Mon. Not. R. astr. Soc. (1971) **151**, 333–350.

ON THE THEORY OF STABILITY OF GALAXIES

P. Bartholomew

(Received 1970 October 9)

SUMMARY

A sufficient condition for determining the stability of stellar systems is obtained. The criterion has direct physical significance, in that it simply requires the energy of the equilibrium to be a local minimum, when compared with relevant perturbed states. The paper discusses invariants of the flow, whose existence restricts those perturbations considered relevant.

The criterion is of such generality as to encompass most of the existing work in the subject. Particular effort is made to establish a relationship with the Hartree–Fock operator obtained by Lynden-Bell.

1. INTRODUCTION

One of the most remarkable things about most observed galaxies is their regularity of shape, and also the simplicity with which their forms may be classified. Because of this, the majority of larger galaxies may be assumed to be in a steady state, and to be attainable that steady state must be stable.

The theory of the stability of such systems was pioneered by Antonov (1), (2). While Antonov's condition should prove a wide range of steady states stable, it involves the use of trial functions over six variables, and in addition it is difficult to interpret physically.

Lynden-Bell & Sanitt (3) and Milder (4) are in a position to show a limited range of steady states stable, by comparison with analogous gaseous equilibria.

Also progress has been made in the study of uniform density distributions, and conditions for local stability obtained (5), (6).

More recently Lynden-Bell (7) has used the concept of marginal stability to obtain a stability criterion, which should prove powerful. This involves whether or not an operator is positive definite with respect to trial functions taken over the three variables defining position.

This paper obtains the same criterion for stability, but uses very different basic concepts. Here canonical transformation theory is used throughout, first to obtain an energy criterion, and then to show it to be equivalent to Lynden-Bell's criterion.

2. EQUILIBRIUM

The galaxy is treated as a system of point masses moving under the influence of the gravitational potential of the whole system. It is described in detail by a distribution function $\mathscr{F}$ over six dimensional phase space of generalized coordinates q_i and conjugate momenta p_i.

This is defined so that $\mathscr{F}(q_i, p_i, t)\, d\tau$ is the total mass of all those stars with coordinate points lying in phase-space volume $d\tau$, where $d\tau = dq_1\, dq_2\, dq_3\, dp_1\, dp_2\, dp_3$ and is of invariant form for any canonical change of variables.

By considering the continuity of the flow through volume $d\tau$ of phase space, the equation

$$\frac{\partial \mathscr{F}}{\partial t}+\frac{\partial}{\partial q_i}\left(\mathscr{F}\frac{dq_i}{dt}\right)+\frac{\partial}{\partial p_i}\left(\mathscr{F}\frac{dp_i}{dt}\right) = 0 \tag{1}$$

is obtained. Here the repeated suffixes denoting arrays also imply summation, and d/dt is the rate of change operator that follows the motion in phase space.

Use of Hamilton's canonical equations lead to the equation

$$\frac{\partial \mathscr{F}}{\partial t}+\frac{\partial}{\partial q_i}\left(\mathscr{F}\frac{\partial H}{\partial p_i}\right)-\frac{\partial}{\partial p_i}\left(\mathscr{F}\frac{\partial H}{\partial q_i}\right) = 0, \tag{2}$$

where H is the Hamiltonian of a unit mass in the gravitational potential of the system. Finally in terms of a Poisson bracket, defined in Appendix I,

$$\frac{\partial \mathscr{F}}{\partial t}+[\mathscr{F}, H] = 0. \tag{3}$$

This form of the Boltzmann–Liouville equation may be shown to be invariant under canonical transformation. In particular it is valid for rotating frames of reference.

The Hamiltonian H normally has two parts, the first simply a function of the generalized coordinates, corresponding to kinetic energy, and the second linearly dependent on the distribution function. In this work the second part is simply the gravitational potential Ψ.

In cartesian coordinates

$$\rho = \int \mathscr{F}\, d^3v \tag{4}$$

and so

$$\begin{aligned}\Psi &= G\int \frac{\rho'\, d^3r'}{|\mathbf{r}-\mathbf{r}'|}\\ &= G\int \frac{\mathscr{F}'\, d\tau'}{|\mathbf{r}-\mathbf{r}'|}.\end{aligned} \tag{5}$$

In the earlier sections of this work, such an explicit formulation in terms of a particular coordinate system is inconvenient. However, it is significant that part of the Hamiltonian is expressible in the form

$$-\int K(q_i, p_i; q_i', p_i')\mathscr{F}(q_i', p_i')\, d\tau', \tag{6}$$

where K is a symmetric kernel involving some, or all of the variables shown.

Proceeding from equation (2) in a slightly different manner, it may be shown that

$$\frac{\partial \mathscr{F}}{\partial t}+\frac{\partial q_i}{dt}\frac{\partial \mathscr{F}}{\partial q_i}+\frac{\partial p_i}{dt}\frac{\partial \mathscr{F}}{\partial p_i} = 0;$$

or expressed in an alternative form

$$\frac{d}{dt}\mathscr{F}(q_i, p_i, t) = 0, \tag{7}$$

where d/dt is defined as above. This states that the distribution function is constant when following the motion along an orbit. Consequently, to represent a steady state, a distribution function F must have the same value at all points of each orbit; and is therefore expressible as a function of five variables ($I_1 \ldots, I_5$) which are themselves constant along each orbit. These 'variables' are integrals of motion for each star, such as energy or perhaps angular momentum.

Alternatively, information about the properties of steady states may be deduced from equation (3). Since, by definition $\partial F/\partial t = 0$, then

$$[F, H] = 0. \tag{8}$$

It can also be shown that the integrals $I_1 \ldots, I_5$ satisfy the condition

$$[I, H] = 0,$$

so any distribution function of the form

$$F(I_1 \ldots, I_5)$$

is a steady state.

Note in particular $[H, H] \equiv 0$, so H is always one of the integrals.

The above discussion needs modifying in two ways. Firstly, the gravitational field which was assumed smooth does have time dependent fluctuations superimposed, due to the passage of individual stars. However, the relaxation caused by these fluctuations occurs on a time scale too long to be of interest (**8**). Secondly several of the integrals $I_1 \ldots, I_5$ may be densely multivalued at each point in phase space, and so would not form suitable arguments for F which must by definition be single valued. Such integrals are called non-isolating (**9**).

3. PERTURBATION

The distribution function $\mathscr{F}$ representing the perturbed system is allowed to differ from the equilibrium F by a small function δF.

$$\mathscr{F} = F + \delta F. \tag{9}$$

Similarly there is a small change in Hamiltonian δH, caused by the redistribution of mass.

From equation (6)

$$\delta H = -\int K\,\delta F'\,d\tau'. \tag{10}$$

The linearized equation governing the change in time of δF is

$$\frac{\partial}{\partial t}(\delta F) + [\delta F, H] + [F, \delta H] = 0. \tag{11}$$

The aim of stability analysis is to determine whether this change can be growth, or whether it is limited to oscillations about equilibrium. An energy based variational principle is derived here to do this.

A sufficient proof of stability would be provided by proving, that for all relevant perturbations, the total energy of the galaxy increases. Then the growth of any perturbation would violate the conservation of energy.

If, on the other hand, there also exist perturbations which give rise to a decrease of energy, then this is sufficient to show that there exist perturbed states which involve no energy change, and so are not debarred from growth by the necessity to conserve energy. However there are many other reasons why the system may find such states inaccessible. For example, no state which involves a change in total mass could be reached by the growth of a perturbation from near equilibrium, since mass cannot be supplied indefinitely even in a perturbed galaxy. Similarly no state which has been reached by growth may involve a change in angular momentum.

In fact any property which is an invariant characteristic throughout the evolution of a perturbation, must also be conserved while deriving the initial perturbation from the equilibrium.

A class of such invariant properties may be shown to exist by first returning to Hamilton's equations,

$$\frac{dq_i}{dt} = [q_i, H] \quad \frac{dp_i}{dt} = [p_i, H]. \tag{12}$$

They may be shown (**10**) to be the equations of an infinitesimal canonical transformation, which gives new values $q_i + dq_i$, $p_i + dp_i$ from the initial q_i, p_i.

Canonical transformations are normally used to discuss the same physical system in terms of more convenient, or fundamental coordinates. However, as in any transformation theory, canonical transformations may also be used to discuss displacement of the physical system in terms of the original coordinates.

It is in this sense that the system may be regarded as evolving under a sequence of infinitesimal canonical transformations, generated by the Hamiltonian through equation (12).

Consequently the state of a system at any time may be obtained from its initial configuration by a single canonical transformation.

This limitation of states which are attainable introduces an entirely new class of invariant properties. These are the Poincaré integral invariants, defined as follows:

$$J_1 = \iint_{S_2} (dq_1 dp_1 + dq_2 dp_2 + dq_3 dp_3)$$

$$J_2 = \iiiint_{S_4} (dq_1\, dq_2\, dp_1\, dp_2 + dq_2\, dq_3\, dp_2\, dp_3 + dq_3\, dq_1\, dp_3\, dp_1)$$

$$J_3 = \int\!\!\int\!\!\int\!\!\int\!\!\int\!\!\int_{S_6} dq_1\, dq_2\, dq_3\, dp_1\, dp_2\, dp_3$$

where the integrals are taken over arbitrary volumes of phase space of the dimension indicated ((**10**) Ch 8). These invariants characterize each particular flow throughout its evolution.

The lowest order invariant J_1 may be written

$$\oint_{\Gamma} (p_1\, dq_1 + p_2\, dq_2 + p_3\, dq_3)$$

where the integral is taken around the boundary of the arbitrary two dimensional surface S_2. This is a generalization of the hydrodynamical quantity circulation, and the sense in which it is invariant is the same. That is, as the flow progresses the curve Γ will distort, but will always have the same circulation.

The invariance of J_3 also leads to some well-known properties. It states that a 6-volume of phase space will distort, but will not alter its volume. Since the volume, from its definition always contains the same stars its phase space density $\mathscr{F}$ remains unaltered. In physical terms this means a system with a high phase space density region is as much labelled as a system with high angular momentum, and such properties must be conserved in the perturbation. Also it is this property that causes the conservation of entropy,

$$\int \mathscr{F} \log \mathscr{F} \, d\tau$$

during the Hamiltonian evolution of the system

However, for the purposes of this paper the exact nature of these invariants is unimportant, it is only important to realize their existence.

As a result of this, the relevant perturbations δF are those which conserve total mass and angular momentum, and may be generated by canonical transformation. For δF to have grown from equilibrium it would also be necessary to conserve energy. However, since energy (being the Hamiltonian function for the entire system), determines the behaviour of the system, it is profitable to make a fuller investigation of the variations of energy with δF, rather than simply insisting on it being zero. This will enable stable modes, which invariably cause changes in energy, to be discussed. Such modes by their very nature do not grow and equally cannot arise spontaneously from equilibrium.

The perturbation is considered as the result of moving stars at (q_i, p_i) to a nearby point $(q_i+\xi_i, p_i+\eta_i)$. This leads to the conservation of mass, since mass has only been redistributed. For this transformation to be represented by an infinitesimal canonical transformation, there must exist some function $\mathscr{G}$ for which

$$\xi_i = \frac{\partial \mathscr{G}}{\partial p_i},$$

and

$$\eta_i = -\frac{\partial \mathscr{G}}{\partial q_i}. \tag{13}$$

To be expressed in such a manner is obviously a severe restriction of the functions ξ_i, η_i. However it should be noted that all perturbations which could result from applying a time dependent potential field to the equilibrium can be expressed in this form.

Equations (13) may be written more elegantly as

$$\xi_i = [q_i, \mathscr{G}], \quad \eta_i = [p_i, \mathscr{G}].$$

The change F which corresponds to this motion may be found by writing the following equation,

$$\delta F + \frac{\partial}{\partial q_i}(F\xi_i) + \frac{\partial}{\partial p_i}(F\eta_i) = 0 \tag{14}$$

which corresponds to equation (1).

By analogy with the first section this leads to

$$\delta F+[F, \mathscr{G}] = 0 \tag{15}$$

which defines δF in terms of G. It may be of interest to notice that by defining a Lagrangian change ΔF, by

$$\begin{aligned}\Delta F &= \mathscr{F}(q_i+\xi_i, p_i+\eta_i)-F(q_i, p_i)\\ &= \delta F+\xi_i \frac{\partial F}{\partial q_i}+\eta_i \frac{\partial F}{\partial p_i}\end{aligned} \tag{16}$$

on expanding to first order in ξ_i, η_i, the equation

$$\Delta F = 0,$$

equivalent to equation (7), is obtained.

4. ESTABLISHING AN EXPRESSION FOR ENERGY

By substituting

$$\delta F = -[F, \mathscr{G}] \quad \text{from (15)}$$

into the equation (11)

$$\frac{\partial}{\partial t}(\delta F)+[\delta F, H]+[F, \delta H] = 0$$

the equation governing the evolution of $\mathscr{G}$ may be obtained. This equation,

$$\left[F, \frac{\partial \mathscr{G}}{\partial t}\right] = -[[F, \mathscr{G}], H]+[F, \delta H], \tag{17}$$

is of central importance. It is first used to prove the constancy of a quadratic functional of $\mathscr{G}$, shown to be energy.

This is done by multiplying equation (17) by $\partial\mathscr{G}/\partial t$ and integrating over all phase space. On using various properties (A 1) to (A 4), given in the mathematical Appendix I to this paper, this leads to

$$\int \frac{\partial \mathscr{G}}{\partial t}\left[F, \frac{\partial \mathscr{G}}{\partial t}\right] d\tau = -\int \left[F, \frac{\partial \mathscr{G}}{\partial t}\right][H, \mathscr{G}]\, d\tau - \iint \left[F, \frac{\partial \mathscr{G}}{\partial t}\right] K[F', \mathscr{G}']\, d\tau'\, d\tau. \tag{18}$$

The left-hand side of this equation may be shown to be zero by virtue of the expression's skew-symmetry; this being in the sense that

$$\int \frac{\partial \mathscr{G}_1}{\partial t}\left[F, \frac{\partial \mathscr{G}_2}{\partial t}\right] d\tau = -\int \frac{\partial \mathscr{G}_2}{\partial t}\left[F, \frac{\partial \mathscr{G}_1}{\partial t}\right] d\tau. \tag{19}$$

Also the right-hand side of equation (18) may be shown to be symmetric in $\mathscr{G}$ and $\partial\mathscr{G}/\partial t$, leading to a new form for equation (18)

$$\frac{d}{dt}\left\{-\tfrac{1}{2}\int [F, \mathscr{G}][H, \mathscr{G}]\, d\tau-\tfrac{1}{2}\iint [F, \mathscr{G}]K[F', \mathscr{G}']\, d\tau\, d\tau'\right\} = 0. \tag{20}$$

The behaviour of the functional of $\mathscr{G}$ in brackets will be used to provide a suitable criterion for determining stability.

The functional is now shown to be the energy change due to the perturbation, evaluated to the second order in $\mathscr{G}$.

In a stationary frame of reference the Hamiltonian H is the energy per unit mass at that point. In a fixed potential field the total energy E would be given by

$$E = \int FH\, d\tau.$$

However, because in this case altering F at one point alters the energy of other material through altering potential, the expression is modified to become

$$E = \int FH\, d\tau + \tfrac{1}{2}\iint FKF'\, d\tau\, d\tau'. \tag{21}$$

Hence the extra energy, if any, involved in the perturbation is given by

$$\begin{aligned}\delta E &= \int (F+\delta F)(H+\delta H)\, d\tau + \tfrac{1}{2}\iint (F+\delta F)K(F'+\delta F')\, d\tau\, d\tau' - E \\ &= \int H(\delta F)\, d\tau - \tfrac{1}{2}\iint (\delta F)K(\delta F)'\, d\tau\, d\tau'.\end{aligned}$$

The first-order change in energy due to an infinitesimal canonical change given by

$$\delta F = -[F, \mathscr{G}],$$

is

$$\begin{aligned}\delta E &= -\int H[F, \mathscr{G}]\, d\tau \\ &= \int [F, H]\mathscr{G}\, d\tau = 0.\end{aligned} \tag{22}$$

So using this form of δF shows that an equilibrium has no first-order change in energy when perturbed. Equation (22) also shows that only equilibria have this property for all $\mathscr{G}$.

This exactly parallels results for simple dynamical systems.

To obtain an expression for the second-order energy change, δF must be known to higher order in $\mathscr{G}$. The expression

$$\delta F = -[F, \mathscr{G}] + \frac{1}{2!}[[F, \mathscr{G}], \mathscr{G}] - \dots$$

is obtained in Appendix II.

Using this

$$\delta E = -\tfrac{1}{2}\int [H, \mathscr{G}][F, \mathscr{G}]\, d\tau - \tfrac{1}{2}\iint [F, \mathscr{G}]K[F', \mathscr{G}']\, d\tau'\, d\tau \tag{23}$$

to second order. This is exactly the functional proved constant in equation (20). It should be noticed that if the frame of reference is rotating then H is no longer energy. E is then the Jacobi integral invariant, and depends on the total angular momentum about the axis of revolution.

In the next section equation (17) is fourier analysed and the mode structure is investigated.

5. MODE STRUCTURE

This section is not vital to the main stream of the paper, however insight may be gained by considering the modes resulting from fourier analysis. In particular it shows how instabilities may be detected by seeking perturbations which lower the total energy of the system, even though the individual unstable modes must cause no energy change in order to grow. However it is also shown to be possible that a stable mode which lowers total energy could occur. Such a system could not be proved stable by the criterion outlined.

Equation (17) is linear in $\mathscr{G}$, and is only dependent on time through $\mathscr{G}$, and so may be fourier analysed. Natural modes for $\mathscr{G}$ of the form

$$\mathscr{G} = \zeta\, e^{i\omega t},$$

are sought.

Using an operator notation equation (17) may be written

$$i\omega \mathbf{A}\zeta = \mathbf{B}\zeta \tag{24}$$

where the operator $\mathbf{A}$ is defined by

$$\mathbf{A}\zeta = [F, \zeta], \tag{25}$$

and the operator $\mathbf{B}$ by

$$\mathbf{B}\zeta = [H, [F, \zeta]] - \int [F, K[F', \zeta']]\, d\tau' \tag{26}$$

Where $\eta^*\mathbf{A}\zeta$ is defined to be

$$\int \eta^*[F, \zeta]\, d\tau$$

the operator $\mathbf{A}$ is real and skew-hermitian in the sense

$$\eta^*\mathbf{A}\zeta = -(\zeta^*\mathbf{A}\eta)^*$$

Similarly $\mathbf{B}$ is a hermitian operator, being real symmetric. A very wide range of stability problems may be expressed in the form of equation (24). For example the Hamiltonian equations governing the motion of a particle in equilibrium at the centre of a rotating field may be expressed

$$i\omega \left(\begin{array}{cc|cc} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) \left(\begin{array}{c} x \\ y \\ \hline u \\ v \end{array}\right) = \left(\begin{array}{cc|cc} \alpha & 0 & 0 & \Omega \\ 0 & \beta & -\Omega & 0 \\ \hline 0 & -\Omega & 1 & 0 \\ \Omega & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c} x \\ y \\ \hline u \\ v \end{array}\right).$$

In every case $\zeta^*\mathbf{B}\zeta$ is closely connected with the energy change caused by the perturbation

$$\delta E = \tfrac{1}{4}\zeta^*\mathbf{B}\zeta. \tag{27}$$

Even though it is frequently impossible to solve equation (24) to determine the natural modes of the system, some knowledge of their frequencies may be obtained by simply investigating the possible values of δE.

From previous work it is known that if

$$\zeta^* \mathbf{B} \zeta$$

is positive definite then the system is stable. This is easily shown from the equation

$$i\omega \mathbf{A}\zeta = \mathbf{B}\zeta$$

by premultiplying by ζ^* to obtain

$$i\omega\zeta^* \mathbf{A}\zeta = \zeta^* \mathbf{B}\zeta.$$

Since $\zeta^*\mathbf{B}\zeta$ is real, and $\zeta^*\mathbf{A}\zeta$ is pure imaginary it follows that ω cannot be complex unless both $\zeta^*\mathbf{A}\zeta$ and $\zeta^*\mathbf{B}\zeta$ are zero. Thus if

$$\zeta^*\mathbf{B}\zeta > 0 \quad \text{for all } \zeta$$

the system is stable. However it is possible to do better than this by showing that where any unstable mode exists, there also exists a ζ (which is not itself a mode) which makes

$$\zeta^*\mathbf{B}\zeta$$

actively negative. This is far easier to detect than the existence of a zero.

Before doing this it is necessary to show that the existence of a mode ζ with frequency ω, i.e. a solution of

$$i\omega\mathbf{A}\zeta = \mathbf{B}\zeta,$$

also implies the existence of other modes with related frequences.

Firstly there is the complex conjugate ζ^* which has frequency $-\omega^*$. This is obtained by taking the complex conjugate of each term throughout equation (24) to obtain

$$-i\omega^*\mathbf{A}\zeta^* = \mathbf{B}\zeta^*.$$

Since $\zeta^* \, e^{-i\omega^* t}$ is obviously the complex conjugate of $\zeta \, e^{i\omega t}$ this merely shows that it is possible to form real combinations suitable for describing motion in a physical system.

However there exist other modes whose presence is far more significant, whenever ω is complex. This arises from realizing that the transpose of equation (24)

$$i\omega^*\zeta^*\mathbf{A} = \zeta^*\mathbf{B}^\star$$

implies the existence of a mode η satisfying

$$i\omega^*\mathbf{A}\eta = \mathbf{B}\eta.$$

This means that all modes with complex ω cannot be decaying modes. If ζ were a decaying mode with frequency ω, then there automatically exists a growing mode η having frequency ω^*. It will be found that the energy changes due to the combinations $(\zeta+\eta)$ and $(\zeta-\eta)$ are of particular interest.

If only stable modes, having real ω, were present then the total energy change would simply be the sum of the energy changes due to individual modes. Mathematically this arises because the modes are orthogonal with respect to the operator **B**.

⋆ This is an operator equation. The two sides produce the same answer when they operate on an arbitrary ξ.

That is, given two modes ξ_1 and ξ_2 having frequencies ω_1 and ω_2

$$\xi_1{}^*\mathbf{B}\xi_2 = 0 \quad \text{unless} \quad \omega_1 = \omega_2{}^* = \omega_2. \tag{28}$$

However when unstable modes are present it is not correct to assess their contribution to the total energy to be zero, even though individually they have no energy. This is because modes having frequencies ω and ω^* are not orthogonal. Such modes ζ and η give rise to an ' interaction energy ' $\frac{1}{2}\eta^*\mathbf{B}\zeta$ when they occur simultaneously. Thus the energy of the perturbation $(\zeta+\eta)$ is:

$$\begin{aligned}\tfrac{1}{2}(\zeta^*+\eta^*)\mathbf{B}(\zeta+\eta) &= \tfrac{1}{2}\eta^*\mathbf{B}\zeta+\tfrac{1}{2}(\eta^*\mathbf{B}\zeta)^* \\ &= -\tfrac{1}{2}(\zeta^*-\eta^*)\mathbf{B}(\zeta-\eta)\end{aligned}$$

which is also the negative of the energy change due to the perturbation $(\zeta-\eta)$. Thus if such modes occur, there exist both perturbations which increase energy, and others which decrease energy. So if there exist no perturbations which lower energy then unstable modes are known to be absent.

Unfortunately, the existence of a perturbation which lowers the energy does not necessarily indicate the existence of an instability. This is because it cannot be proved, in general that stable modes necessarily increase energy. Such ' negative energy modes ' are likely to occur in disk systems which have differential rotation present. Such modes could be of interest if, for example, they could be shown to be related to spiral structure.

In a certain sense, these modes may be regarded as unstable, in spite of having real frequencies ω. This is not as unreasonable as it may at first appear, because high order analyses often show that it is the normal behaviour of such a mode to increase its amplitude by feeding energy into ordinary stable modes.

This section has shown that the issue of stability can be largely resolved simply by seeking perturbations which lower energy.

6. CONNECTION WITH LYNDEN-BELL'S CRITERION

So far the condition obtained to determine stability is that the functional (23) (usually energy),

$$-\tfrac{1}{2}\int [F, \mathscr{G}][H, \mathscr{G}]\,d\tau-\tfrac{1}{2}\iint [F, \mathscr{G}]K[F', \mathscr{G}']\,d\tau\,d\tau',$$

should be positive for all relevant $\mathscr{G}$. $\mathscr{G}$ was introduced to limit the class of perturbations δF to those which are dynamically feasible. However $\mathscr{G}$ itself may be subject to further constraints, which may enable a wider class of steady states to be shown stable. For example, it appears necessary to conserve total angular momentum when considering a rotating steady state; whereas in a non-rotating case this is unnecessary since violating such a constraint, by causing the system to rotate, will cause an increase of kinetic energy and so will not make any stable system appear unstable.

Also it may be noticed that the functional is zero whenever $[F, \mathscr{G}] = 0$. However, this is a trivial case, which does not imply any instability since then $\delta F = 0$, and there exists no perturbation.

So far the condition involves using a trial function defined over six variables. However by minimizing the above functional, it is possible to obtain Lynden-Bell's ' Hartree–Fock exchange operator criterion ' (7), which depends on trial functions spanning three dimensions only.

6.1 *Minimizing energy*

The second term of the functional (23) may be shown to be

$$\int \frac{1}{8\pi G} (\nabla\psi)^2 \, d^3x \tag{29}$$

since ψ, the gravitational potential due to the perturbation, is given by

$$\psi = \int K[F', \mathscr{G}'] \, d\tau' = -\delta H \tag{30}$$

and

$$\int [F, \mathscr{G}] \, d^3v = \frac{1}{4\pi G} \nabla^2\psi.$$

Since (29) is positive, (23) may be shown to be positive definite if the minimum value of λ is positive, where λ is defined by

$$\lambda = \frac{-\int [F, \mathscr{G}][H, \mathscr{G}] \, d\tau - \int [F, \mathscr{G}]K[F', \mathscr{G}'] \, d\tau \, d\tau'}{\int [F, \mathscr{G}]K[F', \mathscr{G}'] \, d\tau \, d\tau'}. \tag{31}$$

By variation of $\mathscr{G}$ it may be shown that the minimum value of λ is also the lowest eigenvalue of

$$[H, [F, \mathscr{G}]] = -(\lambda+1)[F, \psi]. \tag{32}$$

At this juncture some new operations, developed by McNamara & Whiteman (**11**) are used to change the form of (32). The operators required are introduced below and further properties appear in Appendix I.

If H itself is used as a momentum coordinate, the four other coordinates which have zero Poisson brackets with H are the other independent integrals (8), which may be non-isolating.

The coordinate conjugate to H, (the action variable, χ) is of particular interest. χ gives position in an orbit, and increases steadily in time when following the motion of the stars.

The operation of taking a Poisson bracket with H may now be written in terms of χ, as the rate of change following an orbit viz.

$$[Q, H] = \frac{\partial Q}{\partial \chi}. \tag{33}$$

This enables the definition of an operator ' hat ' which is the inverse of [, H]. First the average of Q, Q—bar, is defined so that

$$\bar{Q} = \lim_{T\to\infty} \left\{\frac{1}{2T}\int_{-T}^{T} Q \, d\chi\right\}. \tag{34}$$

Using this, an integral of Q, which contains no secular terms and is independent of the lower limit of integration, may be defined. This is $Q-$hat where

$$\hat{Q} = \int_0^\chi (Q'-\bar{Q}') \, d\chi' - \overline{\int_0^\chi (Q'-\bar{Q}') \, d\chi'}. \tag{35}$$

It should be noticed that

$$\overline{[Q, H]} = \lim_{T\to\infty} \left\{\frac{1}{2T}\int_{-T}^{T} \frac{\partial Q}{\partial \chi}\, d\chi\right\}$$
$$= 0 \tag{36}$$

for any physical quantity Q.

Using the properties of these operators detailed in Appendix I,

$$[H, [F, \mathscr{G}]] = (\lambda+1)[F, \psi],$$
$$= (1+\lambda)[F, \psi]-(1+\lambda)[F, \bar{\psi}]$$

becomes

$$[F, \mathscr{G}] = (\lambda+1)[F, \hat{\psi}]+R(I), \tag{37}$$

where $R(I)$ is a function of integrals. In particular,

$$R(I) = \overline{[F, \mathscr{G}]}.$$

It is now possible to form a new quadratic functional from (37) using Poisson's equation,

$$\nabla^2\psi = 4\pi G \int [F, \mathscr{G}]\, d^3v.$$

This is done by multiplying by ψ and integrating over volume $d^3x = dV$.

$$\int (\nabla\psi)^2\, dV = -4\pi G \int \psi[F, \mathscr{G}]\, d\tau.$$

But

$$\int \psi[F, \mathscr{G}]\, d\tau = -(1+\lambda) \int \psi[F, \hat{\psi}]\, d\tau - \int \psi\overline{[F, \mathscr{G}]}\, d\tau$$
$$= -(1+\lambda) \int (\psi-\bar{\psi})[F, \hat{\psi}]\, d\tau - \int \bar{\psi}[F, \mathscr{G}]\, d\tau$$

using the fact that the bar operation may be performed on the integrands without altering integrals over phase space. Integrating by parts to rearrange the Poisson brackets, and using the knowledge that

$$[F, \bar{\psi}] = 0$$

gives

$$\int \psi[F, \mathscr{G}]\, d\tau = (1+\lambda) \int [H, \hat{\psi}][F, \hat{\psi}]\, d\tau.$$

So

$$\int (\nabla\psi)^2\, dV = -4\pi G(1+\lambda) \int [F, \hat{\psi}][H, \hat{\psi}]\, d\tau. \tag{38}$$

For stability λ is positive, and so

$$\int [F, \hat{\psi}][H, \hat{\psi}]\, d\tau < 0$$

follows from (38). Assuming this to hold, λ can only be positive if

$$\int (\nabla\psi)^2\, dV + 4\pi G \int [F, \hat{\psi}][H, \hat{\psi}]\, d\tau > 0. \tag{39}$$

This is the criterion for stability sought.

7. DISCUSSION

The first criterion obtained, based on (23), shows marked similarities with Antonov's (1) condition for stability. Both hinge on proving a particular functional positive definite, with respect to trial functions taken over six variables, a formidable task! Antonov's condition is written to apply to distribution functions of energy only; in which case the criterion obtained here reduces to

$$-\int \frac{(\delta F)^2\, d\tau}{\{\partial F/\partial H\}} - 4\pi G \int\int \frac{\delta F\, \delta F'}{|x_i - x_i'|} d\tau\, d\tau' > 0, \tag{40}$$

on observing that for $F = F(H)$

$$[F, \mathscr{G}][H, \mathscr{G}] = \frac{(\delta F)^2}{\{\partial F/\partial H\}}.$$

The expression used in (40) is an energy change, whereas Antonov's condition

$$-\int \frac{q^2\, d\tau}{\{\partial F/\partial H\}} - 4\pi G \int\int \frac{qq'}{|x_i - x_i'|} d\tau\, d\tau' > 0 \tag{41}$$

with

$$q = [\delta F, H],$$

has no simple physical interpretation, even though the criterion gives the same results.

The simplification obtained here is a direct result of introducing $\mathscr{G}$, and using it to form a quadratic hermitian operator equation, from the equation governing (20). Otherwise it would be necessary to use $[H, \delta F]$ and Antonov's condition would result.

To obtain the first criterion of this paper it was necessary to limit the perturbations considered to those which are dynamically feasible. This was done by the use of canonical transformation. Lynden-Bell (7) achieved the same end by a very different method. By restricting his attention to slowly changing modes, he was able to solve the equation

$$i\omega(\delta F) + [\delta F, H] + [F, \delta H] = 0 \tag{11}$$

directly, as a power series in ω. This gives

$$\delta F = [F, \hat{\psi}]$$

to first order, or to higher order

$$\delta F = [F, \hat{\psi}] - i\omega[F, \hat{\hat{\psi}}] - \omega^2[F, \hat{\hat{\hat{\psi}}}] + \dots. \tag{42}$$

This is already more restrictive than the condition

$$\delta F = -[F, \mathscr{G}],$$

which applies in general. For some cases Lynden-Bell is then able to obtain a criterion which is both necessary and sufficient. His criterion is exactly that expressed in equation (39) of this paper.

However in this paper criterion (39) is found to provide a sufficient proof of stability in all cases. Thus the proof given here is both more elegant and more general than Lynden-Bell's for proving stability; but as it stands is incapable of proving a system unstable. Here Lynden-Bell's work considerably enhances the value of the criterion by showing that there are cases when it is both necessary and sufficient.

Such cases include the stability of spherical systems to spherically symmetric perturbation, the stability of axisymmetric modes in disks, and the stability of all modes in systems $F(\mathscr{E})$ solely dependent on energy. In each of these cases the criterion reduces to

$$\int (\nabla\psi)^2\, dV + 4\pi G \int \frac{\partial F}{\partial \mathscr{E}} \overline{(\psi^2 - \bar{\psi}^2)}\, d\tau \geqslant 0 \tag{43}$$

since in such circumstances

$$[F, \hat{\psi}] = \frac{\partial F}{\partial H}[H, \hat{\psi}]$$

$$= \frac{\partial F}{\partial \mathscr{E}}(\psi - \bar{\psi}).$$

This is advantageous because the average $\bar{\psi}$ is easier to evaluate than the indefinite integrals $\hat{\psi}$.

It is of interest to notice that these criteria allow any ψ to be taken as a trial function, rather than just those resulting from densities

$$\delta\rho = \int [F, \mathscr{G}]\, d^3v.$$

By applying boundary conditions to ψ the slightly weaker constraint on δF,

$$\int (\delta F)\, d\tau = 0 \tag{44}$$

may be used.

Through using canonical transformation theory, attention has been drawn to a significant limitation of states which a collisionless system may attain. This has made apparent extremely close connections between the stability of stellar systems, and the stability of single particle equilibria as determined from energy principles.

ACKNOWLEDGMENTS

This paper was financed in part by a S.R.C. research grant, and in part by a grant from the Astronomy Centre, University of Sussex. Particular thanks are due to Dr Lynden-Bell who introduced me to the subject, and to Professor Tayler for giving advice on the preparation of this paper. I am also grateful for hospitality received at Q.M.C. London.

University of Sussex

Received in original form 1970 *June* 5

REFERENCES

(1) Antonov, V. A., 1960. *Astr. Zu.*, **37**, 918, (*Soviet Astr.* **4**, 859).
(2) Antonov, V. A., 1962. *Vest. Leningrad. Univ.*, No. **19**, 96.
(3) Lynden-Bell, D. & Sanitt, N., 1969. *Mon. Not. R. astr. Soc.*, **143**, 167.
(4) Milder, M., 1967. Thesis Harvard University.
(5) Toomre, A., 1964. *Astrophys. J.*, **139**, 1217.

(6) Lynden-Bell, D., 1967. *Relativity Theory and Astrophysics Pt II*, Galactic Structure.
(7) Lynden-Bell, D., 1969. *Mon. Not. R. astr. Soc.*, **144**, 189.
(8) Woltjer, L., 1967. *Relativity Theory and Astrophysics* Pt II, Galactic Structure.
(9) Lynden-Bell, D., 1962. *Mon. Not. R. astr. Soc.*, **124**, 1.
(10) Goldstein, H., 1951. *Classical Mechanics*, Addison-Wesley, Reading Massachusetts.
(11) McNamara, B. & Whiteman, K., 1967. *J. math. Phys.*, **8**, 2029.

APPENDIX I

MATHEMATICAL IDENTITIES

The Poisson brackets are defined by

$$[A, B] = \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}, \tag{A 1}$$

where the repeated suffix is used to imply summation. The basic properties

$$[A, B] = -[B, A], \tag{A 2}$$

and

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \tag{A 3}$$

are used frequently together with the less familiar

$$\int A[B, C]\, d\tau = \int [A, B]C\, d\tau. \tag{A 4}$$

This identity is established by use of the divergence theorem both in q and p-space. The surface integrals are discarded because at least one of the functions A, B, C is zero at infinity

$$\begin{aligned}\int A[B, C]\, d\tau &= -\int C\left\{\frac{\partial}{\partial p_i}\left(A\frac{\partial B}{\partial q_i}\right) - \frac{\partial}{\partial q_i}\left(A\frac{\partial B}{\partial p_i}\right)\right\} d\tau \\ &= -\int C\left\{\frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i}\right\} d\tau \\ &= \int [A, B]C\, d\tau.\end{aligned}$$

The operators bar and hat have already been defined in Section 7 by

$$\bar{A} = \lim_{T\to\infty}\left\{\frac{1}{2T}\int_{-T}^{T} A\, d\chi\right\} \tag{A 5}$$

$$\hat{A} = \int_0^\chi (A' - \bar{A}')\, d\chi' - \overline{\int_0^\chi (A' - \bar{A}')\, d\chi'}. \tag{A 6}$$

It is obvious from the definition of $\hat{A}$ that

$$\bar{\hat{A}} = 0. \tag{A 7}$$

Also where F is a function of integrals of the motion,

$$\widehat{[F, A]} = [F, \hat{A}] \tag{A 8}$$

and

$$\overline{[F, A]} = [F, \bar{A}]. \tag{A 9}$$

If the Hamiltonian is the other function

$$\overline{[H, A]} = [H, \bar{A}] \tag{A 10}$$
$$= 0$$

since $\bar{A}$ is a function of integrals.

$$\int \hat{A}B \, d\tau = -\int A\hat{B} \, d\tau \tag{A 11}$$

follows from

$$\overline{\hat{A}B} = -\overline{A\hat{B}} \tag{A 12}$$

since the χ integration is implied in the integral over phase space $d\tau$. (A 12) is proved by integration by parts.

APPENDIX II

CANONICAL CHANGES TO HIGHER ORDER

The normal expression of any canonical transformation involves six extremely unpleasant simultaneous equations, often rendering it impossible to express the new coordinates explicitly in terms of the old.

The simplification obtained by considering infinitesimal transformations are considerable, and well known. Reference (**10**). Unfortunately the transformations under discussion at present are merely small, not infinitesimal. Consequently the problem of determining physical quantities to higher order may well arise.

To do this the small perturbation is considered to be the result of n lesser perturbations each generated by $\mathscr{G}$, and in the limit as n increases, each truly infinitesimal.

So q_i first becomes

$$q_i + \frac{1}{n}[q_i, \mathscr{G}]$$

which is most conveniently written in terms of operations as

$$\left\{\mathbf{I} + \frac{1}{n}[\ ,\mathscr{G}]\right\} q_i.$$

It then becomes

$$\left\{\mathbf{I} + \frac{1}{n}[\ ,\mathscr{G}]\right\}^2 q_i,$$

and finally

$$\left\{\mathbf{I} + \frac{1}{n}[\ ,\mathscr{G}]\right\}^n q_i.$$

This invites the expansion

$$\left\{\mathbf{I} + n.\frac{1}{n}[\ ,\mathscr{G}] + \frac{n(n-1)}{2!}.\frac{1}{n^2}[\ ,\mathscr{G}]^{[2]} + \ldots\right\} q_i.$$

Where $[\ ,\mathscr{G}]^{[2]}$ is used to mean the operation acting twice viz. $[[\ ,\mathscr{G}], \mathscr{G}]$.

Proceeding to the limit as $n \to \infty$ gives

$$\left\{\mathbf{I}+[\ ,\ \mathscr{G}]+\frac{1}{2!}[\ ,\ \mathscr{G}]^{[2]}+\frac{1}{3!}[\ ,\ \mathscr{G}]^{[3]}\right\} q_i.$$

So the new coordinates may be expressed explicitly by

$$Q_i = q_i+[q_i,\ \mathscr{G}]+\frac{1}{2!}[q_i,\ \mathscr{G}]^{[2]}+\dots$$

$$P_i = p_i+[p_i,\ \mathscr{G}]+\frac{1}{2!}[p_i,\ \mathscr{G}]^{[2]}+\dots$$

or perhaps more evocatively by the formal notation

$$\begin{aligned} Q_i &= \{\exp[\ ,\ \mathscr{G}]\}q_i, \\ P_i &= \{\exp[\ ,\ \mathscr{G}]\}p_i. \end{aligned} \tag{A 13}$$

Since the intent was to obtain a canonical transformation, the above expressions ought to be checked, to see whether this aim has been realized.

This may be done by checking the relationships expressed in terms of Poisson brackets,

$$[Q_i, Q_j] = 0, \quad [Q_i, P_j] = \delta_{ij}, \quad [P_i, P_j] = 0,$$

where δ_{ij} the Krönecker delta. The expression

$$[Q_i, P_j] = [\{\exp[\ ,\ \mathscr{G}]\}q_i, \quad \{\exp[\ ,\ \mathscr{G}]\}p_j]$$

is a form of product between two exponential functions, and just as the proof from series of the result

$$\exp\alpha . \exp\beta = \exp(\alpha+\beta)$$

involves recognition of terms $(\alpha+\beta)^r$, so terms of the type

$$[[q_i, p_j],\ \mathscr{G}]^{[r]}$$

need recognizing here.

Jacobi's identity

$$[[A, B], C]+[[C, A], B]+[[B, C], A] \equiv 0,$$

is used to show

$$[[q_i, p_j],\ \mathscr{G}] = [[q_i,\ \mathscr{G}], p_j]+[q_i, [p_j,\ \mathscr{G}]],$$

and higher-order terms, involving binomial coefficients, follow. For example the second-order terms are

$$[[q_i, p_j],\ \mathscr{G}]^{[2]} = [q_i, [p_j,\ \mathscr{G}]^{[2]}]+2[[q_i,\ \mathscr{G}], [p_j,\ \mathscr{G}]]+[[q_i.\ \mathscr{G}]^{[2]}, p_j].$$

Identifying terms shows

$$\begin{aligned} [Q_i, P_j] &= \{\exp[\ ,\ \mathscr{G}]\}.[q_i, p_j] \\ &= \{\exp[\ ,\ \mathscr{G}]\}.\delta_{ij} \\ &= \delta_{ij} \end{aligned}$$

since

$$\frac{\partial}{\partial q_k}\delta_{ij} = \frac{\partial}{\partial p_k}\delta_{ij} = 0.$$

Also

$$[Q_i, Q_j] = [P_i, P_j] = 0.$$

So the transformation (A 13) is canonical. Besides detailing the canonical transformation to higher order it is perhaps more important to use a similar technique to obtain δF to higher order.

From expression (15), valid for infinitesimal changes,

$$\mathscr{F} = \{\mathbf{I} - [\ ,\ \mathscr{G}]\}F$$

becomes

$$\begin{aligned} \mathscr{F} &= \lim_{n\to\infty} \left\{\mathbf{I} - \frac{1}{n}[\ ,\ \mathscr{G}]\right\}^n \\ &= \{\exp(-[\ ,\ \mathscr{G}])\}F. \end{aligned} \tag{A 14}$$

This expression is used to second order in Section 4, in the form

$$\delta F = -[F, \mathscr{G}] + \frac{1}{2!}[[F, \mathscr{G}], \mathscr{G}] - \ldots . \tag{A 15}$$