THE STABILITY OF A MODEL GALAXY

by

THOMAS ARTHUR ZANG

B.S., University of Notre Dame
1971

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF
TECHNOLOGY
June, 1976

Signature of Author . . . . . .
Department of Mathematics
February 17, 1976

Certified by . . . . . .
Thesis Supervisor

Accepted by . . . . . .
Chairman, Departmental Committee
on Graduate Students

THE STABILITY OF A MODEL GALAXY

by

THOMAS ARTHUR ZANG

Submitted to the Department of Mathematics on February 17, 1976 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

This thesis examines the large-scale stability of a class of infinite, self-similar models of disk galaxies characterized by a constant circular velocity. These axisymmetric and self-gravitating disks are presumed to consist only of stars which, in addition to their mean rotation around the center, are endowed with considerable random velocities. The surface density of these equilibrium models varies inversely with radius, but the in-and-out components of the velocities of their stars are assumed distributed in an identical Gaussian manner at all radii.

Our stability analysis consists of determining whether any given member of this family of models admits exponentially growing modes of infinitesimal amplitude. Without loss of generality, we presume such modes to have the angular dependence $\exp(im\theta)$, where θ is the angular coordinate, and we examine separately the appropriate linearized equations of motion for each angular harmonic m. We take advantage of the self-similarity of the models to perform certain of the integrations analytically. Ultimately, for each m, we obtain an exact integral equation of Fredholm type that must be satisfied by the radial Fourier transform of the disturbance density. The solutions of this integral equation are pursued numerically.

We find that the present disks are stable with respect to all axisymmetric ($m = 0$) disturbances provided the velocity dispersion is chosen sufficiently large. Specifically needed for stability is that the r.m.s. random velocity in the radial direction exceed 0.3781 times the circular speed; this result agrees closely with an estimate based on a short-wavelength approximation.

In order to examine non-axisymmetric disturbances, it was necessary here to immobilize the centermost parts of our disks, owing chiefly to the arbitrarily rapid angular motions found therein. This was done by artificially reducing, from unity at large radii to zero at the center, the fraction of the stars that was actually permitted to partake in a disturbance.

Among the non-axisymmetric results so obtained, probably the most interesting concern bisymmetric ($m = 2$) disturbances. Provided the center has been "cut out" fairly gradually, we find invariably that the $m = 2$ modes become stable sooner than the axisymmetric or $m = 0$ modes as the velocity dispersion is increased. This conclusion contrasts markedly with several reports in the recent literature that $m = 2$ modes seem distinctly more troublesome than those of kind $m = 0$; we do not dispute such findings, but only reemphasize that they do not seem to apply here. To be sure, in cases where the central stars were excluded rather sharply with radius, even our models exhibit some growing $m = 2$ modes that persist beyond the axisymmetric stability boundary; however, the pattern speeds then leave little doubt that such instabilities are indebted to the artificially sharp inner edge of the "active" disk.

Similar studies of the $m = 3$ and $m = 4$ disturbances indicate that these always stabilize sooner than the $m = 2$ modes. On the other hand, our analysis of the $m = 1$ modes remains incomplete, inasmuch as we have not yet examined properly the $m = 1$ modes that leave fixed the center of mass of the entire system. Instead, our $m = 1$ studies thus far have been confined largely to the simpler but less realistic problem for which it is the excluded central disk that is held fixed in space. Here we find an instability that persists long after axisymmetric stability has been achieved, but it is not at all clear yet that such behavior would remain if the supposedly rigid center were released and thus were able to displace laterally.

Thesis Supervisor: Alar Toomre
Title: Professor of Applied Mathematics

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Above all, I wish to thank my thesis advisor, Prof. Alar Toomre, for suggesting this problem and for guiding me toward its resolution.

I am especially grateful to Dr. Agris Kalnajs for his many useful suggestions which have been incorporated in this thesis.

It is a pleasure to acknowledge the profitable discussions I have had with Prof. C. C. Lin and Prof. James Mark.

I am indebted to my parents for their encouragement and support throughout my studies.

Finally, I am appreciative of the financial support provided by the National Science Foundation.

I. INTRODUCTION

A good deal of progress has been made during the past decade toward answering the long-standing questions: why are there spiral galaxies? why are there barred galaxies? why indeed do disk galaxies as a class display such a rich variety of forms? The answers, as we now know, involve density waves and related shocks, some Jeans instabilities, objects like Riemann ellipsoids, and once in a while even interactions with neighboring galaxies. All these phenomena are basically gravitational in nature.

All these phenomena also touch sooner or later upon the fundamental issue of the overall stability of a galaxy. Unfortunately, that important topic is still not well understood. No firm theoretical basis yet exists for determining just what it takes to stabilize a disk galaxy against all conceivable large-scale perturbations or, for that matter, for deciding which disturbances are the most difficult to suppress. The empirical evidence suggests that spiral and/or bar-like waves are the perturbations which an almost but not quite stable galaxy is most prone to develop spontaneously. But even this tendency has still not been firmly established from the laws of dynamics.

That most of the grand — and presumably also many of the not so grand — spiral and bar-like structures observed in galaxies indeed represent large-scale "density waves" of the kind first suggested by B. Lindblad (1942) in the context of bars and, more vaguely in his 1963-64 papers, as the sort of "circulation" of material in and out of arms needed to explain normal spirals, seems hardly in dispute any longer. It is, of course, due especially to Lin and Shu (1964, 1966, and subsequent papers) that this wave idea has by now grown into a widely appreciated physical theory. Interestingly, the most crucial contribution of even Lin and Shu consisted, in retrospect, not of inferring from any first principles that certain spiral waves can long endure; rather, their key step was simply to hypothesize boldly that such waves really do exist. From this, as they and others showed, one can not only deduce various dynamical properties of the waves which invite comparison with observations, but one even discovers that the observations seem to agree!

By now, the most impressive area of agreement concerns the gas flow in a galaxy whose force field itself contains some spiral structure. In such circumstances, as Roberts (1969) demonstrated, somewhat following the lead of Fujimoto (1968), a shock wave will form in the gas on the inner edge of a trailing spiral arm if the spiral field is stronger than a few percent of the background or mean gravity. Subsequently the radio astronomers Mathewson, van der Kruit and Brouw (1972)

found compelling evidence that such a shock is indeed present in the galaxy M51: on the inner edge of its spiral arms are located not only the well-known optical dust lanes but also an intense synchrotron emission at radio wavelengths; on the density-wave hypothesis, both are expected to occur in regions of greatest gas compression. There now also appears to be rather good evidence of a shock wave in the galaxy M81: from his own Westerbork radio observations with Shane, Rots (1974) reported that the neutral hydrogen there seems not only concentrated in the major optical arms but it also exhibits velocities to be expected from flows such as studied by Roberts. Altogether, the possibility of gas shocks seems very reassuring: as Lin, Shu and others have remarked, it offers much hope of accounting for the presence of the many bright, young stars seen in the vicinity of dust lanes. As they have argued, it is quite plausible (although the details remain vague) that the sudden gas compression in the shock triggers star formation; the resulting bright stars burn out not long after their birth, and hence remain visible only near their places of origin. Facts like these explain much of the present confidence in the density-wave hypothesis.

Yet even these successes presuppose some understanding of the stability requirements of a galaxy. This is illustrated by two of the main assumptions which Lin and his coworkers felt obliged to make in their calculations of the detailed properties of the waves whose existence they had postulated.

For one thing, as Lin and Shu seemed acutely aware from their 1966 paper onwards, it made sense to contemplate density waves only in circumstances where these were not about to be overwhelmed by some (other) severe instabilities of the system. They and most subsequent workers assumed that the random stellar velocities in various parts of any likely disk have roughly to equal the minimum value which Toomre (1964) had "locally" estimated to be necessary for ensuring axisymmetric stability. This assumption was both reasonable and convenient, but of course it never seemed very certain, even in the 1960's, that such velocities sufficed to ensure complete stability.

The second point is that practically all the Lin-Shu and related analyses assumed the waves to be tightly-wrapped, mainly in order to be mathematically tractable. This wise decision gave rise, among other things, to the useful stellar-dynamical dispersion relation of Lin and Shu (1966). But again it was never very clear, either <u>a priori</u> or <u>a posteriori</u>, that all significant parts of a genuine galactic density wave really obey such a short-wave assumption. At any rate, both the angular velocity and any conceivable growth rate — plus even the trailing versus leading nature — of the wave patterns seemed long to be almost free parameters in the Lin-Shu theory.

In reverse order, those two sources of vague uneasiness with the density-wave theory each turned into a serious issue involving stability, as follows.

Chronologically first came the demonstration by Toomre (1969) that even the tightly-wrapped waves possess a significant group velocity, and that this velocity should cause existing waves to drift away into the vicinity of Lindblad resonances [and there decay, as Mark (1971) later established] within a few "galactic years". Yet the disk galaxies in which grand spiral patterns are observed have themselves existed for tens of rotation periods, and the prevalence of their spiral patterns suggests that even these are hardly transitory phenomena. Hence if the observed spirals involve density waves, a mechanism for replenishing the waves must be found and exhibited. Broadly speaking, what seems needed in the typical isolated galaxy is some further, if relatively mild, instability. Quite possibly, the fundamental scale of such instability may need to be truly of galactic dimensions — and hence well outside the tightly-wrapped approximation — if we are to trust the corollary impression that the group velocity also carries information from the very open toward the moderately tight waves. To this date, no very convincing candidate has yet emerged.

The other and probably even more serious worry is that it has become increasingly evident in recent years that local or short-scale instabilities are really not the most serious ones to which a disk galaxy may be prone. The elimination of all non-axisymmetric instabilities now appears to require, at least in some major regions of a disk, considerably more

velocity dispersion, or "heat", than was estimated by Toomre for axisymmetric reasons alone. The N-body calculations of Miller, Prendergast and Quirk (1970) and Miller (1971) furnished the first clear signals of this worrisome behavior. Subsequent, more extensive N-body studies by Hohl (1971) confirmed it. The disks which started with barely enough velocity dispersion to eliminate Jeans instabilities of the axisymmetric sort were found by both groups of workers to develop very quickly various large-scale and non-axisymmetric inhomogeneities — which in turn soon mixed away into a newly smooth but much hotter disk. The case was strengthened when Kalnajs (1972) reported that a similar fate seemed to befall certain stellar disks which he had studied analytically. Partly on this basis, Ostriker and Peebles (1973) went so far as to suggest that a very massive halo may be required for a galaxy to be stable and yet exhibit random velocities in the disk itself which are as small as those in the solar neighborhood seem to be. Thus, even apart from spiral waves, it is obviously urgent that the dynamical conditions under which disk galaxies are stable be clarified as soundly as possible.

* * * *

Of ccurse, all this has been meant only as a review. The need for comprehensive studies of the global stability and/or modes of specific galactic disk models has been apparent

for quite some time. In principle, of course, such stability questions may be answered in a direct enough manner. Since real galactic disks are quite thin and flat, it has long seemed sensible to idealize these disks as infinitesimally thin; also, since the stars in a galaxy number in many billions, it has seemed only natural to describe them collectively in terms of distribution functions, Vlasov equations, and the like. As is well known, the mathematical problems governing the linearized modes of such idealized stellar disks can then be formulated easily enough. The trouble, however, is that even those systems still depend on two coordinate and momentum variables each, and they do so in both a differential and an integral sense. Hence except for special galaxy models where all or at least many of the computations can be done analytically, the numerical solution of this problem remains simply too formidable at present for results to be obtained abundantly at a reasonable cost.

Certainly a more straightforward approach to the stability of hot stellar disks was taken in the extensive N-body simulations to which we have referred above. Here again, though, the sheer size of the problem is staggering: the computation of billions of individual orbits, with each star influenced by the gravitational attraction of all the other stars in a galaxy, is simply out of the question with present-day (and any foreseeable) machines. However, reasonably honest simulations consisting of roughly 10^5 stars moving in their own

smoothed gravitational field have been produced. These simulations were devised by Miller and Prendergast (1968) and by Hohl and Hockney (1969). Especially Hohl (1973) has recently shown that the relaxation times in such softened N-body simulations seem to be much longer than the evolutionary times typically explored; hence the computed systems seem effectively collisionless, just like actual galaxies.

As mentioned earlier, it was the numerical experiments by Miller et al. that gave the first indications that non-axisymmetric instabilities are much more serious than axisymmetric ones. More specifically, Hohl (1973) later reported that although Toomre's criterion does in fact provide security against axisymmetric and even arbitrary short-scale instabilities, it does not protect a disk against the slow but steady growth of large-scale, and often bar-like, structures. He also noted that even when the high random velocites of the eventual hot disks are reduced artificially in the hope of obtaining cooler stable disks, the resulting models soon again heat up dramatically through non-axisymmetric instabilities.

In addition to suggesting that considerable random velocities are required to stabilize disk galaxies, the numerical simulations have provided some clues to the dominant types of disturbances in moderately unstable systems. Some striking trailing spiral patterns have been observed in the experiments previously cited and in others by Quirk (1971) and by Hockney

and Brownrigg (1974). These patterns were unmistakably wave-like: the stars clearly moved into and then out of the density concentrations. However, none of these wave patterns managed to remain essentially unchanged for more than a couple galactic revolutions; instead, the spirals gradually sheared and/or broke up and reassembled into somewhat different structures. Nothing resembling a true spiral mode can really be said to have been found in any of the N-body experiments. Moreover, some type of non-stellar (or "gaseous") component has frequently appeared necessary for even these transitory spirals to develop. On the other hand, a reasonably steady bar-like structure of finite amplitude had resulted fairly often in these experiments.

A thorough exploration of galaxy models by N-body experiments would almost certainly resolve most of the pressing stability issues involved. Unfortunately, such simulations are still very expensive, despite all the cleverness that already underlies the present schemes. Thus even the most valuable experiments can still yield only fragmentary information. Moreover, although the computational relaxation times seem fairly large, some suspicions remain that certain gross statistical properties of the simulations may still be contaminated by microscopic errors. In essence, though, it is the cost of the N-body computations that provides the truly compelling reason for seeking more efficient alternative analyses of hot galaxy models.

To date, the only thoroughly reported large-scale, or global, analysis of the linear modes of an idealized disk galaxy with non-vanishing random stellar motions has been the study by Kalnajs (1972) referring to a family of uniformly rotating models invented by Freeman (1966). He exhibited certain hot disks which had no growing modes and others which were unstable. Prominent among the latter was one model which had much larger random velocities than are required to suppress all axisymmetric instabilities, and yet it still had many growing non-axisymmetric modes. Its most pronounced instability definitely had a bar-like nature. However, it should be cautioned that the uniformly rotating Freeman models differ in two significant respects from real galaxies: their complete absence of shear, and their rather unusual velocity distributions.

In addition, Kalnajs has long been engaged in studies of large-scale modes in more realistic galaxy models; from those studies, relatively little seems yet to have been published (cf. Kalnajs 1970), although his mathematical methods were described extensively in his 1971 paper.

In contrast to the situation with hot stellar disks, the question of global stability in cold disks (purely circular stellar orbits) is fairly tractable. Obviously these systems are much simpler mathematically, in large part because momenta in them do not need to be treated as independent variables. The investigations have typically searched for modes whose

angular and temporal behavior can be described by the factor $e^{i(m\theta-\omega t)}$, where θ is the angular coordinate, t is the time, m is the angular harmonic number and ω is the frequency. The analyses have sought the characteristic frequencies ω for various angular harmonics. The first global examinations of cold disks were performed by Hunter (1963, 1965), who also studied, among other things, a finite uniformly rotating disk — a cold version of the hot disks later examined by Kalnajs. For axisymmetric perturbations Hunter found a single stable disturbance and an infinity of unstable ones, these having ever-decreasing wavelengths. For each of the non-axisymmetric harmonics he found an infinity of unstable modes. When he examined a few other galaxy models he found that several (but by no means all) of the long-wavelength axisymmetric disturbances could be stabilized by an increase in the central concentration of the surface density of the equilibrium disk. In a separate work, Toomre (1964) looked for the self-consistent axisymmetric perturbations of an infinite cold disk model; both discrete stable modes and a continuum of unstable modes were displayed by him. Miyamoto (1969) later searched for the bisymmetric ($m = 2$) modes of the same model and again found a variety of stable and unstable disturbances.

Intermediate in difficulty between cold disks and hot disks composed only of stars are the gaseous disks with finite pressure. Bardeen (1975) has recently produced the first

global stability analysis of some such systems. The models he studied had a polytropic relation between the density and the pressure and frequently were imbedded in a rigid three-dimensional halo. Focusing on bisymmetric disturbances, Bardeen integrated numerically the equations satisfied by the perturbation variables and was able to extract the most rapidly-growing mode. His results for these systems were consistent with the computer simulations: $m = 2$ instabilities definitely remain in disks which are barely stable to $m = 0$ disturbances. Several of the growing spiral patterns that he found had rather large wavelengths and thus were truly global in nature.

Besides its literal interpretation, a non-zero pressure in a gaseous disk can also be viewed as a rough substitute for some of the effects of random velocities in stars. An alternative modification of cold disks to mimic a true velocity dispersion is to soften the gravitational interaction by changing the potential from its usual form of

$$\phi(r) = -GM/s$$

to

$$\phi(r) = -GM/\sqrt{s^2 + a^2} \ ,$$

where G is the gravitational constant, M is the mass of an attracting body, s is the particle separation and the constant a provides the length scale of the softening. Erickson (1974) and Miller (1974) both employed this idea in

full-scale stability analyses — having noticed that under such a force law the local dispersion relation for short-wavelength waves is qualitatively similar, for both long and short wavelengths, to the Lin-Shu dispersion relation for stellar disks. A very attractive feature of these systems is that all additional complications resulting from initially non-circular stellar orbits are avoided. Unfortunately, although Erickson found that all axisymmetric disturbances can be stabilized in such a modified disk with a Gaussian density by choosing moderate values of the softening length a, he discovered that it was not possible to stabilize the disk against non-axisymmetric disturbances, even when the softening length was chosen comparable to the scale of the disk.

All in all, there are many indications that disk galaxies are more susceptible to large-scale non-axisymmetric disturbances than to any axisymmetric type. Likewise bars and/or trailing spirals seem to emerge often, but not always. However, neither type of evidence can yet be called conclusive in view of the scarcity of thorough stellar dynamical analyses.

* * * *

This thesis is about to study the global stability of a special family of hot stellar disks in the spirit of large-scale perturbation analysis rather than via N-body simulations. As we have remarked, the only previous studies of this sort

appear to be those by Kalnajs, most notably his thorough analysis of various hot Freeman disks characterized by uniform angular velocity. Our family of models, on the other hand, will be characterized by a constant linear velocity of rotation, somewhat resembling fairly large segments of many observed rotation curves. Our objectives, like those of Kalnajs, will be two-fold: (i) to determine how much random motion is required to suppress all instabilities for each angular harmonic; (ii) to explore the instabilities that remain when the velocity dispersion is chosen sufficiently small.

The present models will be described in detail in Chapter II. As we will see, these models suggested by Toomre are attractive primarily because they are self-similar. Their surface density and angular velocity both vary inversely with the radius, whereas the r.m.s. random speeds are independent of location. Hence the intrinsic properties of these disks at one radius turn out to be just scaled versions of the properties at any other, and there is also a close resemblance between the stellar orbits near different radii. This self-similarity of the models enables us to perform exactly several of the required integrations, and thereby to surmount at a reasonable computational cost the notorious analytical difficulties presented by the hot stellar disks.

The mathematical details of our analysis will be presented in Chapter III. As will be seen there, our precise

goal will be to determine whether some given member of our family of models is able to sustain any infinitesimal but exponentially growing perturbations to its equilibrium configuration; we will examine separately each angular harmonic m. The axisymmetric ($m = 0$) stability criteria for these models will be reported in Chapter IV. Detailed results for the bisymmetric ($m = 2$) disturbances will be presented in Chapter V. Results for angular harmonics $m = 1$, 3 and 4 will be discussed, more briefly, in Chapter VI. Finally, some concluding remarks will appear in Chapter VII.

* * * *

We wish to stress that, unlike perhaps Kalnajs with the Freeman disks, we will be unable to draw any rigorous conclusions about the stability of our galaxy models. Conclusions of this sort would require either a complete study of the initial value problem or else a complete study of both the discrete and any continuous modes. Such analyses for the inhomogeneous disks under consideration here are presently beyond us. Furthermore, we are currently even unable to consider any self-consistent perturbations whose growth laws are other than exponential. We cannot, for example, rule out modes (unlikely though they may seem) which instead grow with time like some polynomial multiplied by an exponential. At present we simply have to limit ourselves to drawing plausible inferences about the linear stability of our disks from their

ability or inability, as the case may be, to exhibit exponentially growing modes alone. This slight deficiency, however, seems to be shared by almost all other local or global analyses of disk systems.

II. GALAXY MODELS

The equilibrium galaxy models that will be studied here are all ones with the stars confined to an infinitesimally thin galactic plane referred to as $z = 0$. In this chapter, we describe first our basic self-similar model, then take a detailed look at the stellar orbits in its inverse-first-power force field, and finally we consider some specific ways of truncating or immobilizing this sort of a model gradually near its center, to avoid having to cope in our analysis with the arbitrarily large angular velocities that would otherwise prove a real nuisance.

a) Self-Similar Model

The distinguishing property of the self-similar model examined here is its constant rotational velocity (the speed required by a star moving in a purely circular orbit),

$$V(r) = V_o \ . \tag{2.1}$$

Mestel (1963) appears to have been the first to notice that infinite disks with this characteristic circular speed have the surface mass density

$$\mu_s(r) = \mu_o r_o/r \ , \tag{2.2}$$

where μ_o and r_o are constants which set the density and distance scales and which must be related to the velocity scale V_o via

$$V_o^2 = 2\pi G \mu_o r_o \ . \tag{2.3}$$

Since the circular speed is V_o, the radial force (per unit mass) is clearly given by

$$F_r(r) = -V_o^2/r \ . \tag{2.4}$$

The gravitational potential, then, is simply

$$\phi(r) = V_o^2 \ \ell n(r/r_o) \ . \tag{2.5}$$

Unlike most other models of infinite extent, the gravitational potential in this disk blows up as $r \to \infty$, and so the usual convention, $\phi(\infty) = 0$, cannot be employed. Instead, we adopt $\phi(r_o) = 0$.

Any star moving in this axisymmetric $1/r$ force field has the usual two integrals: the angular momentum J and the energy E, both reckoned per unit mass. If the radial and tangential components of the velocity vector of the star are denoted by u and v, respectively, then

$$J = rv \tag{2.6}$$

and

$$E = (u^2 + v^2)/2 + V_o^2 \, \ell n(r/r_o) \ . \tag{2.7}$$

It is of no immediate concern that the density $\mu_s(r)$ in this model is singular at the origin $r = 0$: after all the disk mass contained inside radius r ,

$$M(r) = 2\pi \int_0^r x \, \mu_s(x) \, dx = 2\pi \mu_o r_o r \ , \tag{2.8}$$

is clearly finite. On the other hand, this model has the unpleasant feature of possessing infinite total mass, for $M(r)$ grows linearly without bound as $r \to \infty$. Even this may not be a serious defect, however: In any reasonable perturbation, the surface density will decay rapidly for large r , and thus one can well imagine that the adoption of some large but finite outer radius R for the disk — as in the finite, Mestel (1963) models with a slightly more complicated density — would scarcely affect either the equilibrium properties or the nature of the perturbations for $r << R$. The use of some finite outer radius R would, of course, result immediately in a finite total mass for the disk. Nevertheless, we much prefer to retain, mathematically if not conceptually, the infinite extent of the model in order to exploit its self-similarity in our ensuing stability analysis.

A couple of other interesting properties of this model are the angular velocity $\Omega(r)$ and the epicyclic frequency $\kappa(r)$. The first of these is simply the circular velocity divided by the radius,

$$\Omega(r) = V_o/r \ ; \tag{2.9}$$

the other is the rate of infinitesimally small oscillations about circular motions and can here be found from the conventional prescription

$$\kappa(r) = 4\ \Omega^2(r)\ [1 + 0.5\ r\Omega^{-1}(r)\ d\Omega(r)/dr]$$

to be

$$\kappa(r) = \sqrt{2}\ V_o/r = \sqrt{2}\ \Omega(r) \ . \tag{2.10}$$

None of these properties — the circular velocity, the surface density, etc. — depend upon the velocity dispersion of the model. In fact, the description furnished so far is more than adequate for a cold stellar disk.

For our purposes, however, a hot model is needed. That, of course, requires a self-consistent equilibrium mass distribution function, $f_s(r,u,v)$ — which we define as usual so that the total mass in the annulus between r and $r + dr$ contributed by stars with velocities in the range $u + du$ and $v + dv$ is $f_s(r,u,v)\ 2\pi r\ dr\ du\ dv$. About a decade ago Toomre realized that a particularly convenient choice for this function is

$$f_s(E,J) = C\, J^a\, e^{-E/\sigma_u^2} \qquad (J \geq 0) \tag{2.11}$$

expressed in terms of the two integrals E and J. In this expression C is a normalization constant, σ_u is the dispersion in the radial velocities and self-consistency demands

$$a = (V_o/\sigma_u)^2 - 1 \; ; \tag{2.12}$$

there are no counter-rotating stars in this model, or $f_s = 0$ for $J < 0$. Since this result was never published by Toomre, a detailed summary of the properties of this hot disk is now provided below.

In terms of the variables r, u and v, the distribution function (2.11) is

$$f_s(r,u,v) = C\, r_o^{a+1} r^{-1} v^a\, e^{-(u^2 + v^2)/2\sigma_u^2} \; . \tag{2.13}$$

It is a simple task to verify that this distribution function indeed implies the surface mass density of eqn. (2.2), for

$$\begin{aligned} \mu_s(r) &= \int_{-\infty}^{\infty} \int_0^{\infty} f_s(r,u,v)\, du\, dv \\ &= \sqrt{\pi}\, 2^{a+2} \sigma_u^{a+2}\, \Gamma[(a+1)/2]\, C\, r_o^{a+1} r^{-1} \; , \end{aligned}$$

where $\Gamma(z)$ is the gamma function. This indicates that the appropriate value of the normalization constant is

$$C = \mu_o r_o^{-1} \{\sqrt{\pi}\, 2^{a/2}\, \sigma_u^{a+2}\, \Gamma[(a+1)/2]\}^{-1} \; . \tag{2.14}$$

With this choice of distribution function this hot version of the infinite Mestel model has the following velocity moments:

$$\begin{aligned} \langle u \rangle &= 0 \\ \langle v \rangle &= \sqrt{2}\ \sigma_u\ \Gamma[(a+2)/2]/\Gamma[(a+1)/2] \\ \langle u^2 \rangle &= \sigma_u^2 \\ \langle v^2 \rangle &= V_o^2 = (a+1)\ \sigma_u^2\ . \end{aligned} \tag{2.15}$$

It follows that the tangential velocity dispersion σ_v is

$$\sigma_v = \{(a+1) - 2\Gamma^2[(a+2)/2]/\Gamma^2[(a+1)/2]\}^{1/2}\ \sigma_u\ . \tag{2.16}$$

As $a \to \infty$ (or $\sigma_u \to 0$), the ratio $\sigma_v/\sigma_u \to \sqrt{2}/2$, which agrees with the estimate provided by the epicyclic approximation [cf. Chandrasekhar(1960), p. 159]. Even when the velocity dispersion is considerable, say $\sigma_u = 0.378\ V_o$ (this choice is explained below), $\sigma_v/\sigma_u = 0.694$, which is still fairly close to $\sqrt{2}/2$.

To estimate the minimum radial velocity dispersion, $\sigma_{u,min}(r)$, needed at some radius r to stabilize the disk against local axisymmetric disturbances, notice that Toomre's (1964) criterion,

$$\sigma_{u,min}(r) \doteq 3.36\ G\mu(r)/\kappa(r)\ , \tag{2.17}$$

in our case reduces simply to a value independent of r :

$$\sigma_{u,min}(r) = 0.378\ V_o\ . \tag{2.18}$$

b) Stellar Orbits

As previously emphasized, the most valuable property of the kind of disk just described is its self-similarity. In a rather obvious sense, this similarity refers already to the simple inverse-first-power dependence of the surface density, radial force, angular speed, and epicyclic frequency upon the radius r. Less obviously, any two stars in this model which have the same radial and the same tangential components of velocity, u and v — even at different radii and instants — have orbits of exactly the same shape. This is very convenient, for it means that the shapes of all unperturbed stellar orbits in this model can be categorized by a single parameter analogous to eccentricity, as we are about to see.

The first part of this claim — that stars with the same u and v have similar orbits — may be verified by referring to the equations governing the stellar motions. Let $r'(t')$ and $\theta'(t')$ denote the coordinates, at time t', of a star which at time t was located at radius r and azimuthal angle θ and moved with radial speed u and tangential speed v. The differential equations and initial conditions appropriate to this star are

$$dr'^2/dt'^2 = J^2/r'^3 - V_o^2/r' \tag{2.19}$$

$$d\theta'/dt' = J/r'^2 \tag{2.20}$$

$$\theta'(t) = \theta\ ,\quad r'(t) = r,\quad dr'(t)/dt' = u\ . \tag{2.21}$$

Now introduce fully dimensionless variables $\hat{r}$, $\hat{\theta}$, and $\hat{t}$ via

$$r' = (rv/V_o)\,\hat{r}$$

$$\hat{\theta}' = \hat{\theta} \tag{2.22}$$

$$t'-t = (V_o/rv)\,\hat{t} \ .$$

In terms of these variables the orbit equations become

$$d^2\hat{r}/d\hat{t}^2 = 1/\hat{r}^3 - 1/\hat{r} \tag{2.23}$$

$$d\hat{\theta}/d\hat{t} = 1/\hat{r} \tag{2.24}$$

$$\hat{\theta}(0) = \theta \ , \quad \hat{r}(0) = (v/V_o)^{-1} \ , \quad d\hat{r}(0)/d\hat{t} = u/V_o \ . \tag{2.25}$$

One thus sees fairly explicitly that if two stars have the same u and v , then their orbits must be similar in shape [described by the functions $\hat{r}(\hat{t})$ and $\hat{\theta}(\hat{t})$], although the sizes of the orbits will be proportional to the initial radius r and the time scales will be inversely proportional to the same. The orbit equations are invariant under time translations; so the comparison can be made at arbitrary instants.

This last remark suggests that we pick some standard position in the orbit of each star at which to compare the radial and tangential speeds. A convenient choice is the point at which the tangential speed is precisely equal to the circular speed, i.e., $v' = V_o$. We call this position the home radius and denote it by r_H . At this location a star

executing a purely circular motion would have the angular momentum $J = rv$. Thus

$$r_H = J/V_o = rv/V_o \ . \tag{2.26}$$

In this thesis we will often use r_H and J interchangeably. The radial velocity at the home radius is termed the <u>eccentric velocity</u> and is denoted by U . It can be found from the first integral

$$u'^2 + J^2/r'^2 + 2\ V_o^2\ \ell n(r/r_o) = \text{constant} \ ,$$

of the radial equation of motion, eqn. (2.19), where $u' = dr'/dt'$. The constant here is determined by evaluating the left-hand side of the equation at $t' = t$, so that

$$u'^2 = u^2 + v^2 - 2\ V_o^2\ \ell n(r'/r) - J^2/r'^2 \ . \tag{2.27}$$

Therefore,

$$U^2 = u^2 + v^2 - V_o^2[1 + 2\ \ell n(v/V_o)] \ , \tag{2.28}$$

or, in terms of the usual constants of motion

$$U^2 = 2E - V_o^2[1 + 2\ \ell n(J/r_o V_o)] \ . \tag{2.29}$$

One can easily show [by differentiating with respect to r' the right-hand side of eqn. (2.27)] that the eccentric velocity U is also the maximum radial velocity that a given

star achieves during its orbit. It seems worth stressing again that in our model this maximum radial speed depends only upon the given velocities u and v and not at all upon the location. To summarize, when a star is located at its home radius, its radial velocity has the value U, and its tangential velocity is V_o by definition. All stars with the same eccentric velocity U thus have <u>similar</u> orbits, and we shall refer to them henceforth as belonging to the same U-class. The home radius serves only to fix the size and time scale of an orbit.

These two new variables, U and r_H, will be useful to us in a number of ways. They will frequently be used in place of u and v to parameterize the velocity space. The transformation given by eqns. (2.26) and (2.28) between (u,v) and (U,r_H) has as its inverse

$$v = V_o(r_H/r) \tag{2.30}$$

$$u^2 = U^2 + V_o^2[1 + 2\ \ell n(r_H/r) - (r/r_H)^{-2}] \ . \tag{2.31}$$

Of course the last equation determines only the absolute value of u. This ambiguity arises because the original transformation to the (U,r_H) description loses track of the sense of radial motion. When resorting to these two new variables care must be exercised to account properly for both ingoing and outgoing stars.

In the remainder of this section we will take a detailed look at the stellar orbits in our model. In order to reduce references to arbitrary scale factors both here and in the ensuing analyses we will often resort to dimensionless variables. The units of distance, velocity, mass and time will be r_o , V_o , $\mu_o r_o$ and $r_o V_o$, respectively. The corresponding dimensionless variables will be denoted by tildas — e.g., $\tilde{U} = U/V_o$ is the dimensionless eccentric velocity.

Our description of the orbits starts by finding the maximum and minimum radii, r_{max} and r_{min} , that a given star will reach. Suppose that the star has eccentric velocity U and home radius r_H . Its instantaneous radial speed, given by eqn. (2.27), can be written

$$u'^2 = U^2 + V_o[1 + 2\,\ell n(r'/r_H) - (r'/r_H)^{-2}] \ . \qquad (2.32)$$

The extreme radii are found by setting $u' = 0$ in this expression. Hence they are given by the roots of

$$\tilde{U}^2 + 1 - 2\,\ell n(r/r_H) - (r/r_H)^{-2} = 0 \ .$$

For each U-class these radii can be standardized as

$$\begin{aligned} r_{max} &= r_H\, x_{max}(\tilde{U}) \\ r_{min} &= r_H\, x_{min}(\tilde{U}) \ , \end{aligned} \qquad (2.33)$$

where $x_{max}(\tilde{U})$ and $x_{min}(\tilde{U})$ are the roots of

$$\tilde{U}^2 + 1 - 2\ \ell n\ x - x^{-2} = 0\ . \tag{2.34}$$

The ratio $x_{max}(\tilde{U})/x_{min}(\tilde{U})$ of maximum-to-minimum radii can be viewed as one measure of the orbital eccentricity.

The radial position of a star will oscillate between its two extreme radii. The period of this radial motion is

$$T(\tilde{U},r_H) = 2\int_{r_{min}}^{r_{max}} dr'/(dr'/dt')\ ,$$

or

$$T(U,r_H) = \frac{2}{V_o}\int_{r_{min}}^{r_{max}} \frac{dr'}{\{1 + \tilde{U}^2 - 2\ \ell n(r'/r_H) - (r'/r_H)^{-2}\}^{1/2}}\ .$$

Upon the substitution $x = r'/r_H$ this becomes

$$T(U,r_H) = 2\frac{r_H}{V_o}\int_{x_{min}}^{x_{max}} \frac{dx}{\{1 + \tilde{U}^2 - 2\ \ell n\ x - x^{-2}\}^{1/2}}\ .$$

This can be expressed in terms of a standard dimensionless period $\tilde{T}(\tilde{U})$ for this $\tilde{U}$-class by

$$T(\tilde{U},r_H) = (r_H/V_o)\ \tilde{T}(\tilde{U})\ , \tag{2.35}$$

where

$$\tilde{T}(\tilde{U}) = 2\int_{x_{min}}^{x_{max}} \frac{dx}{\{1 + \tilde{U}^2 - 2\ \ell n\ x - x^{-2}\}^{1/2}}\ . \tag{2.36}$$

Thus, within this eccentricity or $\tilde{U}$-class, the period of a star with arbitrary home radius can be expressed as a simple

multiple of the period of a star with unit home radius ($r_H/r_o = 1$). Correspondingly, if the radial frequency is denoted by $\kappa(\tilde{U},r_H)$, then

$$\kappa(\tilde{U},r_H) = (V_o/r_H)\ \tilde{\kappa}(\tilde{U}) \ , \tag{2.37}$$

where

$$\tilde{\kappa}(\tilde{U}) = 2\pi/\tilde{T}(\tilde{U}) \ . \tag{2.38}$$

Since the orbits of all stars of a given $\tilde{U}$-class are similar, we will commonly use this practice of referring their properties to the reference orbit of the $\tilde{U}$-class — the orbit of a star with unit home radius.

Obviously the frequency $\kappa(\tilde{U},r_H)$ of radial vibration is a generalization of the usual epicyclic frequency $\kappa(r)$, which is the radial frequency of an orbit that departs only infinitesimally from circular motion. Although the symbol κ has been employed here for both these frequencies, they are distinguished by their arguments. We wish to emphasize that we will always be dealing with the <u>exact</u>, finite amplitude stellar orbits.

The orbit of a star of course also includes an angular motion. During one radial oscillation it moves the angular distance $\theta(\tilde{U},r_H)$, where

$$\theta(\tilde{U},r_H) = \int_0^{T(\tilde{U},r_H)} (d\theta'/dt')\ dt'$$

or

$$\Theta(\tilde{U},r_H) = 2\int_{r_{min}}^{r_{max}} \frac{J^2}{r'^2V_o^2} \frac{dr'}{\{1 + \tilde{U}^2 - 2\,\ell n(r'/r_H) - (r'/r_H)^{-2}\}^{1/2}}$$

$$= 2\int_{x_{min}}^{x_{max}} \frac{dx}{x^2\{1 + \tilde{U}^2 - 2\,\ell n\, x - x^{-2}\}^{1/2}} \quad .$$

This last integral resembles the one in eqn. (2.36); hence the notation

$$I_n(\tilde{U}) = 2\int_{x_{min}}^{x_{max}} \frac{dx}{x^n\{1 + \tilde{U}^2 - 2\,\ell n\, x - x^{-2}\}^{1/2}} \tag{2.39}$$

is convenient. We then have

$$\Theta(\tilde{U},r_H) = I_2(\tilde{U}) \tag{2.40}$$

and

$$\tilde{\kappa}(\tilde{U}) = 2\pi/I_0(\tilde{U}) \quad . \tag{2.41}$$

The angular frequency, defined as the mean angular speed of the star, is

$$\Omega(\tilde{U},r_H) = \Theta(\tilde{U},r_H)/T(\tilde{U},r_H) \quad .$$

This in turn can be expressed as

$$\Omega(\tilde{U},r_H) = (V_o/r_H)\ \tilde{\Omega}(\tilde{U}) \quad , \tag{2.42}$$

where

$$\tilde{\Omega}(\tilde{U}) = I_2(\tilde{U})/I_0(\tilde{U}) \tag{2.43}$$

is the dimensionless angular frequency of the reference star for this $\tilde{U}$-class. Again these generalized angular frequencies are distinguished from the usual angular velocity (i.e., rate of circular orbiting) of the disk, $\Omega(r)$, by the difference in arguments.

The dimensionless orbit parameters for the reference orbits of typical $\tilde{U}$-classes are listed below.

$\tilde{U}$	x_{min}	x_{max}	I_0	I_1	I_2	$\tilde{\kappa}$	$\tilde{\Omega}$
.1	.933	1.075	4.463	4.447	4.441	.9950	1.4078
.2	.874	1.160	4.525	4.458	4.435	.9802	1.3885
.3	.820	1.256	4.630	4.476	4.426	.9559	1.3570
.4	.772	1.366	4.782	4.502	4.413	.9229	1.3139
.5	.729	1.493	4.985	4.536	4.397	.8819	1.2603
.6	.689	1.638	5.246	4.577	4.377	.8342	1.1976
.7	.654	1.808	5.574	4.626	4.353	.7809	1.1272
.8	.621	2.005	5.980	4.682	4.326	.7233	1.0507
.9	.591	2.237	6.479	4.747	4.295	.6629	.9698
1.0	.564	2.511	7.089	4.819	4.261	.6011	.8863
1.1	.539	2.837	7.835	4.899	4.225	.5392	.8019
1.2	.515	3.229	8.749	4.986	4.185	.4784	.7182
1.3	.494	3.701	9.869	5.081	4.144	.4199	.6367
1.4	.474	4.274	11.248	5.183	4.100	.3645	.5586
1.5	.455	4.977	12.955	5.293	4.055	.3130	.4850

(The integral $I_1(\tilde{U})$ is needed later for normalization.) Notice that the ratio x_{max}/x_{min} increases monotonically with increasing $\tilde{U}$, as is to be expected from the interpretation of $\tilde{U}$ as a measure of the orbital eccentricity. For small values of the eccentric velocity the orbits are nearly

circular. The radial frequency, $\kappa(\tilde{U},r_H)$, of such orbits is close to the epicyclic frequency, $\kappa(r_H) = \sqrt{2}\ V_o/r_H$, and the angular frequency, $\Omega(\tilde{U},r_H)$, is close to the angular velocity of the disk, $\Omega(r_H) = V_o/r_H$. (This comparison refers to orbital parameters of a star with given r_H and the overall disk parameters at the radius $r = r_H$.) Observe that $I_0(\tilde{U})$ and $I_2(\tilde{U})$ both tend to $\pi\sqrt{2}$ as $\tilde{U} \to 0$; therefore, just as one would expect from the small-amplitude or epicyclic approximation, $\tilde{\kappa}(\tilde{U}) \to \sqrt{2}$ and $\tilde{\Omega}(\tilde{U}) \to 1$. At the other extreme, as $\tilde{U} \to \infty$, the radial period $I_0(\tilde{U})$ tends monotonically to infinity, but $I_2(\tilde{U})$, or the angular distance traversed in one radial oscillation, tends to π . So orbits with very large eccentric velocities are nearly straight, in-and-out trajectories.

In space, these stellar orbits in general are not closed and instead form rosettes. For some special values of $\tilde{U}$, however, the radial and angular frequencies are commensurable, i.e., $p\ \tilde{\kappa}(\tilde{U}) = q\ \tilde{\Omega}(\tilde{U})$ for some integers p and q . These orbits will close after q radial oscillations and p angular revolutions; for example, a star with $\tilde{U} \simeq 1.1917$ satisfies the above relationship with $p = 2$ and $q = 3$.

Even if a stellar orbit is not closed when viewed from a non-rotating frame, it will appear closed when viewed from a frame that rotates steadily in space with the mean angular speed $\Omega(\tilde{U},r_H)$ of the star; in fact, the stellar orbit then seems to be periodic with the period equal to the time

$T(U,r_H)$ of one radial oscillation. Thus, the stellar orbit as it appears in the non-rotating frame of the disk can be decomposed into a periodic radial and angular motion on top of a mean angular motion, much as recommended in general by Kalnajs (1971).

In order to get specific expressions of this decomposition for future reference, we further specialize the reference orbit for those stars of a specific $\tilde{U}$-class to be not only an orbit with unit home radius, but also one with a particular set of initial conditions: at time $t = 0$ a star in the standard orbit is at its pericenter — $r = r_{min}$ — and at angular position $\theta = 0$. Referring to eqns. (2.23), (2.24) and (2.25) and now using bars instead of hats to denote these particular reference orbit variables, we have

$$d^2\bar{r}/d\bar{t}^2 = 1/\bar{r}^3 - 1/\bar{r} \tag{2.44}$$

$$d\bar{\theta}/d\bar{t} = 1/\bar{r}^2 \tag{2.45}$$

$$\bar{r}(0) = x_{min} \ , \quad d\bar{r}/d\bar{t}(0) = 0 \ , \quad \bar{\theta}(0) = 0 \ , \tag{2.46}$$

as the equations describing the orbit.

In terms of the time variable $\bar{t}$, the radial coordinate $\bar{r}(\bar{t})$ has the period $2\pi/\tilde{\kappa}(\tilde{U})$, which is different for each $\tilde{U}$-class. In terms of the <u>orbital phase</u> ψ defined by

$$\psi = \tilde{\kappa}(\tilde{U})\ \bar{t} \ , \tag{2.47}$$

the period is 2π for each $\tilde{U}$-class. The phase has been chosen so that it has the value 0 when the star is at its pericenter, π when it is at its apocenter and 2π when it returns to the pericenter for the first time. (This phase is ambiguous within an additive constant $2\pi n$ where n is an integer.) Similarly, we can define the phase of the orbital motion for an arbitrary star — one with any home radius and any initial conditions at $t' = t$. In so doing, it is convenient to let t_p stand for the (most recent) time at which the star was located at its pericenter and θ_p for its angular position at that time. (There will not be any need to have explicit expressions for θ_p and t_p.) The phase can now be written

$$\psi = \kappa(\tilde{U}, r_H) \, (t' - t_p) \, , \tag{2.48}$$

and, moreover, the orbit of the star can be expressed in terms of the reference orbit as

$$r'(t') = r_H \, \bar{r}(\psi;\tilde{U}) \tag{2.49}$$

$$\theta'(t') = \theta_p + \bar{\theta}(\psi;\tilde{U}) \, , \tag{2.50}$$

where ψ is related to t' by eqn. (2.48).

We will usually work with the <u>orbital excursions</u> of a star in the logarithm of the radial position and in the deviation from the mean angular motion. For the reference orbit these are defined by

$$X(\psi;\tilde{U}) = \ell n\ \bar{r}(\bar{t};\tilde{U}) \tag{2.51}$$

$$Y(\psi;\tilde{U}) = \bar{\theta}(\bar{t};\tilde{U}) - \Omega(\tilde{U})\ \bar{t}\ , \tag{2.52}$$

with $\bar{t}$ related to ψ by eqn. (2.47). These functions can be used to describe the orbit of an arbitrary star:

$$r'(t') = r_H\ e^{X(\psi;\tilde{U})} \tag{2.53}$$

$$\theta'(t') = Y(\psi;\tilde{U}) + \theta_p + \Omega(\tilde{U},r_H)\ (t' - t_p)\ , \tag{2.54}$$

where eqn. (2.48) connects t' to ψ. The orbital excursions $X(\psi;\tilde{U})$ and $Y(\psi;\tilde{U})$ describe the periodic portions of the orbit. The final term in eqn. (2.54) describes the mean angular motion.

Before leaving this discussion of the stellar orbits one other point needs to be made. A star traversing the reference orbit of any given $\tilde{U}$-class will, along the way, assume all the dimensionless radial and tangential velocity pairs that correspond to that $\tilde{U}$-class. We may therefore use the equations of motion to parameterize velocity space in terms of eccentric velocity and phase instead of by $\tilde{u}$ and $\tilde{v}$. In particular, from eqn. (2.28) we obtain

$$\tilde{U}^2 = \tilde{u}^2 + \tilde{v}^2 - 1 - 2\ \ell n\ \tilde{v}\ , \tag{2.55}$$

whereas the conservation of angular momentum implies

$$\tilde{v}(\xi;\tilde{U}) = 1/\bar{r}(\xi;\tilde{U})\ . \tag{2.56}$$

When used in this context to describe velocity space rather than a stellar orbit, the phase will be denoted by ξ instead of by ψ, and it will be termed the velocity phase to distinguish it from the orbital phase ψ. Any given pair $(\tilde{U},\xi)$ will correspond uniquely to a pair $(\tilde{u},\tilde{v})$: eqn. (2.56) determines $\tilde{v}$ from $\tilde{U}$ and ξ ; eqn. (2.55) then provides the magnitude of $\tilde{u}$; the sign of $\tilde{u}$ is determined by whether $\xi \in [0,\pi]$, in which case $\tilde{u} \geq 0$, or whether $\xi \in [\pi,2\pi]$, in which case $\tilde{u} \leq 0$. Unlike the variables $\tilde{U}$ and r_H, there is no ambiguity about the sign of $\tilde{u}$ when transforming to $(\tilde{u},\tilde{v})$. The stability analysis in Chapter III will soon put this parametrization of velocity space to good use.

c) Centrally Cut-out Models

The entirely self-similar or singular disk described above has one annoying feature: all its rapidly rotating stars near the center require undue care during analysis. To lessen this annoyance, we found it advisable to focus our study on certain simple variants of the basic singular model — ones in which progressively more and more of the innermost stars were imagined "frozen" into their equilibrium orbits, to render them incapable of participating in any disturbance.

These centrally cut-out disks — as we shall call them — can be regarded as composed of two stellar populations.

One is the usual "active" or mobile disk of stars able to respond to all disturbance forces; the other is a "rigid" or totally immobile component, here concentrated near the middle. Like the rigid "halo components" imagined in some other analyses, the latter is of course only fictitious; however, in a loose sense , it perhaps mimics the hot nucleus or nuclear bulge of a galaxy.

Our crucial requirement in concocting these cut-out disks was that their <u>combined</u> equilibrium force from the active and rigid mass distributions must remain exactly of type r^{-1} as before. This was imposed to preserve the very desirable similarity of the unperturbed stellar orbits from radius to radius, just as in the singular disk. Our only other considerations were that the rules of exclusion be sensible and convenient. Obviously we wished the fraction of density identified as the active component to rise smoothly and monotonically from zero at the very center toward unity at large radii; in addition, it seemed wise to avoid classes of functions that seem disagreeable to residue calculus.

These aims were met by simply multiplying the distribution function $f_s(E,J)$ from eqn. (2.11) by a <u>cut-out factor</u>

$$H(J) = \frac{J^N}{J^N + (r_o V_o)^N} \tag{2.57}$$

that depends only upon the specific angular momentum J ,

besides the (positive) integer N , which we refer to as the cut-out index. In detail, the distribution function for the active stars of a cut-out disk is thus

$$f_o(E,J) = H(J)\ f_s(E,J) \ , \tag{2.58}$$

or

$$f_o(E,J) = C\ H(J)\ J^a\ e^{-E/\sigma_u^2} \quad (J \geq 0) \ , \tag{2.59}$$

where, as before, the constant

$$a = (V_o/\sigma_u)^2 - 1 \ ,$$

the normalization constant C is given by eqn. (2.14), and there are no counter-rotating stars.

For future reference, note that the distribution function for the singular disk — and hence also the one for the cut-out disks — can handily be expressed in terms of the eccentric velocity U and the angular momentum J instead of in terms of the energy E and the angular momentum. This is achieved by replacing the energy in eqn. (2.11) with the eccentric velocity by means of eqn. (2.29):

$$f_s(U,J) = C_a\ J^{-1}\ e^{-U^2/2\sigma_u^2} \ , \tag{2.60}$$

where

$$C_a = e^{-(a+1)/2}\ (r_o V_o)^{a+1}\ C \ . \tag{2.61}$$

A dimensionless normalization constant $\tilde{C}_a$ is introduced by

$$C_a = (\mu_o r_o / V_o)\ \tilde{C}_a \ . \tag{2.62}$$

This constant $\tilde{C}_a$ can be found explicitly by combining eqns. (2.12), (2.14), (2.61) and (2.62):

$$\tilde{C}_a^{-1} = \sqrt{\pi}\ [(a+1)/2]^{a/2}\ \Gamma[(a+1)/2]\ e^{(a+1)/2}\ /(a+1) \ . \tag{2.63}$$

A useful alternative expression is obtained by integrating eqn. (2.59) over velocity space in terms of $\tilde{U}$ and J and comparing that density with $\mu_s(r)$. This requires the Jacobian, which is

$$\frac{\partial(u,v)}{\partial(\tilde{U},J)} = \frac{\tilde{U}\ V_o^2}{r\ \{1 + \tilde{U}^2 - 2\ \ell n(r/r_H) - (r/r_H)^{-2}\}^{1/2}} \ ,$$

where $r_H = J/V_o$. This procedure implies that the dimensionless normalization constant must be

$$\tilde{C}_a^{-1} = \int_0^\infty \tilde{U}\ I_1(\tilde{U})\ e^{-(a+1)\ \tilde{U}^2/2}\ d\tilde{U} \ . \tag{2.64}$$

We will use this form rather than eqn. (2.63) when computing $\tilde{C}_a$ numerically.

Our stability analysis requires only that we have a distribution function for the active stars. Therefore, no distribution function for the rigid stellar component will be given. These immobile stars may be thought of as part of a high-velocity disk component or perhaps even, as already

mentioned, as part of a high-velocity central bulge. Such stellar populations would have very little ability to respond to perturbations in a low or moderate-velocity disk component. Because of the current interest in galactic halos and the possible interpretation of the immobile component as such a system, we have supplied, in Appendix B, a closed form expression for the volume density of a suitable spherical halo.

Whatever interpretation is placed upon the rigid material, we note once again that it is essential for this analysis that the combined equilibrium force field of the active and rigid components coincide with the force field given in eqn. (2.4) of the singular disk. Above all, we wish to retain the orbital similarity from radius to radius that characterizes the singular disk, and it is of course the combined equilibrium force field and not the combined projected surface mass density that determines the stellar orbits. Since the total force field of the cut-out disks coincides with that of the singular disk, these models have the same rotation speed, total gravitational potential, angular speed and epicyclic frequency as given in eqns. (2.1), (2.5), (2.9) and (2.10) for the singular disk. Furthermore, the description of the stellar orbits furnished in section (b) applies equally well to the cut-out models. The mean angular speed of a star, given by eqn. (2.42), is of particular interest because it was the stars of low angular momentum that were the

troublesome, rapidly rotating ones. Clearly, our choice of the cut-out factor has greatly reduced the importance of such stars.

Naturally, the surface density $\mu_a(r)$ of the active stars is not the simple r^{-1} density of the singular disk. This active density is given instead by

$$\mu_a(r) = \int_{-\infty}^{\infty} \int_{0}^{\infty} C\, H(J)\, J^a\, e^{-E/\sigma_u^2}\, du\, dv \; . \qquad (2.65)$$

In the absence of random motions — i.e., when $\sigma_u = 0$ — that density is simply

$$\mu_{a,cold}(r) = \frac{\mu_o r_o}{r} \frac{r^N}{r^N + r_o^N} \; ; \qquad (2.66)$$

Otherwise it will be somewhat "smeared" from this exact form. In particular, when $\sigma_u = 0.378\, V_o$ — or in circumstances where the singular disk is expected to become axisymmetrically stable — the densities of disks with cut-out indices $N = 1$, 2 , 3 and 4 are those displayed in Figure 1. The top diagram indicates how the cut-out factor appears in physical space; the active surface density itself is given in the bottom diagram. Note that for $N \geq 2$ this surface density appears to have a "hole" near the center — there is then a reasonably well-defined inner edge to the active material, as well as an intrinsic scale length for the models themselves. As one indication of the degree to which $\mu_a(r)$ departs from $\mu_{a,cold}(r)$, observe that $\mu_{a,cold}(r=r_o) = 0.5\, \mu_o$

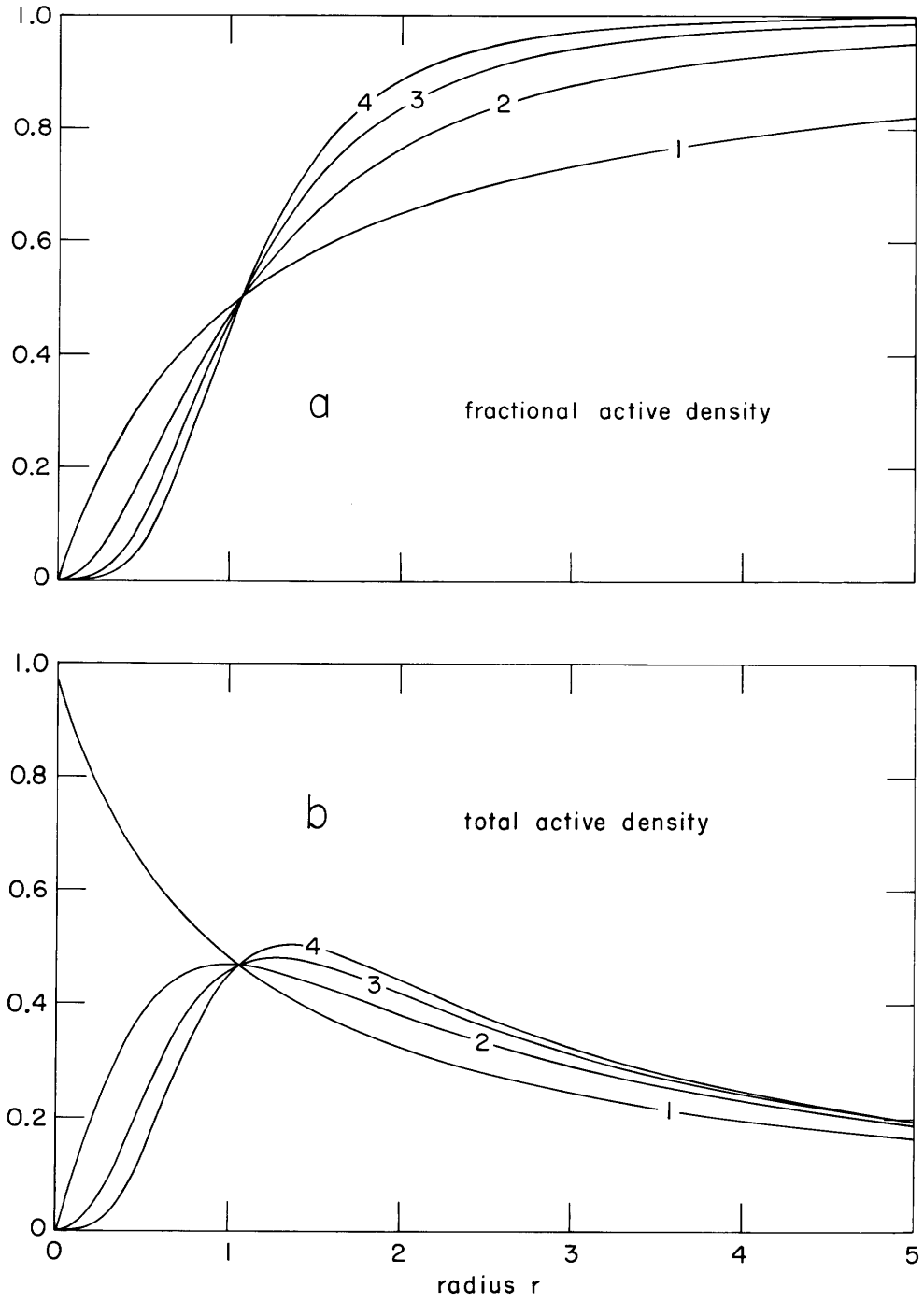

Figure 1. The surface densities of four cut-out models with $\sigma_u = 0.378\ V_o$: (a) fractional active density $r\mu_a(r)$; (b) total active density $\mu_a(r)$. Radius is measured in units of r_o and density in units of μ_o . The cut-out index N is marked on each curve.

whereas $\mu_a(r=r_o) = 0.475\ \mu_o$ when $\sigma_u = 0.378\ V_o$. Evidently, $\mu_{a,cold}(r)$ provides a fairly decent estimate of the actual active surface mass density in a hot disk.

From either eqn. (2.66) or Figure 1 it is apparent that $\mu_a(r) \to \mu_s(r) = \mu_o r_o/r$ as $r \to \infty$. Therefore, the total active mass of the cut-out models remains infinite. As we explained for the singular disk in sub-section (a), this does not seem to be a serious concern. While the total mass of the rigid component depends upon whether it is interpreted as a three-dimensional or a two-dimensional system, it is easy to see that under the latter interpretation, its total mass is finite for $N \geq 2$.

The knowledge of the general behavior of $\mu_a(r)$ and $\kappa(r)$ permits us to use Toomre's criterion to estimate the velocity dispersion needed to stabilize the disk against short-scale axisymmetric disturbances. In the singular disk we saw that this minimum dispersion $\sigma_{u,min}(r)$ is the same at all radii. However, in the cut-out disks it does depend on the radius, since the relevant surface density $\mu(r)$ in eqn. (2.17) is now the active one, whereas $\kappa(r)$ is unchanged from the singular model. Obviously, our $\mu_a(r)$ departs sharply from the inverse-first-power profile near the middle. Consequently, $\sigma_{u,min}(r)$ rises monotonically from a value of zero at the center and approaches the singular disk value $0.378\ V_o$ as $r \to \infty$.

Interpreted locally, the usual stability parameter

$$Q(r) = \sigma_u/\sigma_{u,min}(r) \ , \tag{2.67}$$

referring to these cut-out disks, is therefore infinite at the center and drops toward the constant, singular disk value as $r \to \infty$. For the global stability studies conducted in this thesis, however, we will reserve the symbol

$$Q = \sigma_u/\sigma_{u,min} \tag{2.68}$$

(without any radial argument) to denote the ratio of the actual velocity dispersion σ_u to the velocity dispersion $\sigma_{u,min}$ required to stabilize the disk against <u>all</u> axisymmetric disturbances. The precise value of this overall $\sigma_{u,min}$ will be reported in Chapter IV.

* * * *

To summarize, the basic self-similar and cut-out models share the very convenient property that all stellar orbits in them belong to just a single-parameter family of shapes. As discussed above, the latter models were introduced primarily to reduce the nuisance caused by the rapidly rotating stars found near the middle of the former disk; incidentally, this artifice also provides a natural scale length for the models.

The large angular velocities of the central stars could, of course, have been avoided also by the adoption of other variants of the singular disk, for instance, ones designed to have a finite and nearly constant angular velocity towards the center. One particularly attractive example of that sort, noticed by Rybicki (1974, private communication), has the density law

$$\mu(r) = \mu_o r_o / \sqrt{r^2 + r_o^2}$$

and the rotation curve

$$V(r) = V_o \left[1 - r_o / \sqrt{r^2 + r_o^2}\right]^{1/2} .$$

As described in Appendix A, a Gaussian sort of exact distribution function for the "hot" stars even in such a model can be supplied fairly succinctly, and the complete model can be shown to approach the Mestel-Toomre singular disk as $r \to \infty$. Unfortunately, the self-similarity of the equilibrium orbits is lost. Consequently, the disks suggested by Rybicki are much harder to analyze for stability than our cut-out disks — even though the latter obviously are more artificial.

III. MATHEMATICAL APPARATUS

The aim of this chapter and its related appendices is to provide a detailed description of the mathematical apparatus that enables us to decide whether a given one of our models, specified by its cut-out index N and velocity dispersion σ_u, is able to sustain any infinitesimally small, exponentially growing perturbations in the active surface density of the stars. Any disk that admits exponentially growing modes is clearly unstable, but as discussed at the close of Chapter I, it is conceivable that a disk can be unstable and yet have no exponentially growing modes. Nevertheless, we take the position that it is plausible to call stable those disks without any exponentially growing modes.

a) Dynamics of the Disks

Before getting into the particulars of the stability analysis, a few remarks about the dynamics of our model galaxies seem in order. Both the singular disk and the cut-out models have been assumed to be infinitesimally thin, with all the stellar motions confined to the plane $z = 0$, where z is the vertical coordinate. A full description of the stars in the active component is furnished by their mass distribution function $f(r,\theta,u,v,t)$. As usual, the amount

of mass, at time t, contributed by those stars with coordinates contained in the range $(r,r+dr)$ and $(\theta,\theta+d\theta)$ and velocities in the range $(u,u+du)$ and $(v,v+dv)$ is given by $f(r,\theta,u,v,t)\ r\ dr\ d\theta\ du\ dv$. This function will frequently be referred to as simply the distribution function. Since the rigid stars, if any, are frozen in their equilibrium orbits, there is no need to concern ourselves with their mass distribution function. The stars move in the $z = 0$ plane of the force field described by the total gravitational potential $\phi(r,\theta,z,t)$. This potential includes the effects of both the active and the rigid stars.

The conventional assumption of a collisionless stellar model is made. The dynamics of the mobile stars are then described by the Vlasov equation

$$Df/Dt = 0 , \tag{3.1}$$

where D/Dt represents the convective time derivative following the motion of the star:

$$D/Dt = \partial/\partial t + \underline{u} \cdot \partial/\partial \underline{x} - \nabla\phi \cdot \partial/\partial \underline{u} . \tag{3.2}$$

Here $\underline{x}$ denotes the position vector of the star, $\underline{u}$ its velocity vector and $-\nabla\phi$ its acceleration under the total gravitational potential. A detailed discussion of this equation, as well as the applicability of the assumption of a collisionless model for galaxies such as the Milky Way may be found in Chandrasekhar (1960).

The total surface mass density — due to both the mobile and immobile disk components — is denoted by $\mu(r,\theta,t)$. It is related to the total gravitational potential by Poisson's equation

$$\nabla^2\phi = 4\pi G\ \mu(r,\theta,t)\ \delta(z)\ , \tag{3.3}$$

where ∇^2 represents the (three-dimensional) Laplacian and $\delta(z)$ is the Dirac delta-function. In terms of the distribution function, the active material's contribution to the total surface mass density, $\mu_a(r,\theta,t)$, is

$$\mu_a(r,\theta,t) = \int_{-\infty}^{\infty} \int_{0}^{\infty} f(r,\theta,u,v,t)\ du\ dv\ . \tag{3.4}$$

(Recall that our models contain no counter-rotating stars; therefore, the distribution function vanishes for $v < 0$.) Although the potential ϕ depends on all three spatial coordinates as well as the time, its dependence upon the vertical coordinate z will be suppressed hereafter since the dynamics in the plane of the disk can be described solely in terms of the value of the potential in the plane of the disk. So, by $\phi(r,\theta,t)$ is meant $\phi(r,\theta,z=0,t)$.

In equilibrium these functions are steady, or free from any explicit time dependence. The equilibrium models considered here are all axisymmetric, so that they are also free from any dependence upon the azimuthal angle θ . In general, equilibrium functions will be signified by a subscript o .

It is a simple task to verify that the equilibrium models described in Chapter II satisfy eqns. (3.1) and (3.3).

b) Fourier Representation of Modes

We begin our stability analysis by noting that when the surface density of the active stars is perturbed, it will change from its equilibrium value

$$\mu(r,\theta,t) = \mu_{a,o}(r)$$

to the perturbed value[1]

$$\mu(r,\theta,t) = \mu_{a,o}(r) + \mu_1(r,\theta,t) \ .$$

Perturbation quantities such as $\mu_1(r,\theta,t)$ will be denoted by the subscript 1 . It is convenient to use complex variables to describe the perturbations; in particular, $\mu_1(r,\theta,t)$ is represented as the real part of a complex function, say $q(r,\theta,t)$:

$$\mu_1(r,\theta,t) = \mathrm{Re}\{q(r,\theta,t)\} \ .$$

The physical quantities are all periodic in the angular coordinate θ and so $q(r,\theta,t)$ can be expanded in a Fourier series, namely,

$$q(r,\theta,t) = \sum_{m=-\infty}^{\infty} q_m(r,t)\ e^{im\theta} \ .$$

---

[1]Since only active densities are perturbed, the subscript a is dispensed with for all but the equilibrium density.

Since a linear analysis in the perturbations is being applied, each angular harmonic may be examined separately, and since only real parts have physical significance, the angular harmonic number m may be assumed to be non-negative. Thus we assume that

$$q(r,\theta,t) = q_m(r,t)\ e^{im\theta} \ .$$

The terminology "angular harmonic m " will often be used to refer to disturbances with angular harmonic number m .

We confine ourselves to perturbations whose growth can be represented by an exponential factor e^{st} . Our analysis is restricted to non-negative values of the growth rate s . Temporal oscillations are also allowed for by including a factor $e^{-im\Omega_p t}$ in $q_m(r,t)$, i.e., we further assume that

$$q_m(r,t) = D(r)\ e^{st-im\Omega_p t} \ .$$

The pattern speed Ω_p is the angular frequency at which the disturbance rotates. In short, we look for density perturbations of the sort described by

$$\mu_1(r,\theta,t) = \mathrm{Re}\{D(r)\ e^{i(m\theta-\omega t)}\} \ , \tag{3.5}$$

where the complex temporal frequency of the disturbance is

$$\omega = m\Omega_p + is \ . \tag{3.6}$$

Both Ω_p and s are intrinsically real quantities.

The logarithmic spiral representation introduced by Kalnajs (1965, 1971) is well-suited to the next step of our analysis: connecting the change in the density of the active material, given by eqn. (3.5), with the change in the total gravitational potential. Like Kalnajs, we now represent the radial part of the disturbance density as the Fourier integral

$$D(r) = \mu_p \, (r/r_o)^{-3/2} \int_{-\infty}^{\infty} A(\alpha) \, e^{i\alpha \, \ell n(r/r_o)} \, d\alpha \, , \qquad (3.7)$$

where μ_p carries the dimensions of surface mass density and measures the amplitude of the perturbation (assumed infinitesimally small). The density transform in wavenumber[2] space is denoted here by $A(\alpha)$. In effect, $D(r)$ is Fourier analyzed in the logarithm of the radial coordinate. For non-axisymmetric disturbances, where $m > 0$ is assumed, the positive wavenumbers correspond to trailing components of the perturbation, and the negative wavenumbers to leading components. Under any density perturbation the total gravitational potential will change from its equilibrium value $\phi_o(r)$ [cf. eqn. (2.5); we have here added the subscript o to emphasize that this is the equilibrium potential] to the new value

$$\phi(r,\theta,t) = \phi_o(r) + \phi_1(r,\theta,t) \, .$$

---

[2]Actually, α is a logarithmic wavenumber and not a wavenumber in the usual sense. However, it will be referred to here simply as a wavenumber.

The connection between the perturbation potential $\phi_1(r,\theta,t)$ and the perturbation density of eqn. (3.5) is provided by Poisson's equation:

$$\nabla^2\phi(r,\theta,t) = 4\pi G\ \mu_1(r,\theta,t)\ \delta(z)\ . \tag{3.8}$$

This of course is an exact relationship — there has been no linearization so far. As Kalnajs (1965) noted, a distinct advantage of the logarithmic spiral representation is that when $\phi_1(r,\theta,t)$ has the form described by eqns. (3.5) and (3.7), the perturbation potential implied by eqn. (3.8) is just

$$\phi_1(r,\theta,t) = \mathrm{Re}\{-2\pi G r_o \mu_p\ e^{i(m\theta-\omega t)}\ (r/r_o)^{-1/2} \times \int_{-\infty}^{\infty} K(\alpha,m)\ A(\alpha)\ e^{i\alpha\ \ell n(r/r_o)}\ d\alpha\}\ , \tag{3.9}$$

where

$$K(\alpha,m) = \frac{1}{2}\ \frac{\Gamma[(m+1/2+i\alpha)/2]\ \Gamma[(m+1/2-i\alpha)/2]}{\Gamma[(m+3/2+i\alpha)/2]\ \Gamma[(m+3/2-i\alpha)/2]} \tag{3.10}$$

is a dimensionless function sometimes referred to as the Kalnajs gravity factor. Two useful properties of this function are that it is real whenever α is real, and that $K(-\alpha,m) = K(\alpha,m)$. Another advantage of this representation is that the response of a disk to an elementary forcing in wavenumber space — $A(\alpha) = \delta(\alpha-\alpha_f)$ — is well-behaved, in contrast to the response of a disk to a unit forcing in

physical space at a single radius which exhibits a logarithmic singularity at the forcing radius. Once the disturbance potential is available, the perturbation forces arising from the change in the surface density can be readily found:

$$F_{1,r} = -d\phi_1/dr$$

$$= \mathrm{Re}\{(2\pi G \mu_p) e^{i(m\theta-\omega t)} (r/r_o)^{-3/2}$$

$$\times \int_{-\infty}^{\infty} (i\alpha-1/2) K(\alpha,m) A(\alpha) e^{i\alpha \ell n(r/r_o)} d\alpha\} \qquad (3.11)$$

$$F_{1,\theta} = -r^{-1} d\phi_1/d\theta$$

$$= \mathrm{Re}\{(2\pi G \mu_p) e^{i(m\theta-\omega t)} (r/r_o)^{-3/2}$$

$$\times \int_{-\infty}^{\infty} im K(\alpha,m) A(\alpha) e^{i\alpha \ell n(r/r_o)} d\alpha\} . \qquad (3.12)$$

Here $F_{1,r}$ and $F_{1,\theta}$ denote the disturbance forces in the radial and tangential directions.

Of course, not all functions $D(r)$ can be represented as a superposition of logarithmic spirals. Thus our stability analysis admits only those perturbations for which the Fourier integral in eqn. (3.9) is valid and for which the inverse formula,

$$A(\alpha) = (2\pi)^{-1} \int_0^{\infty} \mu_p^{-1} (r/r_o)^{3/2} D(r) e^{-i\alpha \ell n(r/r_o)} (dr/r) , \qquad (3.13)$$

holds. This requires, among other things, that $(r/r_o)^{3/2} D(r)$ tends to zero as $r \to 0$ and $r \to \infty$. The use of the Fourier

representation of the perturbation density, then, implicitly permits a singularity of the disturbance at the center and requires the density to drop off faster than the equilibrium density as $r \to \infty$. These restrictions, however, are not viewed as too severe. For any $s > 0$ we certainly expect $D(r)$ to decay rapidly as $r \to \infty$ for the simple reason that stars at very large radii move too slowly to respond to the disturbance. We expect most disturbances to decay toward the center also both because of absorption at resonances and because there is little active matter in the middle.

The possibility of a singularity in the disturbance at the origin raises another issue: do perturbations of the sort just described permit a valid linearization of the Vlasov equation? This procedure requires that the perturbation forces be much smaller than the equilibrium forces and that the perturbed distribution function be much smaller than the equilibrium distribution function. A specific examination of the linearization will be given here only for the forces. The distribution function behaves in a similar fashion. Both the singular and cut-out disks have equilibrium forces which vary with radius as $(r/r_o)^{-1}$ [cf. eqn. (2.4)]. As $(r/r_o) \to \infty$ the disturbance forces given explicitly in eqns. (3.11) and (3.12) decay at least as fast as $(r/r_o)^{-3/2}$ and therefore the linearization clearly remains valid at large radii. As $(r/r_o) \to 0$, however, there is cause for concern. The $(r/r_o)^{-3/2}$ factor in the perturbation forces seems to imply

that these forces increase more rapidly than the equilibrium forces. If this growth of the perturbation forces toward the center of the disk is not mitigated in some manner, then no matter how small the amplitude of the perturbation, the forces associated with it will exceed the equilibrium forces at sufficiently small (r/r_o) . There are two reasons for not worrying about this. To begin with, for any small disturbance of magnitude ε , the amount of matter in the regime where the linearization apparently breaks down [for $(r/r_o) \lesssim \varepsilon$] is negligibly small compared to the amount of matter in the rest of the disk inside any moderately-sized outer radius R . We have in mind a radius R sufficiently large to encompass the bulk of the region affected by the perturbation, but not so large that the infinite total mass of the disk becomes an issue. But more importantly, for any imposed density field described by a density transform $A(\alpha)$ with a finite half-width[3], the integrals over the wavenumbers α in eqns. (3.11) and (3.12) will decay rapidly for large $|\ell n(r/r_o)|$. So in fact, as $(r/r_o) \to 0$, the $(r/r_o)^{-3/2}$ growth of the perturbation forces is mitigated by the decay in (r/r_o) caused by the Fourier integral. For example, if $K(\alpha,m)\ A(\alpha)$ is a Gaussian in α , then

$$\int_{-\infty}^{\infty} K(\alpha,m)\ A(\alpha)\ e^{i\alpha\ \ell n(r/r_o)}\ d\alpha\ ,$$

---

[3]This excludes the elementary forcing at a fixed wavenumber α_f for which $A(\alpha)$ is a delta-function in α .

will be a Guassian in $\ell n(r/r_o)$; in this case,

$$(r/r_o)^{-3/2} \int_{-\infty}^{\infty} K(\alpha,m)\ A(\alpha)\ e^{i\alpha\ \ell n(r/r_o)}\ d\alpha$$

actually tends to zero rather than blowing up as feared, when the center is approached.

Although the calculations in the following section are based on the computation of the response to an elementary forcing of the disk, we eventually integrate over all values of α and thus we expect the smoothing effect discussed above to take hold. In fact, looking ahead for a moment, all the modes of the cut-out disks that we specifically located do exhibit this smoothing effect and decay in amplitude toward very small radii.

c) Calculation of the Dynamical Response

To be worthy of its name, any mode must, of course, be self-consistent — that is, the disturbance forces that arise from the density field of a mode must regenerate the same densities back, with due allowance for the temporal growth and rotation of the modal pattern. The task undertaken in this section is the derivation of the specific conditions that must be satisfied by any mode of the singular disk or one of the cut-out disks. The central issue is the dynamical response of one of these model disks to an arbitrary imposed force field.

For the time being imagine that a potential perturbation,

$$\phi_{l,imp}(r,\theta,t) = Re\{-2\pi G\mu_p r_o \, e^{i(m\theta-\omega t)} \, (r/r_o)^{-1/2} \times \int_{-\infty}^{\infty} K(\alpha,m) \, A_{imp}(\alpha) \, e^{i\alpha \, \ell n(r/r_o)} \, d\alpha\} \, , \quad (3.14)$$

has been imposed upon a disk. The forces resulting from this imposed potential will induce a response density, $\mu_{l,res}(r,\theta,t)$, in the disk. This process is governed by the Vlasov equation, eqn. (3.1). At the same time, as noted in the preceding section, this imposed potential corresponds, via Poisson's equation, to the imposed density,

$$\mu_{l,imp}(r,\theta,t) = Re\{\mu_p \, e^{i(m\theta-\omega t)} \, (r/r_o)^{-3/2} \times \int_{-\infty}^{\infty} A_{imp}(\alpha) \, e^{i\alpha \, \ell n(r/r_o)} \, d\alpha\} \, . \quad (3.15)$$

If the density transform $A_{imp}(\alpha)$ represents a mode of angular harmonic number m and frequency ω , then the imposed density will exactly equal the response density, i.e., $\mu_{l,res}(r,\theta,t) = \mu_{l,imp}(r,\theta,t)$.

The specific question now is how the disk will respond to an arbitrary imposed potential of the form expressed by eqn. (3.14) and this in turn centers on how the disk will respond to an elementary forcing,

$$\phi_{l,imp}(r,\theta,t) = Re\{-2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i(m\theta-\omega t)} \times (r/r_o)^{-1/2} \, e^{i\alpha \, \ell n(r/r_o)}\} \, , \quad (3.16)$$

at a single wavenumber. Henceforth the explicit indication that the real parts of these complex-valued functions are intended will be dropped.

In the presence of this elementary forcing the distribution function will change from its equilibrium value

$$f(r,\theta,u,v,t) = f_o(r,u,v)$$

to the new amount

$$f(r,\theta,u,v,t) = f_o(r,u,v) + f_1(r,\theta,u,v,t) \ .$$

Under the assumption of infinitesimally small disturbances the specific value that the distribution function perturbation, $f_1(r,\theta,u,v,t)$, takes is governed by the linearized Vlasov equation,

$$\frac{Df_1}{Dt} = \frac{\partial \phi_{1,imp}}{\partial r} \frac{\partial f_o}{\partial u} + \frac{1}{r} \frac{\partial \phi_{1,imp}}{\partial \theta} \frac{\partial f_o}{\partial v} \ , \qquad (3.17)$$

where the convective derivative D/Dt is given by eqn. (3.2). Since the equilibrium distribution function has the functional dependence $f_o(r,u,v) = f_o(E,J)$, eqn. (3.17) can be rewritten as

$$\frac{Df_1}{Dt} = \frac{\partial f_o}{\partial E} \{u \frac{\partial \phi_{1,imp}}{\partial r} + \frac{v}{r} \frac{\partial \phi_{1,imp}}{\partial \theta}\} + \frac{\partial f_o}{\partial J} \frac{\partial \phi_{1,imp}}{\partial \theta} \ . \qquad (3.18)$$

One standard method for solving this partial differential equation for $f_1(r,\theta,u,v,t)$ is to integrate the right-hand

side along the equilibrium orbit of a star which at time t has positions and velocities (r,θ,u,v) [see, for example, Stix (1962) in a related, plasma context]. In order to do this, the radial and tangential perturbation forces that such a star is subjected to along its unperturbed orbit at all previous times must be computed. We denote these radial and tangential forces at time t' by $F_{1,r}(r,\theta,u,v,t;t')$ and $F_{1,\theta}(r,\theta,u,v,t;t')$, respectively. The solution of the linearized Vlasov equation can now be expressed as

$$f_1(r,\theta,u,v,t) = - (\partial f_o/\partial E)\ \Delta E - (\partial f_o/\partial J)\ \Delta J\ , \tag{3.19}$$

where ΔE and ΔJ are the cumulative changes in the star's energy and angular momentum due to the perturbation:

$$\Delta E = \int_{-\infty}^{t} [u'(t')\ F_{1,r}(r,\theta,u,v,t;t') + v'(t')\ F_{1,\theta}(r,\theta,u,v,t;t')]\ dt' \tag{3.20}$$

$$\Delta J = \int_{-\infty}^{t} r'(t')\ F_{1,\theta}(r,\theta,u,v,t;t')\ dt'\ . \tag{3.21}$$

Here, as in section II.b, primed variables refer to quantities along the unperturbed orbit of the star.

The results of these and other calculations in this subsection can be conveniently expressed in terms of the imposed potential sampled by a star along its equilibrium orbit. This <u>sampled potential</u> is denoted by $\phi_{1,imp}(r,\theta,u,v,t;t')$ and is given by

$$\phi_{1,imp}(r,\theta,u,v,t;t') = -2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i[m\theta'(t') - \omega t']}$$
$$\times \, [r'(t')/r_o]^{-1/2} \, e^{i\alpha \, \ell n[r'(t')/r_o]} \, . \qquad (3.22)$$

Both the sampled potential and the perturbation forces con-contain a dependence upon the angular harmonic number m and the forcing wavenumber α in addition to the dependencies on the variables explicitly listed; these secondary dependencies, however, have been suppressed.

We now shift to the $(\tilde{U},r_H)$ description of velocity space and use the formulas for the stellar orbits developed in section II.b:

$$\psi = \kappa(\tilde{U},r_H) \, (t'-t_p)$$
$$r'(t') = r_H \, e^{iX(\psi,\tilde{U})} \qquad (3.23)$$
$$\theta'(t') = Y(\psi,\tilde{U}) + \Omega(\tilde{U},r_H) \, (t'-t_p) + \theta_p \, .$$

With these expressions in mind, the sampled potential is now rewritten as

$$\phi_{1,imp}(r,\theta,\tilde{U},r_H,t;t') = -2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i(m\theta-\omega t)}$$
$$\times \, e^{(i\alpha-1/2) \, \ell n(r/r_o)} \, e^{i[m\Omega(\tilde{U},r_H)-\omega](t'-t)}$$
$$\times \, \{e^{(i\alpha-1/2) \, \ell n[r'(t')/r_o] + i[m\theta'(t')-\Omega(\tilde{U},r_H)(t'-t)-\theta]}\} \, . \qquad (3.24)$$

At this point we adopt a further suggestion by Kalnajs (1973, private communication) and, noting that the term in braces in

the above equation is periodic in t', expand it into the Fourier series

$$\{e^{(i\alpha-1/2)\ \ell n[r'(t')/r_o] + i[m\theta'(t') - \Omega(\tilde{U},r_H)(t'-t) - \theta]}\}$$

$$= \sum_{\ell=-\infty}^{\infty} P_{\ell m}(\alpha;\tilde{U},r_H)\ e^{i\ell\kappa(\tilde{U},r_H)(t'-t)}\ ,$$

where

$$P_{\ell m}(\alpha;\tilde{U},r_H) = T^{-1}(\tilde{U},r_H) \int_t^{t+T(\tilde{U},r_H)} e^{-i\ell\kappa(\tilde{U},r_H)(t'-t)} \times$$

$$\{e^{(i\alpha-1/2)\ \ell n[r'(t')/r_o]+i[m\theta'(t')-\Omega(\tilde{U},r_H)(t'-t)-\theta]}\}\ . \quad (3.25)$$

These Fourier coefficients can be expressed in simpler terms, owing to the similarity of the stellar orbits. This is achieved by performing the integration in eqn. (3.25) over the orbital phase ψ instead of over the time t' and by replacing the $(\tilde{U},r_H)$ description with the $(\tilde{U},\xi)$ one. As implied in section II.b, the velocity phase ξ has the value

$$\xi = \kappa(\tilde{U},r_H)(t-t_p)\ .$$

Noting that when $t' = t$ in eqns. (3.23),

$$\psi = \xi\ ,$$

$$r = r_H\ e^{X(\xi;\tilde{U})}\ ,$$

and

$$\theta = Y(\xi;\tilde{U}) + \Omega(\tilde{U},r_H)(t-t_p) + \theta_p\ ,$$

we obtain

$$P_{\ell m}(\alpha;\tilde{U},\xi) = e^{-(i\alpha-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})} \times$$
$$(2\pi)^{-1} \int_{\xi}^{2\pi+\xi} e^{(i\alpha-1/2)X(\psi;\tilde{U})+imY(\psi;\tilde{U})-i\ell\psi}\, d\psi \ .$$

Finally, by shifting the integration to the interval $[0,2\pi]$, the factorization

$$P_{\ell m}(\alpha;\tilde{U},\xi) = e^{-(i\alpha-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})+i\ell\xi}\, Q_{\ell m}(\alpha;\tilde{U}) \ , \qquad (3.26)$$

where

$$Q_{\ell m}(\alpha;\tilde{U}) = (2\pi)^{-1} \int_{0}^{2\pi} e^{(i\alpha-1/2)X(\psi;\tilde{U})+imY(\psi;\tilde{U})-i\ell\psi}\, d\psi \ , \qquad (3.27)$$

results. This last expression may be interpreted as a Fourier coefficient for the reference orbit of the stars with eccentric velocity $\tilde{U}$, i.e., an orbit with unit home radius in which the star is initially at the pericenter.

When the Fourier series expansion and the $(\tilde{U},\xi)$ velocity space description are used in the sampled potential, eqn. (3.22), we find that

$$\phi_{1,imp}(r,\theta,\tilde{U},\xi,t;t') = -2\pi G\mu_p r_o\, K(\alpha,m)\, e^{i(m\theta-\omega t)}$$
$$\times\ e^{(i\alpha-1/2)\ \ell n(r/r_o)}\ e^{i[m\Omega(\tilde{U},r_H)-\omega](t'-t)}$$
$$\times \sum_{\ell=-\infty}^{\infty} P_{\ell m}(\alpha;\tilde{U},\xi)\ e^{i\ell\kappa(\tilde{U},r_H)(t'-t)} \ . \qquad (3.28)$$

As Kalnajs pointed out to us, this Fourier series expansion

of the sampled potential enables the orbit integration to be carried out explicitly, even for eqn. (3.20):

$$\Delta E = 2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i(m\theta-\omega t)} \, e^{(i\alpha-1/2)\,\ell n(r/r_o)} \times \sum_{\ell=-\infty}^{\infty} \frac{\ell\kappa(\tilde{U},r_H) + m\Omega(\tilde{U},r_H)}{\ell\kappa(\tilde{U},r_H) + m\Omega(\tilde{U},r_H) - \omega} \, P_{\ell m}(\alpha;\tilde{U},\xi) \, . \tag{3.29}$$

Then also, if more obviously :

$$\Delta J = 2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i(m\theta-\omega t)} \, e^{(i\alpha-1/2)\,\ell n(r/r_o)} \times \sum_{\ell=-\infty}^{\infty} \frac{m \, P_{\ell m}(\alpha;\tilde{U},\xi)}{\ell\kappa(\tilde{U},r_H) + m\Omega(\tilde{U},r_H) - \omega} \, . \tag{3.30}$$

These results may be verified in a straightforward manner, using

$$u'(t') = dr'(t')/dt' \, , \quad v'(t') = J/r'(t') \, ,$$

$$F_{l,r}(r,\theta,u,v,t;t') = - \frac{\partial \phi_{l,imp}}{\partial r}(r'(t'),\theta'(t'),\tilde{U},\xi,t;t') \, ,$$

$$F_{l,\theta}(r,\theta,u,v,t;t') = - \frac{1}{r'(t')} \frac{\partial \phi_{l,imp}}{\partial \theta}(r'(t'),\theta'(t'),\tilde{U},\xi,t;t') \, .$$

We are now able to write down the perturbed distribution function. Employing eqns. (2.11), (2.58), (3.19), (3.29) and (3.30) we find that

$$f_l(r,\theta,\tilde{U},\xi,t) = 2\pi G\mu_p r_o \, K(\alpha,m) \, e^{i(m\theta-\omega t)} e^{(i\alpha-1/2)\ell n(r/r_o)} f_s(E,J) \times \sum_{\ell=-\infty}^{\infty} \frac{\{(a+1)[\ell\kappa(\tilde{U},r_H)+m\Omega(\tilde{U},r_H)]/V_o^2 - am/J\}H(J) - mH'(J)}{\ell\kappa(\tilde{U},r_H) + m\Omega(\tilde{U},r_H) - \omega} \, P_{\ell m}(\alpha;\tilde{U},\xi) \, . \tag{3.31}$$

Now that the perturbed distribution function is available, the density response to the forcing at the single wavenumber α , which is denoted by $\mu_{1,res}(r,\theta,t;\alpha)$, can be obtained by integration:

$$\mu_{1,res}(r,\theta,t;\alpha) = \int_{-\infty}^{\infty} \int_{0}^{\infty} f_1(r,\theta,\tilde{U},\xi,t)\ du\ dv\ .$$

We choose to represent the result of this calculation as

$$\mu_{1,res}(r,\theta,t;\alpha) = \mu_p\ (r/r_o)^{-3/2}\ e^{i(m\theta-\omega t)}\ \hat{S}_m(r;\alpha;\sigma_u,\omega) \qquad (3.32)$$

in terms of the <u>elementary response function</u> $\hat{S}_m(r;\alpha;\sigma_u,\omega)$. In order to perform the integration over velocity space in terms of the variables $(\tilde{U},\xi)$ instead of (u,v) , the Jacobian of this transformation is required:

$$\frac{\partial(u,v)}{\partial(\tilde{U},\xi)} = (2\pi)^{-1}\ I_0(\tilde{U})\ \tilde{U}\ e^{-2X(\xi;\tilde{U})}\ V_o^2\ . \qquad (3.33)$$

The elementary response function is then found to be

$$\begin{aligned}
\hat{S}_m(r;\alpha;\sigma_u,\omega) = {} & 2\pi G r_o\ K(\alpha,m)\ (\mu_o r_o/V_o)\ \tilde{C}_a\ e^{(i\alpha-1/2)\ \ell n(r/r_o)} \\
& \times \int_0^{\infty} \int_0^{2\pi} (2\pi)^{-1}\ I_0(\tilde{U})\ e^{-2X(\xi;\tilde{U})}\ V_o^2\ J^{-1} \\
& \times e^{-\tilde{U}^2/2\tilde{\sigma}_u^2} \sum_{\ell=-\infty}^{\infty} e^{-(i\alpha-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})+i\ell\xi} \\
& \times \frac{\{(a+1)[\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})]/J-am/J\}H(J)\ -\ mH'(J)}{[\ell\tilde{\kappa}(\tilde{U})\ +\ m\tilde{\Omega}(\tilde{U})]V_o^2/J\ -\ \omega} \\
& \times Q_{\ell m}(\alpha;\tilde{U})\ \tilde{U}\ d\tilde{U}\ d\xi\ .
\end{aligned}$$

In this expression we have made extensive use of the dimensionless quantities introduced in section II.b. The variable J must be interpreted as a function of r, $\tilde{U}$ and ξ, namely,

$$J = rV_o\, e^{-X(\xi;\tilde{U})} \quad .$$

This occurs because the tangential velocity v has been parameterized by $\tilde{U}$ and ξ. Eventually, another coordinate transformation will be adopted that permits the integration to occur over the more natural angular momentum variable. The expression for the elementary response function may be simplified by noting that $V_o^2 = 2\pi G\mu_o r_o$ [see eqn. (2.3)] and by introducing the additional dimensionless variables

$$\begin{aligned} \tilde{\omega} &= (r_o/V_o)\ \omega \\ \mathfrak{J} &= (r_o V_o)^{-1}\ J \quad . \end{aligned} \tag{3.34}$$

The cut-out factor $H(J)$ and its derivative are also put in dimensionless form via

$$\begin{aligned} \tilde{H}(\mathfrak{J}) &= H(r_o V_o \mathfrak{J}) \\ \tilde{H}'(\mathfrak{J}) &= r_o V_o\ dH(r_o V_o \mathfrak{J})/dJ \quad . \end{aligned} \tag{3.35}$$

Referring to eqn. (2.57), we have

$$\begin{aligned} \tilde{H}(\mathfrak{J}) &= \mathfrak{J}^N/(1 + \mathfrak{J}^N) \\ \tilde{H}'(\mathfrak{J}) &= N\mathfrak{J}^{N-1}/(1 + \mathfrak{J}^N)^2 \quad . \end{aligned} \tag{3.36}$$

We then obtain

$$\hat{S}_m(r;\alpha;\tilde{\sigma}_u,\tilde{\omega}) = \tilde{C}_a\ K(\alpha,m)\ e^{i\alpha\ \ell n(r/r_o)} \int_0^\infty \{ \int_0^{2\pi} (2\pi)^{-1}\ I_0(\tilde{U})$$
$$\times\ e^{-\tilde{U}^2/2\tilde{\sigma}_u^2} \sum_{\ell=-\infty}^{\infty} e^{(-i\alpha-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})+i\ell\xi}$$
$$\times\ \frac{[(a+1)(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )-am]\tilde{H}(\tilde{J})\ -\ m\tilde{J}\tilde{H}'(\tilde{J})}{\ell\tilde{\kappa}(\tilde{U})\ +\ m\tilde{\Omega}(\tilde{U})\ -\ \tilde{\omega}\tilde{J}}\ d\xi\}$$
$$\times\ Q_{\ell m}(\alpha;\tilde{U})\ \ \tilde{U}\ d\tilde{U}\ . \qquad (3.37)$$

This function measures the density response to forcing at a single wavenumber. The real interest, of course, is in the response to a forcing from an arbitrary imposed density of the form of eqn. (3.15). If this density response is expressed as

$$\mu_{1,res}(r,\theta,t) = D_{res}(r)\ e^{i(m\theta-\omega t)}\ ,$$

then clearly[4]

$$D_{res}(r) = \mu_p\ (r/r_o)^{-3/2} \int_{-\infty}^{\infty} \hat{S}_m(r;\alpha;\tilde{\sigma}_u,\tilde{\omega})\ A_{imp}(\alpha)\ d\alpha\ .$$

The response density itself has the transform $A_{res}(\alpha)$ which

[4]Here and at several subsequent points in this section we have rather freely interchanged the order of integration in a multiple, improper integral containing an infinite sum. These mathematical manipulations must be regarded as merely formal, for no rigorous justification for interchanging the several limiting processes in these expressions is offered. We submit that our cavalier approach to these questions is vindicated in Appendix F, in which we report on an independent confirmation of some of the results obtained on the basis of the analysis of this section.

is determined by

$$A_{res}(\beta) = (2\pi)^{-1}\int_0^\infty \mu_p^{-1}(r/r_o)^{3/2}\, D_{res}(r)\, e^{-i\beta\, \ell n(r/r_o)}\, (dr/r) .$$

We now define the transfer function, $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, which relates the transform of the response density to that of the imposed density:

$$A_{res}(\beta) = \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\, A_{imp}(\alpha)\, d\alpha \quad . \tag{3.38}$$

The transfer function is simply the transform of the elementary response function, i.e.,

$$S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega}) = (2\pi)^{-1}\int_0^\infty S_m(r;\alpha;\tilde{\sigma}_u,\tilde{\omega})\, e^{-i\beta\, \ell n(r/r_o)}\, (dr/r). \tag{3.39}$$

It specifies how the density response to an elementary forcing at wavenumber α spreads itself out over the response wavenumbers β . It is calculated from eqns. (3.39) and (3.37) by interchanging the orders of integration and summation until the integral over r is the innermost operation and then changing this integration variable from r to $\tilde{\jmath}$ where $\tilde{\jmath} = (r/r_o)\, e^{-X(\xi;\tilde{U})}$. The result is

$$\begin{aligned} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega}) = \tilde{C}_a\, K(\alpha,m) \sum_{\ell=-\infty}^{\infty} \int_0^\infty I_0(\tilde{U})\, e^{-\tilde{U}^2/2\tilde{\sigma}_u^2}\, Q_{\ell m}(\alpha;\tilde{U}) \\ \times \{(2\pi)^{-1} \int_0^{2\pi} e^{(-i\beta-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})+i\ell\xi}\, d\xi\} \\ \times \{(2\pi)^{-1} \int_0^\infty \frac{[(a+1)(\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U}))-am]\tilde{H}(\tilde{\jmath})-m\tilde{\jmath}\tilde{H}'(\tilde{\jmath})}{\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) - \tilde{\omega}\tilde{\jmath}} \\ \times e^{i(\alpha-\beta)\, \ell n\, \tilde{\jmath}}\, (d\tilde{\jmath}/\tilde{\jmath})\}\, \tilde{U}\, d\tilde{U} . \end{aligned}$$

But notice that

$$\bar{Q}_{\ell m}(\beta;\tilde{U}) = (2\pi)^{-1} \int_0^{2\pi} e^{(-i\beta-1/2)X(\xi;\tilde{U})-imY(\xi;\tilde{U})+i\ell\xi}\, d\xi$$

[see eqn. (3.27)], where the bar denotes the complex conjugate, and define the <u>angular momentum function</u>, $F_{\ell m}(\eta;\tilde{U},\tilde{\omega})$, by

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = (2\pi)^{-1} \int_0^{\infty} \frac{[(a+1)(\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U}))-am]\tilde{H}(\mathcal{J})-m\mathcal{J}\tilde{H}'(\mathcal{J})}{\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})-\tilde{\omega}\mathcal{J}} \times e^{-i\eta \ell n\, \mathcal{J}}\, (d\mathcal{J}/\mathcal{J}) .$$

The transfer function then is simply

$$S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega}) = \tilde{C}_a\, K(\alpha,m) \sum_{\ell=-\infty}^{\infty} \int_0^{\infty} I_0(\tilde{U})\, e^{-\tilde{U}^2/2\tilde{\sigma}_u^2} \times Q_{\ell m}(\alpha;\tilde{U})\, \bar{Q}_{\ell m}(\beta;\tilde{U})\, F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})\, \tilde{U}\, d\tilde{U} . \quad (3.40)$$

The angular momentum function can be expressed in the alternate form

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = (2\pi)^{-1} \int_{-\infty}^{\infty} \frac{[(a+1)(\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U}))-am]\hat{H}(h)-m\hat{H}'(h)}{\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})-\tilde{\omega}e^{h}} e^{-i\eta h}\, dh \quad (3.41)$$

by making the change of variables $h = \ell n\, \mathcal{J}$ and the definitions $\hat{H}(h) = \tilde{H}(e^h)$ and $\hat{H}'(h) = d\hat{H}/dh$.

Now that all these specific expressions are available we return to eqn. (3.38) and apply the self-consistency requirement — $A_{res}(\alpha) = A_{imp}(\alpha)$ — to obtain the condition

which the growing modes of the singular and cut-out disks must satisfy, namely,

$$\boxed{A(\beta) = \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\ A(\alpha)\ d\alpha\ .} \qquad (3.42)$$

This singular homogenous Fredholm integral equation for the density transform of the modes is central to our subsequent determination of whether exponentially growing modes are indeed possible in any given one of our model galaxies. Before proceeding to a description of the kernel , $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, we wish to emphasize that the angular momentum function has only been derived for growing disturbances, i.e., those with $Im\{\tilde{\omega}\} > 0$.

d) Description of the Integral Equation

The nature of the kernel depends critically upon the behavior of the angular momentum function. The Fourier integral, eqn. (3.41), defining this function may be performed analytically for both the singular and the cut-out disks. For the singular disk, $\hat{H}(h) = 1$, while for the cut-out disks, $\hat{H}(h) = e^{Nh}/(1 + e^{Nh})$. In fact, the particular choice of the cut-out factor in section II.c was guided by the desire to perform these integrations analytically rather than numerically. The trick is to use residue calculus with a contour that includes the lines $Im\{h\} = 0$ and $Im\{h\} = 2\pi$ and closes at infinity. Then, in the case of the singular disk,

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = \frac{(a+1)(\ \ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})\ ) - am}{\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})} \{ \quad (1/2)\ \delta(\eta)$$

$$\times \quad e^{-i\eta\ \ell n[(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )/\tilde{\omega}]}\ /[1 - e^{2\pi\eta}]\ , \qquad (3.43)$$

and in the case of a cut-out disk

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = -\frac{i}{1 - e^{2\pi\eta}} \{ \frac{[\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})]^{N-1}}{[\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})]^{N} + \tilde{\omega}^{N}}$$

$$\times \{ [(a+1)(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )-am] - \frac{Nm\tilde{\omega}^{N}}{[\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})]^{N} + \tilde{\omega}^{N}}\}$$

$$\times \quad e^{-i\eta\ \ell n[(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )/\tilde{\omega}]}$$

$$-\frac{1}{N}\sum_{j=1}^{N} \frac{e^{(2j-1)\pi\eta/N}}{\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U}) - \tilde{\omega}e^{(2j-1)\pi i/N}} \{ \quad -im\eta +$$

$$[(a+1)(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )-am] + \frac{m\tilde{\omega}\ e^{(2j-1)\pi i/N}}{\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U}) - \tilde{\omega}e^{(2j-1)\pi i/N}}\}\} \ . \qquad (3.44)$$

Throughout this thesis the principal branch of the logarithm is chosen, i.e., its imaginary part lies in the range $[0,2\pi]$. Two cases are then distinguished for the logarithm appearing in the previous two formulas:

Case 1: $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) > 0$

$$\ell n[(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )/\tilde{\omega}] = \ell n\left|\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\right| - \ell n\ \tilde{\omega} + 2\pi i \qquad (3.45)$$

Case 2: $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) < 0$

$$\ell n[(\ \ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\ )/\tilde{\omega}] = \ell n\left|\ell\tilde{\kappa}(\tilde{U})+m\tilde{\Omega}(\tilde{U})\right| - \ell n\ \tilde{\omega} + \pi i \ . \qquad (3.46)$$

The sign of $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ determines whether the ℓ'th radial Fourier component of the potential is directly (> 0) or counter (< 0) rotating. Since for moderate random velocity dispersions, the typical values of $\tilde{U}$ are rather small, a good handle on the sign of $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ is provided by its behavior for $\tilde{U} \to 0$. In this case, $\tilde{\kappa}(\tilde{U}) = \sqrt{2}$ and $\tilde{\Omega}(\tilde{U}) = 1$, and hence for $m = 1$, the direct harmonics are $\ell = 0, 1, 2, \cdots$, for $m = 2$, they are $\ell = -1, 0, 1, 2, \cdots$, for $m = 3$ they are $\ell = -2, -1, 0, 1, 2, \cdots$, for $m = 4$ they are $\ell = -2, -1, 0, 1, 2, \cdots$, etc. Referring to the discussion of the orbital parameters in section II.b, we find that as $\tilde{U}$ increases, both $\tilde{\kappa}(\tilde{U})$ and $\tilde{\Omega}(\tilde{U})$ tend to zero but their ratio $\tilde{\kappa}(\tilde{U})/\tilde{\Omega}(\tilde{U}) = 2\pi/I_2(\tilde{U})$ tends to 2 . A specific analysis reveals that for $m = 1, 2$ and 4 , $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ has the same sign for each fixed ℓ , regardless of the value of $\tilde{U}$. For all other $m > 0$, there are some values of ℓ for which $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ changes sign as $\tilde{U}$ increases, and hence for which $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ vanishes. An eccentricity class for which this occurs is one for which the orbits are closed in inertial space. Any rotating disturbance will appear virtually stationary to the rapidly rotating stars in such closed orbits near the center of the disk. Thus, low angular momentum stars in closed orbits are, for all practical purposes, in resonance with the disturbance. In the singular disk there are a large number of active stars in these nearly resonant orbits and the angular momentum function in eqn. (3.43)

blows up as $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) \to 0$. In the cut-out disks, at least when the cut-out index N is greater than or equal to two, there is a very small number of such stars and the troublesome term actually vanishes as $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) \to 0$. In this thesis we are primarily concerned with $m = 0$, 1 and 2 . In these cases, the term $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ never vanishes and so the above expressions for the angular momentum function can be used without concern. For $m = 3$, however, $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) = 0$ when $\ell = -2$ and $\tilde{U} \simeq 1.19$. Extra care is needed here. This point is considered further when the results for $m = 3$ are presented in Chapter VI.

Even when there are no closed orbits in the non-rotating inertial frame, there will still be orbits which are closed in the frame rotating with the angular speed $\tilde{\Omega}_p$ of the disturbance. Stars in such orbits will be in resonance with the disturbance. These resonances surface clearly in the angular momentum function expression in eqn. (3.41): whenever $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) > 0$ there will be some value of the angular momentum for which the real part of the denominator vanishes. This happens whenever $\ell\kappa(\tilde{U},r_H) + m\Omega(\tilde{U},r_H) = m\Omega_p$. But this is precisely the condition that the orbit of a star with eccentric velocity $\tilde{U}$ and home radius r_H (angular momentum $J = V_o r_H$) close in the frame rotating with angular velocity Ω_p . The home radii of such orbits are in the vicinity of the radii for which $\kappa(r) + m\Omega(r) = m\Omega_p$, where $\kappa(r)$ and $\Omega(r)$ now are the epicyclic frequency and angular

velocity of the disk as a whole at radius r . For stars in circular orbits, resonances occur for all the radial harmonics for which $\ell\kappa(r) + m\Omega(r) > 0$. In the density wave theory of spiral structure three of these resonances play a major role: the corotation resonance ($\ell = 0$), where the angular velocity of the disk equals that of the pattern, the inner Lindblad resonance ($\ell = -1$) and the outer Lindblad resonance ($\ell = +1$). In the present models, corotation occurs at

$$r/r_o = 1/\tilde{\Omega}_p \ ,$$

while the Lindblad resonances occur at

$$r/r_o = (1 \pm \sqrt{2}/m)/\tilde{\Omega}_p \ .$$

The explicit expression for the angular momentum function of the singular disk reveals several interesting features. The most significant of these was pointed out by Kalnajs (1974, private communication). The angular momentum function admits the factorization

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = e^{i\eta\ \ell n\ \tilde{\omega}}\ F^*_{\ell m}(\eta;\tilde{U}) \ ,$$

and consequently the kernel itself has the form

$$S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega}) = e^{i(\beta-\alpha)\ \ell n\ \tilde{\omega}}\ \ S^*_m(\beta,\alpha;\tilde{\sigma}_u) \ ,$$

where the functions $F^*_{\ell m}(\eta;\tilde{U})$ and $S^*_m(\beta,\alpha;\tilde{\sigma}_u)$ implicitly defined by these relations are actually independent of the

frequency $\tilde{\omega}$. If the function $A^*(\alpha)$ is now defined by

$$A^*(\alpha) = e^{-i\alpha \ln \tilde{\omega}} A(\alpha) ,$$

then the integral equation for the growing modes, eqn. (3.42), is seen to be equivalent to

$$A^*(\beta) = \int_{-\infty}^{\infty} S_m^*(\beta,\alpha;\tilde{\sigma}_u) A^*(\alpha) \, d\alpha .$$

If, for some value of $\tilde{\sigma}_u$, this equation has a non-trivial solution, then for each frequency in the upper half $\tilde{\omega}$-plane, eqn. (3.42) will have the non-trivial solution

$$A(\alpha) = e^{i\alpha \ln \tilde{\omega}} A^*(\alpha) .$$

In this event the singular disk will have a two-dimensional continuum of growing modes. More likely, this equation will only have the trivial solution, implying that the singular disk admits no exponentially growing modes. The frequency-independent nature of the kernel has removed the customary freedom to adjust the frequency so that the kernel admits non-trivial solutions, or modes.[5] It was especially because of this peculiar behavior that we decided to study the effect of carving out part of the highly active center from the singular disk. The cut-out disks that result have the frequency inextricably contained in their kernels — no similar

[5]Kalnajs also noted that this strange feature of the singular disk occurs regardless of the particular distribution function chosen.

factorization of the frequency occurs. The reason that one occurs in the singular disk is that it has no intrinsic scale — it is "too self-similar". The cut-out disks, on the other hand, do have a length scale — the size of the rigid center. In the latter models we expect to be able to find some solutions of eqn. (3.42) corresponding to growing modes for discrete frequencies, provided, of course, that the velocity dispersion is sufficiently small.

A second, though less disconcerting, feature of the basic self-similar model is that not only is the integral equation singular because the range of wavenumbers is infinite, but it also is so because the kernel itself is singular at $\alpha = \beta$, where it behaves as $(\beta-\alpha)^{-1}$. The integral on the right-hand side of eqn. (3.42) should therefore be interpreted in the principal value sense. The singularity at $\alpha = \beta$ arises only because the orders of integration have been interchanged in deriving the integral equation. No such singularity arises in the more natural procedure for computing the response to an arbitrary disturbance of a given frequency: integrating first over the wavenumbers in the disturbance and only then over the velocity space. In any case, the kernel of the cut-out disks is well-defined at $\alpha = \beta$; this principal value integral is immaterial for them.

One should also take note of the strong trailing bias exhibited by both the singular and cut-out disks. Recall that the transfer function $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$ reports how a

forcing at the wavenumber α spreads itself out over the other wavenumbers β. The angular momentum function dominates this redistribution. The trailing bias can be seen most clearly in the singular disk. Assume that the forcing rotates directly ($\tilde{\Omega}_p > 0$) and that it has a vanishingly small growth rate[6] ($\tilde{s} = 0^+$). Then, for those Fourier components which are directly rotating [$\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) > 0$], the angular momentum function behaves roughly as

$$F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega}) \sim \begin{cases} 1 & \beta > \alpha \\ e^{-2\pi(\alpha-\beta)} & \beta < \alpha \end{cases},$$

while for those components which are counter-rotating,

$$F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega}) \sim \begin{cases} e^{-\pi(\beta-\alpha)} & \beta > \alpha \\ e^{-\pi(\alpha-\beta)} & \beta < \alpha \end{cases},$$

exclusive of multiplicative constants. The behavior of the cut-out disks is analogous. The only substantial difference away from $\alpha = \beta$ is that the exponentially decaying factors are all $e^{-|\beta-\alpha|\pi/N}$. Both direct and counter-rotating components occur for any angular harmonic. In the transfer function itself, which involves a sum over the radial Fourier harmonics and an integration over the eccentric velocities,

---

[6] Although the angular momentum function is only defined by the integral in eqn. (3.41) for $\mathrm{Im}\{\tilde{\omega}\} > 0$, we are certainly entitled to use analytic continuation to obtain the value as $\tilde{s} \to 0^+$ and beyond. Since we have explicit expressions for the result of the integration when $\tilde{s} > 0$, this limiting process is straightforward.

the directly rotating components will be more important for $\beta > \alpha$, whereas the counter-rotating components will be more important for $\beta < \alpha$. Clearly these disks will respond to an arbitrary disturbance in a decidedly one-sided, trailing fashion — most of the response occurs in wavenumbers more positive than the forcing wavenumber. Very little of the total response is fed back into more leading wavenumbers. Figure 2 illustrates a typical angular momentum function of each type for the $N = 2$ disk.

Our searches for modes are confined to the cut-out disks. If their kernels were entirely one-sided, the search would be futile, for Volterra kernels have no eigenfunctions. However, as we shall see, even the slight trickle of feedback into more leading wavenumbers suffices to permit growing modes if the velocity dispersion is low enough. For growing disturbances the kernel is not so one-sided as in the case of vanishingly small growth. On the trailing side the angular momentum function is decreased by a factor of $\exp[-|\beta-\alpha| \times \tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)]$ and on the leading side it is increased to the same extent. This favorable tendency, though, is more than offset by an overall decrease in the size of the transfer function as the growth rate is increased. Naturally enough, slowly growing disturbances are easier to produce than rapidly growing ones.

A few remarks about the behavior of the rest of the kernel are in order. As Kalnajs (1971) has shown, for large

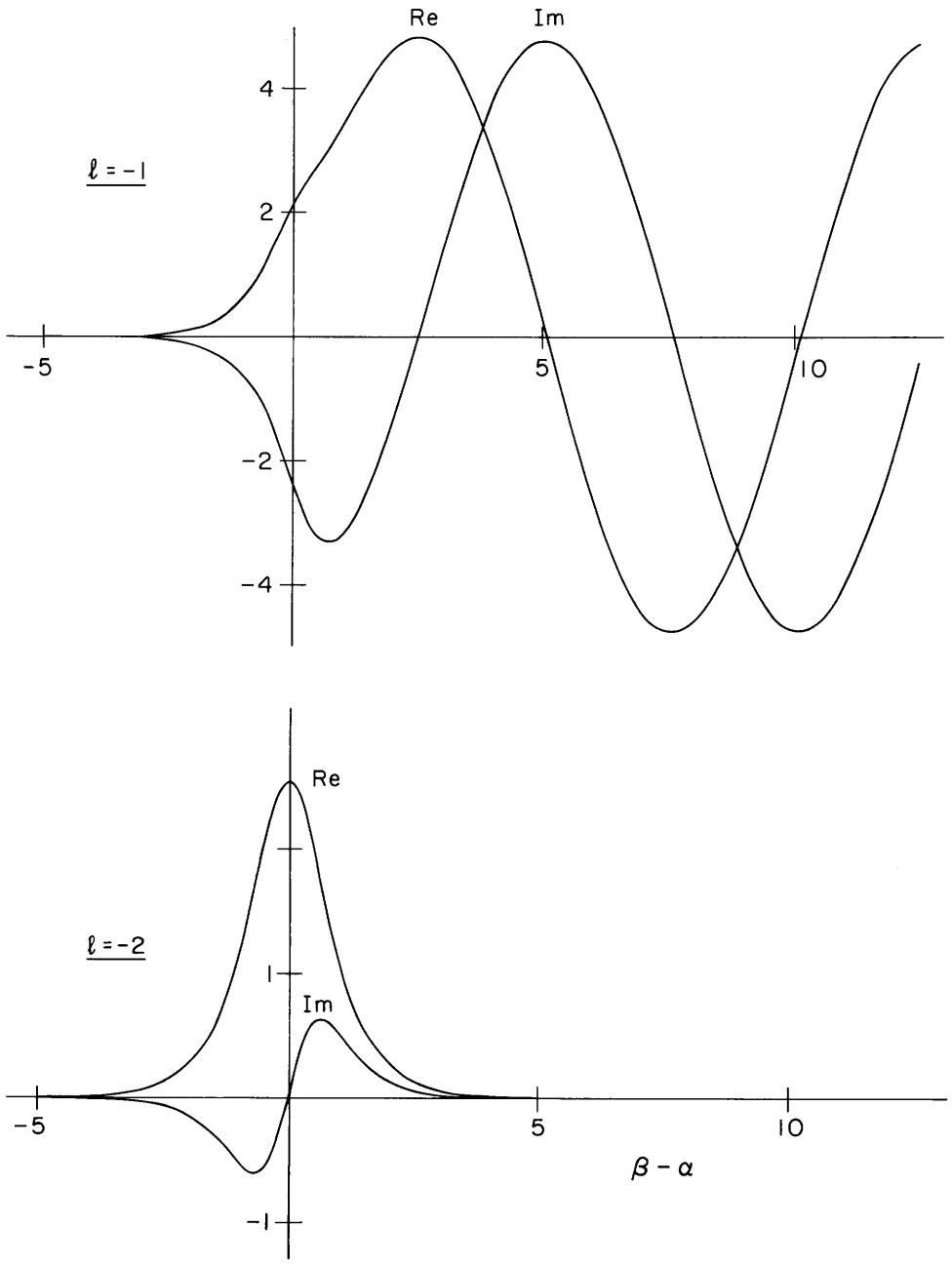

Figure 2. The angular momentum function $F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})$ for the N = 2 disk with $\tilde{\sigma}_u = \tilde{U} = 0.378$, $\tilde{\Omega}_p = 0.5$ and $\tilde{s} = 0^+$. Both curves report m = 2 functions: top curve refers to the case $\ell = -1$, and displays a strong trailing bias; bottom curve shows the Hermitian-symmetric $\ell = -2$ contributions.

$|\alpha|$, $K(\alpha,m) \simeq (\alpha^2 + m^2)^{-1/2}$. For fixed ℓ and $\tilde{U}$, stationary phase arguments imply that $Q_{\ell m}(\alpha;\tilde{U})$ will decay as $|\alpha|^{-1/2}$ for large $|\alpha|$. On the whole then, the transfer function $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$ will exhibit a gradual, algebraic decay along the diagonal $\alpha = \beta$, a marked exponential decay (in α for fixed β) for $\alpha > \beta$, and a mild exponential decay (at least when $\tilde{s} > 0$) for $\alpha < \beta$.

e) Numerical Procedures

The growing modes of a given cut-out disk correspond to the solutions of the integral equation (3.42). We must resort to numerical methods to find what solutions, if any, this equation admits. An essential element of the method employed to solve this equation is the computation of the response due to an imposed density transform $A(\alpha)$ which is known only at, say, $2n+1$ equally-spaced points centered on $\alpha = 0$. This response is given by the integral on the right-hand side of eqn. (3.42). It is sought at the same $2n+1$ locations at which the imposed density is known. Two major numerical tasks are involved in the determination of the response: (i) the computation of the kernel, $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, on a $(2n+1) \times (2n+1)$ grid in (β,α) space, and (ii) the evaluation of the response integral based on these values of the kernel and those of the imposed density.

The computation of the kernel requires numerical procedures for several major elements:

1) The Kalnajs gravity factor $K(\alpha,m)$

The axisymmetric gravity factor $K(\alpha,0)$ can be effectively computed from an infinite product representation. A recursion relation between $K(\alpha,n)$ and $K(\alpha,n+1)$ is then used to obtain $K(\alpha,m)$ from $K(\alpha,0)$.

2) The orbital parameters $x_{min}(\tilde{U})$, $x_{max}(\tilde{U})$ and $I_n(\tilde{U})$

An iterative scheme based on Newton's method is used to obtain $x_{min}(\tilde{U})$ and $x_{max}(\tilde{U})$, the two solutions of eqn. (2.34). The integrals $I_n(\tilde{U})$ are computed from eqn. (2.39) by means of a quadrature designed to handle the singularities in the denominator.

3) The Fourier coefficients $Q_{\ell m}(\alpha;\tilde{U})$

These are computed from eqn. (3.27) by the midpoint rule. The orbital excursions $X(\psi;\tilde{U})$ and $Y(\psi;\tilde{U})$ are found by a Runge-Kutta integration of the unperturbed orbits.

4) The integral over eccentric velocities $\int_0^\infty d\tilde{U}$

This is computed by means of a Gauss-Laguerre quadrature in the variable $\tilde{U}^2$.

5) The normalization constant $\tilde{C}_a$

This, too, is computed by a Gauss-Laguerre quadrature over eccentric velocities.

6) The sum over the radial harmonics $\sum_{\ell=-\infty}^{\infty}$

The results are compared as additional radial harmonics are added.

Further details of the above procedures are furnished in Appendix C. The numerical procedure for computing the response integral itself,

$$A_{res}(\beta) = \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\ A_{imp}(\alpha)\ d\alpha\ ,$$

consists of suitably weighting the computed elements of the kernel so that

$$A_{res}(\beta_j) \simeq \sum_{k=1}^{2n+1} S_{jk}\ A_{imp}(\alpha_k)\ ,$$

where (S_{jk}) is the matrix consisting of the weighted elements of the kernel and β_j and α_k are the sampling wavenumbers. This approach has the advantage of permitting the response to numerous imposed densities to be found efficiently. The basis for the weighting procedure is an 8-point Lagrangian interpolation of both $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$ and $A_{imp}(\alpha)$. This method is described explicitly in Appendix D.

IV. AXISYMMETRIC (m = 0) RESULTS

Now that the mathematical basis of the stability analysis has been presented, we turn to a discussion of the results. In this chapter the simplest disturbances — the axisymmetric (m = 0) ones — are examined. Our primary objective here is to confirm, in the context of the cut-out models, an earlier but less complete calculation by Toomre of the criterion of axisymmetric stability for the singular disk. In the process we will also obtain the corresponding criteria for the cut-out models. Otherwise, we will pay relatively little attention to the axisymmetric modes of the latter disks, since the emphasis in this thesis is on their non-axisymmetric behavior.

a) Singular Disk

The minimum velocity dispersion required to stabilize the singular disk against exponentially growing disturbances was found by Toomre (1970, unpublished) to be remarkably close to the estimate furnished by local theory [cf. eqn. (2.18)]: He found no difference between the global $\tilde{\sigma}_{u,min}$ and the local estimate to within one percent. In 1973 Toomre repeated his computations in double-precision arithmetic, and found that the actual $\tilde{\sigma}_{u,min}$ does in fact differ from the local estimate, but only by a few parts in 10^4. Since this

unpublished work relates to our analysis of the cut-out disks, we will briefly describe Toomre's calculations.

The distribution function of the singular disk meets the requirements of the demonstration by Kalnajs (1971) that there are no overstable large-scale axisymmetric disturbances. Therefore the frequency of conceivable instabilities can be assumed to be purely imaginary, i.e., $\tilde{\omega} = i\tilde{s}$. Toomre's aim was to find the largest velocity dispersion for which a marginally stable disturbance (one with a vanishingly small growth rate $\tilde{s} = 0^+$) can be supported by the singular disk. For any finite growth rate there will be some radius (of order V_o/s in dimensional units) beyond which the natural time scales of the stars [$\sim 1/\kappa(r)$] are too long for them to participate effectively in the disturbance. In this case, since the singular disk is self-similar, the disturbances themselves can be expected to be likewise. For this intuitive reason, Toomre sought the response of the disk to an imposed marginally stable disturbance of the form

$$\mu_{1,\mathrm{imp}}(r) = \mu_p \, (r/r_o)^{-3/2} \cos[\alpha \, \ell n(r/r_o)] \ . \qquad (4.1)$$

He found that the response density does indeed have the same form as the imposed density and can be expressed as

$$\mu_{1,\mathrm{res}}(r) = R(\tilde{\sigma}_u,\alpha) \quad \mu_{1,\mathrm{imp}}(r) \ ,$$

where $R(\tilde{\sigma}_u,\alpha)$ is a real-valued function, which can be

expressed an an integral over radius and over velocity space. The condition

$$R(\tilde{\sigma}_u,\alpha) = 1$$

implicitly relates the velocity dispersions and wavenumbers for which self-consistent, marginally stable disturbances exist. This condition is analogous to the one relating velocity dispersion and wavenumber in the local theory. Toomre's (1964) Figure 5 displays the form that this relationship takes in that approximation.

When working in the variables r , $\tilde{U}$ and r_H there are several integrable singularities to contend with in the numerical evaluation of $R(\tilde{\sigma}_u,\alpha)$. Still, Toomre was numerically able to compute this function to better than 8 digits. He found that the maximum velocity dispersion for which marginally stable modes exist is

$$\tilde{\sigma}_{u,min} = 0.378070 \ , \tag{4.2}$$

in units of the circular speed V_o . For $\tilde{\sigma}_u \geq \tilde{\sigma}_{u,min}$, then, the singular disk is axisymmetrically stable.

When $\tilde{\sigma}_u = \tilde{\sigma}_{u,min}$ exactly, the marginally stable disturbance has the wavenumber

$$\alpha_m = 3.46062 \ . \tag{4.3}$$

When the locally estimated $\tilde{\sigma}_{u,min}$ is computed to the same number of digits, it turns out to be

$$\tilde{\sigma}_{u,min} = 0.377940 \ .$$

In short, the global result seems to be 1.000344 times the local one — an almost embarrassing agreement!

Although the calculation by Toomre seemed sensible, some questions remained about its assumptions. An obvious concern was that one cannot be certain, without a more general investigation, that the marginally stable axisymmetric disturbances indeed have the form specified by eqn. (4.1). In principle this issue can be resolved by solving the integral equation (3.42) for the singular disk when $m = 0$. However, as we saw in section III.d, that formulation suggests another worry: either the singular disk has no growing modes whatsoever, or else a continuum of them. The $\tilde{s} = 0^+$ limit taken by Toomre is apparently some sort of singular process.

Unfortunately, the numerical difficulties caused by the principal value character of the kernel for the singular disk have so far prevented us from developing any reliable procedures for obtaining solutions of eqn. (3.42). Hence we can neither support nor contradict Toomre's estimate directly.

b) Cut-out Disks

On the other hand, the calculations for the cut-out disks are manageable and we can use them to infer the singular disk behavior. As we have frequently noted, any cut-out disk is virtually indistinguishable from the singular one at very large radii. For the present purposes the behavior of the local stability parameter $Q(r)$, defined by eqn. (2.67), is instructive. For the singular disk it is independent of r; for a cut-out disk it has an infinitely large value at the center, and decreases monotonically as r increases, tending axymptotically to the singular disk value as $r \to \infty$. This implies that the cut-out models are increasingly stable as r decreases and that their susceptibility to axisymmetric disturbances is governed by their properties at large radii. Therefore we expect that the axisymmetric stability boundary for any cut-out disk will be exactly the same as that for the singular disk.

For axisymmetric disturbances the integral equation governing the modes is understandably easier to handle than it is for the $m \neq 0$ cases. As was done for the singular disk, the search for growing modes here can be confined to purely imaginary frequencies. The integral equation (3.42) can then be manipulated into an equivalent one with an Hermitian kernel:

$$B(\beta) = \int_{-\infty}^{\infty} T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s})\, B(\alpha)\, d\alpha \,, \qquad (4.4)$$

where

$$B(\alpha) = \sqrt{K(\alpha,0)}\, A(\alpha) \tag{4.5}$$

$$T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s}) = \sqrt{K(\beta,0)\, K(\alpha,0)}\, \tilde{C}_a \sum_{\ell=1}^{\infty} \int_0^{\infty} I_0(\tilde{U})\, e^{-\tilde{U}/2\tilde{\sigma}_u^2}$$

$$\times\, Q_{\ell 0}(\alpha;\tilde{U})\, \bar{Q}_{\ell 0}(\beta;\tilde{U})\, E_\ell(\beta-\alpha;\tilde{U},\tilde{s})\, \tilde{U}\, d\tilde{U} \tag{4.6}$$

and

$$E_\ell(\eta;\tilde{U},\tilde{s}) = F_{\ell 0}(\eta;\tilde{U},i\tilde{s}) + F_{-\ell 0}(\eta;\tilde{U},i\tilde{s}) \ . \tag{4.7}$$

The remaining quantities have all been defined in Chapters II and III. In obtaining this result the property

$$Q_{-\ell 0}(\alpha;\tilde{U}) = Q_{\ell 0}(\alpha;\tilde{U})$$

has been used. This is demonstrated in Appendix C, eqn. (C.3). For any cut-out disk,

$$F_{-\ell 0}(\eta;\tilde{U},i\tilde{s}) = \bar{F}_{\ell 0}(-\eta;\tilde{U},i\tilde{s})$$

[see eqn. (3.41)] and therefore

$$E_\ell(-\eta;\tilde{U},\tilde{s}) = \bar{E}_\ell(\eta;\tilde{U},\tilde{s}) \ .$$

Hence, recalling that $K(\alpha,0)$ is real, we can easily see that the new axisymmetric kernel $T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s})$ is indeed Hermitian. We note here, for future reference, the additional property

$$T_0(-\alpha,-\beta;\tilde{\sigma}_u,\tilde{s}) = T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s}) \ . \tag{4.8}$$

The integral equation (4.4) implicitly determines both the characteristic frequencies and the shapes of the growing axisymmetric modes of a cut-out disk. This equation presents an eigenvalue problem for the growth rates $\tilde{s}$ and the symmetrized density transforms $B(\alpha)$ of the modes. [Henceforth we will refer to $B(\alpha)$ as simply a density transform. It is to be understood that $B(\alpha)$ is related to the actual density transform $A(\alpha)$ by eqn. (4.5).] However, this is not a conventional eigenvalue problem, since the characteristic growth rates $\tilde{s}$ enter into the kernel in a distinctly nonlinear fashion.

Consequently, a useful approach is to consider a more general problem than that of finding the growing modes: for any growth rate $\tilde{s}$ what are the imposed densities, $B_{imp}(\alpha)$, for which the response density, $B_{res}(\alpha)$, has the same shape, differing only by a constant real multiple? We denote this real multiple by $\lambda(\tilde{\sigma}_u,\tilde{s})$. Hence we are looking for those densities for which

$$B_{res}(\alpha) = \lambda(\tilde{\sigma}_u,\tilde{s})\ B_{imp}(\alpha) \ .$$

This more general, but artificial, problem is described by the integral equation

$$\lambda(\tilde{\sigma}_u,\tilde{s})\ B(\beta) = \int_{-\infty}^{\infty} T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s})\ B(\alpha)\ d\alpha \ . \qquad (4.9)$$

This now presents a standard linear eigenvalue problem for

the artificial, purely mathematical eigenvalues $\lambda(\tilde{\sigma}_u,\tilde{s})$ and their corresponding eigenfunctions $B(\alpha)$. To avoid confusion with the physical eigenvalue problem of eqn. (4.4), we will refer to the solutions $\lambda(\tilde{\sigma}_u,\tilde{s})$ of eqn. (4.9) as mathematical or artificial eigenvalues and to the solutions $\tilde{s}$ of eqn. (4.4) as characteristic growth rates.

Since the kernel is Hermitian, all the mathematical eigenvalues of eqn. (4.9) are necessarily real. We expect there to be countably many of them, $\lambda_1(\tilde{\sigma}_u,\tilde{s})$, $\lambda_2(\tilde{\sigma}_u,\tilde{s})$, $\cdots$, which can be ordered so that $|\lambda_1(\tilde{\sigma}_u,\tilde{s})| \geq |\lambda_2(\tilde{\sigma}_u,\tilde{s})| \geq \cdots$, with $|\lambda_k(\tilde{\sigma}_u,\tilde{s})| \to 0$ as $k \to \infty$. We do not attempt to justify this expectation but merely note that the numerical results show no indication of a continuum of mathematical eigenvalues.

Clearly the issue of whether any specific growth rate is a characteristic growth rate reduces to the more manageable question of whether any of the corresponding mathematical eigenvalues are unity. The solutions of eqn. (4.9) can be found by straightforward numerical methods. We use the same basic approach in approximating the response integral in eqn. (4.9) as we use in computing the response integral in eqn. (3.42) for arbitrary non-axisymmetric disturbances. This approach was surveyed in section III.e and is detailed in Appendices C and D.

The general numerical procedures can be streamlined in the axisymmetric case. The Hermitian property of the kernel combined with eqn. (4.8) yield

$$T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s}) = \bar{T}_0(-\beta,-\alpha;\tilde{\sigma}_u,\tilde{s}) \ .$$

It is then easy to show that if $B(\alpha)$ is an eigenfunction of eqn. (4.9) with eigenvalue $\lambda(\tilde{\sigma}_u,\tilde{s})$ then $\bar{B}(-\alpha)$ is also an eigenfunction with the same eigenvalue. (Numerically, we find that the mathematical eigenvalues are in fact degenerate.) Although both

$$B_+(\alpha) = B(\alpha) + \bar{B}(-\alpha)$$

and

$$B_-(\alpha) = B(\alpha) - \bar{B}(-\alpha)$$

are mathematically valid eigenfunctions when $B(\alpha) \neq \bar{B}(-\alpha)$, only $B_+(\alpha)$ is physically relevant. The reason is that the surface density corresponding to $B_-(\alpha)$ is identically zero.[1] Therefore, we may assume that

$$B(\alpha) = \bar{B}(-\alpha) \ , \tag{4.10}$$

and reduce the linear eigenvalue problem to

---

[1] Recall that it is only the real part of the Fourier integral that is physically significant.

$$\lambda(\tilde{\sigma}_u,\tilde{s})\ B(\beta) = \int_0^\infty T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s})\ B(\alpha)\ d\alpha + \int_0^\infty \bar{T}_0(-\beta,\alpha;\tilde{\sigma}_u,\tilde{s})\ \bar{B}(\alpha)\ d\alpha \qquad (\beta \geq 0)\ . \quad (4.11)$$

If this is broken up into its real and imaginary parts and the interpolation scheme described in Appendix D is used to evaluate the integrals, an approximation of the form

$$\lambda(\tilde{\sigma}_u,\tilde{s})\ \underline{B}(\beta_j) = \sum_{k=1}^{n+1} \underline{T}_{jk}\ \underline{B}(\alpha_k) \qquad j = 1,2,\cdots n+1 \quad (4.12)$$

is obtained. Both β_j and α_k are elements of the set of sampling wavenumbers consisting of the $n+1$ equally spaced points between $\alpha_1 = 0$ and $\alpha_{n+1} = n(\Delta\alpha)$. The distance between successive points is represented by $\Delta\alpha$. The two-dimensional vector $\underline{B}(\beta_j)$ consists of the real and imaginary parts of $B(\beta_j)$ and $\underline{T}_{jk}$ is a real 2×2 matrix whose entries are weighted combinations of $T_0(\beta_j,\alpha_k;\tilde{\sigma}_u,\tilde{s})$ and $\bar{T}_0(-\beta_j,\alpha_k;\tilde{\sigma}_u,\tilde{s})$.

The mathematical eigenvalues are approximated by those of the $2(n+1)$-dimensional matrix implied by eqn. (4.12). We have employed the eigenvalue package EISPACK [see Klema, Garbow and Moler (1973)] to compute these numbers. Unfortunately, the interpolation scheme used to obtain the elements $\underline{T}_{jk}$ introduces an artificial asymmetry into the system given by eqn. (4.12). We have found that the same eigenvalues (to better than 1 part in 10^4) are obtained when the system is artificially symmetrized by averaging the appropriate

off-diagonal components. The eigenvalues of the real symmetric matrix that results from this averaging process can be computed much more efficiently than can those of the original, un-symmetric matrix. This is especially true when only a few of the largest eigenvalues are desired.

Another useful consideration is that $F_{\ell 0}(\eta;\tilde{U},i\tilde{s})$ is sharply peaked near $\eta = 0$, decaying as $e^{-\pi|\eta|/N}$ when $N \geq 2$ and as $e^{-\pi|\eta|/2}$ when $N = 1$. Hence the kernel $T_0(\beta,\alpha;\tilde{\sigma}_u,\tilde{s})$ itself decays rapidly away from the diagonal $\beta = \alpha$. Therefore, when the eigenfunction $B(\alpha)$ is sharply peaked near some $\alpha \gtrsim 3$, as happens when $\tilde{s}$ is small, the second integral in eqn.(4.11) can be ignored without significantly altering the computed eigenvalues.

We have found that when the above simplifications are employed, our numerical procedures are capable of producing efficiently the mathematical eigenvalues to an accuracy of one part in ten thousand.[2] Unfortunately, we have not been able to obtain the axisymmetric stability criterion for the cut-out disks to a comparable level of accuracy. The problem is that although $\lambda(\tilde{\sigma}_u,\tilde{s})$ can be computed quite accurately for any given $\tilde{\sigma}_u$ and $\tilde{s} > 0$, the stability boundary can only be approached in an $\tilde{s} \to 0^+$ limit. This limiting process is not an easy matter. The source of the difficulty is

---

[2] In Chapter V some specific examples of the numerical convergence to the mathematical eigenvalues are given, albeit for the general non-axisymmetric routines. The streamlined axisymmetric procedures converge in a similar fashion.

the term $e^{-i(\beta-\alpha)\,\ell n\,\tilde{s}}$ in the kernel of eqn. (4.4). It not only prevents us from computing $\lambda(\tilde{\sigma}_u,0^+)$ directly, but also suggests that, as a function of $\tilde{s}$, $\lambda(\tilde{\sigma}_u,\tilde{s})$ has a singularity at $\tilde{s} = 0$. In the absence of any firm knowledge of the nature of this singularity, no reliable extrapolation from a series of $\tilde{s} > 0$ results to the $\tilde{s} = 0^+$ value can be performed. It is mainly for this reason (we believe) that the present calculation of the axisymmetric stability criterion for one of the cut-out disks can support Toomre's result for the singular disk only to within a few parts in 10^3.

Our efforts to locate the axisymmetric stability boundary have focused on the disk with cut-out index $N = 2$. Two different strategies have been used. The first one leaned heavily on the singular disk result. The aim was to check whether, when $\tilde{\sigma}_u$ has the value 0.378070, the largest mathematical eigenvalue tends, as expected, to 1 as $\tilde{s} \to 0^+$. The following table contains the results.

$\tilde{s}$	$\lambda(0.378070,\tilde{s})$
10^{-1}	0.7558
10^{-2}	0.9421
10^{-3}	0.9796
10^{-4}	0.9903
10^{-5}	0.9944
10^{-6}	0.9963
10^{-7}	0.9974

As one can see, the limiting value (presumably $\lambda = 1$) is approached very slowly. (For the cut-out index $N = 3$ model, the approach to the limit is a little, but not much, faster.)

For the reasons given above, we have not attempted to extrapolate these numbers. By themselves they permit us to conclude only that the stability criteria for the singular disk and this particular cut-out model agree to within a few tenths of one percent.

A couple of interesting features are displayed by the mathematical eigenvalues and eigenfunctions in the $\tilde{s} \to 0^+$ limit. For one, the second and third largest mathematical eigenvalues become larger and larger fractions of the dominant (or largest) eigenvalue as $\tilde{s}$ decreases. Apparently, they too tend to $\lambda = 1$. Quite possibly this is also the case with the fourth, fifth, etc. eigenvalues, but we have not checked, mainly because the smaller eigenvalues are not computed as accurately as the larger ones. This behavior may well be linked to the existence of a continuum of modes in the singular disk for $\tilde{\sigma}_u = \tilde{\sigma}_{u,min}$. Another point is that by examining the eigenfunction $B(\alpha)$ we find that for very small $\tilde{s}$, the peak of the density transform occurs between $\alpha = 3.48$ and $\alpha = 3.49$. For the singular disk, $B(\alpha)$ should be simply a delta-function centered on $\alpha_m = 3.46$ [see eqns. (4.1) and (4.3)]. These features give us further assurance that for very small $\tilde{s}$ the cut-out models are mimicing the singular disk.

The second strategy that we used for locating the stability boundary was less biased toward the singular disk result. The objective was to determine how $\tilde{s}_*$, the growth rate of

the fastest growing mode, varied with $\tilde{\sigma}_u$. This was achieved by picking a value of $\tilde{\sigma}_u$ (< 0.378070) and then iterating in the growth rate $\tilde{s}$ until the largest mathematical eigenvalue equalled one, i.e., $\lambda(\tilde{\sigma}_u, \tilde{s}_*) = 1$. The results are summarized below.

$\tilde{\sigma}$	Q	$\tilde{s}_*$
0.3030	0.8014	0.1007
0.3215	0.8504	0.0619
0.3407	0.9012	0.0298
0.3591	0.9498	0.00829
0.3686	0.9750	0.00176
0.3733	0.9874	0.000247
0.3756	0.9935	0.0000212
0.3768	0.9966	0.00000057

The second column contains the global stability parameter Q [see eqn. (2.68)] where the singular disk result was used as the reference value, i.e.,

$$Q = \tilde{\sigma}_u / 0.378070 \ . \tag{4.13}$$

Ideally, we would like to extrapolate $\tilde{s}_*$ to 0^+ and find the corresponding value of $\tilde{\sigma}_u$. This, of course, would give us the precise minimum velocity dispersion required to suppress all axisymmetric instabilities. Again, though, because of the uncertainties of any such extrapolation we have left things as they stand, coming about as close to the singular disk value here as we did by the other approach.

To summarize, both of the above attempts to locate the axisymmetric stability boundary of this particular cut-out

disk suggest that $\tilde{\sigma}_{u,min} \simeq 0.377$, with some uncertainty in the last digit. Almost certainly the source of this inaccuracy is an irregular behavior of the mathematical eigenvalues as $\tilde{s} \to 0^+$.

Compared with these cut-out disk results, the stability criterion of the singular disk was obtained to twice the accuracy. One reason this was possible was that in Toomre's calculation, the $\tilde{s} \to 0^+$ limit was taken directly, and hence the difficulties we encountered in extrapolating $\tilde{s} > 0$ results were avoided. On the other hand, there was some concern about the assumptions made in the singular disk calculation. One worry centered on the likelihood that exponentially growing modes exist in the singular disk for only one value of $\tilde{\sigma}_u$; this left the $\tilde{s} \to 0^+$ limit somewhat suspect. Furthermore, a specific form was assumed for the marginally stable disturbances. The present, cruder results for the $N = 2$ cut-out disk, though, appear to dispel these suspicions. Since there are solid grounds for expecting the cut-out models to have the same value of $\tilde{\sigma}_{u,min}$ as the singular model, the fact that essentially the same result is obtained for the cut-out disk seems to indicate that the singular disk was indeed properly treated. Thus, we henceforth take the more accurate singular disk result to be the actual $\tilde{\sigma}_{u,min}$ for all the cut-out models; in particular, the axisymmetric stability parameter Q will be based upon $\tilde{\sigma}_{u,min} = 0.378070$.

The last question we consider in this chapter is the shape of the unstable axisymmetric disturbances in a cut-out model. We again use the $N = 2$ disk, this time choosing $\tilde{\sigma}_u = 0.303$ $(Q = 0.801)$. The growth rates of the two most unstable modes are $\tilde{s}_1 = 0.101$ and $\tilde{s}_2 = 0.036$. Their surface densities are illustrated in Figure 3. We have chosen to plot $\tilde{r}\,\tilde{\mu}_1(\tilde{r})$ vs. $\ell n\,\tilde{r}$ rather than $\tilde{\mu}_1(\tilde{r})$ vs. r, where $\tilde{r} = r/r_o$ is the radial coordinate in dimensionless units. This choice enables a direct comparison to be to be made between the form assumed by Toomre in eqn. (4.1) for the marginally stable modes of the singular disk and the form actually taken by the modes in this particular cut-out disk. The former disturbances extend over the entire singular disk whereas the two modes displayed in Figure 3 are effectively confined to an annular region between $\tilde{r} \sim 1$ and $\tilde{r} \sim 1/\tilde{s}$. Both the inner and outer "barriers" have simple explanations. Since most of the material at radial distances less than 1 in a cut-out disk is rigid, very little activity can occur there. At the other extreme, the obstacle to the disturbance is the same one that exists for modes of finite growth rate even in the singular disk — stars beyond $\tilde{r} \sim 1/\tilde{s}$ simply respond too slowly for them to participate in the perturbation. Within the annular region, however, the two modes displayed in Figure 3 are basically sinusoidal. Compared to the more slowly growing disturbance the first mode has a somewhat larger wavelength

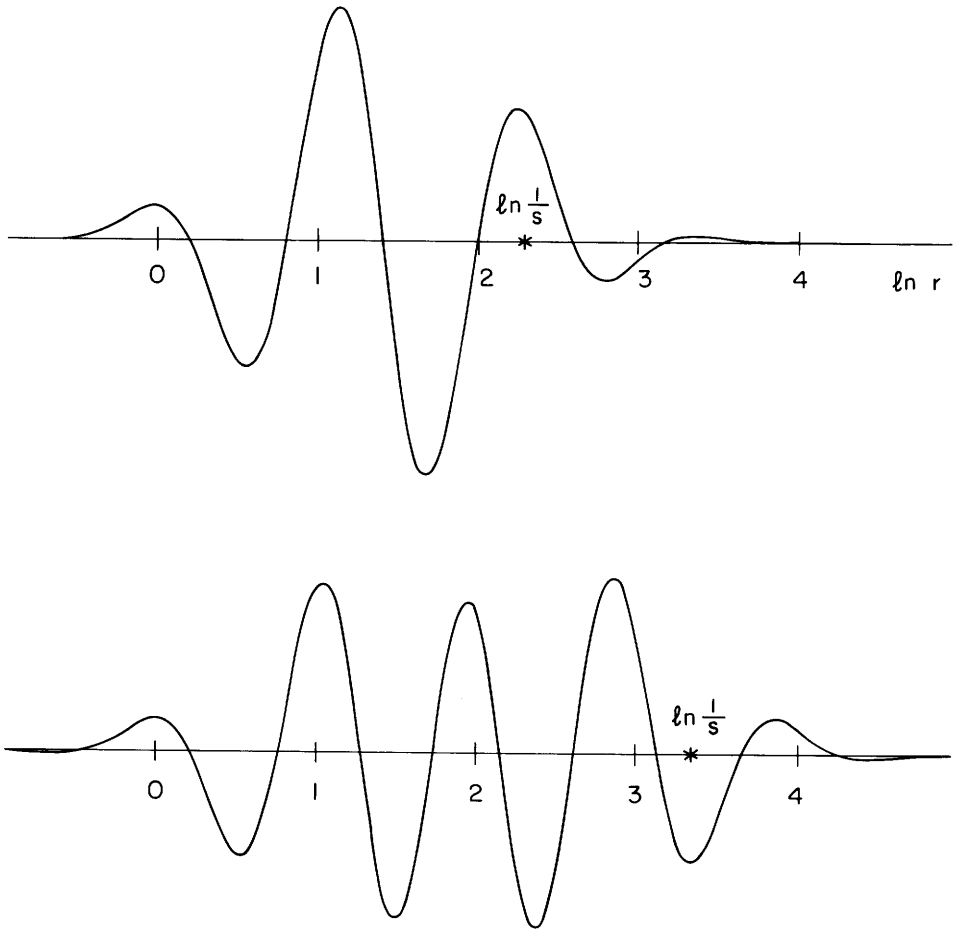

Figure 3. The density fields, shown as $\tilde{r}^{3/2}\,\tilde{\mu}(\tilde{r})$ vs. $\ell n\,\tilde{r}$, of the two most unstable $m = 0$ modes of the $N = 2$ disk with $\tilde{\sigma}_u = 0.303$. The growth rates are $\tilde{s} = 0.101$ (top) and $\tilde{s} = 0.036$ (bottom). Asterisks mark the radii at which $\tilde{r} = 1/\tilde{s}$.

(in $\ell n\ \tilde{r}$) and fewer oscillations in the active region — apparently three fewer "half-wavelengths". In terms of the radius $\tilde{r}$ itself, of course, these modes look somewhat different. The wavelength in physical space is much tighter at smaller radii than at larger ones. Moreover, since the magnitude of the disturbance is scaled by a factor of $\tilde{r}^{-3/2}$ compared to that in Figure 3, the size of the oscillations is much larger at small radii than they appear in that diagram. All in all, however, these instabilities are rather ordinary "ring modes". The fact that there is nothing particularly remarkable about the shapes of these disturbances compared to those intuitively expected by Toomre lends additional support to his singular disk result.

V. BISYMMETRIC (m = 2) RESULTS

Of the various non-axisymmetric disturbances, presumably the most interesting are the two-armed ones. Many of the prominent structures observed in actual galaxies have a distinct bisymmetric shape. In this chapter we will determine the circumstances under which $m = 2$ instabilities are possible in several of the cut-out disks and we will also examine the characteristics of some of those growing disturbances. First, however, we must discuss the strategy we used for solving the crucial integral equation (3.42), in particular how it was possible to decide whether there are any growing disturbances in a given model, and if so, what the characteristic frequencies and shapes are.

a) Use of an Artificial Eigenvalue Problem

In broad terms the strategy is rather similar to that employed in the search for axisymmetric modes. With each frequency $\tilde{\omega}$ we associate the artificial, linear eigenvalue problem

$$\lambda(\sigma_u,\tilde{\omega})\ A(\beta) = \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\ A(\alpha)\ d\alpha\ , \qquad (5.1)$$

and seek those frequencies for which some mathematical

eigenvalue $\lambda(\tilde{\sigma}_u,\tilde{\omega})$ is unity. Compared to the case with the $m = 0$ modes, however, there are additional complications. We must now consider disturbances which not only grow ($\tilde{s} > 0$) but also rotate ($\tilde{\Omega}_p \neq 0$). Moreover, the kernel is no longer Hermitian — not even for $\tilde{s} = 0^+$. Consequently, the mathematical eigenvalues now have both real and imaginary parts. Hence the search for those characteristic frequencies which permit some mathematical eigenvalue to equal unity must take place in two dimensions instead of only one as in the axisymmetric case.

We again use a suitable finite-dimensional matrix equation to approximate the integral equation, namely,

$$\lambda(\tilde{\sigma}_u,\tilde{\omega})\ A(\beta_j) = \sum_{k=1}^{2n+1} S_{jk}\ A(\alpha_k) \qquad (j=1,2,\ldots,2n+1) \qquad (5.2)$$

[see section III.e and Appendix D]. Since (S_{jk}) is not Hermitian, it is now more difficult to compute the eigenvalues. Therefore, we generally rely on an iterative "power method" procedure to find the largest one or two eigenvalues. This method is described in Appendix E.

While the remaining remarks in this chapter are directed specifically to the manner in which we determine whether a given cut-out disk is stable to bisymmetric disturbances, they pertain equally well to the approach used for the other non-axisymmetric harmonics, like $m = 1$ or $m = 3$.

The search for the possible growing modes is guided by the manner in which the largest mathematical eigenvalue (in absolute sense) depends upon the frequency. The two components of the frequency, the pattern speed $\tilde{\Omega}_p$ and the growth rate $\tilde{s}$, play different roles in determining the response of the disk. The pattern speed determines the locations of the corotation and Lindblad resonances. For bisymmetric perturbations the inner Lindblad resonance occurs at $\tilde{r}_{ILR} = 0.293/\tilde{\Omega}_p$, the corotation resonance at $\tilde{r}_{CR} = 1/\tilde{\Omega}_p$, and the outer Lindblad resonance at $\tilde{r}_{OLR} = 1.707/\tilde{\Omega}_p$, at least for particles in nearly circular orbits. In the asymptotic density wave theory of Lin and Shu, the density disturbance is confined to the "principal range" between the two Lindblad resonances. The nature of the response of the disk can be expected to depend sharply upon where the principal range lies with respect to the "edge" in the active stellar component of the cut-out disks. This edge occurs in the vicinity of $\tilde{r} = 1$. Inside this radius the majority of the stars are rigid, whereas outside this point the majority of the stars are mobile. Three ranges of the pattern speed can therefore be distinguished: (i) $\tilde{\Omega}_p \geq 2$: The disturbance lies inside the edge. Since the active density is so low near the center, the response should be rather small. (ii) $2 \geq \tilde{\Omega}_p \geq 0.2$: The disturbance overlaps the edge. The strong gradient in the active density should strongly influence the response. (iii) $\tilde{\Omega}_p \leq 0.2$: The disturbance lies well beyond the edge.

As $\tilde{\Omega}_p \to 0$, the response should approach that of the singular disk. The pattern speed, therefore, can be expected to influence strongly the character of the response. This is measured in large part by the complex phase of the mathematical eigenvalue. The growth rate, on the other hand, affects primarily the strength of the response. It becomes increasingly difficult for a disk to respond effectively to disturbances with larger and larger growth rates. Hence, the magnitudes of the mathematical eigenvalues can be expected to decrease as the growth rate increases.

b) Artificial Eigenvalues of the N = 2 Disk

We now focus on the disk with cut-out index $N = 2$, taking a detailed look at its mathematical eigenvalues. Our starting point is a system which is marginally stable to axisymmetric disturbances. As demonstrated in the previous chapter, a disk with a velocity dispersion of $\tilde{\sigma}_u = 0.378$ will have a global stability parameter Q of unity.

The objective here is to spot general trends. So, for the moment, rough calculations of the mathematical eigenvalues will suffice. The accuracy of the results is controlled mainly by three factors: (i) the fineness of the spacing in wavenumber used in the approximation of the integral equation; (ii) the number of points used in the Gauss-Laguerre integration over the eccentric velocity; and (iii) the number of radial harmonics included. Eigenvalues accurate to 3

or even 4 digits can be obtained with rather coarse wavenumber spacings and only a few eccentric velocity points. (However, this degree of accuracy also requires that a sizeable number of radial harmonics be included, as will be illustrated below.) The convergence to the largest mathematical eigenvalue when $\tilde{\Omega}_p = 0.60$ and $\tilde{s} = 0.20$ is displayed in the table below, where the effect of improving all facets of the approximation is shown.

i)

$\Delta\alpha$	λ_R	λ_I
1.00	0.413338	-0.044536
0.75	0.413807	-0.043120
0.60	0.413921	-0.042928
0.50	0.413957	-0.042897
0.40	0.413974	-0.042892
0.30	0.413980	-0.042894

ii)

$n_{\tilde{U}}$	λ_R	λ_I
4	0.413918	-0.042928
5	0.413950	-0.042903
6	0.413957	-0.042897
7	0.413958	-0.042897
8	0.413958	-0.042897

iii)

ℓ_n	ℓ_p	λ_R	λ_I
-3	6	0.462612	-0.008834
-4	8	0.454289	-0.014609
-5	10	0.452706	-0.015926
-6	12	0.452413	-0.016173
-7	14	0.452405	-0.016196
-8	16	0.452451	-0.016143
-9	18	0.452529	-0.016104
-10	20	0.452579	-0.016068
-11	22	0.452615	-0.016041

The spacing in wavenumber of the approximation to the integral equation is denoted by $\Delta\alpha$. The number of Gauss-Laguerre

points used in the eccentric velocity quadrature is denoted by $n_{\tilde{U}}$. The sum over the radial harmonics extends from $\ell = \ell_n$ to $\ell = \ell_p$. The real and imaginary parts of the eigenvalue have the obvious notation. In each part of this table only one parameter is varied. The nominal values are $\Delta\alpha = 0.50$ over the range $[-12,12]$ (this range is already sufficiently large to obtain 6 digit accuracy in the eigenvalue), $n_{\tilde{U}} = 6$, $\ell_n = -2$ and $\ell_p = 4$. Part (i) shows the convergence as $\Delta\alpha$ is decreased, part (ii) as $n_{\tilde{U}}$ is increased and part (iii) as more radial harmonics are included. [In part (iii) the last three radial harmonics have been given twice their normal weights, for reasons explained in Appendix C. This accounts for the large difference between the $\ell = -2$ to 4 results and the $\ell = -3$ to 6 results.]

For some purposes we settle for results in which only a few radial harmonics are incorporated, typically from $\ell = -2$ to 4 . As the table suggests, this leads to eigenvalues that are individually in error by about 10% . The important consideration is that the relative behavior of the eigenvalues with the frequency is indicated by these crude calculations just as well as by more accurate calculations. Naturally, when precise results are desired, such as the value of the characteristic frequency of a mode, an adequate number of radial harmonics are included.

There are several considerations, besides simple trial and error, that help us to decide how elaborate to make the

numerical approximation in any specific situation. The use of our finite approximation to the artificial eigenvalue problem requires that we limit the range of wavenumbers that are included — we must truncate the infinite range $\alpha \in [-\infty,\infty]$. Just how great this extent needs to be for reliable results depends on several factors. When the velocity dispersion is small, the shorter length scales are more prominent than when the dispersion is large. This suggests that the size of the wavenumber region that we retain must become larger as the velocity dispersion increases. The extent of this region is also influenced by the size of the pattern speed. Ordinarily, a disturbance is largely confined to the portion of the disk between the Lindblad resonances. When $\tilde{\Omega}_p$ is large, however, the hole in the center of a cut-out disk presents a significant, additional inner barrier to the disturbance. So, a perturbation with a small pattern speed will often have a larger width in $\ell n\ \tilde{r}$ than one with a large pattern speed. Since the width of a Fourier transform is roughly inversely proportional to the width of the function in physical space, a broader region of wavenumber space is required to represent adequately the density transform when the pattern speed is large than when it is small. The requirement for a wider wavenumber region for smaller velocity dispersions is somewhat ameliorated by an accompanying reduction in the influence of the higher radial harmonics. This occurs because these radial harmonics are really needed

only for the more eccentric orbits, which are less prevalent for smaller velocity dispersions.

Using coarse results in which only a few radial harmonics are kept, we now proceed to trace the behavior of the dominant eigenvalue of the disk of cut-out index $N = 2$ when its axisymmetric stability parameter Q is unity. Our first step is to pick a growth rate and to find out how the largest eigenvalue varies with the pattern speed. We start with the growth rate $\tilde{s} = 0.40$ and look at the response for the pattern speeds $\tilde{\Omega}_p = 1.00, 0.90, 0.80, 0.70, 0.60, 0.50,$ and 0.40. The results are plotted in Figure 4. For this growth rate they are the points connected by the "left-most" curve in that diagram. Clearly, the eigenvalue will lie on the positive real axis when $\tilde{\Omega}_p \simeq 0.76$. However, it will then have the disappointingly low value of 0.29 — disappointing when the objective is to find a growing mode, since the value $\lambda = 1$ is required for that. Note that none of the eigenvalues along this curve has a modulus even approaching unity. So a mode cannot be found within this range of pattern speeds by examining the smaller eigenvalues. The only chance for a growing mode lies in a growth rate other than $\tilde{s} = 0.40$.

Even there, though, strong grounds for pessimism exist. The use of a finite, numerical approximation to the kernel of the integral equation (5.1) guarantees that the eigenvalues will be analytic in the frequency. Hence the eigenvalue $\lambda(\tilde{\sigma}_u, \tilde{\omega})$ that is computed gives a conformal mapping from the

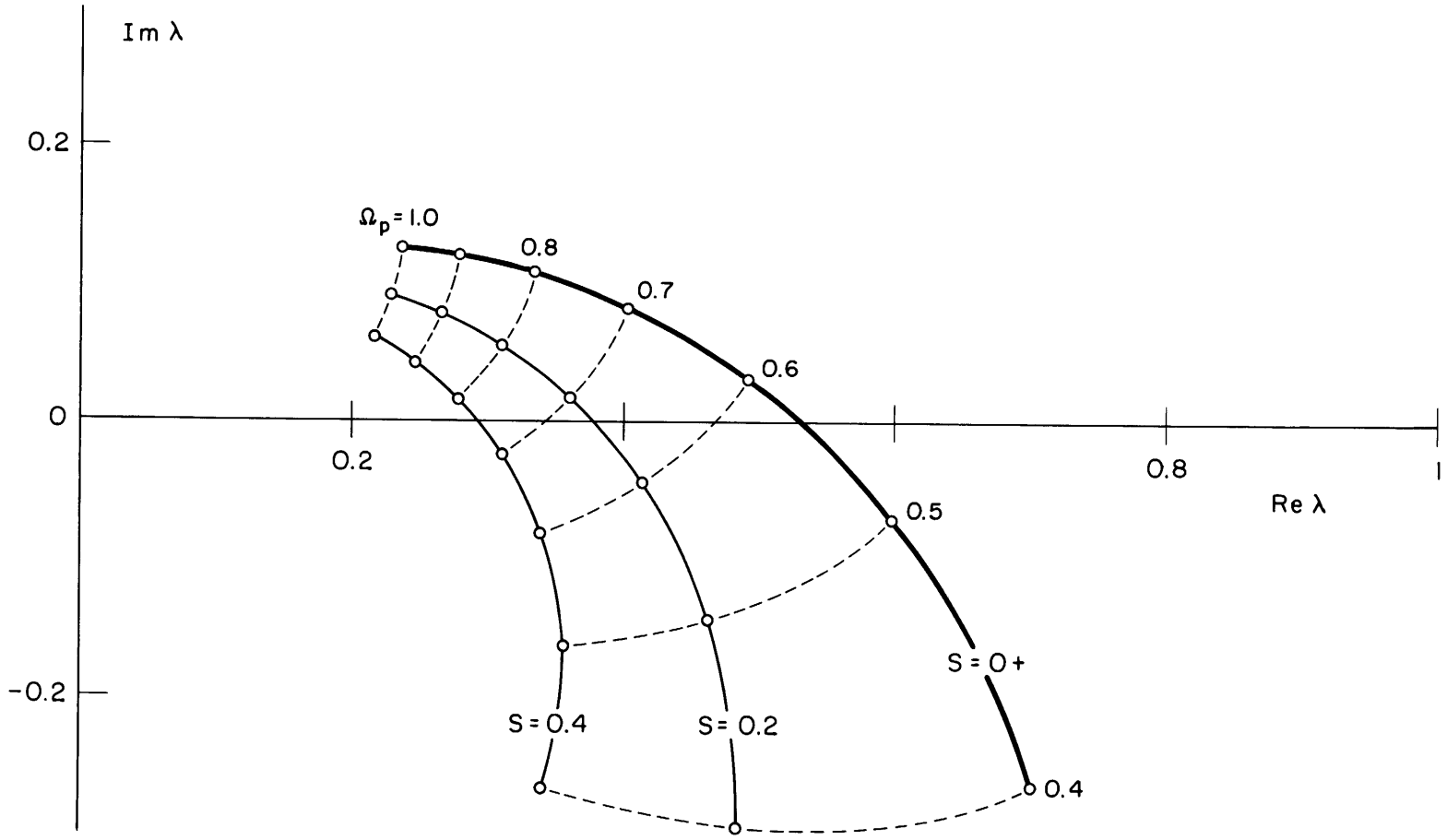

Figure 4. Dependence of largest $m = 2$ eigenvalue, upon the assumed growth rate and pattern speed, for the $N = 2$ disk with velocity dispersion $\tilde{\sigma}_u = 0.378$. Solid curves connect points for the same growth rate $\tilde{s}$; dashed curves connect points for the same pattern speed $\tilde{\Omega}_p$.

upper half of the complex $\tilde{\omega}$-plane into the complex λ-plane. The eigenvalue curve for $\tilde{s} = 0.40$ in Figure 4 is the image of the segment of the line $Im\{\tilde{\omega}\} = 0.40$ between $Re\{\tilde{\omega}\} = 2.00$ and $Re\{\tilde{\omega}\} = 0.80$.[1] Since conformal mappings preserve the orientation of curves, the eigenvalues for growth rates smaller than $\tilde{s} = 0.40$ will lie to the left of the eigenvalue curve for $\tilde{s} = 0.40$ as it is traced in the direction of decreasing pattern speed. In itself, this is a hopeful sign, for at least the eigenvalues for smaller growth rates will lie closer to 1, consistent with the expectation that $|\lambda(\tilde{\sigma}_u, \tilde{\omega})|$ will be larger for smaller growth rates. Unfortunately, analytic mappings also preserve the relative lengths of curves passing through a common point. Merely by examining the distance between the points for $\tilde{\Omega}_p = 0.60$ and $\tilde{\Omega}_p = 0.50$ along the $\tilde{s} = 0.40$ curve in Figure 4, we can estimate that the eigenvalue curve for $\tilde{s} = 0^+$ — which bounds the region of growing disturbances — will cross the real axis in the vicinity of $Re\{\lambda\} = 1/2$. Thus, even the fragmentary evidence of the $\tilde{s} = 0.40$ results suggests that no growing, bisymmetric modes can be supported by the $N = 2$ disk when $\tilde{\sigma}_u = 0.378$.

This suggestion is confirmed by the specific results for the smaller growth rates $\tilde{s} = 0.20$ and $\tilde{s} = 0^+$ included in Figure 4. The latter results represent the limit of

---

[1] Recall that $Re\{\tilde{\omega}\} = m\tilde{\Omega}_p$, so that for $m = 2$, when $\tilde{\Omega}_p = 1.00$, then $Re\{\tilde{\omega}\} = 2.00$.

vanishingly small growth. These eigenvalues for the smaller growth rates lie more or less where expected. From the standpoint of finding possible growing modes, the most significant feature of the results in Figure 4 is that the eigenvalue curve for the limiting case $\tilde{s} = 0^+$ crosses the positive Re{λ}-axis at $\lambda = 0.529$ (when $\tilde{\Omega}_p = 0.562$). Even allowing a 10 percent uncertainty due to neglect of the higher radial harmonics, this limiting eigenvalue curve evidently comes nowhere near the value $\lambda = 1$ required for a bona-fide mode.

The clear implication is that the $N = 2$ cut-out disk, unlike most other galaxy models studied, is not especially prone to $m = 2$ instabilities. This result was so unexpected that, when we first obtained it, we suspected some serious blunder. Perhaps only some numerical constant had been left out inadvertently — after all, if so much as a factor of 2 were missing, our whole conclusion would change, for then the $\tilde{s} = 0^+$ eigenvalue curve would cross the real axis beyond $\lambda = 1$. Consequently, we conducted a number of tests of our procedures.

We first made the rather elementary test of comparing for small $\tilde{\sigma}_u$ the limiting form of the hot disk kernel, $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, calculated in Chapter III on the basis of stellar dynamics, with the strictly cold disk ($\tilde{\sigma}_u = 0$) kernel obtained independently from a fluid description. When $\tilde{\sigma}_u$ had the value 0.01 the hot disk result agreed to several

digits with the cold disk answer. This agreement already appeared to preclude any omission of overall numerical factors.

Next, in the particular case $m = 0$, we compared the kernel computed for a "hot" $\tilde{\sigma}_u = 0.400$ disk by the general non-axisymmetric procedures with the same kernel produced by the specialized axisymmetric routines discussed in Chapter IV. Here again the results were reassuringly close.

The clinching evidence, though, was the verification of our results by explicit orbit integrations. These tests are described in Appendix F. The most basic of these checks involved the response of a cut-out disk to an elementary forcing at a single wavenumber [see eqn. (3.16)]: at several radii the density response implied by the transfer function $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$ is calculated and compared with the density obtained independently from a tedious orbit integration. The earliest such check on the $N = 2$, $m = 2$ results showed agreement to better that 1% when the forcing wavenumber was $\alpha = 5$, the pattern speed $\tilde{\Omega}_p = 0.25$ and the growth rate $\tilde{s} = 0.40$. At the same time, similar checks on the $N = 1$, $m = 2$ results showed even better agreement for both $\alpha = 5$ and $\alpha = 0$. As described in the appendix, we have more recently tested some of the actual, highly accurate solutions $[\lambda(\tilde{\sigma}_u,\tilde{\omega})$, $A(\alpha)]$ of eqn. (5.1) by comparing the imposed density with the response density implied by orbit integrations. In some cases the two densities differ by only a few parts in 10^4.

Thus, there is good reason to believe that our numerical results are reliable. However, we cannot yet firmly conclude that the $N = 2$ cut-out disk is markedly less susceptible to $m = 2$ instabilities than to $m = 0$ ones. Of course, more accurate results than those above are needed, but the real gap in the evidence stated thus far is the behavior of the mathematical eigenvalues for frequencies other than those illustrated in Figure 4. Only moderate pattern speeds — ones for which the disturbance overlaps the edge of the active part of the disk — were considered there. Yet to be examined are eigenvalues for both large and small pattern speeds; likewise we must be careful to pay attention also to the second, third, · · · mathematical eigenvalues admitted in any given situation. Results for a wide range of growth rates are not essential; in fact, the eigenvalues for $\tilde{s} = 0^+$ suffice to determine whether growing modes are possible, since a marginal eigenvalue curve, consisting of mathematical eigenvalues for frequencies of the form $\tilde{\omega} = m\tilde{\Omega}_p + i0^+$, delimits the image in the complex λ-plane of the upper half of the $\tilde{\omega}$-plane. The question of which side of a marginal eigenvalue curve corresponds to this image is resolved through the conformal mapping property of $\lambda(\tilde{\sigma}_u, \tilde{\omega})$ discussed earlier. Clearly, growing modes are possible if and only if the point $\lambda = 1$ lies in the image of the upper half $\tilde{\omega}$-plane of at least one mathematical eigenvalue.

Most of these issues are covered by the more extensive results for the $N = 2$ disk summarized in Figure 5. The marginal eigenvalue curves over a broad range of pattern speeds are displayed there for the two largest mathematical eigenvalues. Sufficient radial harmonics have been included to produce results reliable to within graphical accuracy.

In general terms, observe that for any given (positive) pattern speed the only significant difference between the largest and the second largest eigenvalues involves their moduli; their phases are more or less the same. Here at least, all points inside the second marginal eigenvalue curve are also inside the largest one. Therefore the critical question of the location of the point $\lambda = 1$, which corresponds to a self-consistent disturbance, relative to these two curves is resolved by the largest curve alone.

We focus now on just the dominant, or largest, eigenvalue. For large pattern speeds its behavior is fairly mundane. As expected, $\lambda \to 0$ as $\tilde{\Omega}_p \to \infty$ since the disturbance is then concentrated at the very center of the disk where there is very little mass able to participate in the commotion. Actually, Figure 5 only contains results up to $\tilde{\Omega}_p = 5$. For larger $\tilde{\Omega}_p$, reliable results are difficult to obtain because a very broad range in α is required then to represent $A(\alpha)$ adequately in the approximation to eqn. (5.1)[2].

[2]The reason for this was explained earlier in this sub-section.

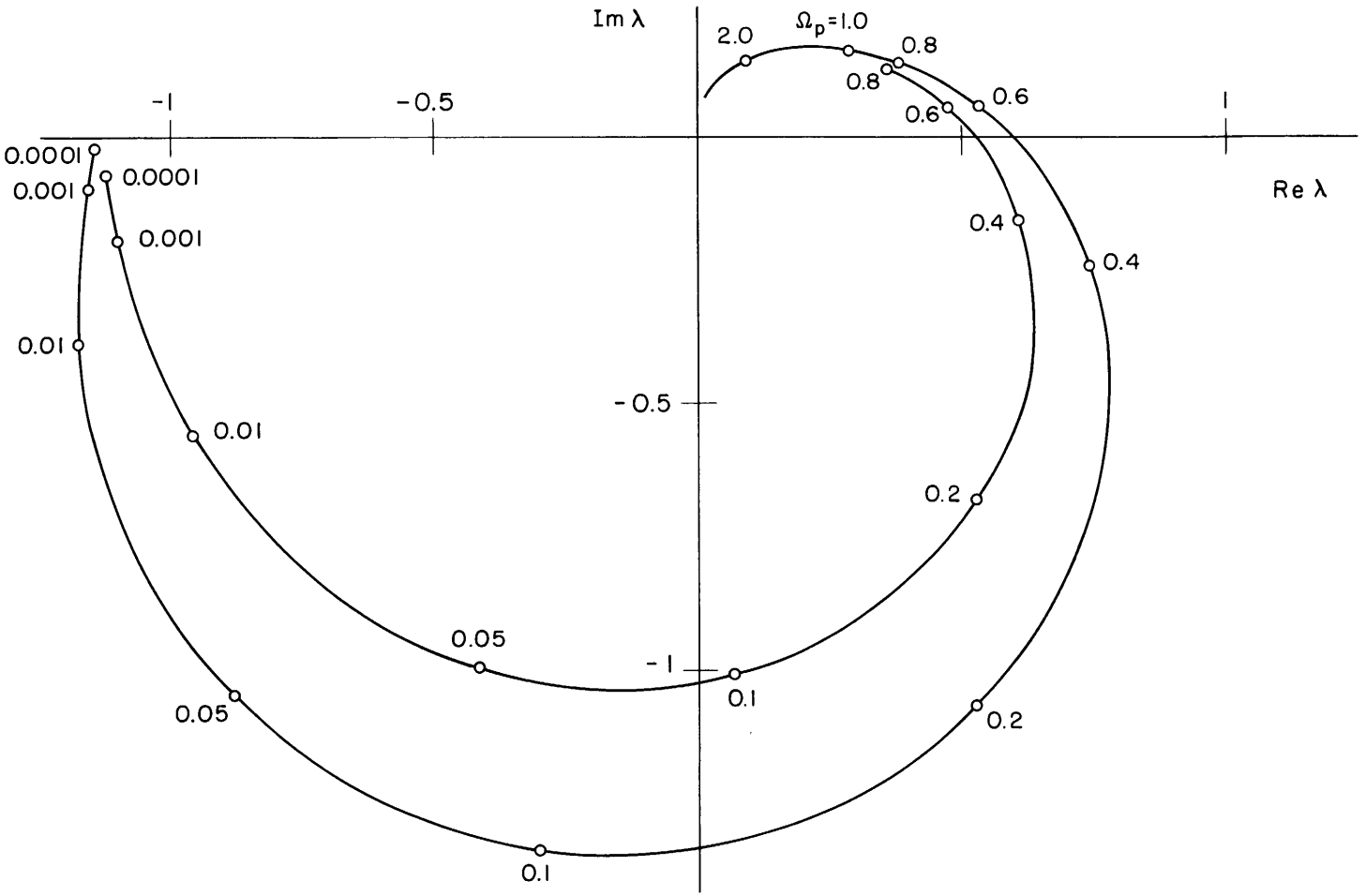

Figure 5. Marginal ($\tilde{s} = 0^+$) eigenvalue curves for the $N = 2$ disk with $\tilde{\sigma}_u = 0.378$. Plotted here are the two largest $m = 2$ eigenvalues as functions of the pattern speed $\tilde{\Omega}_p$; values of the latter are marked next to the data points.

There can be little doubt, though that the dominant eigenvalue curve does approach the origin at this extreme. For moderate pattern speeds the accurate results in Figure 5 differ, as anticipated, by less than 10% from the crude $\tilde{s} = 0^+$ results in Figure 4. The most important feature is that the dominant marginal eigenvalue curve crosses the positive $\mathrm{Re}\{\lambda\}$-axis at $\lambda = 0.595$ when $\tilde{\Omega}_p = 0.538$. This is a slight increase from the crossing at $\lambda = 0.529$ when $\tilde{\Omega}_p = 0.562$ in Figure 4, but hardly enough to alter the significant conclusion that no growing modes with moderate pattern speeds are possible in the $N = 2$ disk. In fact, just from the size of the largest eigenvalue this conclusion can now be extended to rule out growing modes firmly for all pattern speeds greater than 0.284 , since that is the point at which the modulus first exceeds 1 . We observe that by the time $\tilde{\Omega}_p$ is that small this eigenvalue has already swung in phase over 40° clockwise from the positive real axis. As $\tilde{\Omega}_p$ decreases still further the largest eigenvalue remains greater than 1 in modulus and continues swinging in the clockwise direction. It reaches a maximum modulus of 1.39 near $\tilde{\Omega}_p = 0.075$ and thereafter decreases gradually. The results in Figure 5 suggest strongly that in the limit $\tilde{\Omega}_p \to 0^+$, λ tends to a point on the negative real axis in the neighborhood of $\lambda = -1.15$. This apparent tendency has not been more solidly established because of the numerical difficulty of obtaining results for very small pattern speeds. Still, it

appears certain that the largest eigenvalue does not even remotely approach $\lambda = 1$ as $\tilde{\Omega}_p \to 0^+$. To summarize, it seems safe to conclude that the largest eigenvalue cannot equal 1 for $\tilde{s} > 0$ when $\tilde{\Omega}_p > 0$.

Before excluding altogether the possibility of finding any growing modes in this model, both the smaller eigenvalues and the results for $\tilde{\Omega}_p < 0$ must be examined. As already noted, the second marginal eigenvalue curve displayed in Figure 5 behaves much the same as the dominant one. It does not exceed 1 in absolute value until $\tilde{\Omega}_p$ drops below 0.15. The size of the second eigenvalue continually increases as $\tilde{\Omega}_p \to 0^+$, and it seems likely that in this limit it approaches the largest eigenvalue. (This is one reason results are hard to obtain for small $\tilde{\Omega}_p$ — the power method becomes increasingly inefficient.) So now the possibility of growing modes has been restricted to $\Omega_p < 0.15$. Further bounds require an analysis of the third, fourth and even smaller eigenvalues, but it seems likely to us that they too tend to the negative real axis beyond $\lambda = -1$. It has not seemed to us worth the expense to close in this manner an already convincing case against growing modes.

When $\tilde{\Omega}_p < 0$ the eigenvalue curves are more complicated. The difficulty is that there are two distinct types of solutions to eqn. (5.1) for counter-rotating disturbances. One solution corresponds to a distinctly leading pattern — $A(\alpha)$ is most active in the vicinity of $\alpha = -10$ — and the

other to a more or less non-spiral eigenfunction — $A(-\alpha) \simeq \bar{A}(\alpha)$. The first type of solution is the dominant one for $-\infty < \tilde{\Omega}_p < -0.05$; its eigenvalue starts out at the origin in quadrant III for large, negative $\tilde{\Omega}_p$ and circles into quadrant IV as $\tilde{\Omega}_p \to 0^-$, never much exceeding $1/2$ in modulus. Near $\tilde{\Omega}_p = -0.05$ the nearly non-spiral solution becomes the dominant one; its eigenvalue lies a shade below the negative $\mathrm{Re}\{\lambda\}$-axis, drifting out to the left as $\tilde{\Omega}_p \to 0^-$, possibly tending to the $\tilde{\Omega}_p \to 0^+$ limit.[3] We conclude that this $N = 2$ disk does not admit any growing modes with negative pattern speeds.

So far, we have given a fairly complete description of the manner in which the mathematical eigenvalues $\lambda(\tilde{\sigma}_u, \tilde{\omega})$ vary with the frequency. The dependence upon the velocity dispersion is illustrated in Figure 6, where portions of the dominant marginal eigenvalue curves for $Q = 0.80$, 1.00 and 1.20 are displayed. Here, as in Figure 4, we have resorted to coarse results for which only a few radial harmonics are included. (With the exception of Figures 4 and 6 all results cited here are reliable to the number of digits quoted, and certainly those used for the plots are reliable to within graphical accuracy.) Naturally, the eigenvalues increase as $\tilde{\sigma}_u$ decreases, since colder disks respond more than hotter disks to a given imposed potential. In fact, the eigenvalues displayed in Figure 6 vary more or less as

---

[3] Since the kernel has a logarithmic branch point at $\tilde{\omega} = 0$, it is not clear that these two limits coincide.

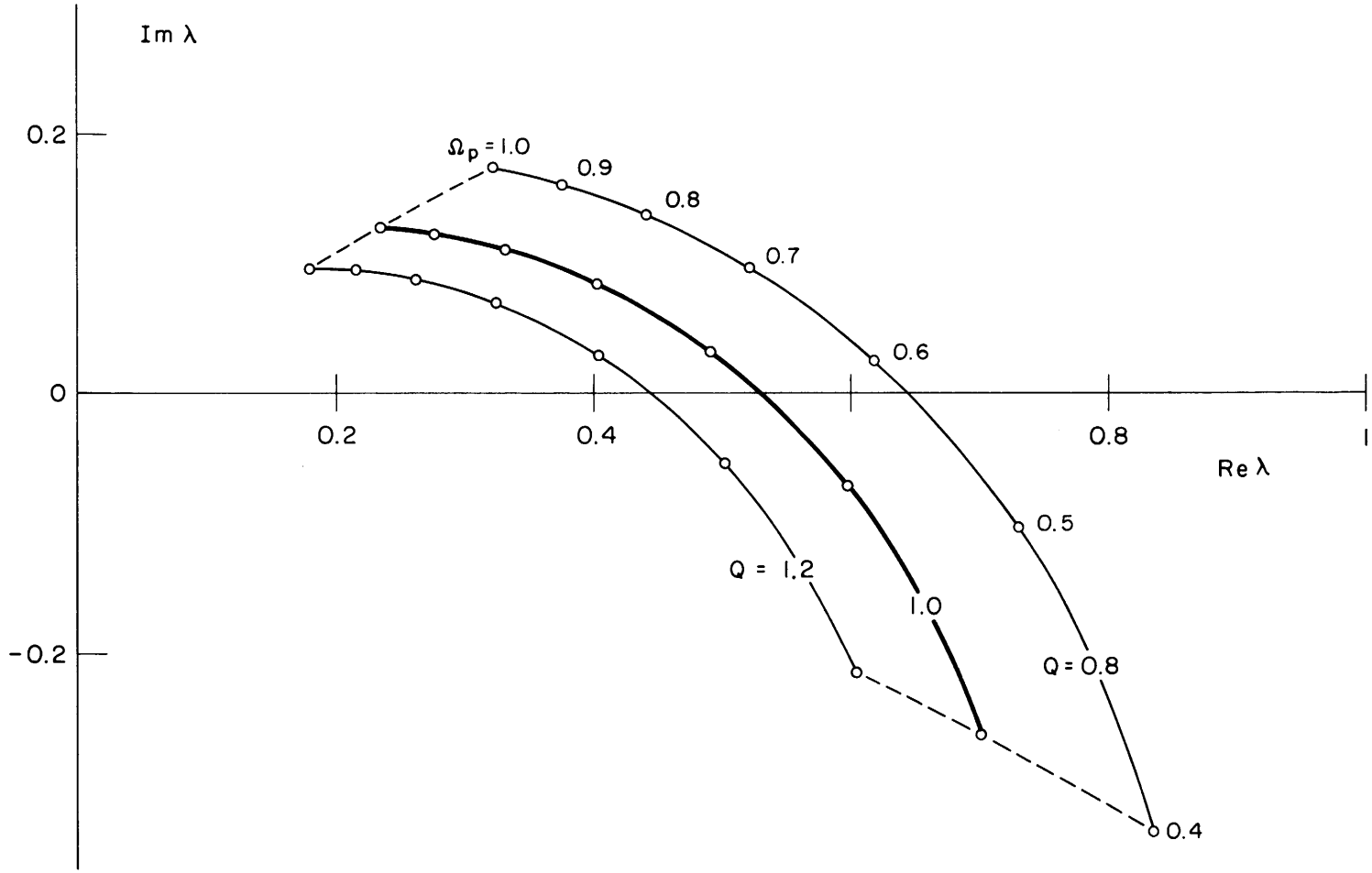

Figure 6. Dependence of largest $m = 2$ eigenvalue, upon the velocity dispersion $\tilde{\sigma}_u = 0.378\,Q$ of the $N = 2$ cut-out disk, for stability parameter $Q = 0.8$, 1.0 and 1.2 . All three curves refer to vanishing growth rates ($\tilde{s} = 0^+$), and to the range $0.4 \leq \tilde{\Omega}_p \leq 1.0$ of pattern speeds as indicated.

$\tilde{\sigma}_u^{-1}$. Probably the most significant feature of these curves is the point at which they cross the Re{λ}-axis:[4]

$$Q = 1.20: \quad \lambda = 0.443 \quad \text{when} \quad \tilde{\Omega}_p = 0.556$$
$$Q = 1.00: \quad \lambda = 0.529 \quad \text{when} \quad \tilde{\Omega}_p = 0.561$$
$$Q = 0.80: \quad \lambda = 0.643 \quad \text{when} \quad \tilde{\Omega}_p = 0.576 \quad .$$

These results suggest that $\lambda = 1$ does not fall within the image of the upper half of the $\tilde{\omega}$-plane (and therefore no $m = 2$ instabilities are present) until $Q \lesssim 1/2$. A more precise stability boundary for the $N = 2$ model as well as that of several other cut-out models will be reported in sub-section (d).

c) Artificial Eigenvalues of Other Models

The issue we address here is the behavior of the eigenvalues of the other cut-out disks. Although our examinations of these eigenvalues were not as detailed as our study of the $N = 2$ model, there seems little doubt that, regardless of the cut-out index N , $\lambda(\tilde{\sigma}_u, \tilde{\omega})$ can assume positive real values only for moderate values of $\tilde{\Omega}_p$. That region of the λ-plane is of course the most relevant one when seeking growing modes. We needn't worry about what happens for large pattern speeds, for surely $\lambda \to 0$ as $\tilde{\Omega}_p \to \infty$. In fact, this must occur more rapidly with increasing N , since there

[4] The curve for $Q = 1.00$ here does not exactly match the corresponding curve in Figure 5 since a slightly better numerical approximation was used here.

is then progressively less active mass at the small radii where disturbances of large speed are likely to be concentrated. At the other extreme, for very small $\tilde{\Omega}_p$, the exact form of the cut-out center should not be very important, for the corresponding density disturbances are concentrated far outside the center. Thus we expect that the eigenvalues for all the cut-out disks should behave similarly for very small $\tilde{\Omega}_p$. The results listed below for $N = 1$, 2, 3 and 4 when $\tilde{\Omega}_p = 0.01$ and 0.001 are consistent with this expectation:

	$\tilde{\Omega}_p = 0.01$		$\tilde{\Omega}_p = 0.001$	
N	λ_R	λ_I	λ_R	λ_I
1	-0.913	-0.496	-1.096	-0.192
2	-1.180	-0.390	-1.160	-0.097
3	-1.241	-0.295	-1.164	-0.066
4	-1.255	-0.244	-1.163	-0.053

The eigenvalues are noticeably closer for the smaller of the two pattern speeds. We saw in the previous section that when $N = 2$, the eigenvalues appear to end up on the negative $\mathrm{Re}\{\lambda\}$-axis. Apparently the eigenvalues of the other models do likewise in the limit $\tilde{\Omega}_p \to 0^+$. Therefore, small pattern speeds are not a serious concern either.

Relying then on the above evidence and the detailed $N = 2$ results, we hereafter restrict our examinations of the cut-out models to the regime of moderate pattern speeds. For the models with cut-out indices $N = 1$, 2, 3 and 4,

the relevant portions of the dominant marginal eigenvalue curves ($\tilde{s} = 0^+$) are displayed in Figure 7. As in the two earliest diagrams of this chapter the velocity dispersion is such that $Q = 1$. From these results alone, we can determine which models are stable by the location of the point $\lambda = 1$: if it lies to the left of the curve as it is traversed in the direction of decreasing $\tilde{\Omega}_p$, then no growing modes are possible since the eigenvalues for $\tilde{s} > 0$ lie to the right of a marginal eigenvalue curve.

On this basis it is clear that the $N = 1$ model is extraordinarily stable to $m = 2$ disturbances when $Q = 1$. Curiously, the eigenvalue which has the largest modulus when $\tilde{\Omega}_p \lesssim 0.65$ is not the same one that is largest when $\tilde{\Omega}_p \gtrsim 0.65$. The first and second eigenvalues, then, switch roles near that pattern speed. For purposes of clarity, only the eigenvalue that dominates for the smaller pattern speeds has been displayed. The eigenvalue which dominates for larger $\tilde{\Omega}_p$ lies roughly $45°$ in phase closer to the Re{λ}-axis than does the eigenvalue plotted in Figure 7 for $N = 1$; moreover, it crosses the real axis at a larger value — at $\lambda = 0.225$ when $\tilde{\Omega}_p = 0.807$. The other cut-out models do not exhibit this interchange of eigenvalue roles. Evidently the $N = 1$ disk would have to be virtually cold ($\tilde{\sigma}_u \lesssim 0.1$) before the marginal eigenvalue curve would cross the real axis beyond $\lambda = 1$ (using the rough rule of thumb that λ varies with the velocity dispersion in an inverse-first-power manner),

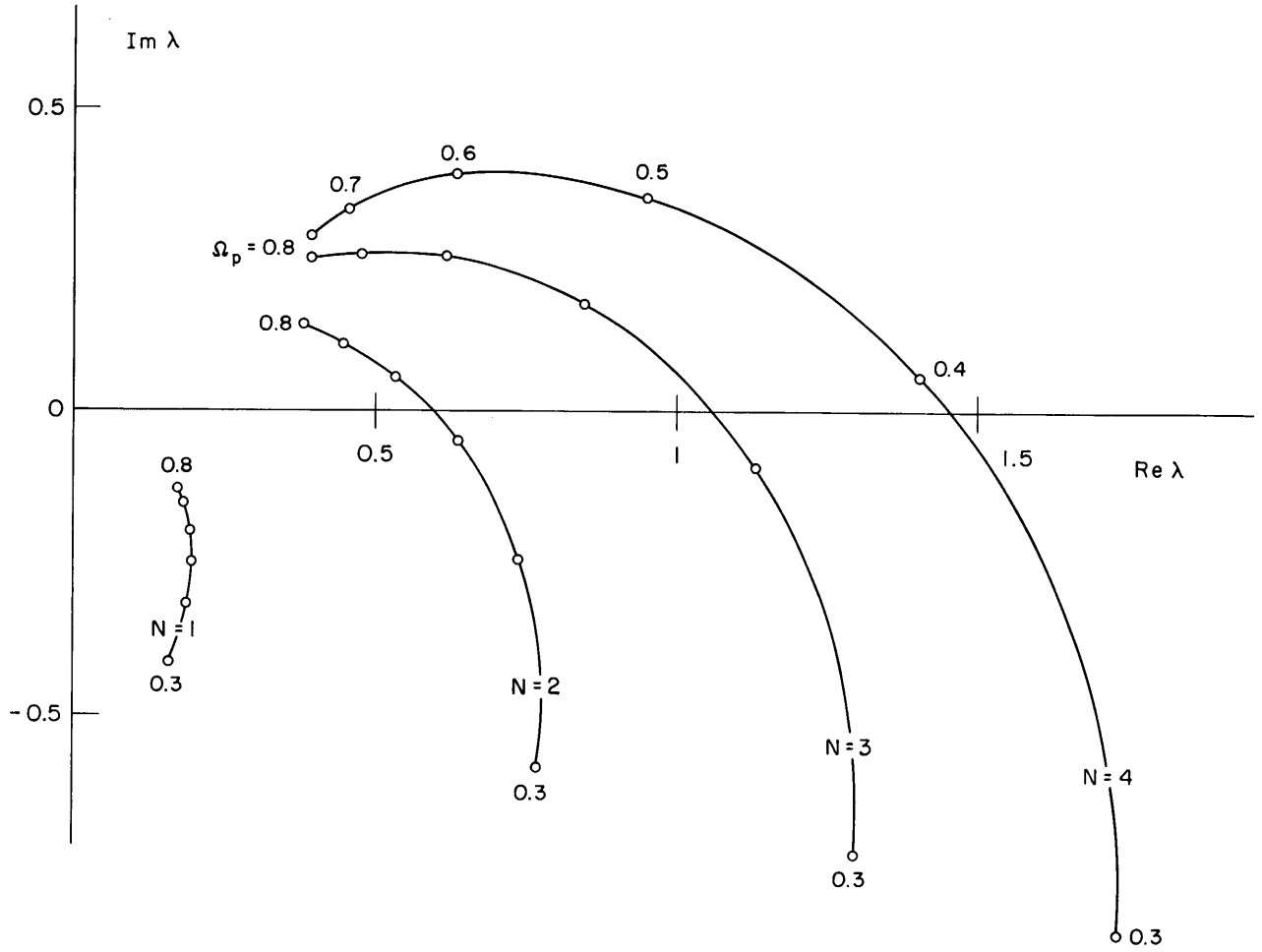

Figure 7. Marginal ($\tilde{s} = 0^+$) eigenvalue curves for the N = 1, 2, 3 and 4 disks with $\tilde{\sigma}_u = 0.378$. Shown here is the largest m = 2 eigenvalue calculated for each disk, assuming pattern speeds $0.3 \leq \tilde{\Omega}_p \leq 0.8$.

or equivalently, before any growing $m = 2$ modes would be possible.

As previously noted, the cut-out index $N = 2$ disk is already well-stabilized to bisymmetric disturbances when $Q = 1$. Recall that its marginal eigenvalue curve crosses the $\mathrm{Re}\{\lambda\}$-axis at $\lambda = 0.595$, and that Q consequently needs to decrease by about a factor of 2 before any $m = 2$ modes become unstable.

The $N = 3$ disk is the first member of this family of cut-out models that admits a growing $m = 2$ mode when Q is 1. As depicted in Figure 7, the point $\lambda = 1$ lies within the marginal eigenvalue curve, so that we are sure to find some pattern speed $\tilde{\Omega}_p$ and some growth rate $\tilde{s} > 0$ for which $\lambda(\tilde{\sigma}_u, \tilde{\omega}) = 1$. The marginal eigenvalue curve intersects the positive $\mathrm{Re}\{\lambda\}$-axis at $\lambda = 1.061$ when $\tilde{\Omega}_p = 0.424$. Note that the point $\lambda = 1$ does not fall within the marginal eigenvalue curve with much room to spare; an increase of $\tilde{\sigma}_u$ by only 10% will surely suffice to suppress this instability.

In contrast, the $N = 4$ model is unstable by a wide margin when the axisymmetric instabilities first disappear. The $\mathrm{Re}\{\lambda\}$-axis is crossed by its marginal eigenvalue curve at $\lambda = 1.453$ when $\tilde{\Omega}_p = 0.390$. We fully expect this disk to remain susceptible to $m = 2$ instabilities until Q reaches the neighborhood of 1.5.

Before proceeding to our report of the more accurate $m = 2$ stability criteria for these cut-out models, we wish to make several remarks. Neither the $N = 3$ nor the $N = 4$ model has a second unstable mode when $Q = 1$. For $N = 3$, the marginal eigenvalue curve for the second eigenvalue intersects the $\mathrm{Re}\{\lambda\}$-axis at $\lambda = 0.782$ for $\tilde{\Omega}_p = 0.430$, while for $N = 4$ it does so at $\lambda = 0.983$ for $\tilde{\Omega}_p = 0.392$. Note that in both cases the pattern speed at which the second marginal eigenvalue is purely real differs by only 1% from the pattern speed at which the marginal eigenvalue curve for the dominant eigenvalue crosses the real axis. This feature reinforces our earlier impression, formed from the specific $N = 2$ results, that the phases of the various mathematical eigenvalues for a given pattern speed are more or less the same.

The foregoing results suggest that the $N = 5, 6, 7, \cdots$ models will be progressively more unstable, i.e., the instabilities will have increasingly large growth rates, when $Q = 1$ (or, for that matter, for any other Q in the vicinity of 1); moreover, it appears certain that for $N \geq 5$, there will be at least two unstable modes when Q is 1. We have not pursued the more sharply cut-out models sufficiently to determine what, if any, limiting behavior they exhibit as $N \to \infty$. In this limit, of course, the cut-out factor $H(J)$ tends to a step function at $J = r_o V_o$, and consequently the active density appears more and more to have simply a sharp hole in the center.

Finally, since the cut-out models in the $\tilde{\Omega}_p \to 0^+$ limit are expected to mimic the singular disk, our results for small $\tilde{\Omega}_p$ suggest that the dominant eigenvalue of the basic self-similar model has a real negative value when $Q = 1$. In fact, Kalnajs had anticipated just this result when warning us about the peculiar frequency dependence of the singular disk kernel. Furthermore, since it seems likely that only the modulus of λ will change as $\tilde{\sigma}_u$ is varied, the implication is that the singular disk admits no exponentially growing modes of type $m = 2$ when Q is in the vicinity of 1 .

d) Stability Criteria

Our detailed $N = 2$ results led us to conclude that the last $m = 2$ instabilities to disappear from a given cut-out disk as $\tilde{\sigma}_u$ increases are indeed those associated with the largest mathematical eigenvalue. Therefore, the minimum velocity dispersion for which one of these disks is $m = 2$ stable is surely that value of $\tilde{\sigma}_u$, denoted here by $\tilde{\sigma}_{2,crit}$, for which the dominant marginal eigenvalue curve passes exactly through the point $\lambda = 1$, i.e., that value of $\tilde{\sigma}_u$ for which $\lambda(\tilde{\sigma}_{2,crit}, m\tilde{\Omega}_p + i0^+) = 1$ for some $\tilde{\Omega}_p$. Thus far we have provided only rough estimates of $\tilde{\sigma}_{2,crit}$ by relying on the $Q = 1$ results and the observed $\lambda \sim \tilde{\sigma}^{-1}$ behavior. In this section we will report our refinements of these estimates, which were obtained by determining, to several digits,

the specific values of $\tilde{\Omega}_p$ and especially $\tilde{\sigma}_u$ for which $\lambda(\tilde{\sigma}_u, m\tilde{\Omega}_p) = 1$, where, of course, by $m\tilde{\Omega}_p$ we actually mean $m\tilde{\Omega}_p + i0^+$.

A two-dimensional iteration in $\tilde{\sigma}_u$ and $\tilde{\Omega}_p$ was used to find the solutions of

$$\lambda_R(\tilde{\sigma}_u, m\tilde{\Omega}_p) = 1$$

$$\lambda_I(\tilde{\sigma}_u, m\tilde{\Omega}_p) = 0 .$$

We employed essentially a Newton-Ralphson method of iteration, using the last three iterates to estimate the partial derivatives.

No attempt was made to find $\tilde{\sigma}_{2,crit}$ for the $N = 1$ model. The primary reason for this omission was the small size of this $\tilde{\sigma}_u$ (as we have seen, it is surely less than 0.10); hence a very broad range in α would have been needed to obtain an adequate approximation to the integral equation (5.1).

For the $N = 2$ model, the minimum velocity dispersion is $\tilde{\sigma}_{2,crit} = 0.21$ ($Q = 0.55$). The marginally stable $m = 2$ mode has $\tilde{\Omega}_p = 0.57$. In the $N = 3$ cut-out disk, $\tilde{\sigma}_{2,crit} = 0.417$ ($Q = 1.10$), with the marginally stable mode having $\tilde{\Omega}_p = 0.417$. The last member of the family of cut-out models for which we tracked down the stability boundary was the $N = 4$ model, where $\tilde{\sigma}_{2,crit} = 0.603$ ($Q = 1.59$). The marginally stable mode occurs when $\tilde{\Omega}_p = 0.363$.

Surely the fact that the $N = 2$ model is $m = 2$ stable for $Q \geq 0.55$ is surprising enough. But even for the other two models, for which $m = 2$ stability is not achieved until $Q > 1$, the ratio of the energy in random motions to that in rotational motion is rather small. At any point in the disk, this ratio, denoted here by R, is given by

$$R = \sigma_u^2 \, [1 + (\sigma_v/\sigma_u)^2] \, .$$

The tangential velocity dispersion can be obtained from eqn. (2.16). Using our above stability criteria we find that, at marginal stability, $R = 0.064$ for the $N = 2$ disk, $R = 0.257$ for the $N = 3$ disk and $R = 0.527$ for the $N = 4$ disk. These values of R disagree sharply with the ratio $R = 2.6$ that Ostriker and Peebles (1973) have suggested characterizes the stability boundary in disk galaxies. (Actually, they phrased their criterion in terms of the ratio of total rotational to total gravitational energy; we have recast it in slightly different terms.) Although their criterion does apply to many systems, the cut-out models are exceptions. Of course, this may be the case only because we have carved out the region where their result can be most expected to hold.

Although there is only a single unstable $m = 2$ mode in the $N = 3$ and $N = 4$ disks when $Q = 1$, we can in principle determine how low $\tilde{\sigma}_u$ must be before other instabilities arise — ones associated with the smaller mathematical

eigenvalues. We have done this only for the $N = 3$ model. We found that a second instability develops when $\tilde{\sigma}_u$ drops below 0.30 ($Q = 0.80$) and a third when the dispersion is lower than $\tilde{\sigma}_u = 0.28$ ($Q = 0.74$). At its point of marginal stability the second mode has a pattern speed of $\tilde{\Omega}_p = 0.46$ while the corresponding pattern speed of the third mode is $\tilde{\Omega}_p = 0.42$. These angular speeds are again in the moderate range.

For more sharply cut out models ($N \geq 5$) we expect that $\tilde{\sigma}_{2,crit}$ will be increasingly large. The determination of the $m = 2$ stability criteria of these models would require computation for very hot disks ($Q \gtrsim 2$). The typical orbits in these systems would have very large eccentricities. We have not attempted these calculations, since an inordinately large number of radial harmonics would then have been needed to get decent results.

e) Growing Modes When $Q = 1$

No doubt the most essential result on the bisymmetric disturbances of the cut-out models is that which has just been reported — the amount of "heat" necessary to stabilize all such modes. But perhaps the most interesting feature of disks which are still unstable to two-armed disturbances ($\tilde{\sigma}_u < \tilde{\sigma}_{2,crit}$) is the nature of the growing modes — their characteristic frequencies and shapes. The former properties are discussed in this sub-section.

We will consider for this purpose disks with the velocity dispersion $\tilde{\sigma}_u = 0.378070$ $(Q = 1)$. These are systems which are barely stable to axisymmetric perturbations. Of the four cut-out models examined in this thesis only the $N = 3$ and $N = 4$ disks have any growing modes at this velocity dispersion.

For this fixed $\tilde{\sigma}_u$, we seek to find the specific pattern speed $\tilde{\Omega}_p$ and growth rate $\tilde{s}$ for which $\lambda(\tilde{\sigma}_u, m\tilde{\Omega}_p + i\tilde{s}) = 1$. As in our earlier search for the marginally stable modes, an iterative procedure is used to find the characteristic frequency $\tilde{\omega} = m\tilde{\Omega}_p + i\tilde{s}$. However, in the present case we do not need to resort to such a clumsy two-dimensional iteration. The reason is that λ is an analytic function of $\tilde{\omega}$ and therefore an iteration in the single variable $\tilde{\omega}$ is possible. An inverse Lagrange interpolation which employs all the previous $[\tilde{\omega}, \lambda(\tilde{\sigma}_u, \tilde{\omega})]$ pairs is easy to implement and was used here to produce the successive estimates of the characteristic frequency.

For the $N = 3$ cut-out disk, the $Q = 1$ unstable mode has a pattern speed of $\tilde{\Omega}_p = 0.433$ and a growth rate of $\tilde{s} = 0.021$. As we had anticipated from the $\tilde{s} = 0^+$ eigenvalues, this is a rather slowly growing disturbance. On the other hand, the $N = 4$ unstable mode grows considerably faster, as expected. Its growth rate is $\tilde{s} = 0.1270$ and its pattern speed is $\tilde{\Omega}_p = 0.4394$.

The near agreement between the pattern speeds of the growing modes in these two disks is striking. The inner Lindblad resonance, corotation radius and outer Lindblad resonance for the mode of the $N = 3$ disk occur at $\tilde{r}_{ILR} = 0.676$, $\tilde{r}_{CR} = 2.31$ and $\tilde{r}_{OLR} = 3.94$, and for the mode of the $N = 4$ model, these resonances are located at $\tilde{r}_{ILR} = 0.667$, $\tilde{r}_{CR} = 2.28$ and $\tilde{r}_{OLR} = 3.89$. These resonance radii, of course, are those for the stars in circular orbits. We will take a closer look at the latter mode in the following sub-section.

For now, note that the corotation circles of these modes — and indeed even those for the marginally stable modes discussed earlier — lie within a factor of 2 or at most 3 from the edge in the active density. Moreover, the inner Lindblad resonance is noticeably inside this edge, but it is still close enough for there to be an appreciable amount of material near the resonance. Most of the activity of the density disturbances themselves can thus be expected to lie in the vicinity of the rigid center. Both this evidence from the pattern speeds themselves and the additional fact that the instability is stronger in the more sharply cut out disks, leave little doubt that these modes are somehow due to the presence of an edge to the active density. Although we have some ideas about the details of this process, we are presently unable to offer a <u>clear</u> physical mechanism for this behavior.

We have one other result of some interest to present in this sub-section. Even though the $N = 2$ disk has no exponentially growing modes when $Q = 1$, we can still search for the least rapidly (exponentially) decaying disturbance. The mathematical apparatus developed in Chapter III was based on the assumption that the perturbation was growing, i.e., $\tilde{s} = \mathrm{Im}\{\tilde{\omega}\} > 0$. By analogy with the Landau treatment of the initial value problem in plasma physics [cf., for example, Stix (1962)] the decaying modes are given by the solutions of eqn. (3.42) when the kernel, eqn. (3.40), is analytically continued into the region $\mathrm{Im}\{\tilde{\omega}\} < 0$. The only portion of the kernel in which the frequency appears is the angular momentum function. Since a specific expression, eqn. (3.44), is available for this function, the continuation process is fairly easy. The only subtle point is the treatment of the term

$$\ell n \left[\frac{\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})}{\tilde{\omega}} \right] .$$

In expressions (3.45) and (3.46) the assumption that $\mathrm{Im}\{\tilde{\omega}\} > 0$ was explicitly used. In those formulas the principal branch of the logarithm was specified. The analytic continuation into the region $\mathrm{Im}\{\tilde{\omega}\} < 0$ is achieved by using, for $\ell n\ \tilde{\omega}$ that branch of the logarithm that enables the argument of the logarithm to vary smoothly as $\tilde{s}$ crosses 0, i.e., if $\tilde{\Omega}_p > 0$, then, whether $\tilde{s} > 0$ or $\tilde{s} < 0$, $\ell n\ \tilde{\omega}$ should be taken to be $\ell n|\tilde{\omega}| + i \tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)$. However, note that

the angular momentum function has a branch point at $\tilde{\omega} = 0$. Therefore, although we may find that the dominant exponential behavior has a "damping rate" $\tilde{s} < 0$, we cannot conclude that all disturbances will decay at least as fast as $e^{\tilde{s}\tilde{t}}$.

With this caution in mind, we report now that for $Q = 1$, the dominant decaying mode of the $N = 2$ cut-out disk has a damping rate $\tilde{s} = -0.220$ and a pattern speed $\tilde{\Omega}_p = 0.414$. In principle, it is possible to compute similar damping rates for the dominant mode of the $N = 1$ disk and even for the modes associated with the sub-dominant mathematical eigenvalues of the $N = 3$ and $N = 4$ disks. However, we have chosen not to pursue this point any further.

f) Illustration of a Growing Mode

We close our discussion of the bisymmetric disturbances with a detailed illustration of one of the growing modes whose characteristic frequency was cited in the previous subsection: the mode of the disk with cut-out index $N = 4$ when the system is marginally stable to axisymmetric disturbances.

The density transform $A(\alpha)$ of this mode is displayed in Figure 8. The real and imaginary parts (solid curves) of this eigenfunction are given along with its magnitude (dotted curve). Of course, this selection of the real and imaginary parts (as well as that of the vertical scale) is arbitrary since any non-zero complex multiple of $A(\alpha)$ is also a mode.

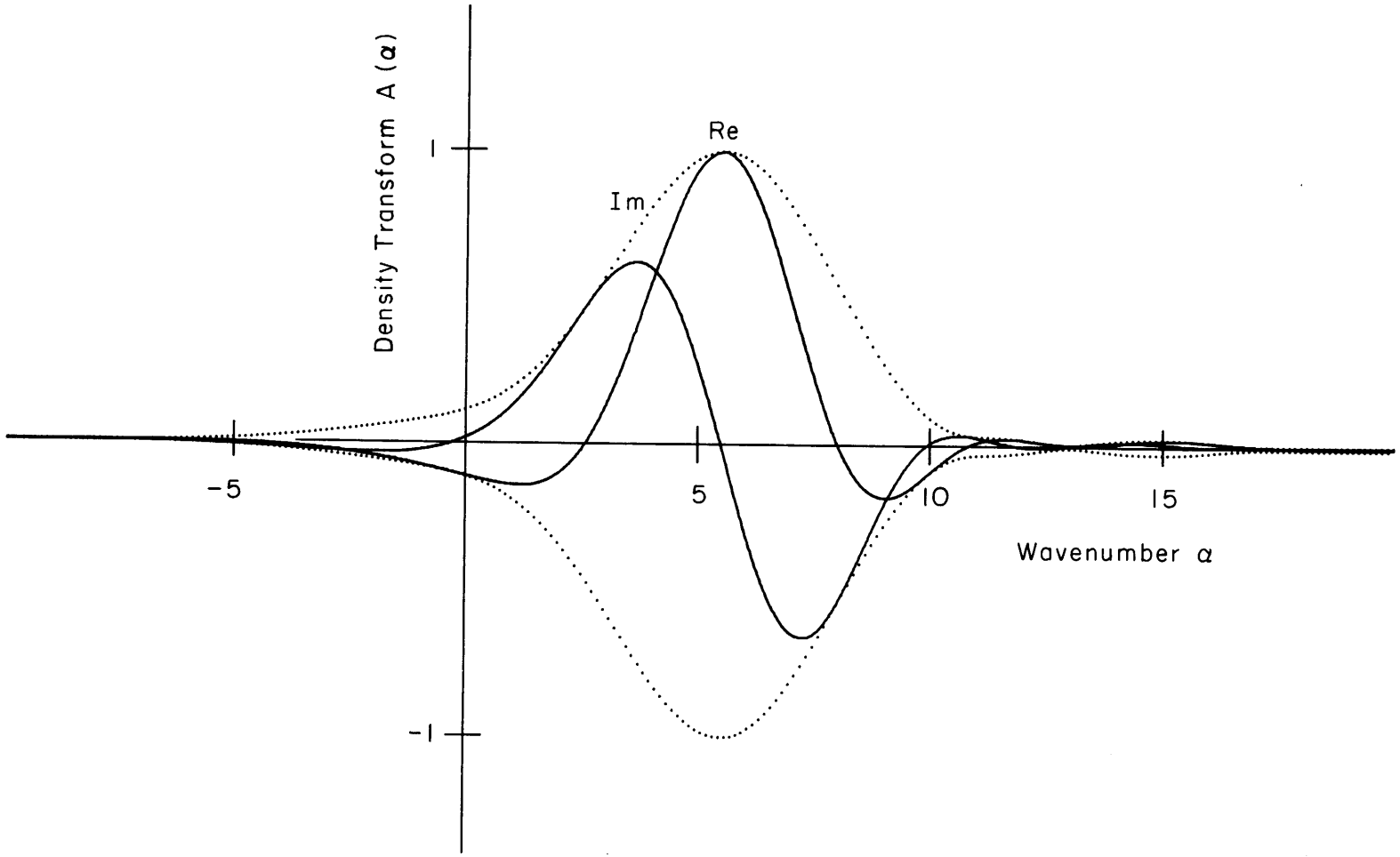

Figure 8. The density transform $A(\alpha)$ for the $m = 2$ mode of the $N = 4$ disk with $Q = 1$. Solid curves denote real and imaginary parts, and dotted curves denote the envelope. For this mode, $\tilde{\Omega}_p = 0.4394$ and $\tilde{s} = 0.1270$.

Recalling that the positive wavenumbers correspond to trailing components of a disturbance density, it is obvious from Figure 8 that this mode is a trailing one — there is very little content in the negative, or leading, wavenumbers. In fact, since the density transform peaks near $\alpha = 5.5$, we can expect the density itself to be roughly characterized by this wavenumber.

The only other comment we have about Figure 8 concerns the high-wavenumber content of $A(\alpha)$. The noteworthy feature is that the density transform decays in amplitude far more rapidly to the left of its peak than it does to the right: by the time $\alpha = -8$, $|A(\alpha)|$ is already down to the 0.1% level, while at an equal distance to the right of the peak, it has only dropped to 0.6% of its maximum; even at $\alpha = +30$, $|A(\alpha)|$ still has 0.3% of its peak value. This can be explained by referring to the transfer function $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, given explicitly by eqn. (3.40). This function describes how the disk responds at the wavenumber β to a forcing at the wavenumber α. For the density disturbance of this mode the forcing wavenumbers are mainly in the neighborhood of $\alpha \sim 5.5$. The term in the transfer function which dominates the redistribution to other wavenumbers is the angular momentum function $F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})$. As we pointed out in section III.d, this function is responsible for a strong trailing bias in the transfer function itself. In fact, as $\beta-\alpha \to +\infty$, the $\ell \geq -1$ angular momentum functions

decay only as $e^{-|\beta-\alpha| \tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)}$. Clearly, the smaller the growth rate, the larger $A(\alpha)$ will be for large positive wavenumbers. Presumably some of this is due to the effects of resonant stars — after all, the $\ell \geq -1$ terms are the resonant ones.

Of course, the density transform is only of secondary importance; the real interest lies in $\tilde{\mu}_1(\tilde{r})$. (We are measuring density in units of μ_o and radius in units of r_o ; the dependence upon θ and t has been suppressed.) This density in physical space is displayed in Figure 9 in terms of a contour map and in Figure 10 in terms of two radial cross sections. The positive contours at 80, 60, 40, 20 and 10% of the peak disturbance density are shown as solid curves in Figure 9 ; the nodes of this density are indicated by the dotted lines; the negative contours are not displayed. The arrows mark the corresponding contour levels on Figure 10. In the latter diagram the designations Re and Im refer to the real and imaginary parts of the Fourier integral

$$\tilde{\mu}_1(\tilde{r}) = \tilde{r}^{-3/2} \int_{-\infty}^{\infty} A(\alpha)\ e^{i\alpha\ \ell n\ \tilde{r}}\ d\alpha\ .$$

The actual positions of these two radial slices are marked in Figure 9. The resonance locations marked on these diagrams are those for stars in circular orbits.

Our first observation is that this density forms a rather striking, trailing spiral pattern which is mildly tightly-wound.

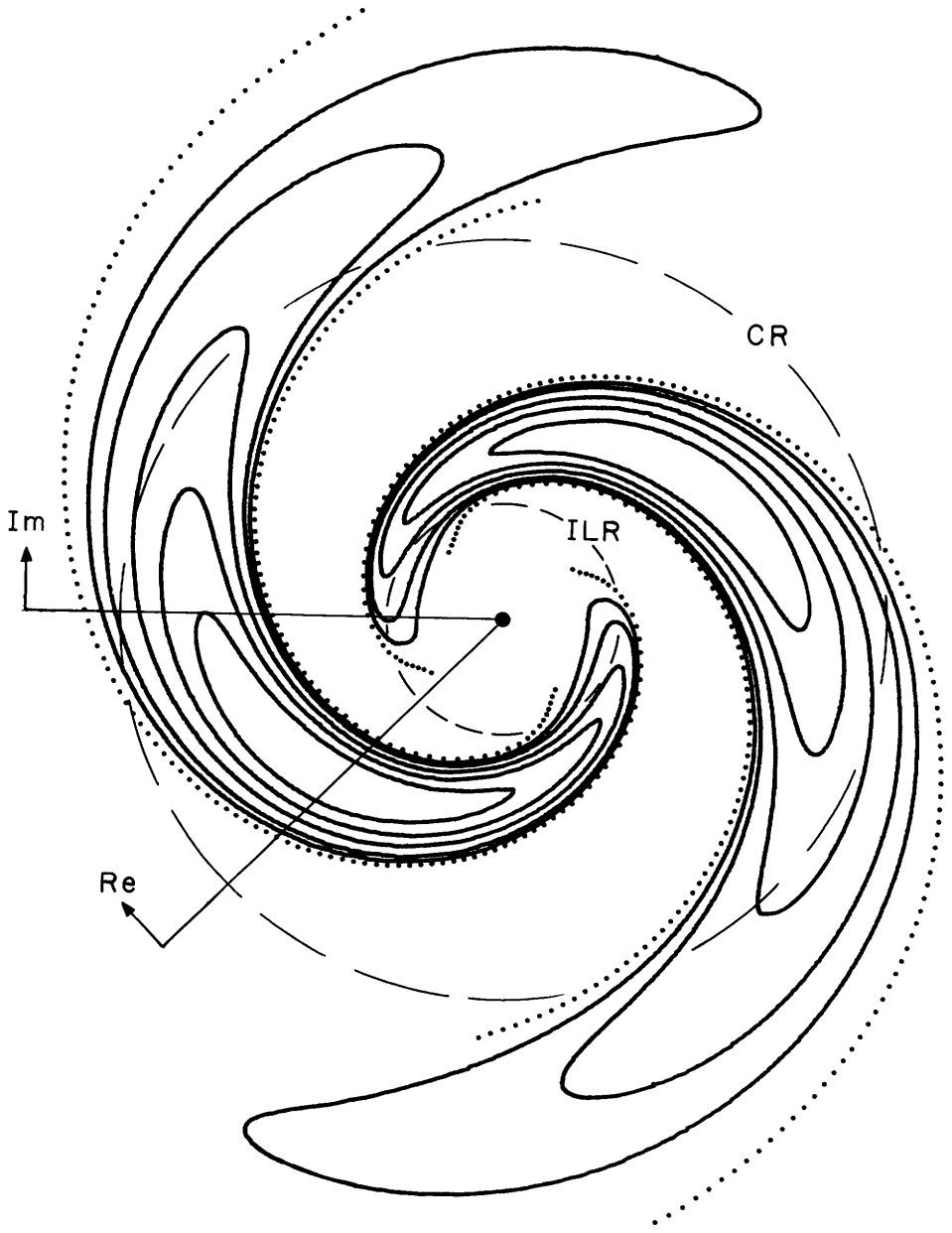

Figure 9. Density contours of the only unstable $m = 2$ mode of the $N = 4$ disk with $Q = 1$. Solid curves denote 80, 60, 40, 20 and 10% of peak density; dotted curves are the nodes. The lines marked Re and Im indicate the locations of the radial slices shown in Figure 10. Both the model and this growing pattern rotate counter-clockwise. Inner Lindblad resonance (circle marked ILR) and outer Lindblad resonance (not shown) are equidistant from corotation resonance (CR).

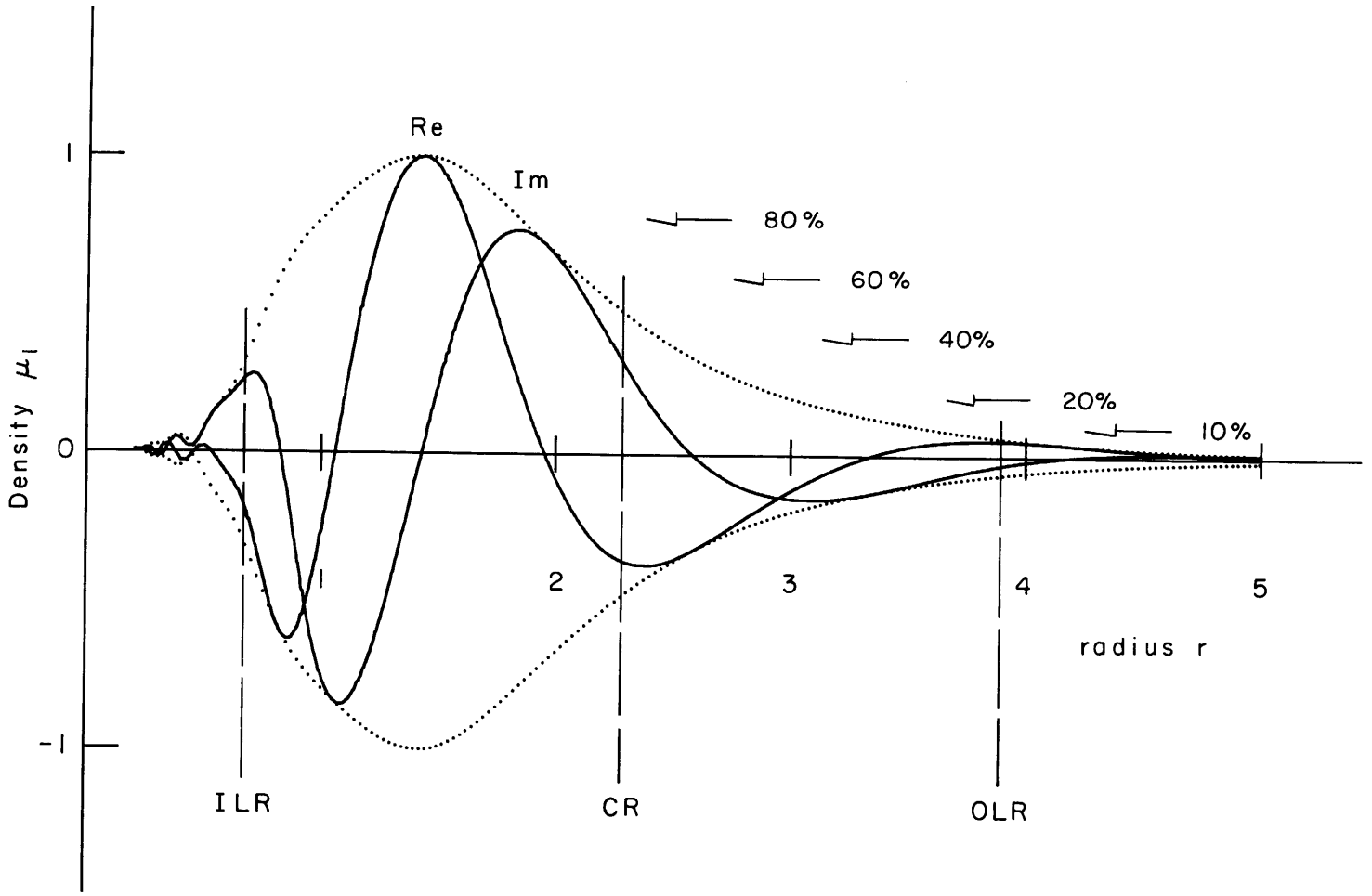

Figure 10. The disturbance density $\mu_1(r)$ for the $m = 2$ mode shown in Figure 9. Solid curves denote real and imaginary parts, dotted curves the envelope. Arrows signify the contour levels plotted in Figure 9. Inner Lindblad resonance (ILR), corotation resonance (CR), and outer Lindblad resonance (OLR) for stars in circular orbits are all marked by dashed lines.

We find that the wavenumber $\alpha = 5.5$ does characterize the density field. To be more specific, by referring to Figure 10 we observe that one complete oscillation of the real part occurs between the nodes at $\tilde{r}_1 = 1.06$ and $\tilde{r}_2 = 3.33$, while for the imaginary part the relevant locations are $\tilde{r}_1 = 0.83$ and $\tilde{r}_2 = 2.58$. The corresponding (logarithmic) wavenumber is $\alpha = 2\pi/\ell n(\tilde{r}_2/\tilde{r}_1)$. For this wavenumber, the arms are inclined by roughly $20°$ to the circular direction.

Secondly, note how well the disturbance appears to be confined to the "principal range" between the two Lindblad resonances. The density peaks roughly halfway between the inner Lindblad resonance and the corotation circle. Just as we had surmised earlier on the basis of the pattern speed alone, this disturbance overlaps the edge in the active density. Note that the inner resonance is situated just barely inside the edge at $\tilde{r} \sim 1$ in the active density. To give some indication of how wide the effective inner Lindblad resonance region is in this hot disk, we note that a resonant star with the typical eccentric velocity $\tilde{U} = 0.4$ oscillates between $\tilde{r}_{min} = 0.467$ and $\tilde{r}_{max} = 0.827$ (cf. table on p. 36). These number should be compared with the nominal resonance location $\tilde{r}_{ILR} = 0.667$. Note also that the resonances are blurred further because the growth rate $\tilde{s}$ is sizeable.

On the whole, then, these illustrations reinforce our earlier impression that the instability is indebted to the sharply cut out center of these models.

VI. RESULTS FOR SEVERAL OTHER ANGULAR HARMONICS

Besides the $m = 0$ and the fairly thorough $m = 2$ results which have already been described, our stability analysis includes results for the $m = 1$, 3 and 4 angular harmonics. The same general approach was taken for these other three non-axisymmetric harmonics as was adopted for the $m = 2$ harmonic. Specifically, we again introduced an artificial, mathematical eigenvalue problem analogous to eqn. (5.1). The presence or absence of exponentially growing modes was then inferred from the numerically obtained mathematical eigenvalues.

The results for the three remaining cases could not be obtained quite as efficiently as in the case $m = 2$. When $m = 1$ or $m = 3$, the magnitude of the largest eigenvalue proved usually to be not much greater than the magnitude of the second largest. For this reason, the iterative procedure described in Appendix E for finding the dominant eigenvalue converged only slowly. Moreover, the convergence of the sums over radial harmonics also seemed markedly slower than when $m = 2$. Hence, and also because there is probably less interest in the $m \neq 2$ results, we draw only broad conclusions about them.

a) m = 1 Results

The $m = 1$ disturbances differ from those of all the other angular harmonics in one very important respect: only then does the center of mass of the perturbed density not necessarily coincide with the center of the unperturbed system; in other words, only when $m = 1$ can the first moment of the disturbance density be non-zero. In deriving the integral equation (3.42) which governs the modes, we assumed that all parts of the equilibrium disk remain fixed. Therefore, any $m = 1$ "modes" computed on the basis of that equation are such that the center of mass of the entire system moves. To be more precise, it spirals out from $r = 0$, increasing in radius at the rate e^{st}, while rotating at the angular speed $\tilde{\Omega}_p$.

Clearly what is called for is a procedure which allows the equilibrium disk the freedom to displace laterally in such a fashion that its own first moment exactly cancels that of the perturbed density. The center of mass will then remain fixed. This approach is complicated by the infinite total mass of the models we are studying. However, only the mobile component has infinite total mass, at least when the cut-out index $N \geq 2$; the total mass of what we called the immobile or rigid component then has a respectable, finite value. Therefore, it is the rigid center alone which needs to be moved in order to leave the center of mass of the entire system fixed under an $m = 1$ disturbance. The movement of this

"plug" will cause a perturbation of the equilibrium $1/r$ force field that is over and above the perturbation caused by the disturbance density itself. In the linear approximation[1], the total disturbance force felt by a mobile star is simply the sum of two separate forces — one as usual from the perturbed density of the active component and another from the motion of the rigid matter.

Although we have developed a specific procedure for including in the integral equation the response of the active stars to the motion of the rigid center, our results for the stability of the $m = 1$ fixed-center-of-mass disturbances are, at present, inconclusive.

In the remainder of this section we therefore describe our results only for the naive $m = 1$ disturbances — those for which the equilibrium disk is held fixed. As we mentioned earlier, when m is equal to 1, the iteration to the largest mathematical eigenvalue converges slowly. Largely for this reason we do not present any graphical illustrations of the manner in which the mathematical eigenvalue varies with $\tilde{\Omega}_p$ and $\tilde{s}$.

We find that the eigenvalue $\lambda(\tilde{\sigma}_u, \tilde{\omega})$ again depends upon the frequency $\tilde{\omega}$ in the expected analytic manner; hence, we

---

[1]The magnitude of the motion of the rigid center is directly proportional to the first moment of the disturbance density and that perturbation we have assumed to be infinitesimally small. Even though the first moment consists of an integral over the infinite extent of the disk, it still has a finite value, at least in all cases that we have encountered.

determine whether or not any growing $m = 1$ modes exist by examining the eigenvalues when the growth rate is vanishingly small. Most of our specific results are for the model with cut-out index $N = 2$. Qualitatively, the marginal eigenvalue curve for large $\tilde{\Omega}_p$ is much like the corresponding curve for $m = 2$: for very large $\tilde{\Omega}_p$, λ tends to the origin. For moderate pattern speeds both the real and imaginary parts of λ are positive. Once the pattern speed drops below $\tilde{\Omega}_p = 1$, then the $m = 1$ marginal eigenvalue curve departs sharply from the $m = 2$ curve. The $m = 1$ curve apparently trails off along, but slightly above, the positive $\mathrm{Re}\{\lambda\}$-axis as $\tilde{\Omega}_p \to 0$, whereas we recall that the $m = 2$ curve wound around toward the negative $\mathrm{Re}\{\lambda\}$-axis. To be specific, when $\tilde{\sigma}_u = 0.378070$ $(Q = 1)$, $\mathrm{Re}\{\lambda\}$ is at least as large as 1.7 for $\tilde{\Omega}_p < 0.001$. We have no evidence that the $m = 1$ curve ever crosses the $\mathrm{Re}\{\lambda\}$-axis.

The $m = 1$ eigenvalues appear to come in pairs: the two largest ones differ from each other by only a few percent; likewise the third and fourth largest eigenvalues are quite close. However, the eigenvalues in the second pair have only about 80% of the magnitude of those in the first pair. The fifth and sixth eigenvalues are similarly related with respect to the third and fourth. This pairing off is especially noticeable for small $\tilde{\Omega}_p$. All of the pairs drift out along the positive $\mathrm{Re}\{\lambda\}$-axis, well beyond $\lambda = 1$.

The most significant and surprising feature that these eigenvalue curves imply is that the cut-out disks must have a large number of "naive" $m = 1$ instabilities, even for very hot models ($Q \gtrsim 2$). This is surely the case when $Q = 1$ because the point $\lambda = 1$ lies in the direction of increasing growth rates from the $\tilde{s} = 0^+$ curves of many of the eigenvalues. The fact that eigenvalues at least as large as 1.7 exist here for small $\tilde{\Omega}_p$ suggests that the instabilities will persist even for substantially larger Q. Our brief examinations of several of the other cut-out models indicate that they too are plagued with these instabilities. We have not tracked down the precise stability criterion for these $m = 1$ disturbances because of the numerical difficulty of obtaining decent results when $\tilde{\sigma}_u$ is large and $\tilde{\Omega}_p$ is small.

However, when $Q = 1$ we have tracked down a number of the characteristic frequencies. These are

$\tilde{\Omega}_p$	$\tilde{s}$
0.1407	0.0659
0.1322	0.0590
0.0647	0.0369
0.0570	0.0290

As one can see, these instabilities come in pairs. The slower growing modes are concentrated farther and farther away from the cut-out center, judging from the progressive increase in their corotation radii. Even for the fastest-growing mode, the corotation circle is a factor of 7 beyond the edge in the active density.

The density transform $A(\alpha)$ of this dominant $m = 1$ mode is illustrated in Figure 11. It differs in several respects from the transform of the $m = 2$ mode, which was displayed in Figure 8. Most noticeably, there is a more pronounced leading ($\alpha < 0$) component in the $m = 1$ disturbance. In addition, $|A(\alpha)|$ drops off much more rapidly as $\alpha \to +\infty$ than the $m = 2$ transform does: for instance, when $\alpha = +20$, $|A(\alpha)|$ is already below 0.01% of its peak value, whereas the $m = 2$ transform is more than an order of magnitude larger. Both these tendencies are understandable. We refer once again to the trailing bias of the angular momentum function: in particular, we note that the factor $\tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)$, which produces the exponential decay toward more positive wavenumbers, and mitigates the decay toward more negative wavenumbers, is 0.44 for the $m = 1$ mode but only 0.14 for the $m = 2$ one. Another obvious difference is that the $m = 1$ transform is more oscillatory than the previous one, i.e., the typical distance between successive nodes of either the real or imaginary part is less than half the corresponding distance for the $m = 2$ transform. This will be rationalized below.

The actual disturbance density of this mode is displayed in Figures 12 and 13. The conventions are the same as those used in the corresponding $m = 2$ graphs. The corotation resonance of the present mode is located at $\tilde{r}_{CR} = 7.11$ and the outer Lindblad resonance at $\tilde{r}_{OLR} = 17.2$ (well beyond

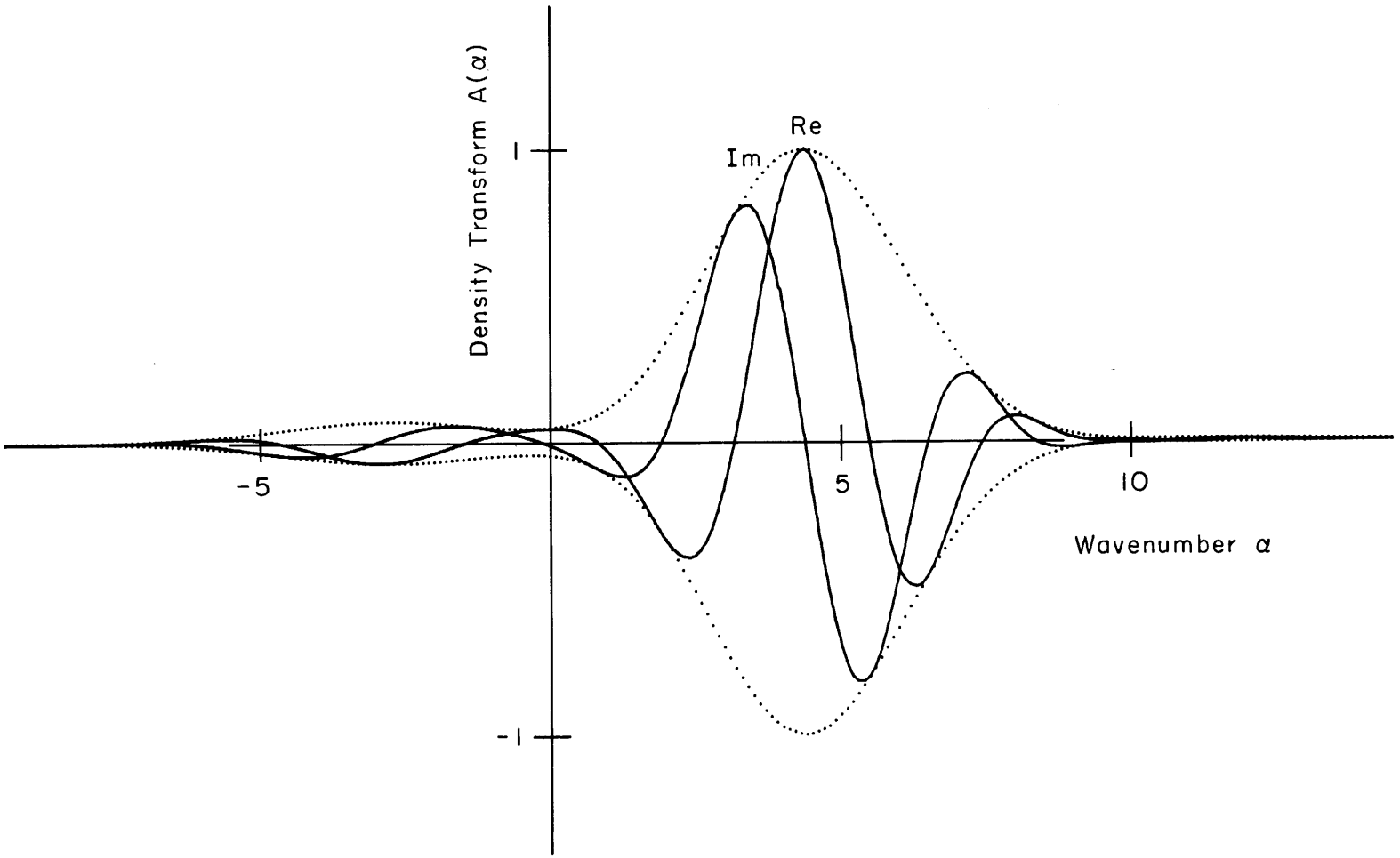

Figure 11. The density transform $A(\alpha)$ for the most unstable $m = 1$ mode of the $N = 2$ disk with $Q = 1$. Solid curves denote real and imaginary parts, and dotted curves the envelope. For this mode, $\tilde{\Omega}_p = 0.1407$ and $\tilde{s} = 0.0659$.

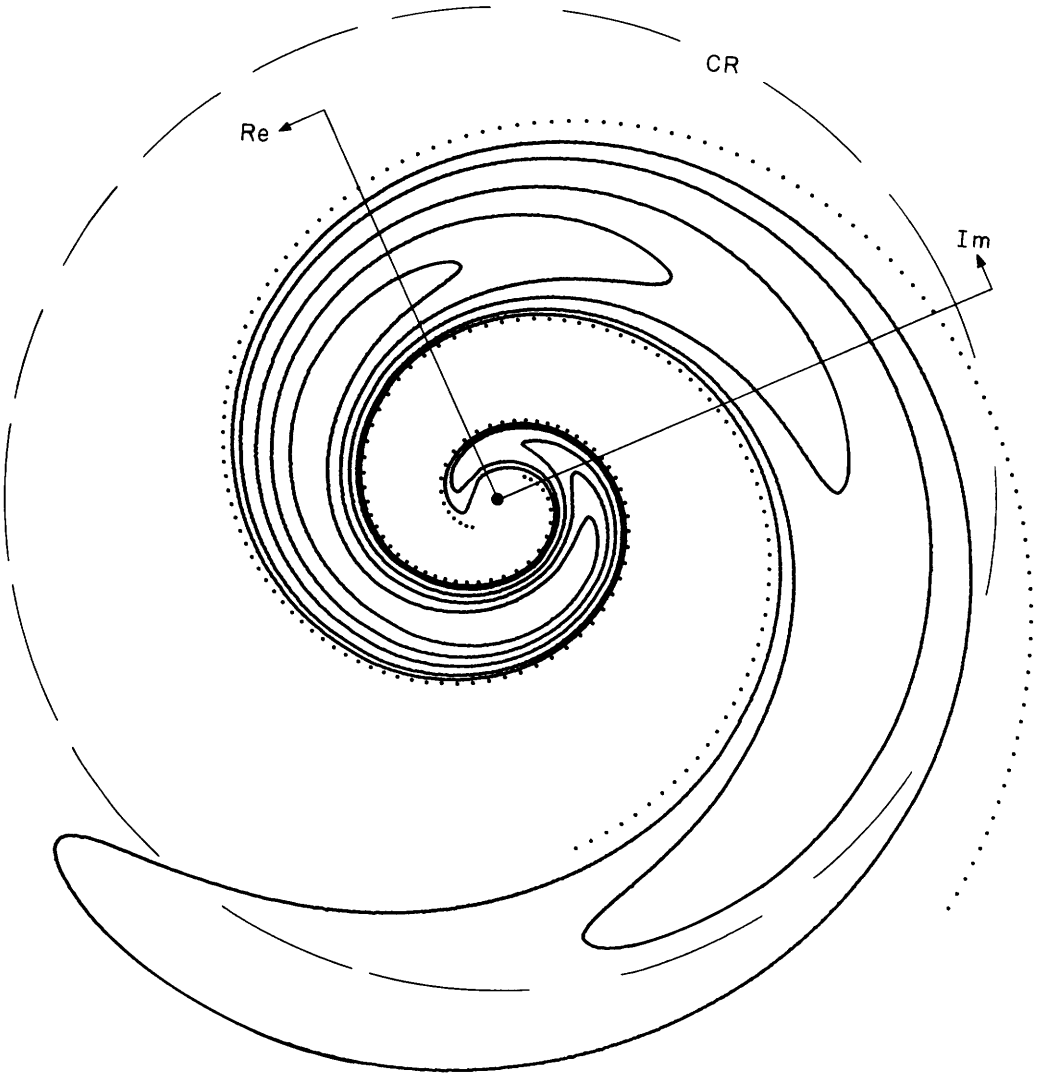

Figure 12. Density contours of the most unstable "naive" $m = 1$ mode of the $N = 2$ disk with $Q = 1$. Solid curves denote 80, 60, 40, 20 and 10% of peak density; dotted curves are the nodes. The lines marked Re and Im indicate the locations of the radial slices shown in Figure 13. Both the model and this growing mode rotate counter-clockwise. Dashed circle marked CR identifies the corotation resonance.

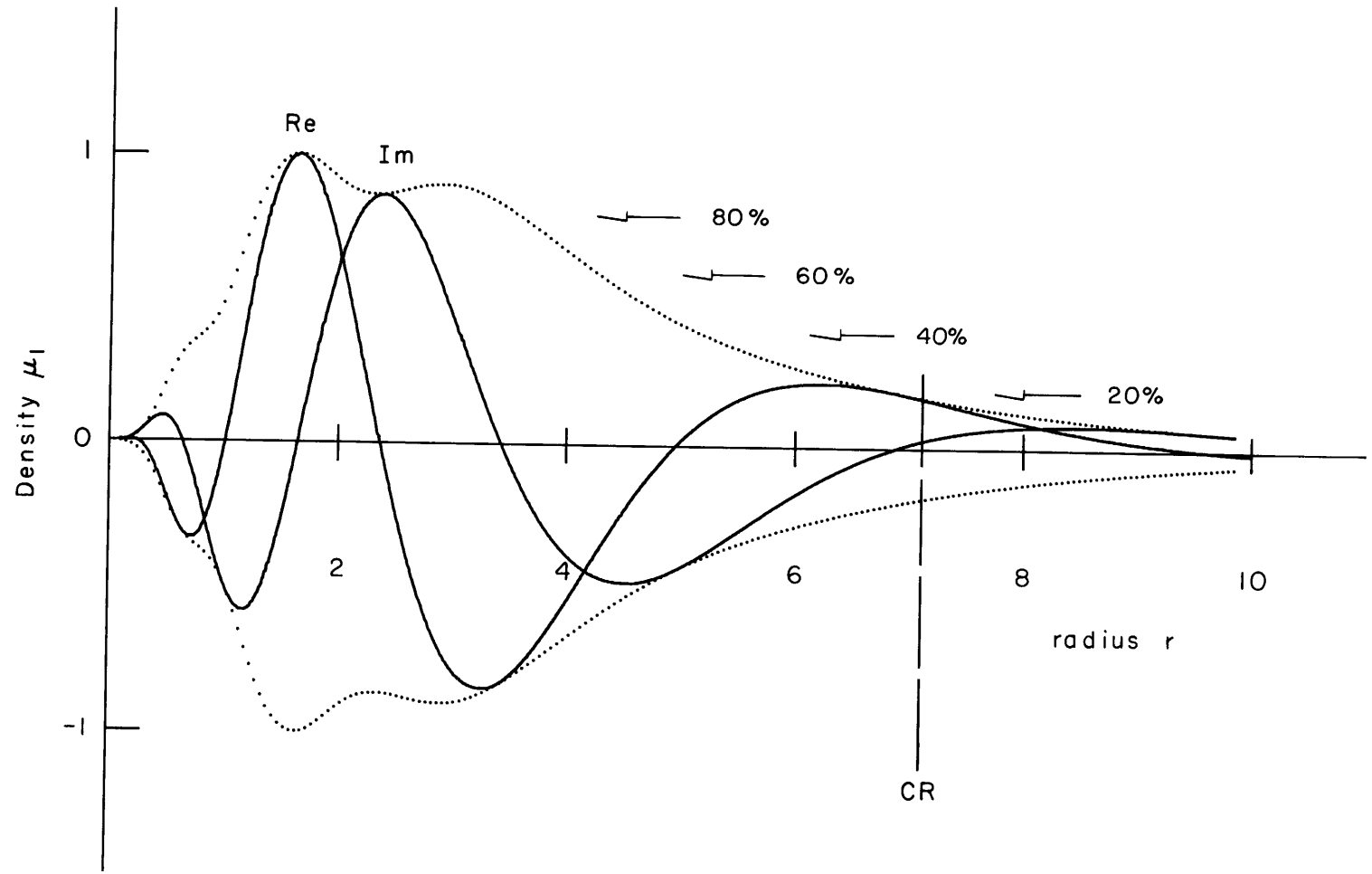

Figure 13. The disturbance density $\mu_1(r)$ for the m = 1 mode shown in Figure 12. Solid curves denote real and imaginary parts, dotted curves the envelope. Arrows signify the contour levels plotted in Figure 12. Corotation resonance (CR) is marked by the dashed line. Outer Lindblad resonance occurs at r = 17.2 ; no inner Lindblad resonance exists for this mode.

the range covered in these diagrams). No inner Lindblad resonance exists for this or any other $m = 1$ mode in our galaxy models, since $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U}) < 0$ for all $\tilde{U}$ when $m = 1$ and $\ell = -1$. Figure 13 shows clearly that even though this $m = 1$ mode has no inner Lindblad resonance, which as we recall from our introductory remarks tends to absorb waves, the disturbance still decays as $r \to 0$. Possibly this is due to the small amount of active mass at the very center. We also observe that the disturbance peaks between $\tilde{r} = 1$ and $\tilde{r} = 4$. The mode here is not so concentrated near the active density edge as the $m = 2$ modes proved to be; we of course expect this simply from the differing pattern speeds. Incidentally, this helps explain why the $m = 1$ density transform is the more oscillatory of the two modes illustrated in this thesis, since even a crude sort of stationary phase argument suggests that the Fourier integral

$$\int_{-\infty}^{\infty} A(\alpha)\ e^{i\alpha\ \ell n\ \tilde{r}}\ d\alpha$$

will be greatest when the integral as a whole is the least oscillatory; near the peak of this integral, typical values of $\ell n\ \tilde{r}$ are larger for the $m = 1$ mode and consequently $A(\alpha)$ can be expected to be more oscillatory.

We close this sub-section with the reminder that these $m = 1$ instabilities may actually be fictitious — we cannot yet say whether they remain when the center of mass is fixed.

b) m = 3 Results

The explorations we have made of the $m = 3$ angular harmonic have been very brief. In short, we first confirmed that our numerical procedures handled this case adequately and then we determined whether or not any $m = 3$ instabilities were present in models with an axisymmetric stability parameter Q of unity.

The reason for the specific check of the numerical procedures here was the concern about the $\ell = -2$ radial harmonic: as we noted in section III.e, when $m = 3$ and $\ell = -2$ the term $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ which occurs in the angular momentum function changes sign as $\tilde{U}$ increases. The worry was that the change of character in the angular momentum function that accompanied the change in sign of $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ [see Appendix D] would create inaccuracies in the Gauss-Laguerre integration over eccentric velocities. Since this term equals zero for $\tilde{U} \simeq 1.19$, the trouble, if any, would occur in the neighborhood of that eccentric velocity. To test this, we compared the largest eigenvalue of an $N = 2$ cut-out disk when it was obtained from a 5-point Gauss-Laguerre scheme with the fourth point situated at $\tilde{U} = 1.18$ with that obtained when the fourth point was located at $\tilde{U} = 1.20$. The results agreed to better than 4 digits. For a 6-point quadrature there was similar agreement between the results as the fourth point straddled $\tilde{U} = 1.19$. Evidently the vanishing of $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ does not cause any serious numerical problems.

As usual we determined whether any growing $m = 3$ modes were possible in a given model by finding where the dominant marginal eigenvalue curve crossed the positive Re{λ}-axis. For the $N = 2$ model with $\tilde{\sigma}_u = 0.378$ ($Q = 1$) this happens at $\lambda = 0.31$ when $\tilde{\Omega}_p = 0.62$. For large $\tilde{\Omega}_p$, λ clearly tends to the origin. At the other extreme, the dominant eigenvalue[2] has been followed down to $\tilde{\Omega}_p$ as small as 0.05. At that point it has a magnitude of roughly 1.5 but a phase of $-50°$ — well into the fourth quadrant. Even though this curve has not been pursued to small enough pattern speeds to determine its destination as $\tilde{\Omega}_p \to 0^+$, there is no indication that it ever re-crosses the positive Re{λ}-axis. Therefore, we submit that no $m = 3$ instabilities are present in this cut-out model when $Q = 1$; moreover, judging from the magnitude of λ at its purely real location, three-armed instabilities are not likely to surface until $\tilde{\sigma}_u \leq 0.1$ (assuming that, as for $m = 2$, $\lambda \sim \tilde{\sigma}^{-1}$).

By comparison, the corresponding $m = 2$ eigenvalue curve has $\lambda = 0.60$ as its value on the positive Re{λ}-axis. Hence the $N = 2$ model is even more stable to $m = 3$ disturbances than it is to $m = 2$ ones.

---

[2] We wish to note here that when $\tilde{\Omega}_p \sim 0.20$, a previously sub-dominant eigenvalue becomes the largest one. (Recall that a similar feature was observed for the $m = 2$ eigenvalues of the $N = 1$ disk.) In fact, a closer examination reveals that many of the $m = 3$ eigenvalue curves cross each other. When $m = 3$, then, the curve consisting of the largest eigenvalue for a given $\tilde{\Omega}_p$ has a discontinuity near $\tilde{\Omega}_p \sim 0.20$. Still, none of the eigenvalue curves comes anywhere close to $\lambda = 1$.

In contrast, the $N = 3$ and $N = 4$ cut-out models are unstable to $m = 2$ disturbances for this $\tilde{\sigma}_u$. But even for the $N = 4$ disk, the $m = 3$ disturbances are still well under control when $Q = 1$ — the dominant eigenvalue has a strictly real value of only $\lambda = 0.61$ (when $\tilde{\Omega}_p = 0.63$).

c) m = 4 Results

The final angular harmonic that we have examined is the $m = 4$ one. Unlike the previous case, $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$ never vanishes when $m = 4$; thus there was no need to check the numerical procedures here. Again the focus was put on the $N = 2$ cut-out model when $Q = 1$. The dominant eigenvalue curve here is similar to that for $m = 2$: for very large $\tilde{\Omega}_p$, $\lambda \to 0$; as $\tilde{\Omega}_p$ decreases, λ swings clockwise from the first quadrant through the fourth one and into the third, possibly also tending to the negative $\mathrm{Re}\{\lambda\}$-axis as $\tilde{\Omega}_p \to 0^+$. The value of λ when it crosses the positive $\mathrm{Re}\{\lambda\}$-axis is 0.22 (at $\tilde{\Omega}_p = 0.65$). By the time $\tilde{\Omega}_p$ has reached 0.10 , $|\lambda|$ has grown as large as 1/2 , but by then its phase is beyond $-90°$. We have no evidence that $|\lambda|$ ever exceeds 1 , let alone approaches the point $\lambda = 1$. Thus we conclude that $m = 4$ instabilities are even harder to obtain than $m = 3$ ones.

The same conclusion arises for the $N = 4$ model when $Q = 1$. There the $\mathrm{Re}\{\lambda\}$-axis is intersected at $\lambda = 0.40$ (when $\tilde{\Omega}_p = 0.54$), compared with $\lambda = 0.63$ when $m = 3$.

Of the five angular harmonics that we have examined, then, the $m = 4$ disturbances are the tamest.

As a final remark, observe that for both $m = 3$ and $m = 4$, the eigenvalues are in the neighborhood of the positive $\mathrm{Re}\{\lambda\}$-axis only for moderate pattern speeds. The implication of this is that even when $\tilde{\sigma}_u$ is low enough for there to be instabilities of these types, they, like the $m = 2$ ones, are surely related in some fashion to the edge in the active density.

VII. SUMMARY AND DISCUSSION

In the preceding chapters we examined a class of galaxy models with uniform circular velocity. In these models the disk component was taken to be increasingly unresponsive toward the center. As we remarked, this rigidity can be interpreted as either a centrally concentrated, high-velocity disk component or else as a high-velocity halo (although Appendix B indicates that disks with cut-out indices $N > 2$ do not admit spherical halos whose volume densities are everywhere non-negative).

Determined as it is by conditions at large radii, the minimum velocity dispersion required to stabilize these models against all axisymmetric disturbances is the same for each cut-out disk — and for the singular disk — specifically, $\sigma_{u,min} \cong 0.3781\ V_o$. Remarkably, this global result differs by only about 3 parts in 10^4 from the estimate provided by local theory. When the velocity dispersion is less than this critical value the disks admit one or more large-scale, exponentially growing modes. The shapes of such axisymmetric modes in a cut-out disk are about what one would expect: the disturbance density is more or less sinusoidal in an annular region bounded on the inside by the edge of the active

density and bounded loosely on the outside by the radius at which the stars no longer have time (in terms of their own "epicyclic clocks") to respond to the perturbation.

Our analysis of the disturbances with angular wavenumber $m = 1$ is still incomplete. It would have been desirable to admit only perturbations which do not shift the center of mass of the system. Thus far, however, we have been able to consider with good numerical accuracy only disturbances which do not shift the rigid center. All of our cut-out disks seem unstable to this sort of disturbance, even for velocity dispersions well in excess of that needed to eliminate axisymmetric instabilities. We have especially examined one such growing, off-center instability in a disk with cut-out index $N = 2$ which was just hot enough to prevent the growth of axisymmetric disturbances. Most of the "activity" in this density perturbation occurs substantially outside the inactive central region. This disturbance can conceivably penetrate to the center of the disk since no inner Lindblad resonance exists to absorb waves at small radii. Unfortunately, it is not clear yet whether analogous disturbances will prove to be unstable for moderate velocity dispersions when a proper analysis with a fixed center of mass (but laterally mobile "rigid" component) has been completed.

Among the remaining non-axisymmetric disturbances, the bisymmetric ($m = 2$) modes are, as expected, the most demanding ones in the cut-out disks; in all of these models, distinctly

more velocity dispersion is required to suppress the last growing $m = 2$ mode than is needed to eliminate the $m = 3$ and $m = 4$ instabilities.

Nevertheless, we were astonished to discover that even these $m = 2$ disturbances in the relatively gently cut-out $N = 1$ and $N = 2$ disks are <u>easier</u> to stabilize than the axisymmetric $m = 0$ perturbations. This contrasts sharply with the marked "two-armed" and/or bar-like instabilities found in many other models, such as the gaseous disks studied by Bardeen, several of the uniformly rotating disks examined by Kalnajs, the disks with modified gravity considered by Erickson, the numerous N-body experiments conducted by Hohl & Hockney and by Miller, Prendergast & Quirk, and even the limited simulations by Ostriker & Peebles of disk-halo systems. In all but one of those models the dominant instability which remained after axisymmetric stability had been achieved was one of type $m = 2$. (The exception was a model studied by Kalnajs in which an $m = 3$ instability was the most persistent.) Bisymmetric instabilities have been particularly evident in the N-body computer simulations cited. Our $N = 1$ and $N = 2$ models, then, are distinct exceptions to this usual behavior.

Admittedly, when the center is more strongly cut out, as in our $N = 3$ and $N = 4$ models, even our dominant instability has a bisymmetric character. However, the spiral density patterns of such growing "two-armed" or $m = 2$ modes

leave little doubt that their unstable behavior is somehow due to the edge of the mobile matter, for the disturbance then is greatest near that edge itself, and its angular speed tends to place the inner Lindblad resonance noticeably within the "hole". We have examined several of these modes in detail and have found them to exhibit a distinctly trailing spiral structure; by-and-large, this structure seems confined to the annular region between the two Lindblad resonances (one of which, as we implied, lies fairly deep within the inactive region). Significantly, however, all four of the cut-out disks specifically examined here require much less random motion relative to their rotational motion to achieve $m = 2$ stability than the amount Ostriker and Peebles have suggested is generally characteristic of galaxy models.

Until we have resolved the question of the $m = 1$, fixed-center-of-mass stability, we cannot assert that our models are really more stable than other current ones. If genuine instabilities analogous to our naive $m = 1$ troubles are in fact present, then our models will have simply exchanged a serious $m = 2$ instability for another of type $m = 1$. So to speak, we may merely have jumped from the frying pan into the fire.

APPENDIX A

AN ALTERNATIVE TO THE CUT-OUT MODELS

i) Description

In section II.c of the text we described the cut-out disks, which were introduced because of the rapid angular motions near the center of the singular disk. This appendix will describe another appealing modification of the basic self-similar model which has the same effect. As in the case with the cut-out models, the differences between the singular disk and the model discussed here are largely confined to the center; elsewhere a strong resemblance to the singular disk prevails.

The connection between the surface density and the rotation law for this model follows from the well-known fact [see, for example, Toomre (1963)] that there is a linear relationship between the surface density $\mu(r)$ and the square of the rotation law $V^2(r)$, i.e., given the pairs $[\mu_1(r), V_1^2(r)]$ and $[\mu_2(r), V_2^2(r)]$ for two galaxy models, other valid models can be obtained by forming linear combinations of these pairs. Recently Rybicki (1974, private communication) noticed, in effect, that one may form the difference between the singular disk which has

$$\mu(r) = \mu_o(r_o/r)$$

$$V^2(r) = V_o^2$$

and Toomre's Model 0 (1963) which has

$$\mu(r) = \mu_o \left[(r_o/r) - r_o/\sqrt{r^2 + r_o^2}\right]$$

$$V^2(r) = V_o^2 \, r_o/\sqrt{r^2 + r_o^2}$$

to achieve a galaxy model with

$$\mu(r) = \mu_o r_o/\sqrt{r^2 + r_o^2} \tag{A.1}$$

$$V^2(r) = V_o^2 \left[1 - r_o/\sqrt{r^2 + r_o^2}\right] . \tag{A.2}$$

This model will be referred to here as the regular disk. The constants are related by $V_o^2 = 2\pi G \mu_o r_o$. The gravitational potential follows immediately upon integration of the radial force $-V^2(r)/r$:

$$\phi(r) = V_o^2 \, \ell n\left[1 + \sqrt{1 + (r/r_o)^2}\right] . \tag{A.3}$$

In this axisymmetric potential the constants of motion are, explicitly, the energy

$$E = (u^2 + v^2)/2 + V_o^2 \, \ell n\left[1 + \sqrt{1 + (r/r_o)^2}\right] \tag{A.4}$$

and the angular momentum

$$J = rv . \tag{A.5}$$

Two other important properties are the angular rotation rate of the disk, given by

$$\Omega(r) = (V_o/r)\ \{1 - r_o/\sqrt{r^2 + r_o^2}\}^{1/2} \qquad (A.6)$$

and the epicyclic frequency, given by

$$\kappa^2(r) = (2V_o^2/r^2)\ [1 - r_o/\sqrt{r^2 + r_o^2}] + V_o^2\ r_o/(r^2 + r_o^2)^{3/2}. \qquad (A.7)$$

Notice that for large values of r the properties of the regular disk approach those of the singular disk, in particular $\mu(r) \to \mu_o(r_o/r)$ and $V(r) \to V_o$. Hence at large radii the two models are indistinguishable. Near the center, however, the two disks differ markedly: the surface density of the singular disk increases without bound, whereas that of the regular disk is finite; at the center of the singular disk, as indeed everywhere else, the rotation speed $V(r)$ is a constant, whereas at the center of the regular disk it is the angular rotation rate that is nearly constant. More to the point, both the angular velocity and the epicyclic frequency of the regular disk are finite at the center — this model does not have the rapidly rotating stars that plague the singular disk at the origin. Note that the regular disk is a more drastic modification of the singular disk than the cut-out models discussed in the text, for the regular disk is not self-similar. Consequently, there are significant differences in the stellar orbits from radius to radius, especially near the origin.

ii) The Distribution Function

Our contribution consists of furnishing an exact distribution function for the regular disk. We looked for one of the form

$$f_o(E,J) = g(J)\ e^{-E/\sigma_u^2} \quad (J \geq 0)\ , \tag{A.8}$$

which satisfies the requirement of self-consistency:

$$\int_{-\infty}^{\infty} \int_{0}^{\infty} f_o(E,J)\ du\ dv = \mu_o\ r_o/\sqrt{r^2 + r_o^2}\ . \tag{A.9}$$

A distribution function which meets these conditions is

$$f_o(E,J) = (\mu_o/\pi\sigma_u)\ e^{-E/\sigma_u^2} \quad \times$$
$$\sum_{k=0}^{\infty} \binom{\lambda}{k} M[-(k-1)/2\ ,\ 1/2\ ;\ -J^2/2\sigma_u^2 r_o^2]\ , \tag{A.10}$$

where

$$\lambda = V_o^2/\sigma_u^2\ , \tag{A.11}$$

$$\binom{\lambda}{k} = \Gamma(\lambda+1)/[\Gamma(k+1)\ \Gamma(\lambda-k+1)]\ , \tag{A.12}$$

and $M[a\ ,\ b\ ;\ x]$ is the confluent hypergeometric function. The verification of this result is delayed until the next subsection of this appendix.

Whenever $\lambda = n$, an integer, the sum in eqn. (A.10) terminates, running from $k = 0$ to $k = n$. Moreover, $\binom{n}{k}$ is then the usual binomial coefficient. For this special

case one can show rather simply that $f_o(E,J) \geq 0$, as required by any physical distribution function. Referring to the identity

$$M[-(k-1)/2 \ , \ 1/2 \ ; \ -x] = e^{-x} \ M[k/2 \ , \ 1/2 \ ; \ x] \qquad (A.13)$$

[Abramowitz & Stegun (1965, p.504)] and to the series expansion for the confluent hypergeometric function

$$M[k/2 \ , \ 1/2 \ ; \ x] = 1 + \frac{(k/2)}{(1/2)1!} + \frac{(k/2)(k/2+1)}{(1/2)(3/2)2!} + \ . \ . \ .$$

[Abramowitz & Stegun (1965,p.504)], as well as to the non-negativeness of the binomial coefficients, it is clear that each term in the sum in eqn. (A.10) from $k = 0$ to $k = n$ is non-negative.

Fortunately the distribution function can be expressed entirely in terms of polynomials, exponentials and error functions. This is achieved by simplifying the two classes of confluent hypergeometric functions that occur in the sum. The first class corresponds to odd values of k in the sum. We use the notation

$$G_q(h) = M[-q \ , \ 1/2 \ ; \ -h^2] \ , \qquad (A.14)$$

where q is a non-negative integer, to denote these terms. The second class corresponds to even values of k and its members are denoted $F_p(h)$, where

$$F_p(h) = M[-p+1/2 \ , \ 1/2 \ ; \ -h^2] \ , \tag{A.15}$$

and p is also a non-negative integer. The distribution function is now written as

$$f_o(E,J) = (\mu_o/\pi\sigma_u^2) \ e^{-E/\sigma_u^2} \ \{ \sum_{p=0}^{\infty} \binom{\lambda}{2p} \ F_p(J/\sqrt{2} \ \sigma_u r_o) + \sum_{q=0}^{\infty} \binom{\lambda}{2q+1} \ G_q(J/\sqrt{2} \ \sigma_u r_o) \} \ . \tag{A.16}$$

It is shown in the next section that

$$G_q(h) = \sum_{j=0}^{q} \binom{q}{j} \ \frac{\Gamma(1/2)}{\Gamma(j+1/2)} \ h^{2j} \tag{A.17}$$

and that $F_p(h)$ can be found from the recurrence relation

$$F_{p+1}(h) = \frac{4p-1+2h^2}{2p} \ F_p(h) - \frac{2p-1}{2p} \ F_{p-1}(h) \qquad (p \geq 1) \ , \tag{A.18}$$

with

$$F_0(h) = e^{-h^2} \tag{A.19}$$

$$F_1(h) = e^{-h^2} + \sqrt{\pi} \ h \ \mathrm{erf}(h) \ , \tag{A.20}$$

where $\mathrm{erf}(h)$ is the error function:

$$\mathrm{erf}(h) = (2/\sqrt{\pi}) \int_0^h e^{-t^2} \ dt \ .$$

Indeed, when $\lambda = n$ the distribution function has the convenient form

$$f_o(E,J) = (\mu_o/\pi\sigma_u)\ e^{-E/\sigma_u^2}\ \{P_1(h^2) + P_2(h^2)\ e^{-h^2} + \sqrt{\pi}\ h\ P_3(h^2)\ \mathrm{erf}(h)\}\ , \qquad (A.21)$$

where $h = J/\sqrt{2}\ \sigma_u r_o$, $P_1(x)$ is a rational polynomial of degree $\leq (n-1)/2$ and $P_2(x)$ and $P_3(x)$ are rational polynomials of degree $\leq (n-2)/2$. Several examples are given in the last section of this appendix.

Thus, whenever $\lambda = n$ the distribution function may be expressed in a comparatively simple closed form, albeit of increasing complexity as n increases. Since $\lambda = V_o^2/\sigma_u^2$, a sequence of models with diminishing velocity dispersion is obtained as n runs through the integers 1, 2, 3, ... The first member of this sequence — corresponding to $n = 1$ — is the very hot model with $\sigma_u = V_o$. As n increases the corresponding models in the sequence approach ever closer to the cold disk ($\sigma_u = 0$) .

The velocity dispersions of the $n = 6$ and $n = 7$ models straddle the estimate of that minimum dispersion needed for the disk to be everywhere locally stable to axisymmetric disturbances: Using Toomre's criterion, this minimum dispersion is estimated to be

$$\sigma_{u,\min}(r) = \frac{(r/r_o)\ [1 + (r/r_o)^2]^{1/4}}{\{[1 + (r/r_o)^2]^{3/2} - (r/r_o)^2/2 - 1\}^{1/2}}\ \sigma_{u,s}\ ,$$

where $\sigma_{u,s} = 0.3779\ V_o$ is the local estimate for the singular disk. For $r << r_o$, the local estimate for the regular

disk is

$$\sigma_{u,min}(r) \doteq [1 + (r/r_o)^2/4]\ \sigma_{u,s}$$

and for $r >> r_o$ it is

$$\sigma_{u,min}(r) \doteq [1 + (r/r_o)^{-2}/4]\ \sigma_{u,s} .$$

The ratio $\sigma_{u,min}(r)/\sigma_{u,s}$ rises gently from unity at $r = 0$ to a value of 1.032 for $r = r_o$, reaches a maximum value of 1.046 when $r = 2.197\ r_o$ and thereafter decreases asymptotically to unity as r tends to infinity. Thus, for $n = 7$, or $\sigma_u = V_o/\sqrt{7} = 0.378\ V_o$, the regular disk is everywhere close to marginal axisymmetric stability. For $n = 6$, or $\sigma_u = 0.408\ V_o$, the regular disk is already everywhere marginally stable in the local axisymmetric sense. The distribution functions for these two cases are included among the examples at the end of this appendix.

iii) Verification

The first task of this section is the confirmation of the distribution function given in eqn. (A.10) for the regular disk. Inserting the functional form of eqn. (A.8) into the requirement of eqn. (A.9) and performing the integration over u, we find that $g(J)$ must satisfy

$$\int_0^\infty g(rv)\ e^{-v^2/2\sigma_u^2}\ e^{\lambda\ \ell n[1 + \sqrt{1 + (r/r_o)^2}]}\ dv = (\mu_o/\sqrt{2\pi}\sigma_u)/\sqrt{1 + (r/r_o)^2} ,$$

or

$$\int_0^\infty g(rv)\ e^{-v^2/2\sigma_u^2}\ dv = \frac{\mu_o}{\sqrt{2\pi}\sigma_u}\ \frac{[1 + \sqrt{1 + (r/r_o)^2}]^\lambda}{\sqrt{1 + (r/r_o)^2}} \ . \quad (A.22)$$

The solution of this integral equation for $g(J)$ was found by using Mellin transforms. This is a very tedious calculation and here we will be content simply to confirm that

$$g(J) = (\mu_o/\pi\sigma_u) \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)} \times M[-(k-1)/2\ ,\ 1/2\ ;\ -J^2/2\sigma_u^2 r_o^2] \quad (A.23)$$

does indeed satisfy eqn. (A.22).

Use the identity of eqn.(A.13) to write eqn. (A.23) as

$$g(J) = (\mu_o/\pi\sigma_u)\ e^{-J^2/2\sigma_u^2 r_o^2} \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)} \times M[k/2\ ,\ 1/2\ ;\ J^2/2\sigma_u^2 r_o^2] \ . \quad (A.24)$$

Denote the left-hand side of eqn. (A.22) by I_λ . Using the preceding form for $g(J)$ in eqn. (A.22) and then interchanging the order of integration and summation[1] we find

$$I_\lambda = (\mu_o/\pi\sigma_u) \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)} \int_0^\infty e^{-v^2(r^2 + r_o^2)/2\sigma_u^2 r_o^2} \times M[k/2\ ,\ 1/2\ ;\ v^2r^2/2\sigma_u^2 r_o^2]\ dv \ .$$

[1] This is surely valid for integral values of λ since the sum is then finite. For non-integral values of λ we need to demonstrate that the sum in eqn. (A.24) is uniformly convergent in order to justify this operation. But since a demonstration that $f_o(E,J) \geq 0$ is lacking for this general case, and moreover, since a convenient sequence of fully justified models is given by the models with integral λ we have not given a general justification for this operation.

The substitution $x = v^2r^2/2\sigma_u^2r_o^2$ leads to

$$I_\lambda = (\mu_o/\pi\sigma_u)\ (\sigma_u r_o/\sqrt{2}r) \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)} \int_0^\infty \sqrt{x}$$
$$\times\ e^{-x(r^2+ r_o^2)/r_o^2}\ M[k/2\ ,\ 1/2\ ;\ x]\ dx\ .$$

According to Gradshteyn and Ryzhik (1965, p.860),

$$\int_0^\infty e^{-sx}\ x^{c-1}\ M[a\ ,\ c\ ;\ x]\ dx = \Gamma(c)\ s^{-c}\ (1 - s^{-1})^{-a}\ .$$

Therefore, since $\Gamma(1/2) = \sqrt{\pi}$

$$I_\lambda = (\mu_o r_o/\sqrt{2}\pi\sigma_u r) \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)} \sqrt{\pi}\ \ [(r^2 + r_o^2)/r^2]$$
$$\times\ [1 - r^2/(r^2 + r_o^2)]^{-k/2}$$
$$= (\mu_o/\sqrt{2\pi}\sigma_u)\ [1 + (r/r_o)^2]^{-1/2}$$
$$\times \sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)}\ [1 + (r/r_o)^2]^{k/2}\ .$$

But

$$\sum_{k=0}^{\infty} \frac{\Gamma(\lambda+1)}{k!\ \Gamma(\lambda-k+1)}\ [1 + (r/r_o)^2]^{k/2} = [1 + \sqrt{1 + (r/r_o)^2}]^\lambda$$

(this is the general binomial expansion). Thus,

$$I_\lambda = (\mu_o/\sqrt{2\pi}\sigma_u)\ [1 + \sqrt{1 + (r/r_o)^2}]^\lambda/\sqrt{1 + (r/r_o)^2}\ ,$$

and the solution in eqn. (A.23) is confirmed.

We now show that the functions $G_q(h)$ and $F_p(h)$ introduced in eqns. (A.14) and (A.15) have the properties cited in eqns. (A.17) - (A.20). The first case is rather simple. The series expansion for $M[-q\ ,\ 1/2\ ;\ -h^2]$ terminates after $(q+1)$ terms [see Abramowitz and Stegun (1965, p. 504)]:

$$M[-q\ ,\ 1/2\ ;\ -h^2] = 1 + \frac{(-q)}{(1/2)\ 1!}(-h^2) + \frac{(-q)(-q+1)}{(1/2)(3/2)\ 2!}(-h^2)^2 + \cdots + \frac{(-q)(-q+1)\cdots(-1)}{(1/2)(3/2)\cdots(1/2+q-1)q!}(-h^2)^q .$$

So,

$$G_q(h) = \sum_{j=0}^{q} \binom{q}{j} \frac{\Gamma(1/2)}{\Gamma(1/2+j)} h^{2j} \tag{A.25}$$

since $(1/2)(3/2)\cdots(1/2+j-1) = \Gamma(1/2+j)/\Gamma(1/2)$ and

$$\frac{(-q)(-q+1)\cdots(-q+j-1)}{j!} = (-1)^q \frac{(q)(q-1)\cdots(q-j+1)}{j!} = (-1)^q \binom{q}{j} .$$

The second case is more complicated. The function $F_0(h)$ is obtained by consulting the identity (A.13) which implies

$$M[1/2\ ,\ 1/2\ ;\ -h^2] = e^{-h^2}\ M[0\ ,\ 1/2\ ;\ h^2] = e^{-h^2},$$

since

$$M[0\ ,\ 1/2\ ;\ x] = 1 .$$

Therefore,

$$F_0(h) = e^{-h^2} . \tag{A.26}$$

We now refer to Abramowitz and Stegun (1965, p. 509) to find that

$$M[1/2 , 3/2 ; -h^2] = (\sqrt{\pi}/2h)\ \mathrm{erf}(h) ,$$

where $\mathrm{erf}(h)$ is the error function

$$\mathrm{erf}(h) = (2/\sqrt{\pi}) \int_0^h e^{-t^2}\ dt .$$

Next, use the recurrence relation

$$M[1/2 , 1/2 ; -h^2] - M[-1/2 , 1/2 ; -h^2] + 2h^2\ M[1/2 , 3/2 ; -h^2] = 0$$

[Abramowitz and Stegun (1965, p. 506, #13.4.4)], which implies that

$$M[-1/2 , 1/2 ; -h^2] = M[1/2 , 1/2 ; -h^2] + 2h^2\ M[1/2 , 3/2 ; -h^2] ,$$

or

$$F_1(h) = e^{-h^2} + \sqrt{\pi}\ h\ \mathrm{erf}(h) . \qquad \text{(A.27)}$$

To find the remaining $F_p(h)$ we use the recurrence relation

$$p\ M[-p-1/2 , 1/2 ; -h^2] + (-2p + 1/2 - h^2) \times M[-p-1/2 , 1/2 ; -h^2] + (p - 1/2)\ M[-p+3/2 , 1/2 ; -h^2] = 0$$

[Abramowitz and Stegun (1965, p. 506, #13.4.3)] to deduce that

$$F_{p+1}(h) = \frac{4p-1+2h^2}{2p} F_p(h) - \frac{2p-1}{2p} F_{p-1}(h) \qquad (p \geq 1) \ . \qquad (A.28)$$

Thus, eqns. (A.26) - (A.28) enable us to find $F_p(h)$ for all integers $p \geq 0$ in terms of polynomials, exponentials and error functions, whereas $G_q(h)$ is given by eqn. (A.25) as a polynomial alone.

ii) Examples

Finally, we give several examples of these distribution functions. Let $f_n(E,J)$ denote the distribution function for the regular disk for which $(V_o/\sigma_u)^2 = n$ and let $h = J/\sqrt{2}\sigma_u r_o$. Then

$$f_1(E,J) = (\mu_o/\pi\sigma_u^2) \{1 + e^{-h^2}\} e^{-E/\sigma_u^2}$$

$$f_2(E,J) = (\mu_o/\pi\sigma_u^2) \{2 + 2e^{-h^2} + \sqrt{\pi}\, h\, \mathrm{erf}(h)\} e^{-E/\sigma_u^2}$$

$$f_4(E,J) = (\mu_o/\pi\sigma_u^2) \{[8 + 8h^2] + [8 + h^2] e^{-h^2} + \sqrt{\pi}\, h\, [15/2 + h^2]\, \mathrm{erf}(h)\} e^{-E/\sigma_u^2}$$

$$f_6(E,J) = (\mu_o/\pi\sigma_u^2) \{[32 + 64h^2 + 8h^4] + [32 + 69/4\, h^2 + 1/2\, h^4] e^{-h} + \sqrt{\pi}\, h\, [315/8 + 35/2\, h^2 + h^4]\, \mathrm{erf}(h)\} e^{-E/\sigma_u^2}$$

$$f_7(E,J) = (\mu_o/\pi\sigma_u^2) \{[64 + 160h^2 + 32h^4 + 8/15\, h^6] + [64 + 203/4\, h^2 + 7/2\, h^4] e^{-h^2} + \sqrt{\pi}\, h\, [693/8 + 105/2\, h^2 + 7/2\, h^4]\, \mathrm{erf}(h)\} e^{-E/\sigma_u^2}$$

APPENDIX B

DENSITY OF A SPHERICAL HALO EQUIVALENT TO A RIGID DISK COMPONENT

In section II.c of the text we offered two interpretations of the rigid stellar component in a cut-out disk: a two-dimensional disk component or a three-dimensional "halo". If the first interpretation is adopted, the immobile surface density $\mu_i(r)$ can be found by subtracting the active density $\mu_a(r)$ from the density of the singular disk $\mu_s(r)$. The active surface density must be found numerically since the integration of the active distribution function over velocity space [see eqn. (2.65)] cannot be performed analytically, at least not by this writer. Alternatively, if the second interpretation — that the active disk is immersed in a three-dimensional halo — is adopted, the rigid volume density must be so chosen that the forces it generates in the plane of the active material are precisely those that a rigid disk component would generate. We furnish here a convenient closed form expression for a spherically symmetric halo that accomplishes this.

We start by finding an expression for the forces in the plane of symmetry due to the rigid disk component. The logarithmic spiral representation of Kalnajs is a convenient tool for this task (see section III.b). It enables us to relate the rigid surface density easily to the gravitational

potential from the rigid stars (in the plane of the disk). This potential is denoted by $\phi_i(r)$. The radial force that we are after, $F_{i,r}(r)$, is simply $-d\phi_i/dr$. The rigid surface density can be expressed as an integral over velocity space:

$$\mu_i(r) = \int_{-\infty}^{\infty} \int_{0}^{\infty} [1 - H(J)]\, f_s(E,J)\, du\, dv \ , \qquad (B.1)$$

where the singular disk distribution function $f_s(E,J)$ is explained in section II.a and the cut-out factor $H(J)$ in section II.c. Whenever the cut-out index N is greater than or equal to 2 , the rigid surface density has the logarithmic spiral representation:

$$\mu_i(r) = \mu_o\, (r/r_o)^{-3/2} \int_{-\infty}^{\infty} B(\alpha)\, e^{i\alpha\, \ell n(r/r_o)}\, d\alpha \ , \qquad (B.2)$$

where

$$B(\alpha) = (2\pi)^{-1} \int_{0}^{\infty} (r/r_o)^{3/2} [\mu_i(r)/\mu_o]\, e^{-i\alpha\, \ell n(r/r_o)}\, (dr/r) \ . \qquad (B.3)$$

The integral over radial velocities in eqn. (B.1) is trivial. Even though the integral over tangential velocities in eqn. (B.1) is apparently intractable, the combination in eqn. (B.3) of an integral over radii and integral over tangential velocities is manageable, provided that the order of integration is reversed so that the integral over r is performed first.

The result is[1]

$$B(\alpha) = (2N)^{-1} \left[(a+1)/2\right]^{-(i\alpha-1/2)/2} \csc\left[(1/2-i\alpha)/N\right] \times \Gamma\left[(i\alpha-1/2)/2 + (a+1)/2\right] / \Gamma\left[(a+1)/2\right] . \quad (B.4)$$

The rigid surface density is axisymmetric. The potential that corresponds to it is

$$\phi_i(r) = -2\pi G\mu_o r_o \, (r/r_o)^{-1/2} \int_{-\infty}^{\infty} K(\alpha,0) \, B(\alpha) \, e^{i\alpha \, \ell n(r/r_o)} \, d\alpha ,$$

where $K(\alpha,0)$ is the Kalnajs gravity factor, eqn. (3.10), when $m = 0$. Hence the radial forces that the rigid disk component generates are given by

$$F_{i,r}(r) = 2\pi G\mu_o \, (r/r_o)^{-3/2} \int_{-\infty}^{\infty} (i\alpha-1/2) \, K(\alpha,0) \, B(\alpha) \times e^{i\alpha \, \ell n(r/r_o)} \, d\alpha .$$

A spherical halo with volume density $\rho_H(r)$ will produce the radial forces

$$F_{H,r}(r) = -4\pi G \, r^{-2} \int_0^r \rho_H(r') \, r'^2 \, dr' .$$

---

[1] To verify this computation, and to remove any lingering doubts that the reversal of the order of integration in eqn. (B.3) is legitimate, we have compared the rigid surface density that this value of $B(\alpha)$ implies with the rigid surface density obtained by subtracting $\mu_a(r)$ from $\mu_s(r)$. To perform this comparison both eqn. (2.65) for $\mu_a(r)$ and eqn. (B.2) for $\mu_i(r)$ were integrated numerically. Happily, the results are in agreement.

Thus, the halo density must be

$$\rho_H(r) = \mu_o/r_o \, (r/r_o)^{-5/2} \, \mathrm{Re} \{ \int_0^\infty (\alpha^2+1/4) \, K(\alpha,0) \times B(\alpha) \, e^{i\alpha \, \ell n(r/r_o)} \, d\alpha \} . \qquad (B.5)$$

In order for this halo density to have a clear physical interpretation, it must be non-negative at all radii. We have numerically computed $\rho_H(r)$ from eqn. (B.5) over the range $|\ell n(r/r_o)| < 7$ when $N = 2$, 3 and 4 for the velocity dispersions 0.200, 0.378, 0.600 and 1.000 . Only for the cut-out index 2 disk is the halo density always non-negative. When $N = 3$ and $N = 4$, $\rho_H(r)$ is negative for large radii, at least for $\tilde{\sigma}_u = 0.200$, 0.378 and 0.600 . As the velocity dispersion is increased, the radius at which $\rho_H(r)$ first turns negative increases, for example, when $N = 3$ and $\tilde{\sigma}_u = 0.200$, the halo density turns negative at $(r/r_o) \sim 2.5$, but when $N = 3$ and $\tilde{\sigma}_u = 0.600$, it turns negative at $(r/r_o) \sim 9$. For $\tilde{\sigma}_u = 1.000$, the halo density is non-negative over the entire range $|\ell n(r/r_o)| < 7$.

APPENDIX C

NUMERICAL COMPUTATION OF THE KERNEL

In section III.e of the text the procedures used to evaluate the kernel, $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, were outlined. This appendix provides more details for each of the essential elements.

1) Kalnajs gravity factor

The basis of our computation of the Kalnajs gravity factor, given by eqn.(3.10), is Euler's formula,

$$\Gamma(z+1) = \lim_{n\to\infty} \{ n!\, n^z / \prod_{k=1}^{n} (z+k) \} .$$

This enables us to write the axisymmetric (m = 0) gravity factor as the limit

$$K(\alpha,0) = \lim_{n\to\infty} \{ \frac{1}{2n} \prod_{k=0}^{n} \frac{(4k+3)^2 + 4\alpha^2}{(4k+1)^2 + 4\alpha^2} \} .$$

Once the axisymmetric gravity factor is available, the corresponding factor for an arbitrary angular harmonic m can be obtained by successively applying the recurrence relation

$$K(\alpha,m)\, K(\alpha,m+1) = [(m + 1/2)^2 + \alpha^2]^{-1}$$

[Kalnajs (1971)]. The infinite product for the axisymmetric gravity factor converges rather slowly. However, it is easily

shown that the error after n terms in the product have been used has the form

$$c_1 n^{-1} + c_2 n^{-2} + \cdots ,$$

where c_1 and c_2 are constants. Therefore, the use of several Richardson extrapolations substantially improves the convergence, to the point that 12 digit accuracy is readily achieved.

2) orbital parameters

The maximum and minimum radii reached by a star with eccentric velocity $\tilde{U}$ and unit angular momentum, denoted by $x_{max}(\tilde{U})$ and $x_{min}(\tilde{U})$, respectively, are the two roots of eqn. (2.34):

$$f(x) = 1 + \tilde{U} - 2 \ln x - x^{-2} = 0 .$$

The function $f(x)$ is simply the square of the radial velocity of this star at radius x. Clearly, $f(x) \geq 0$ only for $x \in [x_{min}(\tilde{U}) , x_{max}(\tilde{U})]$. Furthermore, $f'(x) > 0$ for $x < 1$ and $f'(x) < 0$ for $x > 1$, while $f(x) \to -\infty$ both as $x \to 0$ and as $x \to \infty$. Therefore, an iterative procedure based on Newton's method will converge to $x_{max}(\tilde{U})$ so long as the initial guess x_o is greater than 1. Since an iteration in the variable x may lead to an iterate x_n which is less than 0 when $x_o < 1$, it is preferable to

find $x_{min}(\tilde{U})$ by iterating in the variable $y = 1/x$ with $y_o > 1$. In practice, a half dozen or so iterations will yield $x_{min}(\tilde{U})$ and $x_{max}(\tilde{U})$ to at least a dozen places.

The integrals $I_n(U)$ defined in eqn. (2.39),

$$I_n(\tilde{U}) = 2 \int_{x_{min}(\tilde{U})}^{x_{max}(\tilde{U})} \frac{dx}{x^n \{1 + \tilde{U} - 2 \ln x - x^{-2}\}^{1/2}} ,$$

are needed when $n = 0$, 1 and 2 . The argument of the square root is just the function $f(x)$ above, which has simple zeros at $x_{min}(\tilde{U})$ and $x_{max}(\tilde{U})$ when $\tilde{U} > 0$. The integrands, then, have square root singularities at the end points of integration. Furthermore, the nature of the integrand suggests that the regions $[x_{min}(\tilde{U}) , 1]$ and $[1 , x_{max}(\tilde{U})]$ should have nearly equal weights. An integration in the variable x itself not only has to contend with the singularities in the integrand but also gives a disproportionately large weight to the latter region. These difficulties can be eliminated by transforming first to the variable $y = \ln x$ and then to the variable θ , where

$$y = (y_{max} + y_{min})/2 - (y_{max} - y_{min}) \cos \theta /2 ,$$

with $y_{min} = \ln x_{min}(\tilde{U})$ and $y_{max} = \ln x_{max}(\tilde{U})$. Then

$$I_n(\tilde{U}) = (y_{max} - y_{min}) \int_0^{\pi} e^{-(n-1)y} \{f(x)\}^{-1/2} \sin \theta \, d\theta .$$

Notice that as $\theta \to 0$, $f(x) \sim \theta^2$. So the integrand is now regular at the lower endpoint. Similarly, the singularity

at $\theta = \pi$ also disappears. Since the integrand is periodic in θ, a quadrature by means of the midpoint rule is particularly appropriate. In fact, as few as 10 sampling points in θ suffice to give $I_n(\tilde{U})$ to 12 digits when $\tilde{U} < 2$.

3) Fourier coefficients

The Fourier coefficients are defined by eqn. (3,27),

$$Q_{\ell m}(\alpha;\tilde{U}) = (2\pi)^{-1} \int_0^{2\pi} e^{(i\alpha-1/2)X(\psi;\tilde{U})+imY(\psi;\tilde{U})-i\ell\psi}\, d\psi ,$$

and the orbital excursions can be found from eqns. (2.44) - (2.47), (2.51) and (2.52). The radial excursion $X(\psi;\tilde{U})$ is even in ψ about its apocenter ($\psi = \pi$) while the angular excursion $Y(\psi;\tilde{U})$ is odd about $\psi = \pi$. Hence, only half of the $[0,2\pi]$ range is actually needed:

$$Q_{\ell m}(\alpha;\tilde{U}) = \pi^{-1} \int_0^{\pi} e^{(i\alpha-1/2)X(\psi;\tilde{U})} \cos[mY(\psi;\tilde{U})-\ell\psi]\, d\psi . \qquad (C.1)$$

This integrand is periodic in ψ and the midpoint formula again produces an adequate approximation. We typically use the rule

$$n_\psi = 15\, e^{1.2\,\tilde{U}}$$

to determine how many sampling points in ψ to use. (It is prudent to insure that n_ψ is at least twice as large as the number of Fourier harmonics desired in order to avoid aliasing.) A sufficiently accurate Runge-Kutta integration of the equilibrium orbit is obtained by performing $2n_\psi$ time steps

between the pericenter and the apocenter. These criteria yield Fourier coefficients which are certainly accurate to 6 digits for small $\tilde{U}$, but to only 3 or 4 digits for large $\tilde{U}$. Very accurate Fourier coefficients are not needed when $\tilde{U}$ is large, since the weights assigned to the Fourier coefficients in eqn. (3.40) decrease exponentially with $\tilde{U}$.

We record here several useful symmetry properties of the Fourier coefficients:

$$Q_{\ell m}(-\alpha;\tilde{U}) = \bar{Q}_{\ell m}(\alpha;\tilde{U}) \tag{C.2}$$

$$Q_{-\ell 0}(\alpha;\tilde{U}) = Q_{\ell 0}(\alpha;\tilde{U}) \ . \tag{C.3}$$

These follow directly from eqn. (C.1).

4) integration over eccentric velocities

An inspection of the integrand (in $\tilde{U}$) of the kernel, eqn. (3.40), reveals that all the terms are regular functions of $\tilde{U}^2$ and that $\tilde{U}\, d\tilde{U} = d(\tilde{U}^2)/2$. Therefore, a Gauss-Laguerre quadrature in the variable $x = \tilde{U}^2/2\tilde{\sigma}_u^2$ can be expected to work well. In fact, it works surprisingly well, especially when the sample eccentric velocities are shifted inward by choosing the sample $\tilde{U}_i$ according to $x = \tilde{U}_i^2/2\tilde{\sigma}_n^2$, where x_i is one of the Gauss-Laguerre abscissas and $\tilde{\sigma}_n$ is roughly 80% of $\tilde{\sigma}_u$. The table entries on p. 109 in section V.b illustrate how few points are actually required.

5) normalization constant

The expression for the normalization constant given in eqn. (2.63) also lends itself to a Gauss-Laguerre quadrature. Exactly the same sample velocities are used here as are used in the computation of the kernel.

6) sum over radial harmonics

The greatest nuisance in the computation of the kernel is the sum over the radial harmonics. The dependence upon ℓ is contained in the terms

$$Q_{\ell m}(\alpha;\tilde{U})\ \bar{Q}_{\ell m}(\alpha;\tilde{U})\ F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})\ .$$

The explicit formula for the angular momentum function, eqn. (3.44), indicates that $|F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})|$ tends to a constant value as $\ell \to +\infty$ and to a different constant value as $\ell \to -\infty$. Numerical evaluation of the Fourier coefficients reveals that they decrease fairly quickly as $\ell \to -\infty$ but rather slowly as $\ell \to +\infty$. In fact, in order to obtain the kernel to even 5 digit accuracy, upwards of 50 radial harmonics are frequently required. Fortunately we are actually interested not so much in computing the kernel accurately as we are in obtaining the response accurately, in particular, with finding the mathematical eigenvalues, $\lambda(\tilde{\sigma}_u,\tilde{\omega})$, of the kernel (see section V.a). Since integration tends to smooth out irregularities, our hope is that as additional radial harmonics are added to the kernel, the eigenvalues

converge faster than the kernel itself. This, indeed, is the case. The convergence of the eigenvalues can be improved by giving the last few radial harmonics added weight. This lets them, in some sense, represent the terms left out of the sum. Typically the last three radial harmonics in the sum are given twice their normal value. The convergence of this procedure is illustrated in section V.b of the text.

APPENDIX D

NUMERICAL COMPUTATION OF THE RESPONSE INTEGRAL

The numerical solution of the integral equation (3.42) hinges on the calculation of the response of a cut-out disk to an imposed density with density transform $A(\alpha)$. In our approximation the imposed density is known only at $2n+1$ equally-spaced locations. These sample wavenumbers are denoted by α_j , where $\alpha_j = (j-n+1)\Delta\alpha$. The spacing between successive wavenumbers is denoted by $\Delta\alpha$. We seek to find the response density transform $B(\alpha)$ at these same locations.[1]

In the previous appendix we described how the kernel, $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$, can be found at the $(2n+1)\times(2n+1)$ locations (β_j,α_k) . The issue here is the manner in which the integral

$$B(\beta) = \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\ A(\alpha)\ d\alpha \qquad (D.1)$$

is computed.

For any specific β_j , the most pronounced α-dependence in the kernel is displayed by the angular momentum function; in particular, for $\beta-\alpha \to -\infty$, $F_{\ell m}(\beta-\alpha;\tilde{U},\tilde{\omega})$ decays as $e^{(\beta-\alpha)\pi/N}$, whereas for $\beta-\alpha \to +\infty$, it decays as $e^{-(\beta-\alpha)\tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)}$. Our integration scheme is tailored to this behavior of the angular momentum function.

---

[1] In this appendix we use $B(\alpha)$ instead of $A_{res}(\alpha)$ to denote the response density and simply $A(\alpha)$ instead of $A_{imp}(\alpha)$ to denote the imposed density.

In the limit $\tilde{s} = 0+$, $F_{\ell m}(\eta;\tilde{U},\tilde{\omega})$ has a convenient decomposition into an Hermitian portion and a Volterra portion,

$$F_{\ell m}(\eta;\tilde{U},\tilde{\omega}) = H_{\ell m}(\eta) + V_{\ell m}(\eta) \ ,$$

where $H_{\ell m}(-\eta) = \bar{H}_{\ell m}(\eta)$ and $V_{\ell m}(\eta) = 0$ for $\eta < 0$.[2] The specific formulas for $H_{\ell m}(\eta)$ and $V_{\ell m}(\eta)$ depend on the sign of $\ell\tilde{\kappa}(\tilde{U}) + m\tilde{\Omega}(\tilde{U})$. In the remainder of this appendix this term will be abbreviated by $\ell\tilde{\kappa}+m\tilde{\Omega}$. In detail, we have

$\ell\tilde{\kappa}+m\tilde{\Omega} < 0$

$$H_{\ell m}(\eta) = \frac{-i}{2 \sinh \pi\eta} \{ \frac{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N-1}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}} e^{-i\eta \ln|(\ell\tilde{\kappa}+m\tilde{\Omega})/m\tilde{\Omega}_p|}$$

$$\times \{\frac{Nm(m\tilde{\Omega}_p)^{N}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}} - [(a+1)(\ell\tilde{\kappa}+m\tilde{\Omega}) - am]\}$$

$$+ \frac{1}{N} \sum_{j=1}^{N} \frac{e^{-\pi\eta} e^{(2j-1)\pi\eta/N}}{\ell\tilde{\kappa}+m\tilde{\Omega} - m\tilde{\Omega}_p e^{(2j-1)\pi i/N}} \{-im\eta +$$

$$[(a+1)(\ell\tilde{\kappa}+m\tilde{\Omega}) - am] + \frac{m(m\tilde{\Omega}_p) e^{(2j-1)\pi i/N}}{\ell\tilde{\kappa}+m\tilde{\Omega} - m\tilde{\Omega}_p e^{(2j-1)\pi i/N}} \}\}$$

$$V_{\ell m}(\eta) = 0 \qquad \text{(for all } \eta \text{)}$$

---

[2] We assume throughout this appendix that $\tilde{\Omega}_p > 0$. When $\tilde{\Omega}_p < 0$, the same type of decomposition still exists except that $V_{\ell m}(\eta) = 0$ for $\eta > 0$.

$\ell\tilde{\kappa}+m\tilde{\Omega} > 0$

$$H_{\ell m}(\eta) = \frac{-i}{2\sinh \pi\eta}\left\{ \frac{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N-1}\, e^{-\pi|\eta|}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}}\, e^{-i\eta\, \ell n|(\ell\tilde{\kappa}+m\tilde{\Omega})/m\tilde{\Omega}_p|} \right.$$

$$\times \left\{ \frac{Nm(m\tilde{\Omega}_p)^{N}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}} - [(a+1)(\ell\tilde{\kappa}+m\tilde{\Omega}) - am] \right\}$$

$$+ \frac{1}{N}\sum_{j=1}^{N} \frac{e^{-\pi\eta}\, e^{(2j-1)\pi\eta/N}}{\ell\tilde{\kappa}+m\tilde{\Omega} - m\tilde{\Omega}_p\, e^{(2j-1)\pi i/N}} \left\{ -im\eta + \right.$$

$$\left. \left. [(a+1)(\ell\tilde{\kappa}+m\tilde{\Omega}) - am] + \frac{m(m\tilde{\Omega}_p)\, e^{(2j-1)\pi i/N}}{\ell\tilde{\kappa}+m\tilde{\Omega} - m\tilde{\Omega}_p\, e^{(2j-1)\pi i/N}} \right\} \right\}$$

$$V_{\ell m}(\eta) = 0 \qquad (\eta < 0)$$

$$V_{\ell m}(\eta) = \frac{-i\,(\ell\tilde{\kappa}+m\tilde{\Omega})^{N-1}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}}\, e^{-i\eta\, \ell n|(\ell\tilde{\kappa}+m\tilde{\Omega})/m\tilde{\Omega}_p|}$$

$$\times \left\{ \frac{Nm(m\tilde{\Omega}_p)^{N}}{(\ell\tilde{\kappa}+m\tilde{\Omega})^{N} + (m\tilde{\Omega}_p)^{N}} - [(a+1)(\ell\tilde{\kappa}+m\tilde{\Omega}) - am] \right\}$$

$$(\eta > 0)$$

The corresponding decomposition of the kernel itself has the form

$$S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega}) = K(\alpha,m)\ [H(\beta,\alpha) + V(\beta,\alpha)]\ ,$$

where $H(\alpha,\beta) = \bar{H}(\beta,\alpha)$ and $V(\beta,\alpha) = 0$ for $\alpha > \beta$. In order to perform the numerical integration in eqn. (D.1) we approximate both the kernel and the density transform by an 8-point Lagrange interpolating polynomial. In general, if $f(\alpha)$ is any function, and f_i denotes $f(\alpha_i)$ with $f_i = 0$

when $i < 0$ or $i > 2n+1$, then for $\alpha_1 \leq \alpha < \alpha_{2n+1}$ the approximation is

$$f(\alpha) \simeq \sum_{p=-3}^{4} L_p[(\alpha-\alpha_r)/\Delta\alpha]\ f_{p+r}\ ,$$

where r is chosen so that $\alpha_r \leq \alpha < \alpha_{r+1}$. The Lagrange interpolating polynomials, $L_p[x]$, are given explicitly by

$$L_p[x] = \frac{(-1)^p \prod_{i=1}^{8} (x+4-i)}{(3+p)!\ (4-p)!\ (x-p)}$$

[Abramowitz and Stegun (1965, p. 878)]. Taking into consideration the behavior of the Hermitian and Volterra portions of the kernel, we write

$$H(\beta,\alpha) = e^{-(\alpha-\beta)\pi/N}\ \hat{H}(\beta,\alpha) \qquad \text{for} \quad \alpha > \beta$$

$$H(\beta,\alpha) = e^{(\alpha-\beta)\pi/N}\ \hat{W}(\beta,\alpha) \qquad \text{for} \quad \alpha < \beta$$

$$V(\beta,\alpha) = \hat{V}(\beta,\alpha) \qquad \text{for} \quad \alpha < \beta$$

and approximate the smoother functions $\hat{H}$, $\hat{W}$ and $\hat{V}$ by Lagrange interpolation. We also write the density transforms as

$$A(\alpha) = \hat{A}(\alpha)/K(\alpha,m)$$

$$B(\alpha) = \hat{B}(\alpha)/K(\alpha,m)$$

and represent $\hat{A}(\alpha)$ and $\hat{B}(\alpha)$ by Lagrange interpolation.

Then the response at $\beta = \beta_j$ can be expressed as

$$\hat{B}(\beta_j) = K(\beta_j,m)\ [I_H + I_W + I_V] \ ,$$

where I_H , I_W and I_V are the integrals involving the three portions of the kernel. For example,

$$I_H = \sum_{r=j}^{2n} \int_{\alpha_r}^{\alpha_{r+1}} \sum_{p=-3}^{4} \sum_{q=-3}^{4} e^{-(\alpha-\beta_j)\pi/N}\ L_p[(\alpha-\alpha_r)/\Delta\alpha]$$

$$\times\ L_q[(\alpha-\alpha_r)/\Delta\alpha]\ \hat{H}_{j,q+r}\ \hat{A}_{p+r}\ d\alpha$$

$$= \sum_{r=j}^{2n} e^{-(\alpha_r-\beta_j)\pi/N} \sum_{p=-3}^{4} \sum_{q=-3}^{4} C^H_{pq}\ \hat{H}_{j,q+r}\ \hat{A}_{p+r} \ ,$$

where the weighting coefficients are

$$C^H_{pq} = \Delta\alpha \int_0^1 e^{-\pi x(\Delta\alpha)/N}\ L_p[x]\ L_q[x]\ dx \ .$$

We combine this result with the analogous results for I_W and I_V to obtain a matrix (S_{jk}) which expresses the density response in terms of the imposed density:

$$\hat{B}(\beta_j) = \sum_{k=1}^{2n+1} S_{jk}\ \hat{A}(\alpha_k) \ .$$

For given $\tilde{\sigma}_u$ and $\tilde{\omega}$ this matrix need only be computed once. We are then in a position to find quickly the response to any trial density. In discussions of this approximation elsewhere we simply ignore the slight difference between $\hat{A}(\alpha)$ and $A(\alpha)$.

Although the kernel cannot be decomposed into a Volterra portion and an Hermitian portion when $\tilde{s} > 0$, it still

exhibits the strongly one-sided behavior discussed in section III.d: for $\beta-\alpha < 0$ it decays rapidly as

$$e^{-[\pi/N - \tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)]|\alpha-\beta|} ,$$

whereas for $\beta-\alpha > 0$ it decays slowly as

$$e^{-\tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)\ |\alpha-\beta|} .$$

So the response integrals can be profitably separated into an $\alpha > \beta$ part and an $\alpha < \beta$ part. The interpolation scheme is now tailored to the exponential behavior on each side of $\alpha = \beta$. We find that the response can be obtained nearly as accurately when $\tilde{s} > 0$ as it can be for a comparable approximation when $\tilde{s} = 0^+$. However, a typical $\tilde{s} > 0$ computation takes 30-40% longer than a comparable $\tilde{s} = 0^+$ calculation. Moreover, the weighting coefficients of the $\tilde{s} = 0^+$ scheme, such as C^H_{pq}, are independent of $\tilde{\Omega}_p$, whereas the corresponding weighting coefficients of the $\tilde{s} > 0$ scheme depend upon $\tan^{-1}(\tilde{s}/m\tilde{\Omega}_p)$. This gives a further advantage to the $\tilde{s} = 0^+$ scheme when calculations are desired for a variety of frequencies. Fortunately, as described in section V.b, the question of whether a given disk is stable or unstable can be settled entirely on the basis of the results for vanishingly small growth rates.

APPENDIX E

ITERATIVE PROCEDURE FOR COMPUTING EIGENVALUES

The results of the previous appendix lead to a numerical approximation to the mathematical eigenvalues of eqn. (5.1) in the form

$$\lambda(\tilde{\sigma}_u,\tilde{\omega})\ A(\beta_j) = \sum_{j=1}^{2n+1} S_{jk}\ A(\alpha_k) \qquad j=1,2,\cdots,2n+1. \quad \text{(E.1)}$$

Since only the largest one or two eigenvalues and their eigenfunctions are usually desired, we often resort to an iterative procedure to find them rather than using the general eigenvalue package EISPACK to find all the eigenvalues.

Even though we have no guarantee that the non-Hermitian matrix (S_{jk}) has a complete set of eigenfunctions, the power method has, in practice, produced the largest eigenvalue of eqn. (E.1) along with its eigenfunction. The power method is described in most numerical analysis texts [see, for instance, Ralston (1965)]. In our adaptation of this procedure, the estimate after the p-th iteration used for the largest eigenvalue, $\lambda_1^{(p)}$, is obtained by taking the ratio of one of the components of the response vector, $B^{(p+1)}(\beta_r)$, to the same component of the imposed vector, $A^{(p)}(\alpha_r)$, where

$$B^{(p+1)}(\beta_j) = \sum_{k=1}^{2n+1} S_{jk}\ A^{(p)}(\alpha_k)\ .$$

The best choice for this component is the largest component

of the current response vector. We check about a half dozen representative components of this vector and use the largest of these in forming the ratio. The next imposed vector, $A^{(p+1)}(\alpha_j)$ is obtained from $B^{(p+1)}(\beta_j)$ by renormalizing it so that $A^{(p+1)}(\alpha_{n+1}) = 1$. The iteration converges geometrically, in the fashion

$$\lambda_1^{(p)} = \lambda_1 + r_2(\lambda_2/\lambda_1)^p + \cdots + r_{2n+1}(\lambda_{2n+1}/\lambda_1)^p \ .$$

In order to improve the convergence of this sequence $\{\lambda_1^{(p)}\}$ we form an extrapolated sequence. In general, if a sequence $\{a_p\}$ converges to the limit a_∞ in such a manner that

$$a_p = a_\infty + r_1\Lambda_1^p + r_2\Lambda_2^p + \cdots \ ,$$

then the extrapolated sequence $\{b_p\}$, where

$$b_p = \frac{a_{p-1}a_{p+1} - a_p^2}{a_{p-1} + a_{p+1} - 2a_p} \ ,$$

will converge as

$$b_p = a_\infty + r_2\Lambda_2^p + r_3\Lambda_3^p + \cdots \ .$$

So this extrapolation removes the dominant transient term. Of course it may be applied more than once. We typically form two such extrapolated sequences from the iterates $\{\lambda_1^{(p)}\}$. The iteration was usually stopped once three or more

successive extrapolated estimates of λ_1 agreed to within 1 part in 10^5. Occasionally the criterion for stopping the iteration was more demanding.

Once the largest eigenvalue and its eigenfunction, $A_1(\alpha_j)$, are known, the second largest eigenvalue can be obtained by using the power method on the matrix (W_{jk}), where

$$W_{jk} = S_{jk} - [A_1(\alpha_j)/A_1(\alpha_q)]\, S_{q,k} \quad ,$$

and q denotes some component of the largest eigenvector. (We usually chose $q = n+1$.) The eigenvalues of (W_{jk}) are $0,\lambda_2,\lambda_3, \cdots ,\lambda_{2n+1}$ and the q-th row of (W_{jk}) is zero. When the q-th row and column are dropped from (W_{jk}) the eigenvalues of the resulting $2n\times 2n$ matrix are $\lambda_2,\lambda_3, \cdots ,\lambda_{2n+1}$. This procedure is Wielandt's deflation and it is also described in most numerical analysis texts.

Actually, the extrapolation scheme that accelerates the power method can also be used to estimate both r_2 and (λ_2/λ_1). Hence, a fairly decent estimate of λ_2 (but not the second eigenfunction) can be obtained without resorting to the matrix deflation.

The power method is inefficient whenever λ_2 is close in size to λ_1. In such cases we must resort to EISPACK. This has never been a problem when $m = 2$ or 4. However, it has been a problem on occasion when $m = 1$ or 3.

APPENDIX F

TESTS VIA EXPLICIT ORBIT INTEGRATIONS

In order to corroborate both the general stability analysis presented in Chapter III and the specific numerical procedures described there and in previous appendices, we have performed an independent check of some of the results reported in Chapters V and VI. The checks were performed on two levels: (i) to confirm that the density response to an elementary forcing at a single wavenumber was computed correctly and (ii) to confirm that the numerically obtained solutions of the integral equation (3.42) — both the characteristic frequency α and the associated eigenfunction $A(\alpha)$ — did in fact correspond to self-consistent density perturbations (or modes). In both of these areas the standard for the comparison was a brute force integration of the linearized Vlasov equation. In this appendix we will first describe how the explicit orbit integrations were performed and then report on the outcome of the comparisons.

i) Method of Explicit Orbit Integration

In general the density response is given by the integral over velocity space of the perturbed distribution function $f_1(r,\theta,u,v,t)$ caused by the forcing potential. In turn, $f_1(r,\theta,u,v,t)$ is available from an integration in time over the unperturbed stellar orbits. In the text the orbital

similarity in our models and the Fourier series expansion of the sampled potential were both utilized in streamlining these calculations. Here, though, none of these short cuts were taken. The integration over velocity space was performed in the normal radial and tangential velocity components u and v, while the integration along the equilibrium orbits was done directly.

Consider first an elementary forcing of the form given by eqn. (3.16). The induced change in the distribution function is given by eqn. (3.19):

$$f_1(r,\theta,u,v,t) = - (\partial f_o/\partial E)\ \Delta E - (\partial f_o/\partial J)\ \Delta J\ .$$

We choose, without loss of generality, to concentrate on the response at $t = 0$ and $\theta = 0$. In detail, eqns. (3.20) and (3.20) are then

$$\Delta E = 2\pi G\mu_p\ K(\alpha,m) \int_{-\infty}^{0} [(i\alpha-1/2)\ u'(t') + im\ v'(t')]$$

$$\times\ e^{(i\alpha-3/2)\ \ell n[r'(t')/r_o]}\ e^{im\theta'(t') - \omega t'}\ dt'$$

$$\Delta J = 2\pi G\mu_p\ K(\alpha,m) \int_{-\infty}^{0} im\ e^{(i\alpha-1/2)\ \ell n[r'(t')/r_o]}$$

$$\times\ e^{im\theta'(t') - \omega t'}\ dt'\ .$$

These time integrals are convergent so long as the growth rate s is positive. The backwards time integration can be stopped once $e^{st'}$ becomes negligibly small. A

better approach to this infinite integral, and the one we have adopted in most applications, is to exploit the periodicity of the stellar orbits: Let $T(u,v)$ denote the radial period of the star in question and $\Omega(u,v)$ denote its angular frequency[1]. Then it is easy to see that the contribution to the integral from each complete radial oscillation is merely $e^{-iT(u,v)\ [m\Omega(u,v)-\omega]}$ times the contribution from the previous oscillation. Hence, all that is required is an integral over a single radial oscillation; the total change in E or J is simply the change in one period times the geometric series

$$\sum_{n=0}^{\infty} e^{-inT(u,v)\ [m\Omega(u,v)-\omega]} ,$$

e.g.,

$$\Delta J = \frac{2\pi G\mu_p\ K(\alpha,m)}{1 - e^{-iT(u,v)\ [m\Omega(u,v)-\omega]}} \int_{-T(u,v)}^{0} im \times e^{(i\alpha-1/2)\ \ell n[r'(t')/r_o]}\ e^{im\theta'(t') - \omega t'}\ dt' .$$

In practice these orbit integrals can be computed more efficiently by integrating in angle rather than time. The reason is that an integration in time gives relatively little weight to contributions to the integral from small radii, where the

---

[1] With the single exception of our computation of $T(u,v)$ and $\Omega(u,v)$, the numerical procedures used in the orbit integrations were entirely independent of the methods used for finding the solutions of eqn. (3.40).

integrand is largest, since the star moves very swiftly near its pericenter; in contrast, the integration in angle gives a more sensible weight to that portion of the orbit. We used Simpson's rule to perform this quadrature in angle. The orbit variables r' , u' , v' and t' as a function of θ' were computed by a Runge-Kutta integration of eqns. (2.19), (2.20) and (2.21), modified so that θ' rather than t' was the independent variable.

These procedures enabled us to compute $f_1(r,\theta=0,u,v,t=0)$ for given r , u and v . We found that the perturbed distribution function could be adequately integrated over velocity space (to produce the actual response density at radius r) by performing the crudest possible integration — we simply employed a finite rectangular grid in the two-dimensional velocity space and used the mid-point rule in each direction.

The response densities calculated in this manner were then compared with those computed from the transfer function by means of the Fourier integral

$$\mu_{1,res}(r) = \mu_p(r/r_o)^{3/2} \int_{-\infty}^{\infty} S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})\ e^{i\beta\ \ell n(r/r_o)}\ d\beta\ .$$

(We have suppressed the dependence upon angle and time.) This, too, measures the response to a forcing at a single wavenumber α . The numerical procedures described in Appendix C were used to compute $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$ over a finite range of β's , while the Fourier integral was evaluated

by means of Simpson's rule. The elementary response calculations tested how well we had computed the transfer function $S_m(\beta,\alpha;\tilde{\sigma}_u,\tilde{\omega})$. The results will be described shortly.

Our more ambitious orbit integrations were designed to test how reliably our numerical methods produced the solutions, $[\tilde{\omega},A(\alpha)]$, of eqn. (3.40); alternatively, these are solutions of eqn. (5.1) with the property that $\lambda(\tilde{\sigma}_u,\tilde{\omega}) = 1$. The orbit integrations were performed in a manner analogous to those for the elementary response tests. The difference was that here an added Fourier integral appeared in the expressions for ΔE and ΔJ , e.g.,

$$\Delta J = 2\pi G\mu_p \int_{-\infty}^{0} dt' \int_{-\infty}^{\infty} d\alpha\, A(\alpha)\, K(\alpha,m)\, im \times e^{(i\alpha-1/2)\, \ell n[r'(t')/r_o]}\, e^{im\theta'(t') - \omega t'} .$$

Of course, our computed eigenfunctions $A(\alpha)$ were only available over a finite range, usually for $\alpha \in [-20,20]$. Simpson's rule was used to approximate the truncated Fourier integral. The resulting response density was then compared with the imposed density

$$\mu_{1,imp}(r) = \mu_p\, (r/r_o)^{-3/2} \int_{-\infty}^{\infty} A(\alpha)\, e^{i\alpha\, \ell n(r/r_o)}\, d\alpha .$$

Simpson's rule was used for this integral also. These checks tested the full range of our numerical procedures — those described in Appendices D and E as well as those described in Appendix C.

ii) Results of Explicit Orbit Integrations

As mentioned in the text, our first orbit integration tests were of the elementary response variety. We specifically tested the response of the $N = 1$ and $N = 2$ disks with $\tilde{\sigma}_u = 0.400$ to an imposed density with $m = 2$, $\tilde{\Omega}_p = 0.25$ and $\tilde{s} = 0.40$. (We cite these results in terms of dimensionless variables.) At that time the densities computed via orbit integrations were not correctly normalized, although the ones computed via the transfer function were. The following table contains the results:

	N = 1		α = 0	
r	$\lvert\mu\rvert_{oi}$	$\lvert\mu\rvert_{tf}$	ϕ_{oi}	ϕ_{tf}
1	1.491	1.023	105.60	105.64
2	0.4511	0.3086	42.90	42.81
$\lvert\mu_1/\mu_2\rvert$ =	3.305	0.3315		

	N = 1		α = 5	
r	$\lvert\mu\rvert_{oi}$	$\lvert\mu\rvert_{tf}$	ϕ_{oi}	ϕ_{tf}
1	1.758	0.4455	72.36	73.05
2	0.7117	0.1792	31.69	32.12
$\lvert\mu_1/\mu_2\rvert$ =	2.470	2.486		

	N = 2		α = 5	
r	$\lvert\mu\rvert_{oi}$	$\lvert\mu\rvert_{tf}$	ϕ_{oi}	ϕ_{tf}
1	1.831	0.4611	72.36	73.48
2	0.8386	0.2111	32.13	32.57
$\lvert\mu_1/\mu_2\rvert$ =	2.171	2.184		

The subscript oi denotes an orbit integration result; tf denotes a response computed via the transfer function. The magnitude of the density response is denoted by $|\mu|$; the difference in phase between the imposed density and the response density is denoted by ϕ . As usual, α denotes the wavenumber of the imposed density. Since the orbit integration results were not normalized, we can only compare the ratio of the density at $r = 1$ to that at $r = 2$. This ratio is given in the third line of each part of the table. By comparing the entries in the foregoing table, we find that the two methods of computing the elementary response agree to within 1% . We subsequently normalized the orbit integration results correctly, and confirmed that aspect of our calculations also.

We now turn to the more important tests, the ones of the computed solutions of eqn. (3.40). This has been done for both the $m = 2$ and $m = 1$ modes which are illustrated in the text. The results are

$m = 2$ mode

r	$\mu_{res,R}$	$\mu_{imp,R}$	$\mu_{res,I}$	$\mu_{imp,I}$
1	2.36463	2.35772	0.03700	0.04183
2	-1.74684	-1.73769	-0.90462	-0.90845

$m = 1$ mode

r	$\mu_{res,R}$	$\mu_{imp,R}$	$\mu_{res,I}$	$\mu_{imp,I}$
1	-0.71650	-0.71654	0.15281	0.15253
2	0.50735	0.50722	-0.18220	-0.18214

Here the comparison is made between the real and imaginary parts of the response density, computed via orbit integrations, and those of the imposed density. If $[\tilde{\omega},A(\alpha)]$ truly represents a mode, then these two densities should agree. As one can see, the $m = 2$ results agree to within 0.5% and the $m = 1$ results to within 0.04% . We submit that this agreement is satisfactory confirmation of our methods.

Clearly, the $m = 1$ results agree better than the $m = 2$ ones. We suspect two factors for this difference in accuracy: (i) The larger angular harmonic number m of the latter mode means that the disturbance forces are more oscillatory in the angular direction; consequently, the orbit integrals are more difficult to compute. (ii) For the $m = 2$ mode, the density transform is still alive at large positive α , whereas the $m = 1$ transform is sensibly zero there: at the largest α for which $A(\alpha)$ is available for the $m = 2$ mode, $|A(\alpha)| \sim 0.001$; for the $m = 1$ mode $|A(\alpha)| \sim 0.00001$ at the corresponding point. Therefore our finite approximation to the Fourier integral is worse in the $m = 2$ case.

We should point out that although $A(\alpha)$ is not known beyond $\alpha = +30$ for the $m = 2$ mode, the eigenvalue $\lambda(\tilde{\sigma}_u,\tilde{\omega})$ itself is already well-determined by the values of the transform for $|\alpha| \leq 12$.

REFERENCES

Abramowitz, M. and Stegun, I. 1965, Handbook of Mathematical Functions (New York: Dover).

Bardeen, J.M. 1975, in Proc. I.A.U. Symp. No. 69 (Dordrecht: Reidel), p. 297.

Chandrasekhar, S. 1960, Principles of Stellar Dynamics (New York: Dover).

Erickson, S.A. 1974, unpublished Ph.D. dissertation, M.I.T.

Freeman, K.C. 1966, Monthly Notices Roy. Astron. Soc., 134, p. 15.

Fujimoto, M. 1968, in Proc. I.A.U. Symp. No. 29 (Erevan: Publishing House Academy Sci. Armenian SSR), p. 453.

Gradshteyn, I.S. and Ryzhik, I.M. 1965, Table of Integrals, Series and Products (New York: Academic Press).

Hockney, R.W. and Brownrigg, D.R.K. 1974, Monthly Notices Roy. Astron. Soc., 167, p. 351.

Hohl, F. 1971, Astrophys. J., 168, p. 343.

_____. 1973, Astrophys. J., 184, p. 353.

Hohl, F. and Hockney, R.W. 1969, J. Comp. Phys., 4, p. 304.

Hunter, C. 1963, Monthly Notices Roy. Astron. Soc., 126, p. 299.

_____. 1965, Monthly Notices Roy. Astron. Soc., 129, p. 321.

Kalnajs, A.J. 1965, unpublished Ph.D. dissertation, Harvard University.

_____. 1970, in Proc. I.A.U. Symp. No. 38 (New York: Springer-Verlag), p. 318.

_____. 1971, Astrophys. J., 166, p. 275.

_____. 1972, Astrophys. J., 175, p. 63.

Klema, V.K., Garbow, B.S. and Moler, C.B. 1973, EISPACK, User's Information, Argonne National Laboratory, Applied Mathematics Division.

Lin, C.C. and Shu, F.H. 1964, Astrophys. J., 140, p. 646.

_____. 1966, Proc. Nat. Acad. Sci., 56, p. 229.

Lindblad, B. 1942, Stockholm Obs. Ann., 14, No. 1.

_____. 1963, Stockholm Obs. Ann., 22, No. 5.

_____. 1964, Astrophysica Norvegica, 9, No. 12.

Mark, J. W.-K. 1971, Proc. Nat. Acad. Sci., 68, 2095.

Mathewson, D.S., van der Kruit, P.C. and Brouw, W.N. 1972, Astron. & Astrophys., 17, p. 468.

Mestel, L. 1963, Monthly Notices Roy. Astron. Soc., 126, p. 553.

Miller, R.H. 1971, Astrophys. & Sp. Sci., 14, p. 73.

_____. 1974, Astrophys. J., 190, p. 539.

Miller, R.H. and Prendergast, K.H. 1968, Astrophys. J., 151, p. 699.

Miller, R.H., Prendergast, K.H. and Quirk, W.J. 1970, Astrophys. J., 161, 903.

Miyamoto, M. 1969, Pub. Astron. Soc. Japan, 21, p. 319.

Ostriker, J.P. and Peebles, P.J.E. 1973, Astrophys. J., 186, 467.

Quirk, W.J. 1971, Astrophys. J., 167, p. 7.

Ralston, A. 1965, A First Course in Numerical Analysis (New York: McGraw-Hill).

Roberts, W.W. 1969, Astrophys. J., 158, 123.

Rots, A.H. 1974, Ph.D. dissertation, University of Groningen (also Astron. & Astrophys., 31, p. 245, 45, p. 25 and 45, p. 43).

Stix, T.H. 1962, The Theory of Plasma Waves (New York: McGraw-Hill).

Toomre, A. 1963, Astrophys. J., 138, 385.

_____. 1964, Astrophys. J., 139, 1217.

_____. 1969, Astrophys. J., 158, 899.