saturate when the secondary instability comes into play that directly leads to the breakaway to turbulence. Clearly, the time of such a transition can be as long as hundreds of characteristic shear times; nevertheless, this does not affect the properties and power of turbulent motions later. Although we presently have only the results of studies of planar flows, in the future they can be obtained for quasi-Keplerian flows with high Reynolds numbers, because locally such flows differ from planar ones only by the presence of the Coriolis force stabilizing the flow. A useful illustration here could be simple finite-difference dynamical models of nonnormal systems with positive feedback that illustrate basic properties of the transition to turbulence in spectrally stable shear flows (see [48, 142]). For example, in Fig. 10 in [48], it can be seen that the time for such a simplified scenario to reach the same 'turbulent' state increases with decreasing the perturbation amplitude and ultimately becomes infinite.

To conclude, we note once again that here we have not considered three-dimensional perturbation dynamics. There are indications that taking the natural inhomogeneity of the disk due to vertical density and pressure gradients into account gives a qualitatively new picture of both the transient growth of perturbations and the subsequent transition to turbulence (see [143]). Here, the perturbation dynamics are essentially three-dimensional, and it can be shown that for three-dimensional transient vortices, there is a time-conserved analog of the invariant of motion I (see [60, 63]). New numerical calculations carried out in [144] also point out that taking the vertical inhomogeneity of the disk into account can result in its destabilization in the subcritical regime at high Reynolds numbers, in contrast to the case observed in a homogeneous flow (see [25]).

The work by D N Razdoburdin was supported by the RSF grant 14-12-00146 for writing Section 3 of this review. The work by V V Zhuravlev was partially supported by the RFBR grants 14-02-91172 and 15-02-08476, and also by Program 9 of the Praesidium of the RAS.

References

1. Armitage P J *Astrophysics of Planet Formation* (Cambridge: Cambridge Univ. Press, 2010)
2. Youdin A N, Kenyon S J, in *Planets, Stars and Stellar Systems* Vol. 3 *Solar and Stellar Planetary Systems* (Eds T D Oswalt, L M French, P Kalas) (Berlin: Springer, 2013) p. 1
3. Remillard R A, McClintock J E *Annu. Rev. Astron. Astrophys.* **44** 49 (2006)
4. Monin A S, Yaglom A M *Statistical Fluid Mechanics* Vol. 1 (Cambridge, Mass.: MIT Press, 1971); Translated from Russian: *Statisticheskaya Gidromekhanika* Pt. 1 (Moscow: Nauka, 1965)
5. Menou K *Science* **288** 2022 (2000)
6. Schmid P J, Henningson D S *Stability and Transition in Shear Flows* (New York: Springer, 2001)
7. Henningson D S, Reddy S C *Phys. Fluids* **6** 1396 (1994)
8. Balbus S A, Hawley J F *Astrophys. J.* **376** 214 (1991)
9. Hawley J F, Gammie S F, Balbus S A *Astrophys. J.* **440** 742 (1995)
10. Stone J M et al. *Astrophys. J.* **463** 656 (1996)
11. Balbus S A, Hawley J F *Rev. Mod. Phys.* **70** 1 (1998)
12. Balbus S A *Annu. Rev. Astron. Astrophys.* **41** 555 (2003)
13. Fridman A M, Bisikalo D V *Phys. Usp.* **51** 551 (2008); *Usp. Fiz. Nauk* **178** 577 (2008)
14. Kurbatov E P, Bisikalo D V, Kaigorodov P V *Phys. Usp.* **57** 787 (2014); *Usp. Fiz. Nauk* **184** 851 (2014)
15. Lovelace R V E et al. *Astrophys. J.* **513** 805 (1999)
16. Klahr H, Hubbard A *Astrophys. J.* **788** 21 (2014)
17. Rayleigh Lord "On the dynamics of revolving fluids", in *Scientific Papers* Vol. 6 (Cambridge: The Univ. Press, 1916) p. 447
18. Landau L D, Lifshitz E M *Fluid Mechanics* (Oxford: Pergamon Press, 1987); Translated from Russian of the previous ed.: *Gidrodinamika* (Moscow: Fizmatlit, 2006)
19. Rayleigh Lord "On the stability or instability of certain fluid motions", in *Scientific Papers* Vol. 1 (Cambridge: The Univ. Press, 1880) p. 474
20. Charney J G, Fjörtoft R, von Neumann J *Tellus* **2** 237 (1950)
21. Drazin P G, Reid W H *NASA STI/Recon Tech. Rep. A* **82** 17950 (1981)
22. Joseph D D *NASA STI/Recon Tech. Rep. A* **77** 287 (1976)
23. Balbus S A, Hawley J F, Stone J M *Astrophys. J* **467** 76 (1996)
24. Hawley J F, Balbus S A, Winters W F *Astrophys. J* **518** 394 (1999)
25. Shen Y, Stone J M, Gardiner T A *Astrophys. J* **653** 513 (2006)
26. Ji H et al. *Nature* **444** 343 (2006)
27. Schartman E, Ji H, Burin M J *Rev. Sci. Instrum.* **80** 024501 (2009)
28. Schartman E et al. *Astron. Astrophys.* **543** 13 (2012)
29. Lesur G, Longaretti P Y *Astron. Astrophys.* **444** 25 (2005)
30. Taylor G I *Proc. R. Soc. London A* **157** 546 (1936)
31. Wendt G *Ann. Physik* **409** 445 (1933)
32. Zeldovich Y B *Proc. R. Soc. London A* **374** 299 (1981)
33. Richard D, Zahn J-P *Astron. Astrophys.* **347** 734 (1999)
34. Longaretti P-Y *Astrophys. J.* **576** 587 (2002)
35. Paoletti M S, Lathrop D P *Phys. Rev. Lett.* **106** 024501 (2011)
36. Paoletti M S et al. *Astron. Astrophys.* **547** 11 (2012)
37. Avila M *Phys. Rev. Lett.* **108** 124501 (2012)
38. Edlund E M, Ji H *Phys. Rev. Lett.* **89** 021004 (2014)
39. Kelvin L *Phil. Mag.* **23** 459 (1887)
40. Orr W McF *Proc. R. Irish Acad. A* **27** 9 (1907)
41. Orr W McF *Proc. R. Irish Acad. A* **27** 69 (1907)
42. Goldreich P, Lynden-Bell P *Mon. Not. R. Astron. Soc.* **130** 125 (1965)
43. Julian W H, Toomre A *Astrophys. J.* **146** 810 (1966)
44. Farrell B F *Phys. Fluids* **31** 2093 (1988)
45. Butler K M, Farrell B F *Phys. Fluids A* **4** 1637 (1992)
46. Reddy S C, Schmid P J, Henningson D S *SIAM J. Appl. Math.* **53** 15 (1993)
47. Reddy S C, Henningson D S *J. Fluid Mech.* **252** 209 (1993)
48. Trefethen L N et al. *Science* **261** 578 (1993)
49. Schmid P J *Annu. Rev. Fluid Mech.* **39** 129 (2007)
50. Ioannou P J, Kakouris A *Astrophys. J.* **550** 931 (2001)
51. Yecko P A *Astron. Astrophys.* **425** 385 (2004)
52. Mukhopadhyay B, Afshordi N, Narayan R *Astrophys. J.* **629** 383 (2005)
53. Meseguer A *Phys. Fluids* **14** 1655 (2002)
54. Maretzke S, Hof B, Avila M *J. Fluid Mech.* **742** 254 (2014)
55. Coles D *J. Fluid Mech.* **21** 385 (1965)
56. Lominadze D G, Chagelishvili G D, Chanishvili R G *Sov. Astron. Lett.* **14** 364 (1988); *Pis'ma Astron. Zh.* **14** 856 (1988)
57. Fridman A M *Sov. Astron. Lett.* **15** 487 (1989); *Pis'ma Astron. Zh.* **15** 1122 (1989)
58. Chagelishvili G D et al. *Phys. Rev. Lett.* **79** 3178 (1997)
59. Chagelishvili G D et al. *Astron. Astrophys.* **402** 401 (2003)
60. Tevzadze A G et al. *Astron. Astrophys.* **407** 779 (2003)
61. Afshordi N, Mukhopadhyay B, Narayan R *Astrophys. J.* **629** 373 (2005)
62. Bodo G et al. *Astron. Astrophys.* **437** 9 (2005)
63. Tevzadze A G, Chagelishvili G D, Zahn J P *Astron. Astrophys.* **478** 9 (2008)
64. Heinemann T, Papaloizou J C B *Mon. Not. R. Astron. Soc.* **397** 52 (2009)
65. Heinemann T, Papaloizou J C B *Mon. Not. R. Astron. Soc.* **397** 64 (2009)
66. Tevzadze A G et al. *Mon. Not. R. Astron. Soc.* **401** 901 (2010)
67. Volponi F *Mon. Not. R. Astron. Soc.* **406** 551 (2010)
68. Salhi A, Pieri A B *Phys. Rev. E* **90** 043003 (2014)

69. Umurhan O M et al. *Astron. Astrophys.* **446** 1 (2006)
70. Rebusco P et al. *Phys. Fluids* **21** 076601 (2009)
71. Zhuravlev V V, Razdoburdin D N *Mon. Not. R. Astron. Soc.* **442** 870 (2014)
72. Pringle J E, King A R *Astrophysical Flows* (Cambridge: Cambridge Univ. Press, 2007)
73. Landau L D, Lifshitz E M *Statistical Physics* Vol. 1 (Oxford: Pegramon Press, 1980); Translated from Russian: *Statisticheskaya Fizika* Pt. 1 (Moscow: Nauka, 1976)
74. Tassoul J-L *Theory of Rotating Stars* (Princeton: Princeton Univ. Press, 1978); Translated into Russian: *Teoriya Vrashchayushchikhsya Zvezd* (Moscow: Mir, 1982)
75. Okazaki F T, Kato S, Fukue J *Astron. Soc. Jpn.* **39** 457 (1987)
76. Frank J, Robertson J A *Mon. Not. R. Astron. Soc.* **232** 1 (1988)
77. Kojima Y *Mon. Not. R. Astron. Soc.* **236** 589 (1989)
78. Kojima Y, Miyama S M, Kubotani H *Mon. Not. R. Astron. Soc.* **238** 753 (1989)
79. Churilov S M, Shuhman I G *Astron. Tsirk.* (1157) 1 (1981)
80. Blaes O V, Glatzel W *Mon. Not. R. Astron. Soc.* **220** 253 (1986)
81. Glatzel W *Mon. Not. R. Astron. Soc.* **225** 227 (1987)
82. Glatzel W *Mon. Not. R. Astron. Soc.* **228** 77 (1987)
83. Narayan R, Goodman J, in *Theory of Accretion Disks. Proc. of a NATO Advanced Research Workshop, Garching, Germany, March 6–10, 1989* (NATO ASI Ser. C, Vol. 290, Ed. F Meyer) (Dordrecht: Kluwer, 1989) p. 231
84. Narayan R, in *Structure and Emission Properties of Accretion Disks. Proc. of IAU Colloq. 129, the 6th Institute d'Astrophysique de Paris* (*IAP*) *Meeting, Paris, France, 2–6 July, 1990* (Eds C Bertout, S Collin-Souffrin, J P Lasota) (Gif-sur-Yvette: Editions Frontieres, 1991) p. 153
85. Papaloizou J C B, Pringle J E *Mon. Not. R. Astron. Soc.* **208** 721 (1984)
86. Papaloizou J C B, Pringle J E *Mon. Not. R. Astron. Soc.* **213** 799 (1985)
87. Papaloizou J C B, Pringle J E *Mon. Not. R. Astron. Soc.* **225** 267 (1987)
88. Goldreich P, Goodman J, Narayan R *Mon. Not. R. Astron. Soc.* **221** 339 (1986)
89. Kojima Y *Prog. Theor. Phys.* **75** 251 (1986)
90. Kato S *Publ. Astron. Soc. Jpn.* **39** 645 (1987)
91. Sekiya M, Miyama S M *Mon. Not. R. Astron. Soc.* **234** 107 (1988)
92. Friedman J L, Schutz B F *Astrophys. J.* **221** 937 (1978)
93. Goldreich P, Narayan R *Mon. Not. R. Astron. Soc.* **213** 7P (1985)
94. Drury L O *Mon. Not. R. Astron. Soc.* **217** 821 (1985)
95. Narayan R, Goldreich P, Goodman J *Mon. Not. R. Astron. Soc.* **228** 1 (1987)
96. Glatzel W *Mon. Not. R. Astron. Soc.* **231** 795 (1988)
97. Savonije G J, Heemskerk M H M *Astron. Astrophys.* **240** 191 (1990)
98. Stepanyants Yu A, Fabrikant A L *Sov. Phys. Usp.* **32** 783 (1989); *Usp. Fiz. Nauk* **159** 83 (1989)
99. Zhuravlev V V, Shakura N I *Astron. Lett.* **33** 536 (2007); *Pis'ma Astron. Zh.* **33** 604 (2007)
100. Zhuravlev V V, Shakura N I *Astron. Lett.* **33** 673 (2007); *Pis'ma Astron. Zh.* **33** 754 (2007)
101. Hill G W *Am. J. Math.* **1** 5 (1878)
102. Regev O, Umurhan O M *Astron. Astrophys.* **481** 21 (2008)
103. Binney J, Tremaine S *Galactic Dynamics* 2nd ed. (Princeton: Princeton Univ. Press, 2008)
104. Johnson B M, Gammie C F *Astrophys. J.* **626** 978 (2005)
105. Chagelishvili G D, Rogava A D, Segal I N *Phys. Rev. E* **50** R4283 (1994)
106. Brekhovskikh L M, Goncharov V *Mechanics of Continua and Wave Dynamics* (Berlin: Springer-Verlag, 1985); *Vvedenie v Mekhaniku Sploshnykh Sred* (*v Prilozhenii k Teorii Voln*) (Moscow: Nauka, 1982)
107. Chagelishvili G D, Chanishvili R G, Lominadze D G *JETP Lett.* **63** 543 (1996); *Pis'ma Zh. Eksp. Teor. Fiz.* **63** 517 (1996)
108. Vilenkin N Ya et al. *Functional Analysis* (Ed. S G Krein) (Groningen: Wolters-Noordhoff, 1972); *Funktsional'nyi Analiz* (Ed. S G Krein) (Moscow: Nauka, 1972)
109. Kolmogorov A N, Fomin S V *Introductory Real Analysis* (New York: Dover Publ., 1975); Translated from Russian: *Elementy Teorii Funktsii i Fundamental'nogo Analiza* (Moscow: Nauka, 1968)
110. Il'in V A, Sadovnichii V A, Sendov Bl H *Matematicheskii Analiz* (Mathematical Analysis) Vol. 2 *Prodolzhenie Kursa* (Continuation of the Course) (Moscow: Izd. MGU, 1987)
111. Korn G A, Korn T M *Mathematical Handbook for Scientists and Engineers* (New York: McGraw-Hill, 1968); Translated from Russian: *Spravochnik po Matematike dlya Nauchnykh Rabotnikov i Inzhenerov* (Moscow: Nauka, 1984)
112. Farrell B F, Ioannou P J *J. Atmos. Sci.* **53** 2025 (1996)
113. Razdoburdin D N, Zhuravlev V V *Astron. Lett.* **38** 117 (2012); *Pis'ma Astron. Zh.* **38** 138 (2012)
114. Golub G H, Reinsch C *Numer. Math.* **14** 403 (1970)
115. Golub G H, Van Loan C F *Matrix Computations* (Baltimore: Johns Hopkins Univ. Press, 1996)
116. Hanifi A, Schmid P J, Henningson D S *Phys. Fluids* **8** 826 (1996)
117. Malik M, Alam M, Dey J *Phys. Fluids* **18** 034103 (2006)
118. Zhuravlev V V, Shakura N I *Astron. Nachr.* **330** 84 (2009)
119. Kato S *Publ. Astron. Soc. Jpn.* **53** 1 (2001)
120. Heading J *An Introduction to Phase-Integral Methods* (Mineola: Dover Publ., 2013); Translated into Russian: *Vvedenie v Metod Fazovykh Integralov*. (*Metod VKB*) (Moscow: Mir, 1965)
121. Luchini P *J. Fluid Mech.* **404** 289 (2000)
122. Lin C *The Theory of Hydrodynamic Stability* (Cambridge: Univ. Press, 1955); Translated into Russian: *Teoriya Gidrodinamicheskoi Ustoichivosti* (Moscow: IL, 1958)
123. Andersson P, Berggren M, Henningson D S *Phys. Fluids* **11** 134 (1999)
124. Farrell B F, Ioannou P J *J. Atmos. Sci.* **53** 2041 (1996)
125. Corbett P, Bottaro A *J. Fluid Mech.* **435** 1 (2001)
126. Guégan A, Schmid P J, Huerre P *J. Fluid Mech.* **566** 11 (2006)
127. Gunzburger M D *Perspectives in Flow Control and Optimization* (Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003)
128. Marchuk G I *Sbornik Mathematics* **189** 1505 (1998); *Matem. Sb.* **189** (10) 75 (1998)
129. Cherubini S et al. *J. Fluid Mech.* **656** 231 (2010)
130. Cherubini S et al. *J. Fluid Mech.* **689** 221 (2011)
131. Cherubini S, De Palma P *J. Fluid Mech.* **716** 251 (2013)
132. Richtmyer R D *Principles of Advanced Mathematical Physics* Vol. 2 (New York: Springer-Verlag, 1981); Translated into Russian: *Printsipy Sovremennoi Matematicheskoi Fiziki* Vol. 2 (Moscow: Mir, 1984)
133. Horton W et al. *Phys. Rev. E* **81** 066304 (2010)
134. Canuto V M, Goldman I, Hubickyj O *Astrophys. J. Lett.* **280** L55 (1984)
135. Simon J B, Hawley J F, Beckwith K *Astrophys. J.* **690** 974 (2009)
136. Davis S W, Stone J M, Pessah M E *Astrophys. J.* **713** 52 (2010)
137. Squire J, Bhattacharjee A *Astrophys. J.* **797** 15 (2014)
138. Squire J, Bhattacharjee A *Phys. Rev. Lett.* **113** 025006 (2014)
139. Mamatsashvili G R et al. *Phys. Rev. E* **89** 043101 (2014)
140. Schmid P J, Reddy S C, Henningson D S, in *Advances in Turbulence VI* (Eds S Gavrilakis, L Machiels, P A Monkewitz) (Dordrecht: Kluwer Acad. Publ., 1996) p. 381
141. Reddy S C et al. *J. Fluid Mech.* **365** 269 (1998)
142. Waleffe F *Phys. Fluids* **7** 3060 (1995)
143. Lominadze J G, in *Advances in Plasma Astrophysics. Proc. of the Intern. Astronomical Union, IAU Symp.* Vol. 274 (Eds A Bonanno, E de Gouveia Dal Pino, A G Kosovichev) (Cambridge: Cambridge Univ. Press, 2011) p. 318
144. Marcus P et al. *Astrophys. J.* **808** 87 (2015); arXiv:1410.8143