55

AARHUS-ASTRO 93-?
SW 9341

IFA - 93/34

Sep. 1993

AARHUS-ASTRO-1993-8

ISSN 0906-3870

CERN LIBRARIES, GENEVA

CM-P00068035

STABILITY AGAINST PHASE MIXING OF COLLISIONLESS SELF- GRAVITATING MATTER

Jens Hjorth

Inst. of Physics and Astronomy
Aarhus University
DK-8000 Aarhus C
Denmark

To appear in: The Astrophysical Journal

Stability against Phase Mixing of Collisionless Self-Gravitating Matter

Jens Hjorth

Institute of Physics and Astronomy, Aarhus University, DK-8000 Århus C, Denmark

It is suggested how to define a macroscopic steady state of a collisionless self-gravitating system with Newtonian interactions in terms of H -functions. A new condition for stability is formulated as a consequence thereof. Any single-variable distribution function, $f = f(Q)$, which is a stationary point of some entropy-like functional, must have $df/dQ \leq 0$ to be stable against phase mixing. For the special class of Osipkov–Merritt models, $Q = E + L^2/2r_a^2$, this is found to agree with results of published numerical experiments. The stability criterion may have important implications for the equilibria of galaxies and dark-matter halos. One consequence is that stable spherical galaxies apparently have an anisotropy radius, r_a , greater than $\sim 40\%$ of the half-mass radius. This finding is consistent with dissipationless-collapse simulations.

Subject headings: dark matter — galaxies: kinematics and dynamics — instabilities

Typeset Using *REVTeX*

1. INTRODUCTION

This paper deals with collisionless stellar systems (such as galaxies). These are Newtonian N -body systems for which the binary relaxation time exceeds the time scale of interest, e.g., the age of the Universe. Using the concept of H -functions I shall present a new definition of a steady state for such systems which leads to a general and simple criterion for stability against evolution by phase mixing. A test of this criterion is carried out, and the stability concept and its possible applications are briefly commented on.

The collisional relaxation time associated with binary collisions in a many-body self-gravitating system can be estimated as $\tau_{\text{bin}} \approx 0.1 N \tau_{\text{cross}} / \ln N$ (Binney & Tremaine 1987) where N is the number of particles and τ_{cross} is the typical crossing time. For $N \approx 10^{11}$ and $\tau_{\text{cross}} \approx 10^8$ yr, which are representative values for a galaxy, the binary relaxation time is several orders of magnitude larger than the age of the Universe. Thus, galaxies do not relax via binary collisions. Another important astrophysical application is the elementary-particle dark matter which may form, e.g., the halos in which galaxies and clusters of galaxies are embedded and which presumably constitutes most of the matter in the Universe.

Collisionless systems satisfy the collisionless Boltzmann equation,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{x}} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0, \quad (1)$$

which describes the evolution of the phase-space density (distribution function— df), $f(\mathbf{x}, \mathbf{v}, t)$, as the flow of an incompressible fluid in six-dimensional $(\mathbf{x}, \mathbf{v})$ phase space. The gravitational potential, $\Phi(\mathbf{x}, t) = -G \int \frac{f(\mathbf{x}', \mathbf{v}', t)}{|\mathbf{x} - \mathbf{x}'|} d^3\mathbf{x}' d^3\mathbf{v}'$, is generated by self-gravity. In the following $E = v^2/2 + \Phi(\mathbf{x})$ denotes the energy per unit mass, and $L = |\mathbf{x} \times \mathbf{v}|$ denotes the magnitude of the angular momentum per unit mass.

The conventional definition of a steady state (Binney & Tremaine 1987), $\partial f / \partial t = 0$, means that the distribution function has no explicit time dependence, i.e., $f = f(\mathbf{x}, \mathbf{v})$. This implies that f is an integral of motion [according to Jeans' theorem (Binney & Tremaine 1987)]. For example, spherically symmetric systems supporting

regular orbits are described by $f = f(E, L)$ according to Jeans' strong theorem (Binney & Tremaine 1987). Unfortunately, however, it is in general an unresolved issue whether such a steady state is stable:

For isotropic systems, $f = f(E)$, it is known that $df/dE < 0$ is a sufficient condition for stability (Sygnet et al. 1984; Kandrup & Sygnet 1985) but not a necessary one (Hénon 1973). For anisotropic systems, $f = f(E, L)$, the condition $\partial f/\partial E < 0$ ensures stability only to radial perturbations (Lynden-Bell & Sanitt 1969; Dorémus & Feix 1973). Antonov (1973) showed that a system with purely radial orbits is unstable, and more recently it has been demonstrated that any spherical system characterized by a distribution function, that diverges as $f \sim g(E)L^{-s}$ ($0 < s < 1$; $g(E)$ bounded) when $L \rightarrow 0$, is unstable (Palmer & Papaloizou 1987). This radial-orbit instability leads to formation of an aspherical system, e.g., the formation of a bar (e.g., Londrillo, Messina, & Stiavelli 1991). Fridman & Polyachenko (1984) have proposed that the value $2T_r/T_t = 1.75 \pm 0.25$, for the ratio of kinetic energy in radial and tangential motion, is a discriminator for the onset of instability, but it has been demonstrated that this is not always a useful criterion (Palmer & Papaloizou 1987; Aguilar & Merritt 1990). Gillon, Dorémus, and Baumann (1976) have reported an analytic proof that the conditions $\partial f/\partial E < 0$ and $\partial f/\partial L < 0$ ensure stability to all perturbations, but numerical simulations (Barnes, Goodman, & Hut 1986) seem to disagree with this conclusion. Recent reviews on the subject can be found in Binney & Tremaine (1987), Merritt (1987), and Bertin & Stiavelli (1993).

Here, an alternative approach to studying the stability of collisionless stellar systems is suggested. This approach is based on the fact that instabilities imply an evolution of the system and that evolution of collisionless systems takes place through phase mixing. The relation between phase mixing and instabilities is discussed and illustrated by, e.g., Barnes et al. (1986).

When a system, which is not an exact solution to the stationary collisionless Boltzmann equation, $\partial f/\partial t = 0$, relaxes and approaches an equilibrium the phase

fluid filaments and becomes wound up on smaller and smaller scales. In the end it reaches what one would call a non-evolving state, but there will, however, be small fluctuations on very small scales. [For an interesting discussion of this, including a numerical illustration of phase mixing, see Funato, Makino, & Ebisuzaki (1992).] Therefore, a DF obeying equation (1) never reaches a steady state on the microscopic level; on a microscopic level there is no steady state but on a macroscopic level there is. Thus, a macroscopic steady state can be defined as a state where no more phase mixing will occur. Moreover, since there is a difference in the definition of microscopic and macroscopic steady states one may envisage that, at least some, instabilities can be seen as an evolution of a non-steady (macroscopic) state.

Evolution by phase mixing of collisionless self-gravitating matter is conveniently studied using H -functions (Tremaine, Hénon, & Lynden-Bell 1986; Tolman 1938). This approach, which relies on coarse-graining the phase-space density, encompasses many of the problems met using, e.g., linear-stability analysis or entropy arguments, but it will be stressed that it also leads to a somewhat different concept of stability.

The paper is organized as follows: In Section 2 the theoretical basis of H -functions is introduced, and the idea behind this paper is outlined and put into a mathematical framework. A new definition of a macroscopic steady state is proposed, and a general theorem on the existence of such states is proven. Section 3 is devoted to a discussion of the constraints, such as conservation of energy, which are necessarily imposed on the system. In Section 4 the developed theory is used to derive a new stability criterion, the consequences of which are explored in Section 5 with a special emphasis on the case of spherically symmetric systems, including a preliminary test. Section 6 contains a discussion of the limitations of the approach and some conclusions.

2. MACROSCOPIC STEADY STATES

2.1 H -functions

A short introduction to H -functions is given below. Following Tremaine et al.

(1986) an H -function is defined as

$$H[F] = - \int C(F) d\tau, \quad (2)$$

where $d\tau = d^3x d^3v$. Here $C = C(F)$ is any convex function depending on the coarse-grained distribution function, F , defined as

$$F(\mathbf{x}, \mathbf{v}, t) = \int f(\mathbf{x}', \mathbf{v}', t) K(\mathbf{x}, \mathbf{v}, \mathbf{x}', \mathbf{v}') d\tau' \quad (3)$$

where K is a kernel,

$$\int K(\mathbf{x}, \mathbf{v}, \mathbf{x}', \mathbf{v}') d\tau' = \int K(\mathbf{x}, \mathbf{v}, \mathbf{x}', \mathbf{v}') d\tau' = 1. \quad (4)$$

For example one may have

$$F(\mathbf{x}, \mathbf{v}, t) = \frac{1}{\Delta\tau} \int_{\Delta\tau} f(\mathbf{x}, \mathbf{v}, t) d\tau, \quad (5)$$

where $\Delta\tau$ is a small phase-space volume centered on $(\mathbf{x}, \mathbf{v})$. F may be considered a macroscopic distribution function, f a microscopic (fine-grained) distribution function.

If $C(F)$ is twice differentiable convexity implies

$$C''(F) \geq 0 \quad \forall F, \quad (6)$$

where the prime denotes differentiation with respect to F , i.e., $C'(F) = dC/dF$. To calculate the change in time of H one lets $F \equiv f$ at a given initial time t_0 . At any other time t_1 (for our purposes $t_1 > t_0$) it is straightforward to show that

$$H(t_1) \geq H(t_0) \quad (7)$$

(Thermaine et al. 1986) or, equivalently,

$$H[F] \geq H[f]. \quad (8)$$

Thus, an H -function can never decrease below its initial value. Notice that this is a purely mathematical result which only exploits the properties of convex functions, the definition of a coarse-grained dF , and equation (1). A monotonic increase of H is not

implied; only close to the initial state the second law of thermodynamics, $dH/dt \geq 0$, is obeyed (Kandrup 1987, Soker 1990) because of the special initial condition $F = f$. As the mechanism, which underlies inequality (8), is phase mixing the final state is said to be more *mixed* than the initial state (Wehrli 1978).

2.2 Outline of the Idea

Consider the evolution of a fine-grained dF , f , from $f(t_0)$ to $f(t_1)$. It trivially follows from the collisionless Boltzmann equation (1) that $H[f(t_0)] = H[f(t_1)]$, i.e., without coarse-graining H is independent of time. But for the observable F evolution leads to phase mixing and by equation (8) $H[F(t_1)] \geq H[F(t_0)]$ for any choice of H .

Problems of evolution vs. stability can thus be studied through a discussion based on comparison of a macroscopic steady state, $F_{ss} = f(t_0)$, with a state close to it, F_1 , arising from coarse-graining $f_1 \equiv f(t_1)$. For small (linear) fine-grained changes (e.g., after a short while) one may have $F_1 = f_1$ in which case H is conserved as $H[F_1] = H[f(t_1)] = H[f(t_0)] = H[F_{ss}]$. $H[F]$ will, however, change at some point during an instability. And if linear evolution of f is not going to change it, the first small change in H will come from non-linear evolution of f : A large (non-linear) fine-grained change may therefore give a small (linear) coarse-grained change.

For a system to be in a macroscopic steady state non-evolution of F_{ss} is required. To prevent evolution, it is therefore demanded that all H -functions stay constant in time, i.e., $\forall H : H[F_1] = H[F_{ss}]$.

However, since one is in general not interested in (or capable of) solving the equations of motion, the present approach will allow F_1 to be *any* conceivable function, which is consistent with the global constraints of the system. In other words, it will be assumed that evolution will occur provided coarse graining is consistent with equation (8), i.e., that all phase-mixed states are available (this is formalized as Hypothesis 1 below).

In this case the condition for evolution still is that $\forall H : H[F_1] \geq H[F_{ss}]$. However, since F_1 may now be any function the converse requirement for non-evolution

becomes that at least one H -function will be smaller, i.e., $\exists H : H[F_1] < H[F_{ss}]$ (Definition 3 below).

2.3 Mathematical Details

The fine-grained distribution f_1 , and thereby F_1 [eq. (3)], is in principle uniquely determined by equation (1). However, information of the specific form of f_1 is not used in the present approach:

Hypothesis 1.— F_1 may be any function, which preserves the given constraints.

This is a reasonable hypothesis; if not obeyed more constraints must be implemented. The constraints are, e.g., conservation of total energy. The class of functions, $F_1 \neq f_1$, sufficiently close to F_{ss} that variational calculus can be used, i.e., $|F_1 - F_{ss}| \ll \epsilon F_{ss}$; $0 < \epsilon \ll 1$, which obey the constraints, will be labeled $\mathcal{D}$.

The first theorem to be proven rests on two definitions:

Definition 2.—A stationary point (of an H -function), F_0 , is defined as

$$\exists H : H[F_1] = H[F_0] \quad \forall F_1 \in \mathcal{D}. \quad (\text{SP})$$

When calculating a stationary point of an H -function the constraints imposed by the conserved quantities must be taken into account. This is most elegantly accomplished by using Lagrange multipliers. Hence, the relevant configurational integral to be varied can be written in a general form as the functional $W = H - X$, where X is the sum of the conserved quantities multiplied by Lagrange multipliers. [In the simplest case only the total mass and energy are conserved. Then $X = \alpha M + \beta \mathcal{E}$. For a more elaborate discussion, see Section 3.] It follows from the definition that the stationary points obey

$$\delta W = \delta H - \delta X = 0, \quad (9)$$

where

$$\delta H = - \int C'(F_0) \delta F d\tau. \quad (10)$$

For later reference the second term of equation (9) can be formally written as

$$\delta X = \int Q \delta F d\tau, \quad (11)$$

where Q is a function that remains to be determined (cf. Section 3). Note that $\delta X = 0$ for variations, δF , for which $F_0 + \delta F \in \mathcal{D}$.

Definition 3.—A macroscopic steady state, F_{ss} , of a collisionless system is defined by

$$\forall F_1 \in \mathcal{D} \quad \exists H : H[F_1] < H[F_{ss}]. \quad (\text{SS1})$$

This is the important new definition of a macroscopic steady state; it means that a steady state is a system which, when left unperturbed, cannot evolve by phase mixing since this would entail the decline of at least one H -function.

Note that the converse interpretation is not viable: The definition does not imply that in the neighborhood of F_{ss} all F_1 mix to F_{ss} . For this to be the case F_1 would have to be a fine-grained DF, cf. equation (8), which it is not.

The reason why a sharp inequality is needed is that evolution would not be prevented if there were only an equal sign in the definition. However, from the definition follows

Lemma 4.—A system is in a macroscopic steady state if and only if

$$\forall F_1 \in \mathcal{D} \quad \exists H : H[F_1] = H[F_{ss}]. \quad (\text{SS2})$$

Proof.—Let an H -function which satisfies (SS1) be labeled H_- . Equation (SS1) leads to first order to

$$H_-[F_1] - H_-[F_{ss}] = \delta H[\delta F] = - \int C'(F_{ss}) \delta F d\tau < 0, \quad (12)$$

where $F_1 = F_{ss} + \delta F \in \mathcal{D}$. According to equation (11) one also has $F_{ss} - \delta F \in \mathcal{D}$, and from equation (12) $\delta H[-\delta F] = -\delta H[\delta F] > 0$. Hence, (SS1) implies that

$$\forall F_1 \in \mathcal{D} \quad \exists H_+ : H_+[F_1] > H_+[F_{ss}], \quad (13)$$

i.e., a steady state cannot have all H -functions decreasing. By linear combination of H_+ and H_- one may obtain a new H -function which obeys (SS2).

That (SS2) implies (SS1) is shown in the following way: Define

$$\hat{C}_{\pm}(F) = \hat{C}(F) \pm c(F); \quad c(F) = \begin{cases} \epsilon F^2, & \text{for } F \in I; \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

where the hat denotes a function of a form which may depend on F_1 . Here $\epsilon \ll 1$ is a small positive constant and the interval, I , is defined by requiring that

$$\hat{C}''_{-}(F) \geq 0 \quad \forall F \in I, \quad (15)$$

i.e., convexity. This is possible provided the uninteresting case $\hat{C}'' \equiv 0$ is excluded. From equations (12), (SS2), and (14) it follows that the variation around the steady state is

$$\delta \hat{H}_{\pm} = \delta \hat{H} \mp \int c'(F_{ss}) \delta F d\tau = \mp 2\epsilon \int F_{ss} \delta F d\tau. \quad (16)$$

The condition (SS1) follows by noting that $\delta \hat{H}_{+}$ and $\delta \hat{H}_{-}$ are non-zero and have opposite signs if $F_1 \neq f_1$. Q.E.D.

A consequence of the Lemma is that a steady state can be regarded as "fully phase mixed" and therefore arbitrarily close to a fine-grained distribution function. Since F_{ss} is unchanged in time $H[F_{ss}]$ is unchanged, which is a characteristic of fine-grained distribution functions.

A central result of this paper may now be stated:

Theorem 5.—(i) Any distribution function, F_0 , that is a stationary point of an H -function is a macroscopic steady state. (ii) Any distribution function, F_0 , that is a stationary point of a function, which is not an H -function, is not a macroscopic steady state.

Proof.—Any F_0 which obeys (SP) also obeys (SS2), i.e., a stationary point of an H -function is a steady state. Note, however, that this argument does not prove that any macroscopic steady state is a stationary point for an H -function.

To prove the second part of the theorem, one must demonstrate that the condition

$$\hat{H}[F_1] - \hat{H}[F_0] = 0 \quad \forall F_1 \in \mathcal{D}, \quad (17)$$

[cf. (SP)], where $\hat{H}$ is not an H -function, i.e.,

$$\hat{H}[F] = - \int \tilde{C}(F) d\tau; \quad \exists F : \quad \tilde{C}''(F) < 0, \quad (18)$$

is inconsistent with (SS2)

$$\hat{H}[F_1] - \hat{H}[F_0] = 0 \quad \forall F_1 \in \mathcal{D}. \quad (19)$$

Suppose that the two conditions are consistent, i.e.,

$$- \int \tilde{C}'(F_0) \delta F d\tau = - \int \tilde{C}'(F_0) \delta F d\tau = 0. \quad (20)$$

Choose now δF of the specific form $\delta F = \delta F_A - \eta \delta F_B$, where

$$\delta F_A = \delta^{(e)}(\tau - \tau_A); \quad F_0(\tau_A) = A; \quad (21)$$

$$\delta F_B = \delta^{(e)}(\tau - \tau_B); \quad F_0(\tau_B) = B, \quad (22)$$

and η , A , and B are positive constants; η is determined by the constraint equation (11), $\delta X = Q_A - \eta Q_B = 0$, i.e., $\eta = Q_A/Q_B$. Then

$$\frac{\tilde{C}'(A)}{\tilde{C}''(B)} = \frac{\tilde{C}''(A)}{\tilde{C}''(B)} = \eta. \quad (23)$$

Since $\tilde{C}'$ is at some point a decreasing function of F , $\tilde{C}'$ must be decreasing at some point. But this is not possible since $\tilde{C}$ is convex. Thus, the supposition (20) is wrong and a stationary point of a non- H -function cannot be a steady state. Q.E.D.

The theorem is equivalent to the following statement.—Let F_0 be a stationary point of a functional $-\int \psi(F) d\tau$. Then F_0 is a macroscopic steady state iff ψ is convex.

3. CONSTRAINTS

To be able to compute the stationary points of Theorem 5 it is necessary to incorporate the constraints discussed following Definition 2. Equation (9), $\delta W = 0$, leads, by combining equations (10) and (11), to

$$C'(F_0) = -Q. \quad (24)$$

via independence of integrals. This equation defines F_0 ; furthermore, one obtains

$$C''(F_0) = -\frac{dQ}{dF_0}. \quad (25)$$

From this follows [using convexity, eq. (6)] that the requirement for a stationary point of an H -function is

$$F_0 = F_0(Q); \quad \frac{dF_0}{dQ} \leq 0. \quad (26)$$

Hence, a stationary point of an H -function can be described by a monotonic single-variable DF.

Determining Q via X involves comparing all possible physical realizations of systems consistent with all *a priori* constraints. In general, the total mass, $M = \int F d\tau$ and the total energy, $\mathcal{E} = \int (v^2/2 + \Phi(x)/2) F d\tau$, must be conserved. In addition there may be other constraints:

If the system has a total angular momentum around the z -axis, $\mathcal{L}_z = \int L_z F d\tau$, the constraint function takes the form

$$X = \alpha M + \beta \mathcal{E} + \gamma \mathcal{L}_z, \quad (27)$$

where α , β , and γ are Lagrange multipliers. Then $\delta W = 0$ implies, following, e.g., Ipsier & Horwitz (1979), that

$$Q = \alpha + \beta E + \gamma L_z, \quad (28)$$

α , β and γ being any numbers. Thus the distribution function depends on Jacobi's integral. In a related approach Wiechen & Ziegler (1993), minimizing a coarse-grained energy, calculated macroscopic steady states which are globally stable in a Lyapounov sense. For the case of a non-vanishing total angular momentum they also found the distribution function to depend on Jacobi's integral.

If instead the total angular momentum is zero but the distribution of angular momenta is conserved one gets

$$X = \alpha M + \beta \mathcal{E} + \int \gamma(L') F \delta(L - L') d\tau dL', \quad (29)$$

where α , β , and $\gamma(L')$ are Lagrange multipliers. Constraining the distribution of angular momenta is relevant when the angular momentum of each particle is conserved. This is the case when purely central forces are involved, i.e., the constraint can be used when comparing states, which can be reached by applying radial changes to the density distribution. This is the case for spherically symmetric stellar systems with an anisotropic velocity distribution. Thus, for such systems,

$$Q = \alpha + \beta E + \gamma(L), \quad (30)$$

which implies that the usual angular momentum dependences of the distribution function in terms of L or L^2 are allowed, but that many other choices are possible.

Similarly one may have that the distribution of angular momenta in a given direction is conserved in which case one gets

$$Q = \alpha + \beta E + \gamma(L_z). \quad (31)$$

In general, one may get a combination of many different constraints. In all cases, however, the predicted distribution only depends on a single variable, Q .

4. STABILITY

Theorem 5 obviously entails a *necessary* condition for a system to be stable: Since a stable system must necessarily be in a steady state from the outset, it follows that if F_0 is a single-variable DF corresponding to the extremum of some entropy-like functional, then it must be monotonic to be stable according to equation (26). However, to look deeper into the problem of stability, in particular the question whether *sufficient* conditions can be derived, one must consider the system's response to a perturbation of f_0 . This is the goal of this section, where in fact, the following important result is demonstrated:

Theorem 6.—A stationary point of an H -function is stable against phase mixing.

Proof.—Given a stationary point for an H -function [cf. (SP)]:

$$H[F_1] = H[F_0]; \quad F_1 \in \mathcal{D} \quad (32)$$

and $F_0 = F_0(Q)$. Consider also a stationary point for the same H -function subject to different constraints, i.e.,

$$H[F_P] = H[F_{0P}]; \quad F_P \in \mathcal{D}_P \quad (33)$$

and $F_{0P} = F_{0P}(Q_P)$. The existence of F_{0P} is ensured by stationarity [eq. (24)], i.e., it exists if F_0 exists (as assumed). Moreover, let the constraints be identical except for a small extra term in X , i.e., $X \approx X_P$ which implies that $Q \approx Q_P$. Provided that F_0 does not display any sensitive dependence of the exact form of the constraints this implies that a small change in the constraints leads to a small change in the distribution function, and thus $F_0 \approx F_{0P}$. Explicitly, one then has

$$X_P = X + \epsilon X_\epsilon \quad Q_P = Q + \epsilon Q_\epsilon \quad F_P = F_0 + \epsilon F_\epsilon, \quad (34)$$

$$|\epsilon| \ll 1.$$

Let now δf represent a small perturbation to the steady fine-grained state, f_0 , and let $f_P = f_0 + \delta f$ be a perturbed state, the coarse-grained analog of which is F_P .

Any F_P belongs to the class $\mathcal{D}_P$. Thus F_P is a steady state [eq. (33)]. Furthermore, it is very similar to the original state, F_0 , [eq. (34)]. This implies that the perturbation leads to a slightly different DF, which is not going to evolve. In this sense, F_0 can be considered stable. Q.E.D.

The physical reason behind the stability of stationary points is simple. By Hypothesis 1 time-fluctuations in F_0 generated by the internal dynamics may have given rise to a small change yielding F_1 but this is unimportant since $H[F_1] = H[F_0]$, by (5F). Hypothesis 1 therefore ensures that all states, F_1 , in the vicinity of a stationary point, which obey the constraints, are in principle reachable and in a sense "belong to" the steady state. However, the argument is directly applicable also to perturbations, not only to time-fluctuations. Therefore, a stationary point is stable against constraint-preserving perturbations. And as shown above, the argument can be generalized to perturbations which do not preserve the constraints.

As argued following the proof of Lemma 4, a "fully phase mixed" DF is for practical purposes indistinguishable from a fine-grained DF; in the following f is therefore used to express an observable DF.

For isotropic systems, $f = f(E)$, one has the following corollary:

Stability criterion 7.—A necessary and sufficient condition for the stability of an isotropic system is that $f(E)$ is monotonic.

Proof.—A necessary condition for a system to be stable is that it is in a steady state. Any $f = f(E)$ is a stationary point of an H -function provided $df/d(\beta E) \leq 0$. The corollary then follows from the Theorems. Q.E.D.

For systems with a non-zero total angular momentum or for spherical systems with an anisotropic velocity distribution one has the following criterion which follows from the Theorems and equation (26).

Stability criterion 8.—A necessary and sufficient condition for the stability of any distribution function, $f = f(Q)$, is that $df/dQ \leq 0$.

5. TEST OF STABILITY CRITERIA—EXAMPLES AND CONSEQUENCES

5.1 Isotropic Systems; $f = f(E)$

To avoid a discontinuous behavior of the system at a certain radius it is natural to require that f vanishes continuously at the escape energy ($df/dE < 0$; $E \rightarrow 0$), i.e., $\beta > 0$. Stability criterion 7 then leads to the well-known sufficient condition for stability,

$$\frac{df}{dE} \leq 0, \quad (35)$$

in excellent agreement with previous work (Antonov 1962; Dorémus, Feix, & Baumann 1971; Sygnet et al. 1984; Kandrup & Sygnet 1985).

It should, however, be noted that the construction of discontinuous models ($\beta < 0$) is feasible analytically as well as in numerical simulations. And, although such unphysical realizations are probably not relevant to real stellar systems, they could be stable steady-state systems. Numerical studies by Hénon (1973) and Barnes et al.

(1986) indeed indicated stability of polytropes with $df/dE > 0$. Thus, for isotropic systems, the stability criterion is consistent with all analytical and numerical works that the author is aware of.

5.2 Radially Anisotropic Spherical Systems; $f = f(E, L)$

For anisotropic systems, $f = f(E, L)$, it is reasonable to require that f also vanishes at large values of L [$\partial f / \partial L < 0$; $L \rightarrow L_c(E)$ where $L_c(E)$ is the angular momentum of a circular orbit with energy E]. This implies that $d\gamma(L)/dL > 0$. As a consequence, generic stable spherical systems with an anisotropic velocity distribution are expected to have isotropic cores and radial orbits in the outer regions.

As described in the previous section, the stability criteria only apply to single-variable distribution functions, $f = f(Q)$. A particularly well-motivated choice of Q is

$$Q_{\text{OM}} \equiv E + \frac{L^2}{2r_a^2}, \quad (36)$$

which yields the class of systems known as (radially anisotropic) Osipkov-Merritt DFs (Osipkov 1979; Merritt 1985a, b; Binney & Tremaine 1987). The anisotropy radius, r_a , indicates where the tangential and the radial velocity dispersions become equal. The DF must be of this form if, at every point in space, the density in velocity space is constant on ellipsoidal figures of rotation (Binney & Tremaine 1987, problem 4.23). Another advantage is that the resulting density structure is completely decoupled from the velocity distribution making it possible to make a model with any degree of radial anisotropy while keeping the density profile fixed.

Below the predictions of Stability criterion 8 are compared with the results of numerical tests for a series of previously studied Osipkov-Merritt models. The results are summed up in Figure 1.

The stability of anisotropic Jaffe-models (Merritt 1985b), which have a density profile similar to that of spherical galaxies, has been investigated in N -body simulations by Merritt & Aguilar (1985) who found that instability sets in when the

anisotropy radius, r_a , is lowered to between 0.2 and 0.3 times the half-mass radius, r_h . Merritt & Aguilar (1985) did note, however, that the instability threshold could be somewhat higher, possibly up to $0.5r_h$. The stability criterion leads to $r_a \geq 0.34r_h$ for stability which is in excellent agreement with the instability threshold found numerically.

In a study of the Osipkov-Merritt King-model Weinberg (1991) found, using linear-stability analysis, that $r_a \approx 1.2r_0$ is the instability threshold for the $\Psi_0 = 5$ case [r_0 is the King radius and Ψ_0 is the dimensionless central binding-energy of the King model]. This is in perfect agreement with the stability criterion, which predicts $r_a = 1.28r_0$ (this corresponds to $0.64r_h$).

Using N -body methods Casertano, Phinney, & Villumsen (1987) have found that the $\Psi_0 = 6$ Osipkov-Merritt King-model is marginally unstable for $r_a = 2r_0$. This implies that the instability limit is a bit higher than $0.75r_h$; the criterion predicts $0.93r_h$.

A test by Palmer & Papaloizou (1987) indicates stability of models with $f \sim E_0 - E - L^2/(2r_a^2)$ for any value of r_a . This is consistent with the stability criterion putting $f = -Q_{\text{OM}}$; $df/dQ_{\text{OM}} = -1$.

In a numerical study of the Plummer model, Dejonghe & Merritt (1988) found instability for $r_a \lesssim 1.1r_p$. Analytically one rather finds $0.90r_p$ using the stability criterion (this corresponds to $0.69r_h$). It is difficult to judge whether this is a significant discrepancy. Whereas the inflection point for $f(Q_{\text{OM}})$ for the models discussed up to this point lies slightly below $Q_{\text{OM}} = 0$, it is situated at the very bottom of the potential in the Plummer model, and for values of r_a above the threshold value one has a steeply rising DF. The global dynamics in numerical simulations is known to be quite sensitive to small errors in the representation of the central potential due to the small number of particles (Merritt & Aguilar 1985). There is therefore a danger that if this steep rise is not correctly caught in the numerical implementation, an artificial turning point could be generated. In that case the threshold would be observed to

be too high, consistent with what Dejonghe & Merritt (1988) found for the Plummer model. Models where the turning point occurs in the center leading to potential errors in the generated df , e.g., due to poor sampling in N -body experiments, are therefore disregarded. Unfortunately, one is faced with such a problem in Saha's (1991) careful numerical study of the anisotropic isochrone sphere which accordingly is left out of the present discussion.

In Figure 1 is plotted the predicted instability threshold (smallest possible value of r_a for a stable model) versus that observed in the four useful numerical tests discussed above. It is evident that there is no contradiction between experiment and theory and that Stability criterion 8 appears to be a useful stability criterion. It would be very interesting, however, to test this numerically for a larger series of realistic models, e.g., King models.

A limitation of the approach is that the theorems do not address the question of the stability of systems which do not have $f = f(Q)$. For example, models with $f(E, L) = g(E)h(L)$ are abundant in the literature and the existence of steady-state systems with $f \neq f(Q)$ has not been ruled out. However, besides the instability taking place for systems with $f = f(Q)$; $df/dQ > 0$, one could imagine another possible instability taking place for systems with $f \neq f(Q)$ which are not in a steady state.

Palmer & Papaloizou (1987) studied analytically df_a of the form $f(E, L) \sim g(E)L^{-\alpha}$ and found them to be unstable (the radial-orbit instability). Generalized polytropes of the form $f(E, L) \sim (E_1 - E)^{\alpha-3/2}L^{2\alpha}$, introduced by Hénon (1973), were studied by Barnes et al. (1986) who concluded from their simulations: "of all the models we have studied, only those with nearly isotropic distribution functions appear to be stable". This conclusion indicates that these (anisotropic) models are simply not steady states from the outset.

Most of the galaxy models, $f(E, L) \sim |E|^{3/2} \exp(-\beta E - \gamma L^2)$, studied by Bertin & Stiavelli (1989) using N -body simulations, were found to have stable density profiles,

but oscillations were found to occur. This was interpreted as a numerical artifact in which case the models are an example of a df with $f \neq f(Q)$, which is a steady state. Alternatively, one may speculate that the oscillations could indicate some evolution of the system, perhaps a reorganization in velocity space.

5.3 General Single-Variable Distribution Functions

The tests discussed up to now have mainly dealt with spherically symmetric systems. However, the theorems are quite general and are applicable to any single-variable distribution function of the type discussed in Section 3. Rowley (1988) has studied distribution functions with $Q = E + \Omega L_z - L_z^2/2r_0^2$, and Arnold (1990) has introduced a set of extended Oipkov-Merritt models, $Q = E \pm L^2/2r_a^2 \pm L_z^2/2r_b^2$. The models used by these authors were both monotonic, and thus presumably stable according to Stability criterion 8. Single-variable distribution functions for triaxial systems have also been advocated by Dejonghe (1993).

5.4 Violent Relaxation, Dissipationless Collapse, and the Radial-Orbit Instability

We have recently proposed a simple isotropic model, based on a violent relaxation scenario, which reproduces de Vaucouleur's $R^{1/4}$ empirical law for the surface-brightness distribution in elliptical galaxies (Hjorth & Madsen 1991). This model may also be relevant for dark-matter halos. The central binding energy, Ψ_0 , is the only free shape parameter. Adopting the view that violent relaxation leads to the extremum of some functional (Lynden-Bell 1967; Tremaine et al. 1986; Stiavelli & Bertin 1987; Spiegel & Henning 1992; Hjorth & Madsen 1993a) it is likely that real stellar systems have $f = f(Q)$. Violent relaxation predicts systems with an anisotropic velocity distribution with isotropic cores and radial orbits in the outer regions (White 1987; Hjorth & Madsen 1993b). In Section 5.2 it was argued that this is a general feature of stable single-variable distribution functions. Below the Oipkov-Merritt extension of the isotropic model is discussed, and its stability threshold is derived and compared with results of numerical simulations.

The inflection point occurs for values of Q_{OM} close to 0 (Hjorth & Madsen 1993b), just as for the exact $R^{1/4}$ law (Merritt 1985a) and for Jaffe's model (Merritt 1985b). In Figure 2 is shown the instability threshold for various values of Ψ_0 . In the diagram are also plotted the results of the dissipationless-collapse simulations of Londrillo et al. (1991). It is interesting that the curve lies right in the relevant range of values of τ_a .

According to the stability criterion those systems represented by points lying below the curve cannot be stable spherical systems described by the Osipkov-Merritt model introduced here. It is thus likely that they have undergone an instability. If this instability is the radial-orbit (bar-forming) instability these objects are expected to be non-spherical. This is confirmed in Figure 3 in which the points are plotted in a minor-axis to major-axis ratio diagram. A striking feature of this figure is that those points which lie below the curve in Figure 2 are well-separated in the non-spherical part of the diagram. It thus seems that the stability criterion combined with the Osipkov-Merritt model introduced here predicts the instability limit for dissipationless-collapse simulations.

6. DISCUSSION AND CONCLUSION

Given that the some of the arguments presented in this paper are somewhat subtle, in particular the problem of imposing the right set of constraints, the possibility of restrictions on the range of applicability of the theorems derived must be scrutinized in future work:

As the entire approach rests on Hypothesis 1 and a somewhat imprecise definition of the class of functions close to the steady state, $\mathcal{D}$, one will have to establish whether one can rigorously rule out the existence of an interesting class of F_1 close to F_{SS} , which does not obey Theorem 5.

Similarly, the proof of Theorem 6 rests on the supposition that all states close to F_{SS} can be reached by changing the constraints by δX . The range of applicability of the theorem would be restricted by the possible existence of a class of functions that

cannot be reached by δX , but which are nevertheless, in some sense, close to F_{SS} .

A few notes on the difference between instability to phase mixing and instability as it is normally perceived are also appropriate. A genuine instability undoubtedly implies phase mixing; the converse—that phase mixing occurs in a Hubble time whenever it is dynamically allowed—is less evident. The theorems state nothing about whether the phase-mixing evolution is actually possible, and certainly nothing about the time scale and types of instability one may encounter. A “phase-mixing instability” could take infinitely long time to develop and need not do so exponentially fast, i.e., no growing modes.

With these cautionary remarks in mind, the derived stability criteria quite successfully predicts the instability thresholds for different models of the form $f = f(E)$ or $f = f[E + L^2/(2\sigma_a^2)]$ obtained using a variety of methods (Figure 1).

Support to the criteria also comes from the proof, based on the multiple “water-bag” approach, that $\partial f/\partial E < 0$ and $\partial f/\partial L < 0$ represent necessary and sufficient conditions for stability (Gillon et al. 1976). Provided the steady state df respects $f = f(Q)$ this proof leads to Stability criterion 8. The numerical simulations which conflict with the proof were performed on models with $f \neq f(Q)$, e.g., generalized polytropes (Barnes et al. 1986). It may be a clue to the controversies described in the introduction that the present approach restricts attention to single-variable systems, $f = f(Q)$. For example, the proof of Gillon et al. may be correct but less general than claimed and only valid under restricted circumstances.

All Osipkov-Merritt models giving a de Vaucouleurs-like density profile (Merritt 1985a, b; Hernquist 1990) yield an instability threshold of τ_a in the range $0.2\tau_h - 0.4\tau_h$. Thus, if the Osipkov-Merritt procedure is appropriate spherical galaxies undergo an instability if $\tau_a \lesssim 0.4\tau_h$.

The criteria should be tested more extensively using numerical simulations. For example, it would be interesting to map the stability boundary in the King-model (and possibly of globular clusters) as a function of the central potential (or concentration)

using numerical simulations and compare with the theoretical prediction. It turns out the the predicted stability threshold is very sensitive to the central potential, so it would appear to be a more sensitive test than, e.g., the model used in Figure 2. It would also be useful to study the evolution of the "mixing function", $M(V)$, as discussed by Tremaine et al. (1986), to quantify the mixing occurring during an instability.

In conclusion, two powerful theorems with a broad range of astrophysical implications have been presented. The H -function formalism has thus been demonstrated useful, but the approach should only be seen as complementary to more conventional ones. The relation to other stability analyses remains to be established to get information on time scales and types of instability. As mentioned in Section 3, Wiechen & Ziegler (1993) also have developed the idea of macroscopic steady states based on similar ideas as expressed in this paper. It would be an idea to clarify the connection to this formalism. Finally, it should be kept in mind that there may exist auxiliary steady states besides the stationary points. On the other hand, real stellar systems may well obey some "maximum-entropy like" principle, i.e., be stationary points of some H -function.

ACKNOWLEDGMENTS

This work grew out of my collaboration with Jes Madsen on the dynamics of stellar systems. It is a great pleasure to thank him for the many discussions we have had on this subject, for his constant interest, and for encouraging me to write up the ideas presented here. I also wish to thank an anonymous referee for a careful reading of the manuscript and some insightful comments, Kent Bakgaard, Michael Olesen, Bjarne Thomsen, and in particular Fernando Atrio for discussions, and the Carlsberg Foundation for financial support. The project was supported in part by the Danish Natural Science Research Council.

REFERENCES

- Aguilar, L. A., & Merritt, D. 1990, *ApJ*, 354, 33
- Antonov, V. A. 1962, *Vestnik Leningradskogo Univ.* 19, 96. [English translation in *Structure and Dynamics of Elliptical Galaxies*, IAU Symposium No. 127 ed. T. de Zeeuw (Dordrecht: Reidel), 531]
- Antonov, V. A. 1973, in *The Dynamics of Galaxies and Star Clusters*, ed. G. B. Omarov (Alma Ata: Nauka), 139. [English translation in *Structure and Dynamics of Elliptical Galaxies*, IAU Symposium No. 127, ed. T. de Zeeuw (Dordrecht: Reidel), 549]
- Arnold, R. 1990, *MNRAS*, 244, 465
- Barnes, J., Goodman, J., & Hut, P. 1986, *ApJ*, 300, 112
- Bertin, G., & Stiavelli, M. 1989, *ApJ*, 338, 723
- Bertin, G. & Stiavelli, M. 1993, *Rep. Prog. Phys.*, 56, 493
- Binney, J., & Tremaine, S. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)
- Casertano, S., Pinney, E. S., & Villumsen, J. V. 1987, in *Structure and Dynamics of Elliptical Galaxies*, IAU Symposium No. 127, ed. T. de Zeeuw (Dordrecht: Reidel), 475
- Dejonghe, H. 1993, in *Structure, Dynamics and Chemical Evolution of Elliptical Galaxies*, Proc. ESO/EIPC workshop, eds. I. J. Danziger, W. W. Zeilinger & K. Kjr, ESO Conference and Workshop Proceedings No. 45 (Garching: ESO), 337
- Dejonghe, H., & Merritt, D. 1988, *ApJ*, 328, 93

- Dorémus, J.-P., & Feix, M. R. 1973, *A&A*, 29, 401
- Dorémus, J.-P., Feix, M. R., & Baumann, G. 1971, *Phys. Rev. Lett.*, 26, 725
- Fridman, A. M., & Polyachenko, V. L. 1984, *Physics of Gravitating Systems*, Vols. 1 and 2 (New York: Springer)
- Funato, Y., Makino, J., & Ebisuzaki, T. 1992, *PASJ*, 44, 613
- Gillon, D., Dorémus, J.-P., & Baumann, G. 1976, *A&A*, 48, 467
- Hénon, M. 1973, *A&A*, 24, 229
- Hernquist, L. 1990, *ApJ*, 356, 359
- Hjorth, J., & Madsen, J. 1991, *MNRAS*, 253, 703
- Hjorth, J., & Madsen, J. 1993a, *MNRAS*, in press.
- Hjorth, J., & Madsen, J. 1993b, in *Structure, Dynamics and Chemical Evolution of Elliptical Galaxies*, Proc. ESO/EIPC workshop, eds. I. J. Danziger, W. W. Zeilinger & K. Kjær, ESO Conference and Workshop Proceedings No. 45 (Garching: ESO), 263
- Ipsier, J. R., & Horwitz, G. 1979, *ApJ*, 232, 863
- Kandrup, H. E. 1987, *MNRAS*, 225, 995
- Kandrup, H. E., & Sygnet, J. F. 1985, *ApJ*, 298, 27
- Londrillo, P., Messina, A., & Stiavelli, M. 1991, *MNRAS*, 250, 54
- Lynden-Bell, D. 1967, *MNRAS*, 136, 101
- Lynden-Bell, D., & Sanitt, N. 1969, *MNRAS*, 143, 167
- Merritt, D. 1985a, *AJ*, 90, 1027
- Merritt, D. 1985b, *MNRAS*, 214, 25p
- Merritt, D. 1987, in *Structure and Dynamics of Elliptical Galaxies*, IAU Symposium No. 127, ed. T. de Zeeuw (Dordrecht: Reidel), 315
- Merritt, D., & Aguilar, L. A. 1985, *MNRAS*, 217, 787
- Osipkov, L. P. 1979, *Sov. Astr. Lett.*, 5, 42
- Palmer, P. L., & Papaloizou, J. 1987, *MNRAS*, 224, 1043
- Rowley, G. 1988, *ApJ*, 331, 124
- Saha, P. 1991, *MNRAS*, 248, 494
- Soker, N. 1990, *PASP*, 102, 639
- Spergel, D. N., & Hernquist, L. 1992, *ApJ*, 397, L75
- Stiavelli, M., & Bertin, G. 1987, *MNRAS*, 229, 61
- Syget, J. F., Des Forêts, G., Lachèze-Rey, M., & Pellat, R. 1984, *ApJ*, 276, 737
- Tolman, R. C. 1938, *The Principles of Statistical Mechanics* (Oxford: Oxford University Press)
- Tremaine, S., Hénon, M., & Lynden-Bell, D. 1986, *MNRAS*, 219, 285
- Wehrl, A. 1978, *Rev. Mod. Phys.*, 50, 221
- Weinberg, M. D. 1991, *ApJ*, 368, 66
- White, S. D. M. 1987, in *Structure and Dynamics of Elliptical Galaxies*, IAU Symposium No. 127, ed. T. de Zeeuw (Dordrecht: Reidel), 339
- Wiechen, H., & Ziegler, H. J. 1993, *ApJ*, 408, 293

FIGURES

FIG. 1. A test of Stability criterion 8. On the abscissa are plotted the predicted instability limits (the smallest possible r_a) for the four Osipkov-Merritt models described in the text. Against these are plotted the instability limits derived numerically. The diamond is a lower limit. All four points are consistent with the prediction of Stability criterion 8.

FIG. 2. The instability threshold as a function of ψ_0 . Circles represent the end-products of the dissipationless-collapse simulations of Londrillo et al. (1991). These values have been scaled as described in Fig. 5 of Hjorth & Madsen (1991). According to Stability criterion 8 the filled symbols below the curve represent systems that are expected to have undergone an instability.

FIG. 3. Minor-axis (b and c) to major-axis (a) ratios for the points in Fig. 2 as given by Londrillo et al. (1991). Unstable systems are seen to be considerably more flattened than stable systems, except for the two points in the upper right corner.

FIGURE 1