THE ASTROPHYSICAL JOURNAL, 409:75–90, 1993 May 20

MAPPING SPHERICAL POTENTIALS WITH DISCRETE RADIAL VELOCITIES

DAVID MERRITT AND PRASENJIT SAHA
Department of Physics and Astronomy, Serin Physics Laboratories, Rutgers University, Piscataway, NJ 08855
Received 1991 December 26; accepted 1992 November 13

ABSTRACT

We present a technique for inferring the gravitational potential $\Phi(r)$ of a hot stellar system from line-of-sight velocity data, without any assumptions about the relative distribution of dark and luminous components. We overcome the indeterminacy of mass estimates based on line-of-sight velocity dispersions by fitting the full distribution of radial velocities. The phase-space distribution function $f(E, L^2)$ of the test sample is recovered as well. Unlike parametric mass estimation methods based on the virial theorem or the core-fitting formula, our technique does not require that the spatial distribution of the test sample bear any relation to that of the matter determining the potential; that is, the mass-to-light ratio may be an arbitrary function of radius. The technique also does not require binning, or the computation of velocity moments. Even for relatively modest samples ($N \approx 300$), the technique places narrower constraints on the matter distribution than can be inferred from the projected velocity dispersion and number density profiles alone. Strong, model-independent constraints on the form of the potential require somewhat larger samples, of order $N = 10^3$ or more.

We use our algorithm to infer, in a nearly model-independent way, the distribution of dark matter near the center of the Coma galaxy cluster. We find that the core radius of the dark matter is very unlikely to exceed about 500 kpc ($H_0 = 100$) but could be much smaller. The matter distribution at larger radii is poorly constrained.

Subject headings: celestial mechanics, stellar dynamics — galaxies: kinematics and dynamics — techniques: radial velocities

1. INTRODUCTION

The realization that most of the matter in the universe is dark has led to renewed interest in techniques for estimating the gravitational potentials of stellar and galactic systems. In disk galaxies, the distribution of matter is well constrained by the rotation curve, and, at least in regions near our Sun, by the variation of stellar number density and velocity dispersion with distance above and below the Galactic plane. As is well known, data of the first sort provide strong evidence for dark matter at large ($\gtrsim 10$ kpc) distances from the centers of spiral galaxies (Casertano & van Albada 1990), while data of the second sort imply that little if any of this dark matter is present in the immediate solar neighborhood (Kuijken & Gilmore 1989). Much less is known about the distribution of mass in spherical or ellipsoidal systems. Dark matter is clearly present in galaxy clusters, based on their high internal velocity dispersions and the high temperature of their intracluster gas (Rood 1981). Massive dark halos have been provisionally detected around some giant elliptical galaxies using X-ray data (Fabian et al. 1986); many dwarf elliptical galaxies appear to have much higher dynamical masses than their stellar content would imply (Pryor 1992). Galactic and globular clusters seem, for the most part, to have masses consistent with their luminous components (Pryor et al. 1988). However, in none of these systems has the *distribution* of gravitating matter been dynamically well constrained. In the case of masses inferred from X-ray data, the usual limitation is the lack of spatially resolved temperature data, which prohibits an unambiguous calculation of the radial variation of the gas temperature. In the case of masses inferred from kinematical data, the most serious problem is the unknown form of the orbital distribution of the observed sample, which makes it impossible to infer a unique potential from line-of-sight velocity dispersions. As a result of these limitations, only crude constraints can currently be placed on, for instance, the central density of dark matter in clusters of galaxies, or the radial form of dark matter halos around elliptical galaxies, or even the mass-to-light ratios of globular clusters outside of their cores (e.g., Peebles 1984; Merritt 1987; Pryor & Kormendy 1990).

The indeterminacy of current methods of mass estimation in "hot" stellar systems may eventually be overcome by the availability of new sorts of data. For instance, the next generation of X-ray observatories should provide enough spectral information about the hot gas in galaxies and galaxy clusters to yield very accurate measurements of their temperature profiles, and hence of their potentials. For relatively nearby systems, such as Galactic and (some) globular clusters, proper motions of individual stars will eventually provide two independent components of their velocity ellipsoids. However, for many stellar systems, data like these will probably never be available; the only useful constraints on their gravitational potentials will continue to be those derived from line-of-sight velocities. The purpose of the present paper is to demonstrate that such data can yield much more information about the dynamical state of hot stellar systems than has heretofore been widely appreciated.

The theoretical motivation for this paper is the result of Dejonghe & Merritt (1992; hereafter Paper I) that, in spherical nonrotating systems, the distribution function $f(r, v_r, v_t)$ is in principle known uniquely if the potential $\Phi(r)$, the projected densities, and the line-of-sight velocities are given exactly. In the spirit of this result, we take the following approach: given a discrete set of projected radii and line-of-sight velocities (r_p, v_p), we calculate the likelihood for observing the given values from an underlying distribution function f in a potential Φ, and then maximize the likelihood with respect to f and Φ. (It should be noted that f corresponds only to a luminous tracer population,

but Φ is the total gravitational potential; in other words, mass and light are completely uncoupled.)

There are many schemes in the literature that try to extract information on f and Φ from projected densities and velocities. The simplest of these are the virial theorem (which estimates the total mass) and the core-fitting formula (which estimates the central density), and a number of alternatives with better statistical properties (e.g., Heisler, Tremaine, & Bahcall 1985; Little & Tremaine 1987). Another approach is to assume some functional form for $f(E, L^2)$ and fit the observed density and velocity dispersion profiles, with Φ derived from f via Poisson's equation (e.g., King 1966; Gunn & Griffin 1979; Kent & Gunn 1982; Bertin, Saglia & Stiavelli 1988). Yet another class of methods (e.g., Newton & Binney 1984; Richstone & Tremaine 1984) starts by assuming Φ and tries to find an f that fits the projected profiles. Dejonghe (1989) adopts such an approach and in addition varies the Φ to achieve an optimal fit. While useful, these methods suffer from one or both of two rather serious disadvantages. The first is the need to make restrictive assumptions, e.g., that mass follows light, or that the functional form of f is known. For instance, the core-fitting formula is valid only when the kinematical tracers are distributed like the mass and have an isotropic velocity distribution everywhere (Merritt 1988). The other disadvantage is that most methods do not attempt to fit all the line-of-sight velocity data, but only the velocity dispersions, and therefore f is not uniquely constrained even if Φ is known exactly. Thus, even with an infinite amount of error-free data, the methods in the literature would require additional ad hoc assumptions (e.g., that f maximizes the collisionless entropy; Richstone & Tremaine 1988) to determine f or Φ uniquely. Such methods are likely to have limited usefulness in cases where the dark matter is distributed differently from the luminous matter. At best, they can sometimes be used to rule out particular, extreme forms for Φ (e.g., Merritt 1987), but they can never make statements about the relative likelihood of different potentials that are consistent with the data.

With a finite amount of data our approach also requires additional assumptions, which appear as a finite set of basic functions for f and some adjustable parameters in Φ; but the size of the basis set and the number of adjustable parameters can be varied according to the amount and quality of data without any change in the method. Briefly, the advantage of our new approach over previous work is that it uses all the available information, and thus can discriminate between models that are equally likely on the basis of velocity moments alone, and does this without imposing unjustified constraints on the form of Φ or f.

In §§ 2 and 3 we describe the important features of the technique, and is § 4 we present the results of extensive testing of the algorithm with simulated data sets. Section 5 presents an application of the technique to the Coma galaxy cluster, in which we obtain the first, reasonably model-independent limits on the central matter density in this system. Section 6 sums up.

2. OVERVIEW OF THE TECHNIQUE

Consider a system of stars (or other objects) distributed according to a spherical and unchanging distribution function $f(E, L^2)$, in a fixed spherical potential $\Phi(r)$. (As usual, E is the orbital energy and L^2 the orbital angular momentum squared, both per unit mass; the potential will always be normalized to zero at infinity.) Since the directly observable quantities are the projected radius r_p and the line-of-sight velocity v_p, we define the "projected distribution function" $\nu_p(r_p, v_p)$ such that $\nu_p(r_p, v_p)dv_p$ is the surface density at r_p of stars with line-of-sight velocities in the interval v_p to $v_p + dv_p$. As shown in Paper I, the relation between ν_p, f, and Φ is

$$\nu_p(r_p, v_p) = 2\int_{r_p}^{\infty} \frac{r\,dr}{\sqrt{r^2 - r_p^2}} \iint dv_{r_p}\, dv_\theta f(E, L^2)\,, \qquad (1)$$

where v_{r_p}, v_θ are orthogonal velocity components along the polar coordinates $\{r_p, \theta\}$ in the plane of the sky; the inner two integrals are over all velocities for which f is nonzero. Note that Φ appears in equation (1) implicitly, through $E = v^2/2 + \Phi(r)$.

Now suppose that $N_{\rm data}$ objects have been drawn from the projected distribution function, i.e., $N_{\rm data}$ pairs (r_p, v_p) have been measured. The problem is to infer the "most likely" f and Φ, given these data. (For the moment, we ignore measurement errors.) If the kinematical sample were very large—of order 10^4 or 10^5 positions and velocities—it would be appropriate to approximate the two-dimensional function $\nu_p(r_p, v_p)$ by constructing a histogram from the data, and to solve equation (1) as an integral equation for f and Φ, via numerical inversion. The same approach might be justified if accurate, line-of-sight velocity profiles were available at a variety of radii, from deconvolution of integrated stellar spectra, for instance (e.g., Bender 1990). However, we are concerned here with the more typical situation in which the kinematical data are limited to, at most, several hundred discrete positions and velocities, too few to permit an accurate reconstruction of $\nu_p(r_p, v_p)$. We will therefore not attempt to solve equation (1) directly. Instead, we seek to maximize the likelihood $\mathscr{L}$, where

$$\mathscr{L} = \prod_{i=1}^{N_{\rm data}} \nu_p(r_p^i, v_p^i)\,, \qquad (2)$$

the product of the probabilities of observing each pair (r_p, v_p). The "most likely" functions Φ and f may be defined as those that maximize $\mathscr{L}$, subject to the constraint that $f > 0$ for every E and L^2 allowed by Φ, and the constraint that the integrated number corresponding to f is $N_{\rm data}$. The numerical problem we are faced with is therefore not one of *inversion*, but rather one of *optimization*. Furthermore, as optimization problems go, this is a rather difficult one: first, because $\mathscr{L}$ depends on f and Φ through a triple integral; second, because the optimization must be carried out subject to the constraint $f > 0$; and third, because the functional forms of f and Φ are unknown a priori, especially in cases where dark matter is thought to be present, which means that their numerical representations should ideally be nonparametric.

Before describing a practical algorithm for solving this problem, we need to ask whether the solutions are likely to be unique, or whether there will always be a family of f's and Φ's equally consistent with any given data set. There are really two issues here. The first is a purely mathematical, but essential, one. Are f and Φ uniquely determined by the projected distribution function? The general answer to this question is, unfortunately, unknown. In a *specified* potential $\Phi(r)$, there is a unique $f(r, v_r, v_t)$ corresponding to a given $\nu_p(r_p, v_p)$ (Paper I). But it is conceivable that more than one Φ could be consistent with a nonnegative f that generates a given ν_p; if that is so, then we can never hope to find a single, most likely, solution from any observed sample, no matter how large. However, a number of arguments suggest that the set of potentials consistent with a given $\nu_p(r_p, v_p)$ is small, perhaps vanishingly so, especially if Φ is constrained to be derivable from a nonnegative and declining mass density. For instance, the requirement that the greatest v_p at every r_p be less than the escape velocity at $r = r_p$ puts a

strong lower limit on Φ at every r (Dejonghe 1987). Among the set of potentials for which all stars are bound, many functions Φ will be in violation of the virial theorem, or its higher order analogs (Kent 1991). Even potentials that satisfy some set of these virial constraints may still be ruled out, if it can be shown that the implied f is negative at some point in phase space (Merritt 1987). Explicit construction of the "most extreme" potentials consistent with a given set of velocity moment profiles—that is, moments over v_p of v_p—suggests, in fact, that the range of allowed potentials drops rapidly as knowledge of $v_p(r_p, v_p)$ increases (Paper I). Thus, although we cannot prove uniqueness, it seems likely that the information contained within the full set of line-of-sight velocity distributions is sufficient to constrain the potential, and thus the distribution function, rather tightly, at least in a spherical system.

Even if the mathematical problem has a unique solution, there is another set of questions to be answered: namely, to what extent does there exist a single pair of functions f and Φ that are "most likely" to have generated a *finite* sample drawn from $v_p(r_p, v_p)$, and how easily can we recover these functions numerically? The answer to these questions must depend on the freedom that we allow for the numerical construction of f and Φ. Suppose, for instance, that our numerical algorithm is capable only of representing a very restricted set of functions $f(E, L^2)$. Such an algorithm might well single out one potential as being most likely, only because the f corresponding to that potential happens to be easily representable; another potential—although equally consistent with the limited data—might be judged less likely, because the algorithm has difficulty representing the corresponding, most likely f. At the other extreme, if f is constructed from a basis set that is truly complete, then the "most likely" solution will always be one in which the projected density is very high (in fact, infinite) at the observed points (r_p, v_p), and very low elsewhere; in other words, the algorithm will start fitting the sampling noise. The standard resolution to this common problem (e.g., Tapia & Thompson 1978) is to allow great freedom in the construction of the unknown function, but to somehow "penalize" solutions that are not sufficiently smooth. Here, we take a slightly different approach. Following Dejonghe (1989), we represent f via a truncated set of basis functions. Retaining too few terms in this truncation would lead to estimates of the potential that are strongly influenced by the character of the terms retained. Including too many terms would lead to solutions for f that are unphysically unsmooth (and would also greatly increase the complexity of the optimization problem). If the truncation of the basis set used to represent f is done "correctly," we would expect to find that any finite data set is equally consistent with a range of potentials, and that this range is not too strongly dependent on the degree to which the basis set is truncated. This is in fact what we find, as described below. Nevertheless, it is important to keep in mind that nonparametric function estimation is not always a well-defined procedure in the absence of some a priori knowledge about the form of the functions to be estimated, and that the range of potentials and distribution functions found by the algorithm described here must be influenced, to some degree, by the choice we make for the numerical representations of f and Φ. (We note in passing that any of the classical mass estimation formulae—including the virial theorem, the "core-fitting" formula, the "projected mass" method, etc.—suffer much more strongly from this problem, since they are all based on stringent, ad hoc assumptions about the form of Φ (e.g., mass follows light) and/or f [e.g., $f = f(E)$].

In spite of these uncertainties, there are two reasons for preferring the approach described here to previous ones. First, by basing the estimation of f and Φ on a likelihood function, we avoid the need to bin the data or to compute moments. Functional estimation based on maximum likelihood is well known to be superior in almost every respect to estimation based on moments, at least when the parameterization of the unknown functions is suitably chosen, and the sample size is sufficiently large; in particular, it is less sensitive to contamination. Second—and more fundamentally—by using *all* the information contained within a kinematical data set, we can greatly reduce the mathematical indeterminacy associated with estimates based on velocity dispersions (or any finite set of velocity moment profiles) alone.

Our algorithm is most similar in spirit to that of Kuijken (1991), who analyzed the matter distribution near the plane of the Galaxy using the density profile and solar-neighborhood kinematics of a sample of K dwarf stars. The spherical problem considered here differs from the planar one treated by Kuijken in two respects. First, our positional data are projected onto the plane of the sky, while Kuijken had access to approximate distances for solar neighborhood stars; and second, the distribution function of a spherical system is not fixed uniquely by the potential and the density profile of a set of tracers, as in the planar case, because of the extra integral of motion in a spherical potential. These two factors make the spherical problem somewhat more complicated than the planar one, as discussed below. Furthermore, whereas Kuijken restricted the stellar distribution function to a particular family, we represent f in a manner that is nearly nonparametric, in order to avoid biasing the inferred Φ. Finally, it is worth noting that the potential in the planar problem can, in principle, be found from a simple point-by-point measurement of the gradient of the stellar pressure, without the construction of a complete dynamical model (Oort 1932). No such approach is possible in the spherical case: the observable moments of the stellar distribution function are insufficient, generally by a large margin, to determine $\Phi(r)$. The only way to make progress in the spherical case is to require at the outset that the observed velocities are chosen from an f that solves Boltzmann's equation, that is, to construct a complete dynamical model. While Kuijken and most earlier authors also followed this route, their motivations were computational or statistical, and had nothing to do with the need to render the problem *mathematically* determinate.

3. NUMERICAL IMPLEMENTATION

We found that the difficulties associated with numerical implementation of the technique described above fell into the following categories.

1. Representation of f and Φ;
2. Computation of the $v_p(r_p^i, v_p^i)$;
3. Optimization of f and Φ, subject to constraints.

We discuss these difficulties in turn.

1. *Representation of f and Φ.*—We would like to choose a parametrization of both functions that is sufficiently general to include practically any form, yet sufficiently compact to permit efficient optimization. One way to achieve this is by expanding Φ and f in complete sets of basis functions $\Phi_j, f_{m,n}$:

$$\Phi(r) = \sum_j a_j \Phi_j(r) ,$$

$$f(E, L^2) = \sum_{m,n} c_{m,n} f_{m,n}(E, L^2) . \quad (3)$$

The algorithm should then determine the most likely values for the coefficients a_j and $c_{m,n}$, given the data; retaining a sufficiently large number of terms in the expansions guarantees that the representations of Φ and f, and hence v_p, will be essentially nonparametric. However, it is well known that this approach has limitations: if the number of coefficients approaches or exceeds the number of data points, then the optimal solution will be very unsmooth, with sharp peaks at the observed values of r_p and v_p. Furthermore, maximization of a function of many variables can be very time consuming. In practice, we need to choose a number of coefficients that is sufficiently large that a wide range of functions can be represented, but sufficiently small that the optimal solution remains smooth and well-defined.

In what follows, we chose for the parametrization of f the form first suggested (in the context of axisymmetric systems) by Fricke (1952), and most recently employed by Dejonghe (1989):

$$f_{m,n}(E, L^2) = (-E)^{n-1/2} L^{2m} . \quad (4)$$

Here it is assumed that $\Phi(r) \to 0$ for $r \to +\infty$, so that $f = 0$ at energies corresponding to stars that are just unbound. The density profile $\nu_{m,n}$ generated by component $f_{m,n}$ in potential $\Phi(r)$ is

$$\nu_{m,n}(r) = 2^m (2\pi)^{3/2} \frac{\Gamma(m+1)\Gamma(n+\frac{1}{2})}{\Gamma(n+m+2)} r^{2m} (-\Phi)^{n+m+1} , \quad (5)$$

and the one-dimensional, radial, and tangential velocity dispersions are

$$\sigma_r^2(r) = \frac{-\Phi(r)}{m+n+2} , \quad \sigma_t^2(r) = \frac{-(m+1)\Phi(r)}{m+n+2} . \quad (6)$$

The distribution function (4) is valid for $m > -1$; the corresponding density profile has finite integrated mass for $n \geq m + 3$. We therefore restricted m and n to the (discrete) values

$$m = 0, 1, 2, \ldots, n-3 , \quad n = 3, 4, 5, \ldots, N .$$

The total number of basis functions corresponding to a given N is then $N_{\rm basis} = \frac{1}{2}(N-2)(N-1)$. This restriction in allowed values of m and n is clearly permissible when describing an f whose series expansion contains only powers of E and L^2 from the restricted set. However, many simple and reasonable forms for f cannot be so expanded, for instance, $f = (-E)^3 \exp^{-L^2}$. On the other hand, there is a theorem of Weierstrass (Rice 1964) that any continuous function in a finite interval can be approximated by means of a polynomial. Accordingly, we fitted the above function in (E, L^2) space, via a least-squares algorithm, with small sets of terms from our basis. We found that both f and v_p could be very well approximated by only approximately five terms, and that the goodness of fit increased rapidly with N. In what follows, we generally took $N_{\rm basis} = 10$ (corresponding to $N = 6$).

The choice of a power-law basis for f was made for two reasons: first—as emphasized by Dejonghe (1989)—because the resulting distribution function is smooth and analytical, unlike in most earlier treatments that approximated f by discrete "lumps" in integral space (e.g., Schwarzschild 1979; Richstone & Tremaine 1984); and second, because such a basis considerably simplifies the computation of $v_p(r_p, v_p)$, as discussed below. However, it is well known (e.g., Dahlquist, Björck, & Anderson 1974) that the use of polynomials to approximate functions has certain limitations, and these limitations certainly affect the behavior of the algorithm discussed here. For instance, functions with sharp rises followed by weakly curved stretches are not well suited for approximation by polynomials. Furthermore, since power-law basis functions are not orthogonal, the "best-fit" coefficients in a polynomial expansion can be strongly interdependent, a fact which complicates the optimization. Finally, restriction of the basis set to positive powers of L^2 means that radially anisotropic f's must be represented, inefficiently, by a sum of terms with alternating signs. In future work, it would be worthwhile to explore other ways of representing f, e.g., via splines.

We note that f appears linearly in equation (1). This means that the dependence of $v_p(r_p, v_p)$ on the parameters (m, n) will also be linear; thus, for a given potential, one need compute $v_p(r_p, v_p)$ only once for each of the $f_{m,n}$, which greatly reduces the amount of computation during the optimization of f. The same simplification is not possible for the potential, since v_p depends nonlinearly on Φ. Furthermore, as discussed below, we found it computationally most convenient to perform the optimization in two stages, by fixing Φ and finding the optimum f for that Φ, rather than solving simultaneously for the optimum f and Φ. For these reasons, a completely general parametrization of Φ was not deemed necessary or desirable. Instead, we typically represented Φ as a function, chosen on physical grounds, with at most two undetermined parameters $\{a_1, a_2\}$. By trying a number of such functions, one could, in principle, find the single, "most optimum" form.

2. *Computation of the $v_p(r_p^i, v_p^i)$.*—In a given potential $\Phi(r)$, the projected distribution function may be written

$$v_p(r_p, v_p) = \sum_{m,n} c_{m,n} v_{p_{m,n}}(r_p, v_p) , \quad (7)$$

where $v_{p_{m,n}}$ is the contribution to v_p from the distribution function component $f_{m,n}$. Using equations (1) and (3), we find

$$v_{p_{m,n}}(r_p, v_p) = \frac{1}{2^{n+1/2}} \int_{r_p^2}^{r_{\rm max}^2(v_p)} \frac{r^{2m}\, dr^2}{\sqrt{r^2 - r_p^2}} \times \int_0^{-2\Phi(r) - v_p^2} [-v^2 - v_p^2 - 2\Phi(r)]^{n-1/2}\, dv^2 \times \int_0^{2\pi} [v^2 \cos^2\theta + (v\cos\phi\sin\theta - v_p \sin\phi)^2]^m\, d\theta . \quad (8)$$

Here $\sin\phi = r_p/r$, $\cos\phi = (r^2 - r_p^2)^{1/2}/r$, and $r_{\rm max}$ is the greatest radius at which a star could have the line-of-sight velocity v_p; thus $\Phi(r_{\rm max}) = -\frac{1}{2}v_p^2$. The choice of a power-law dependence of $f_{m,n}$ on L^2 can now be seen to be a felicitous one from a computational point of view, since the final integral in equation (8) can be expressed in terms of elementary functions. For any Φ, evaluation of $v_{p_{m,n}}$ for each data point r_p^i, v_p^i thus required a two-dimensional integral to be done numerically. We found that Gauss quadrature on a 12×12 grid was sufficiently accurate. Computation time for the $v_{p_{m,n}}$ was generally found to be comparable to the time required to optimize $f(E, L^2)$ in that potential.

3. *Optimization of f and Φ subject to constraints.*—At each iteration in the optimization of f and Φ, one must recompute the predicted $v_p(r_p, v_p)$ at each of the observed points (r_p^i, v_p^i), in order to calculate the current value of the likelihood $\mathcal{L}$. Because of the linear dependence of v_p on the $c_{m,n}$, one need carry out this computation only once (for each of the basis functions $f_{m,n}$, at each of the data points r_p^i, v_p^i) for each poten-

tial. If computing time were not a factor, one could imagine simultaneously optimizing f and Φ, a procedure that would require recomputation of the $v_{p_{m,n}}$'s at every iteration (i.e., for every new Φ). We chose instead to carry out the optimization of f and Φ separately: that is, we first chose a set of values for the parameters (typically only two) that defined the potential; next, we computed the $v_{p_{m,n}}$ in this potential; next, we found the set of coefficients $c_{m,n}$ that maximized the likelihood; and then we repeated the procedure with a new Φ. As long as the basis we choose to construct f is essentially complete, we are then fully justified in making statements about the *relative* likelihoods of the different potentials examined. We cannot, however, claim to have found in this way the *single* most likely potential, unless we try a very large number of different forms $\Phi(r)$. In practice, however, this is not a severe limitation, since modest data sets are not likely to define a unique Φ.

Given this simplification, the numerically most taxing aspect of the optimization was found to be the imposition of the constraints $f > 0$. One avenue would be to include in the objective function a quantity, such as the Boltzmann entropy, that goes strongly negative as f approaches zero at any point. This approach is used, for instance, by Richstone & Tremaine (1988), but was avoided here since it amounts to imposing an additional constraint on f whose physical justification is doubtful. (Such an approach is also inconsistent with our philosophy of inferring f and Φ from the data alone.) Our method for ensuring that $f > 0$ was simply to test the distribution function on a grid in phase space $\{E_j, L_k^2\}$; that is, we required

$$\sum_{m,n} c_{m,n} f_{m,n}(E_j, L_k^2) \geq 0 \quad \text{for all} \quad j, k$$

(Dejonghe 1989). While these constraints are linear in the $c_{m,n}$, the likelihood (2) is not. Thus, standard linear- or quadratic-programming routines—which require objective functions that are linear or quadratic functions of the parameters—will not work. Instead, an attempt was made to perform the optimization using a general, "nonlinearly constrained minimization" routine (NCONF of IMSL, based on the routine NLPQL of Schittkowski 1985). In general, this routine, and a similar one in the NAG subroutine library, did not perform well: both refused to give solutions when the number of $\{E, L^2\}$ grid points on which the constraints were imposed exceeded about 100—an insufficient number with which to ensure positivity of f. Attempts to improve the performance of the routines by rescaling, changing variables, etc., were unsuccessful. Much greater success was achieved using a standard, *un*constrained maximization routine (BCONF of IMSL), with the positivity constraint enforced by an "exterior penalty function" designed to "penalize" solutions in which f is anywhere negative (e.g., Martin 1971). The quantity to be maximized is then

$$\sum_{i=1}^{N_{\text{data}}} \log\left[\sum_{m,n} c_{m,n} v_{p_{m,n}}(r_p^i, v_p^i)\right] - \sum_{j,k} \lambda_{j,k} H(f_{j,k}) f_{j,k}^2 , \qquad (9)$$

with

$$f_{j,k} = \sum_{m,n} C_{m,n} f_{m,n}(E_j, L_k^2) ,$$

the value of f at the grid point (E_j, L_k^2), and

$$H(q) = \begin{cases} 0 , & q \geq 0 \\ 1 , & q < 0 , \end{cases}$$

the step function. The $\lambda_{j,k}$ are positive scaling factors, chosen so that different points in phase space contribute roughly equally to the penalty. Note that, for solutions with $f > 0$ everywhere, the penalty function does not appear in the objective function. However, in practice, the optimum solution always contained at least one phase-space point at which $f = 0$. Because of this, the optimum solution always depended somewhat on the choice of penalty function. In most of the tests described below, the penalty function was computed on a 50×50 grid spaced uniformly in E and $L/L_c(E)$, where $L_c(E)$ is the angular momentum of a circular orbit of energy E. Once an optimum solution was obtained, f was computed on a finer grid (typically 100×100) to check that the nonnegativity constraint was well satisfied; if not, the penalty function was made more stringent, and the optimization repeated.

We note that no nonlinear optimization routine is guaranteed to locate a global maximum in every case. We found that, when the number of basis functions was small ($\lesssim 6$), the optimum solution returned by the routine was very insensitive to the starting vector; while for larger N_{basis}, it sometimes happened that the routine terminated at a local—not global—maximum. Two techniques were found to be effective in distinguishing local from global maxima: first, repeating the optimization from different starting vectors; and second, observing the behavior of the optimum solution as N_{basis} was increased.

4. PERFORMANCE OF THE ALGORITHM

4.1. *Estimation of f Given* Φ

In a specified potential $\Phi(r)$, the projected distribution function $\nu_p(r_p, v_p)$ determines the intrinsic distribution function $f(r, v_r, v_t)$ uniquely (Paper I). Here we evaluate the ability of our algorithm to recover f from a discrete data set, assuming that Φ is known. We note in passing that there may be circumstances where estimation of f with Φ fixed is justified: for instance, when the potential can be determined independently, through hydrostatic arguments, or when direct measurement of the stellar mass function is possible and dark matter is not believed to be present.

We generated simulated data from two distribution functions:

Model 1:

$$f_1(E, L^2) = (-E)^{7/2} L^2 , \qquad (10a)$$

Model 2:

$$f_2(E, L^2) = (-E)^{7/2} e^{-L^2/L_0^2} , \quad L_0 = 0.5 , \qquad (10b)$$

in the Plummer potential

$$\Phi(r) = -\frac{1}{\sqrt{1 + r^2}} . \qquad (11)$$

The first distribution function generates a model in which tangential motions dominate radial ones, with constant velocity anisotropy, $\sigma_t^2(r) = 2\sigma_r^2(r)$. This distribution function is, in fact, a member of our basis set (4), with $n = 4$ and $m = 1$. The density profile is rather peculiar: according to equation (5), $\nu_{1,4}(r) \propto r^2/(1 + r^2)^3$, so that the central density is zero. The second distribution function generates a radially anisotropic model, with a density profile that is strongly centrally peaked. It cannot be exactly represented using the basis (4). Note that both distribution functions imply density profiles different

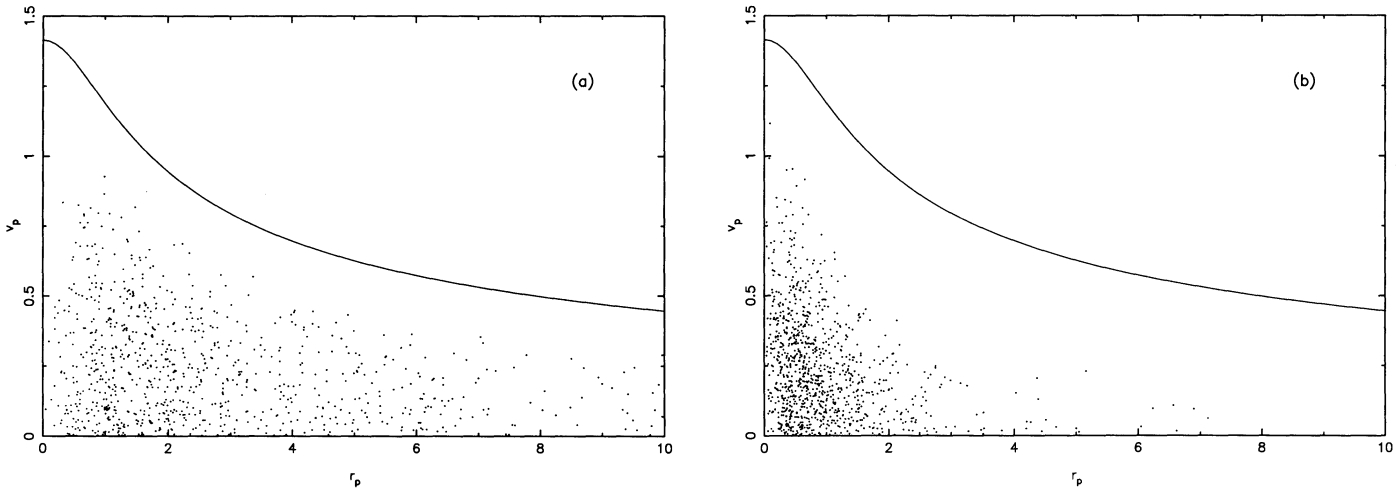


FIG. 1.—Monte Carlo data samples generated from distribution functions f_1 and f_2, with $N_{\rm data} = 1000$, in the Plummer potential. (*a*) model 1; (*b*) model 2.

from that required to generate the Plummer potential (11). Thus "mass does not follow light" in these examples.

We generated Monte-Carlo data sets containing $N_{\rm data} = 100$, 300, and 1000 projected positions and velocities from each distribution function, and carried out the optimization of f with a variety of choices for the number of basis functions $N_{\rm basis}$. Figure 1 shows the 1000 particle data sets generated from the two distribution functions. Figure 2 illustrates the ability of the routine to determine f given data generated from model 1, with $N_{\rm basis} = 10$. Even with 100 data points, the algorithm recovers the number density and velocity dispersion profiles reasonably well, except near the center where the density of particles falls to zero. The distribution function is less well reproduced with this small sample, but the fit improves considerably as N increases from 100 to 1000. Interestingly, even though f_1 was chosen from the basis set (4), the optimal solutions were not found to be strongly peaked around $m = 1$, $n = 4$; in fact, the coefficient $c_{1,4}$ was often negative. This behavior is common when solving integral equations with basis-function expansions, particularly when the basis set is not orthogonal (e.g., Miller 1974; Turchin, Kozlov, & Malkevich 1971), and does not necessarily imply that the basis set is poorly chosen. Figure 3 shows that the algorithm has a slightly harder time recovering the distribution function f_2. Again, the qualitative features of the phase-space distribution are reproduced already for $N_{\rm data} = 100$, but the optimum solutions approach the correct one more slowly as $N_{\rm data}$ is increased. Nevertheless the agreement for $N_{\rm data} = 1000$ is quite good. Note that, in this case, the agreement is best at small radii/low energies, presumably because the density profile places most of the stars there; the algorithm shows no tendency to converge on the correct solution outside of $r \approx 5$, where the number of "stars" is extremely small. Although we do not illustrate it here, comparison of the optimal solutions for $N_{\rm basis} = 3$, 6, 10, and 15 suggests that the goodness of fit—as measured, for instance, by the likelihood—increases only slowly for $N_{\rm basis} \gtrsim 10$. Thus, in most of what follows, we will restrict outselves to 10 basis functions.

4.2. *Estimation of f and Φ*

Next we evaluate the ability of our algorithm to choose the correct form of the potential. For each of the six data sets described above, we computed the most likely distribution function in a two-parameter grid of assumed potentials:

$$\Phi(r; a_1, a_2) = -\frac{a_1}{\sqrt{1 + r^2/a_2^2}}. \tag{12}$$

Figures 4 and 5 are contours of constant $\mathscr{L}$ in (a_1, a_2) space. These figures were constructed by recording the $\mathscr{L}$ corresponding to the most likely distribution function for each choice of a_1 and a_2; thus, each contour may be interpreted as the projection of a 12-dimensional, constant-likelihood region (10 basis functions for f, plus two parameters for Φ) onto a plane. Figures 4 and 5 have a number of features in common. Each exhibits an extended region where the likelihood is roughly constant; outside of this plateau, the likelihood falls off, either gradually—in directions where the potential is becoming deeper; or rapidly—in directions where the potential is becoming shallower. The rapid decrease in likelihood toward shallower potentials is just a consequence of the finite escape velocity; clearly, the likelihood of a potential in which $v_p^2 > -2\Phi(r_p)$, for any single particle, is zero. We interpret the plateau as the region in which there exists a distribution function f whose projected value v_p, at the $N_{\rm data}$ points (r_p, v_p), is essentially the same for each (a_1, a_2) within the region. That is, there appears to exist a range of potentials in which one can construct a nonnegative f which has nearly the same projected value at each of the data points. Because our algorithm is not completely flexible—that is, because the number of basis functions defining f is not infinite—we always expect to find a single, most likely potential (indicated by the dots in Figs. 4 and 5); nevertheless, it is clear that this maximum is ill-defined, standing out only slightly from the surrounding region of parameter space.

As discussed above, the structure in Figures 4 and 5 must depend, to a certain extent, on the selection of basis functions used in the numerical representation of f. Basis functions with much more flexibility would result in plateaus that are larger than the ones found here. In the limit that the basis set consisted of delta functions in (E, L) space, the corresponding projected densities would also be divergent; thus, every potential (among the subset for which all the observed stars are bound) would permit a distribution function whose projected value

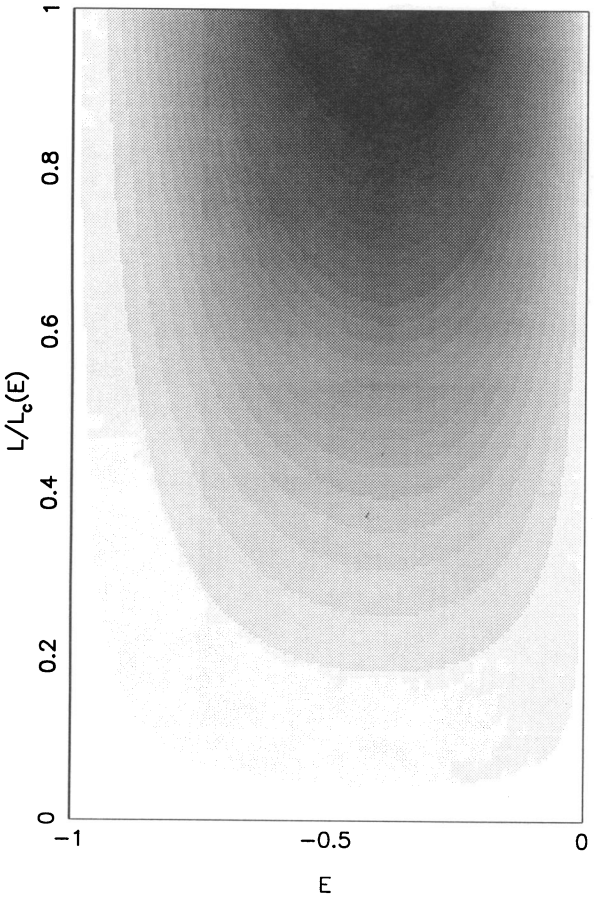
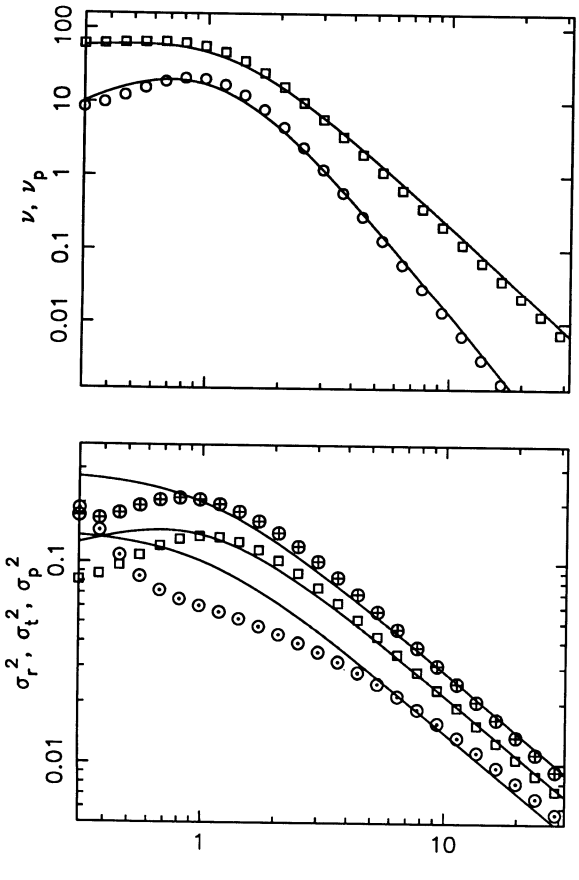
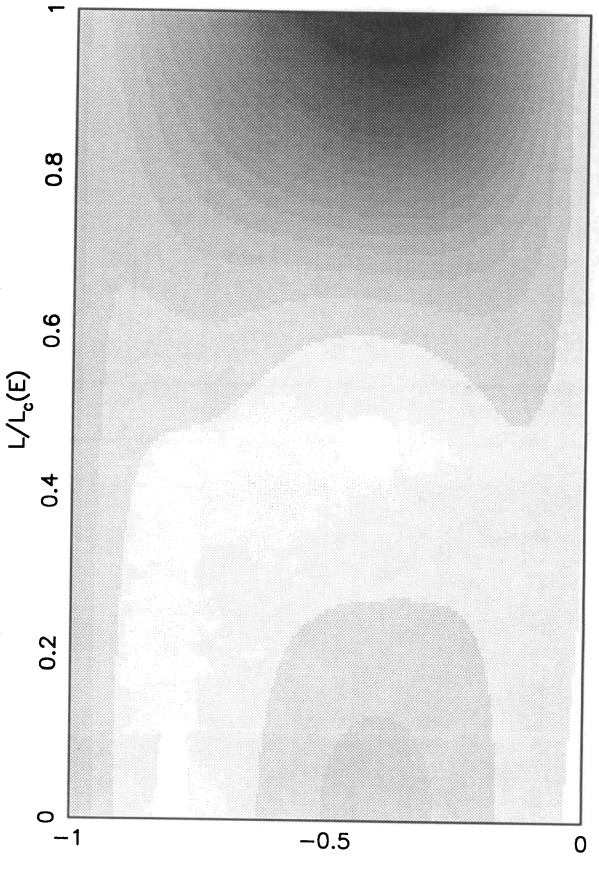


FIG. 2a

FIG. 2.—Ability of our algorithm to recover f given samples generated from distribution function f_1, assuming perfect knowledge of the potential $\Phi(r)$. Right panels are gray-scale plots of the optimal f; the darkness is proportional to $\log [1 + f(E, L^2)E^{-3/2}]$, where the E factor was included to give a better representation of the number density in (E, L)-space. Left panels show low-order moments, intrinsic and projected, of the optimal f: analytic profiles (*solid curves*); intrinsic (◯) and projected (□) densities; intrinsic radial (⊙) and tangential (⊕) velocity dispersions; and projected (□) velocity dispersions. (*a*) Exact solution; (*b*) $N_{\rm data} = 100$; (*c*) $N_{\rm data} = 300$; (*d*) $N_{\rm data} = 1000$.





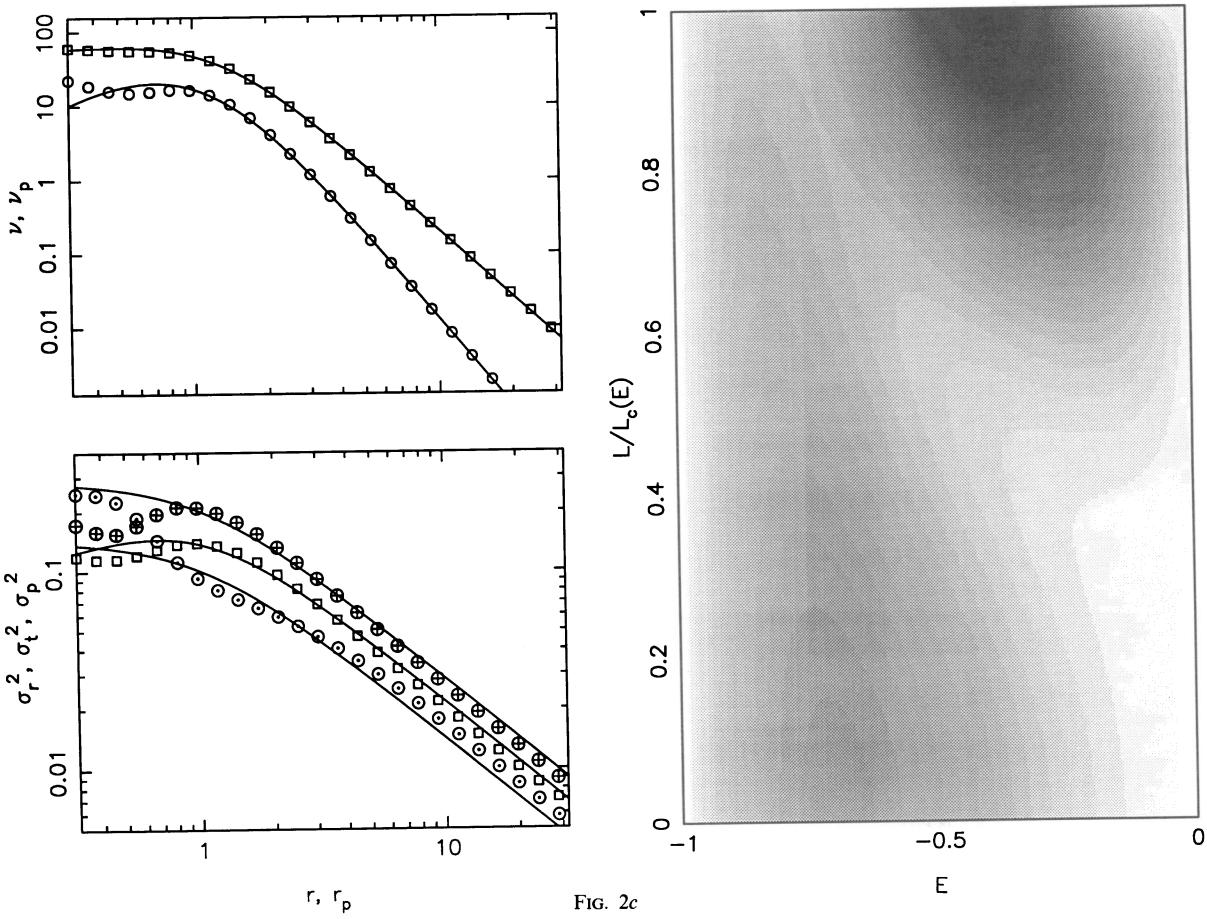


FIG. 2c

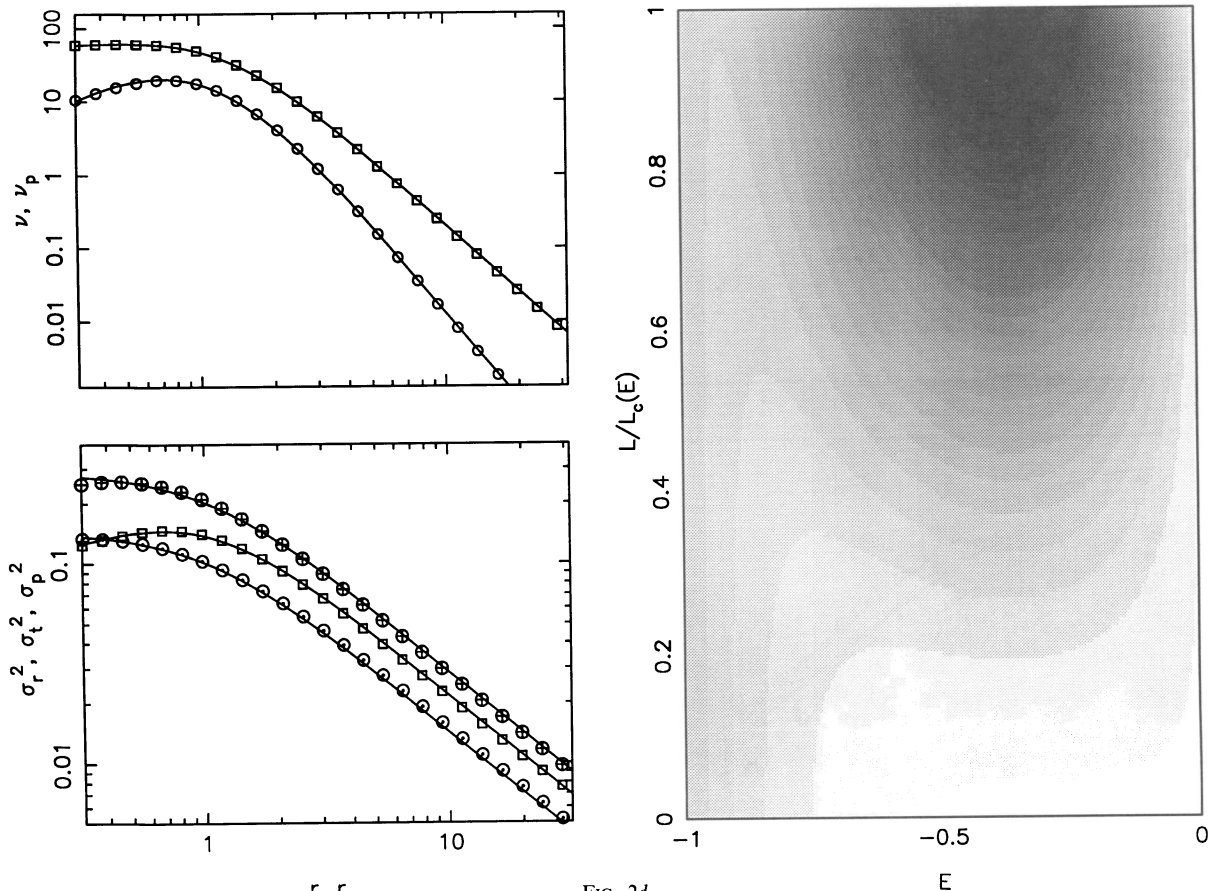


FIG. 2d

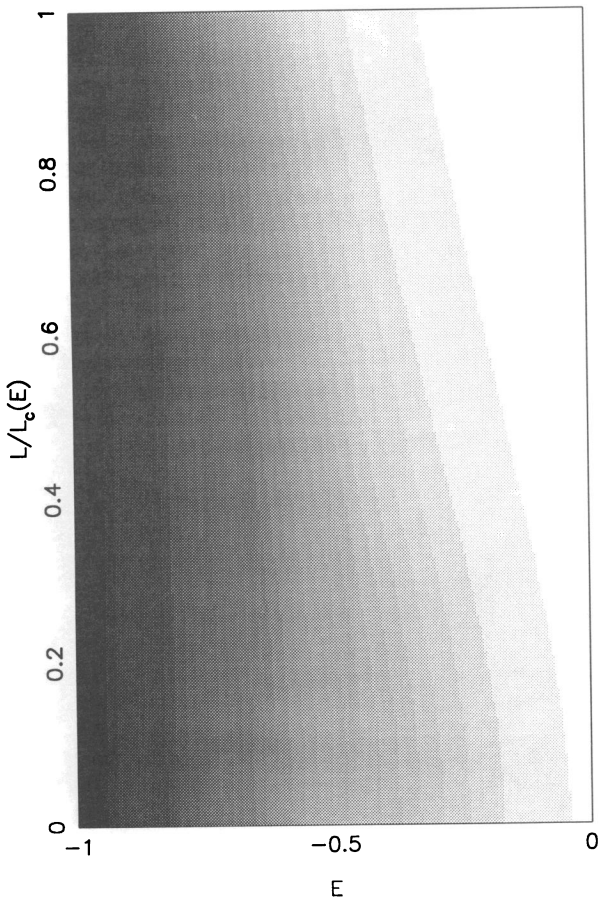
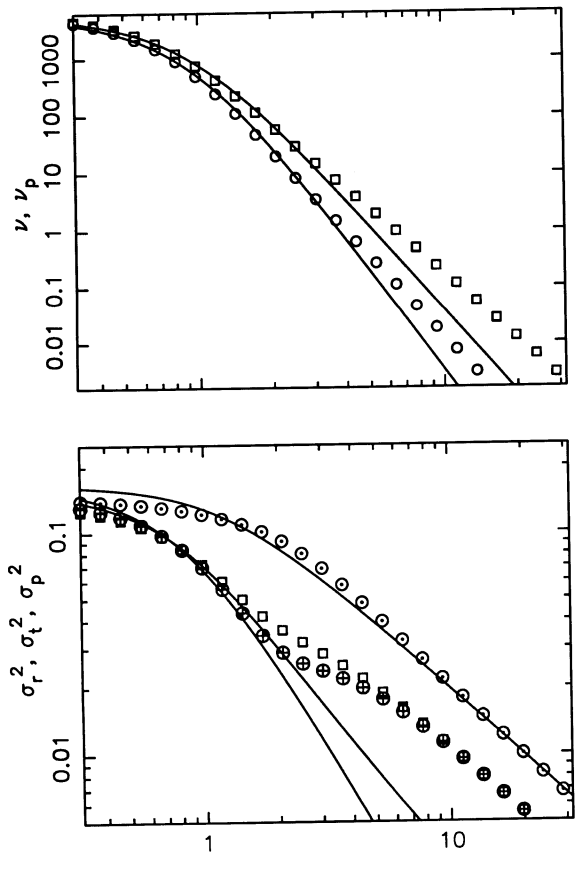


FIG. 3a

FIG. 3.—Like Fig. 2, for samples generated from distribution function f_2.



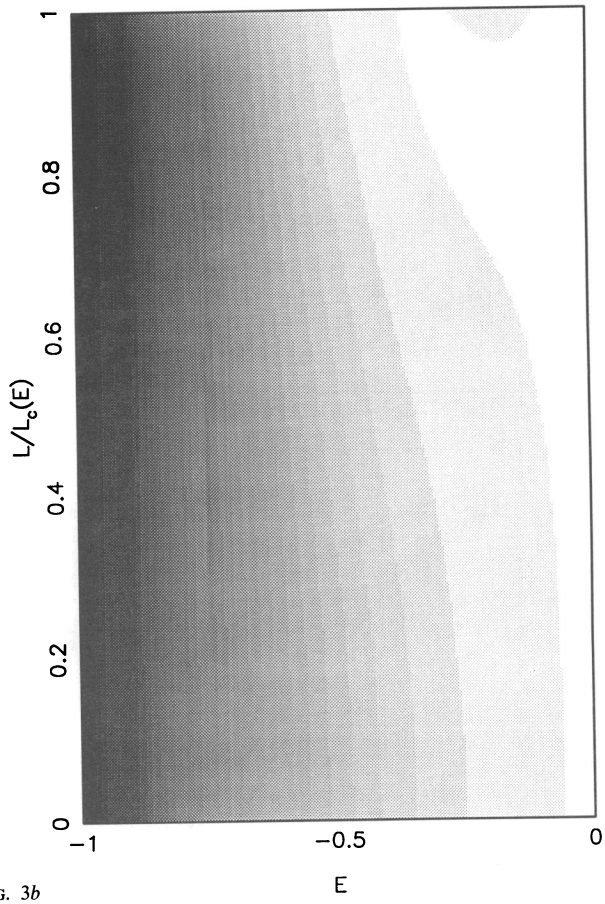


FIG. 3b

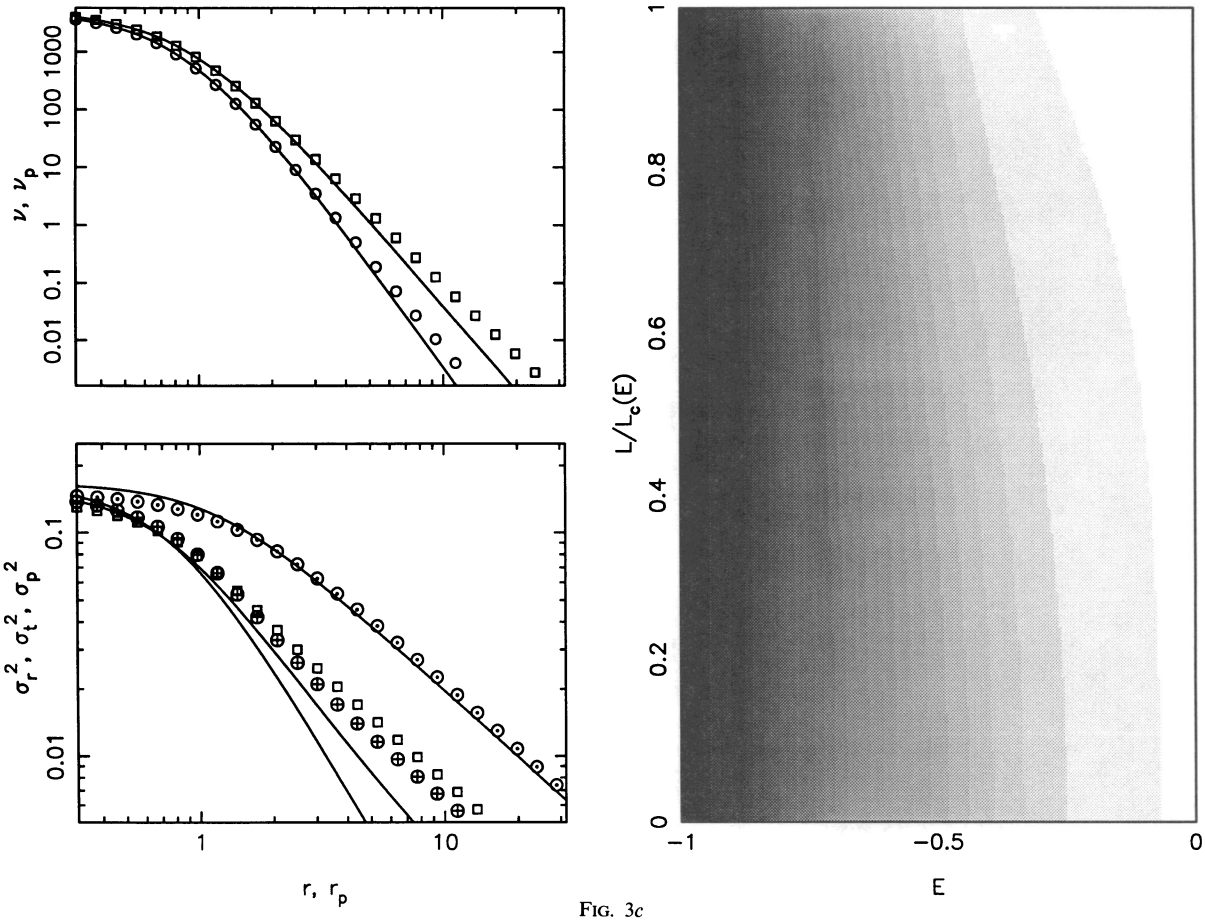


FIG. 3c

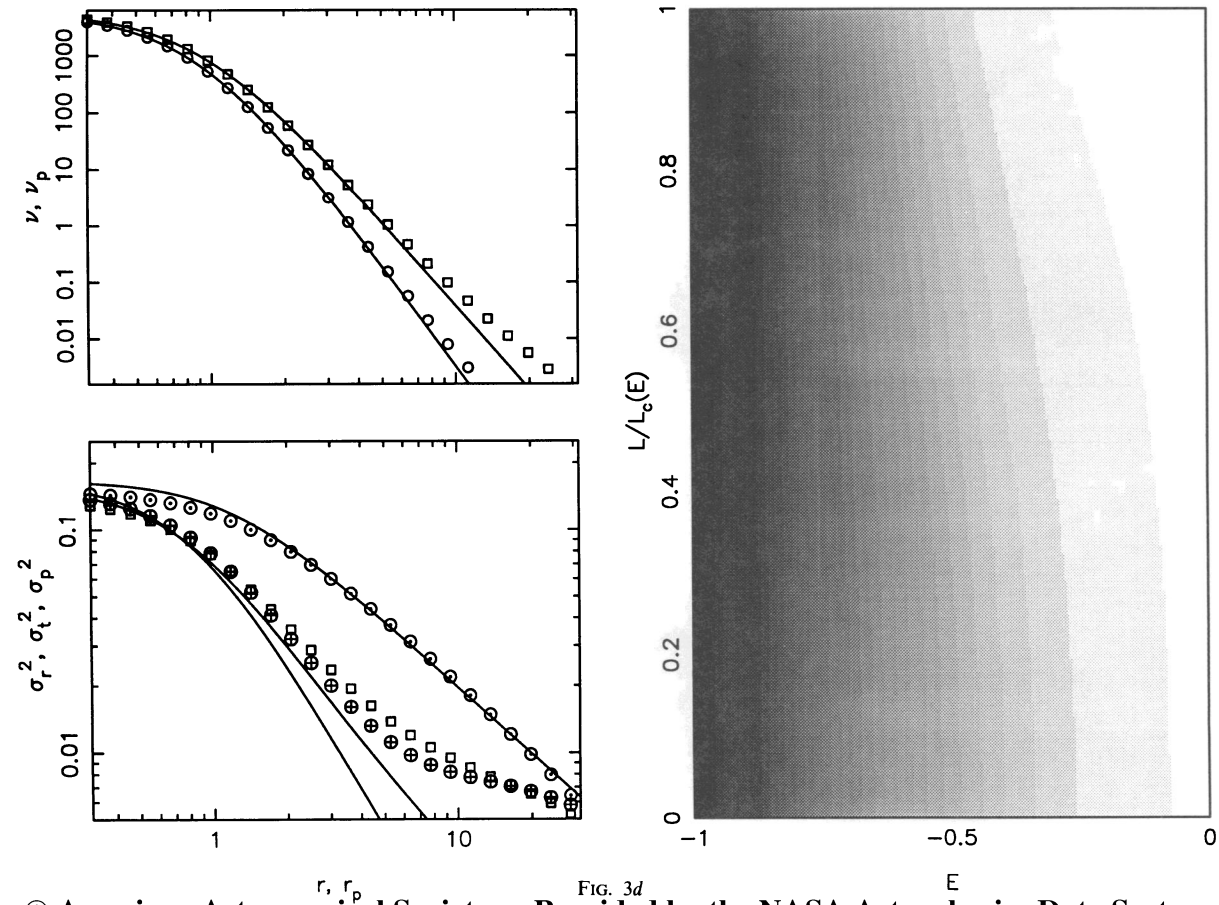


FIG. 3d

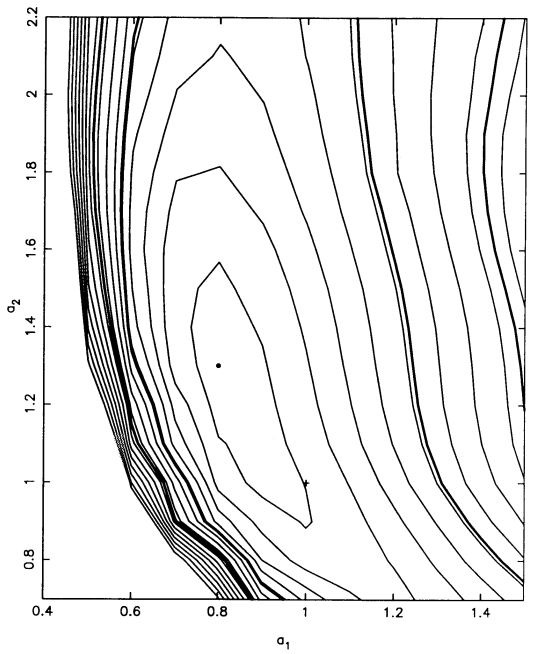


FIG. 4*a*

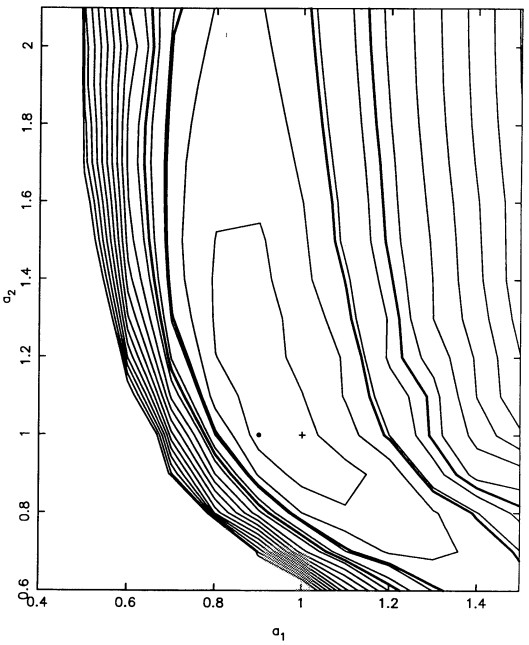


FIG. 4*b*

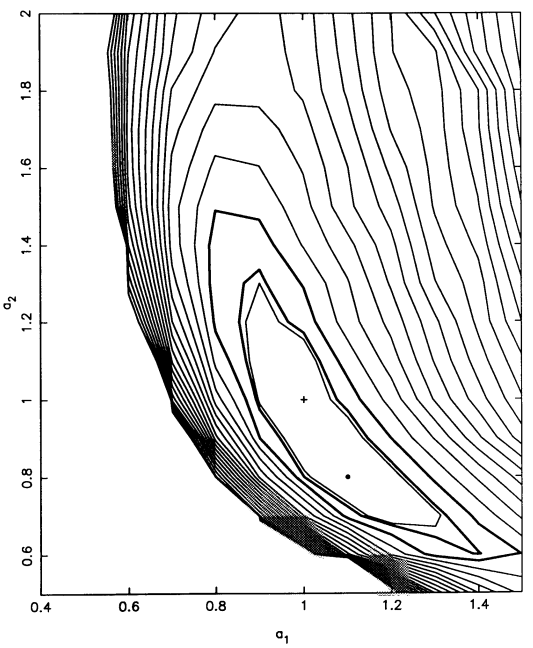


FIG. 4*c*

FIG. 4.—Contours of constant likelihood in the space of parameters (a_1, a_2) that define the potential, for three samples generated from distribution function f_1. The cross identifies the correct solution $(a_1 = 1, a_2 = 1)$, and the dot is the most likely solution found by the algorithm. Heavy curves denote the approximate 50% and 90% confidence regions, as defined in the text. (*a*) $N_{\rm data} = 100$; contours are separated by one in log $\mathscr{L}$; (*b*) $N_{\rm data} = 300$; contours are separated by two in log $\mathscr{L}$; (*c*) $N_{\rm data} = 1000$; contours are separated by four in log $\mathscr{L}$.

was very large at some of the (r_p, v_p), implying a divergent likelihood. Our hope—as in all nonparametric estimation problems—is that, for "reasonably" smooth representations of f, the likelihood contours in Figures 4 and 5 are fairly independent of the precise representation of f. The amount of computational time required to construct Figures 4 and 5 (several hundred hours per figure on a Sparcstation 1) discouraged us from pursuing this question very far. Nevertheless, the fact that the regions of highest likelihood remain centered on the correct solution as the number of particles is increased suggests that our algorithm is estimating likelihoods correctly; if the algorithm were strongly limited in its ability to represent f, we would expect the "most likely" potential to be one for which the corresponding f is easily represented by the code, rather than the correct one, even for very large data sets. Furthermore, we verified that using 15 basis functions, rather than 10, made very little difference in the form of Figure 4*b*. Our tentative conclusion—which ought to be checked more carefully, using a radically different representation of f, if possible—is that our algorithm is doing a fairly good job of delineating the range of "most likely" potentials, rather than simply those potentials for which the corresponding distribution function is well suited for approximation by our particular basis functions.

The estimation of confidence regions for our problem is complicated by the nonnegativity constraints on f. The most secure approach would be to repeat our analysis for many different samples drawn from the same distribution functions, then compute the mean and variance of the potential parameters derived for the different realizations. The available computing resources precluded this approach. Instead, we used the approximate method advocated by Eadie et al. (1971), valid in the limit of large samples: the true potential lies within the region defined by $\log \mathscr{L} > \log \mathscr{L}_{\rm max} - \frac{1}{2}\chi^2_p(\alpha)$ with confidence $C = 1 - \alpha$, where p is the number of degrees of freedom. The heavy contours in Figures 4 and 5 define the 50% and 90% confidence regions, so defined, for a_1 and a_2. For $N_{\rm data} = 100$ and 300, these confidence regions appear to be somewhat too large, based on comparison with the actual errors in a_1 and a_2 for the four samples. For $N_{\rm data} = 1000$, the confidence regions appear more reasonable. The area corresponding to these

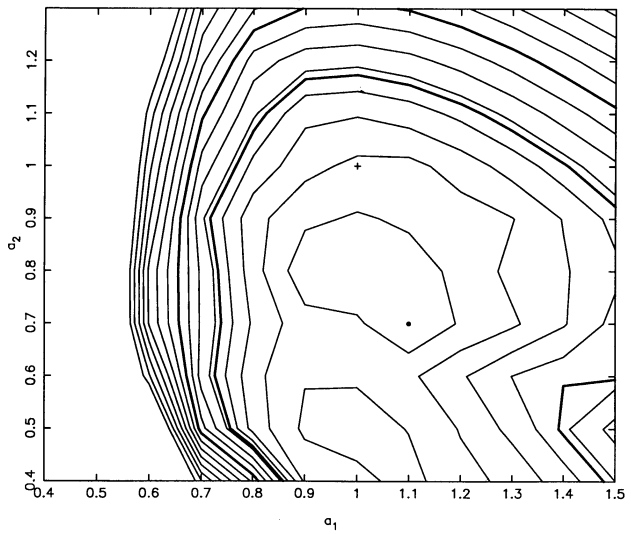


FIG. 5*a*

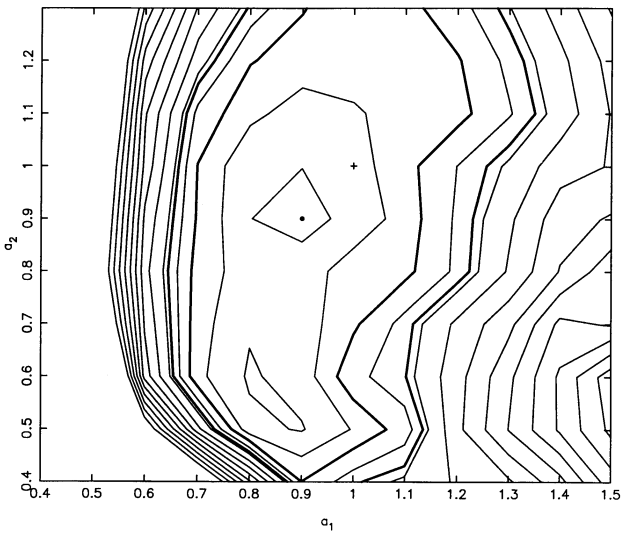


FIG. 5*b*

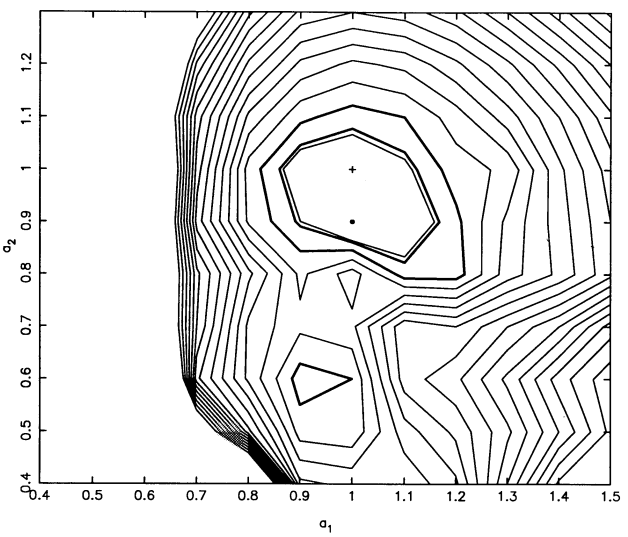


FIG. 5*c*

FIG. 5.—Like Fig. 4, for samples generated from distribution function f_2

regions appears to be roughly the same for the two, $N_{\rm data} = 1000$ samples, although the shapes of the constant-likelihood regions are quite different.

A more intuitive understanding of the shapes of the likelihood contours is provided by Figures 6 and 7. In these figures, contours were plotted in terms of the more physical quantities $a_1 a_2$ and $3a_1/4\pi a_2^2$, which are the total mass and central density, respectively, of the matter producing the potential. The solid curves are the relations between a_1 and a_2 implied by the virial theorem: that is,

$$\langle v^2 \rangle = \langle r\, d\Phi/dr \rangle = a_1 a_2 \langle r^2/(a_2^2 + r^2)^{3/2} \rangle , \quad (13)$$

where brackets indicate averages weighted by number. It is clear that our algorithm for inferring the potential "respects" the virial theorem: in both cases, the contours of constant likelihood bend in such a way as to follow the virial constraint, and the best-fit solution always lies close to the virial theorem curves. Of course, the virial theorem itself gives no clue about where along these curves the most likely solution lies.

It is hardly surprising that the curves of constant likelihood have very different shapes for samples generated from the two distribution functions. However, this fact makes it difficult to estimate the sample size required to give an estimate of, say, the central density with a specified degree of confidence. Roughly speaking, Figures 6 and 7 suggest that samples containing a thousand positions and velocities yield estimates of the central density that are accurate to within a factor of a few, at least when the unknown potential is characterized by only two free parameters. These are substantially tighter limits than can be set using only the virial theorem, or the number density and velocity dispersion profiles (Paper I).

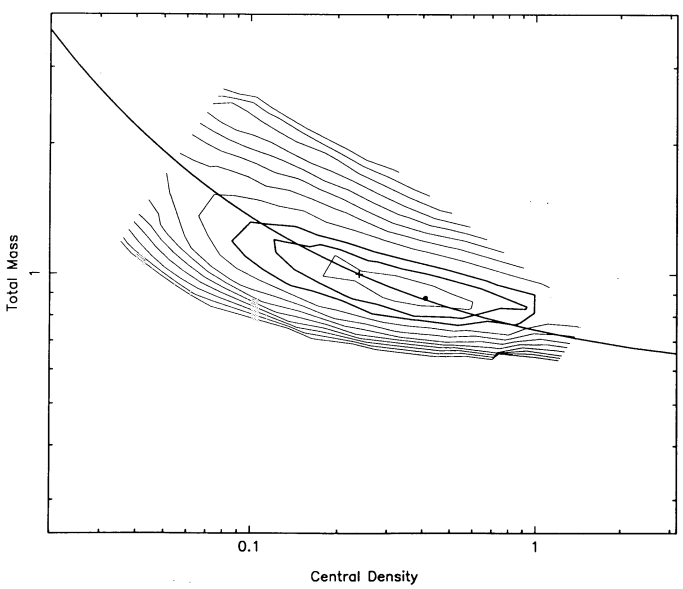


FIG. 6.—Contours of constant likelihood in the space $(a_1 a_2, 3/4\pi a_1/a_2^2)$ for the $N_{\rm data} = 1000$ sample generated from distribution function f_1. Solid curve is the locus of points consistent with the virial theorem, as discussed in the text.

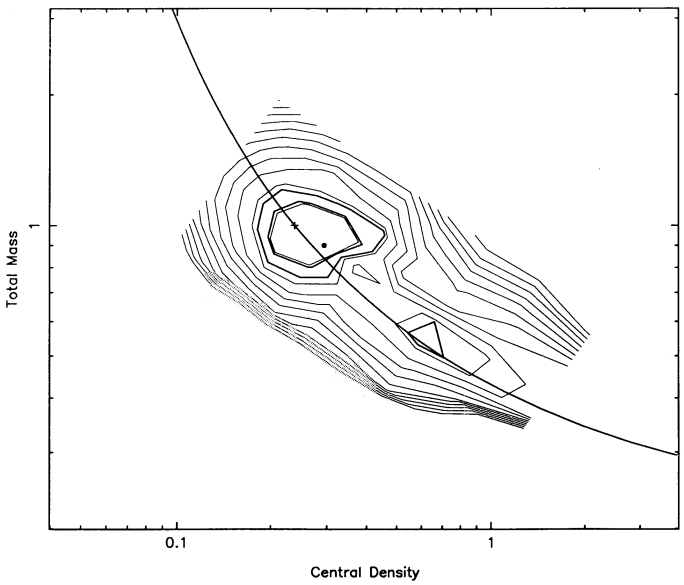


FIG. 7.—Like Fig. 6, for the $N_{data} = 1000$ sample generated from distribution function f_2.

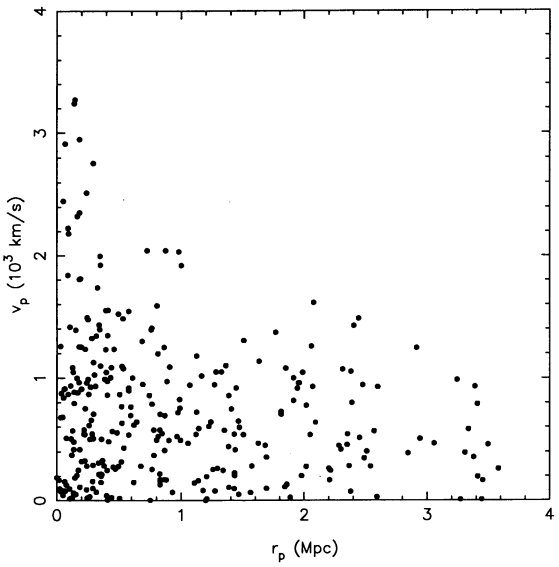


FIG. 8.—Coma Cluster data, from Kent & Gunn (1982)

5. THE COMA CLUSTER

The algorithm described here is best suited to systems in which accurate velocities exist for every object identified as a member, i.e., in which all the useful information is contained within the kinematical sample. Galaxy clusters fall into this category, since membership in a galaxy cluster is determined largely through velocity. (Higher surface brightness objects, such as globular clusters, generally do not, since velocities are often available for only a small fraction of those objects that are known to be members. In these cases, one should impose the additional constraint that f reproduce the known density profile.) Here we apply our algorithm to the Coma galaxy cluster. The data, shown in Figure 8, are taken from the compilation of Kent & Gunn (1982). They comprise 296 galaxies with measured velocities within 3 degrees of the cluster center; membership criteria are those of Kent & Gunn (1982). In what follows, the unit of length is 50 minutes of arc, or 1.0 Mpc for a Hubble constant of 100 km s^{-1} Mpc^{-1}; the unit of velocity is 10^3 km s^{-1}; and $G = 1$. We investigated three parameterized forms for the cluster potential, corresponding to three functional forms for the (presumably dark) matter density. The first potential corresponds to a matter density that is slightly more centrally concentrated than the galaxies, whose number density falls off roughly as r^{-3} at large radii:

$$\rho_1(r) = \frac{a_1}{2\pi a_2^3}\left(1 + \frac{r^2}{a_2^2}\right)^{-2},$$
$$\Phi_1(r) = -\frac{Ga_1}{r}\tan^{-1}\left(\frac{r}{a_2}\right). \tag{14}$$

Figure 9*a* shows contours of constant likelihood in (a_1, a_2) space. The most likely potential from this family has

$$a_1 \approx 4.0\,, \quad a_2 \approx 0.3\,,$$

corresponding to a total mass of $1.5 \times 10^{15}\ M_\odot$ and a central density of $5.5 \times 10^{-3}\ M_\odot\ pc^{-3}$. The core radius of the matter determining the potential—defined as the radius at which the projected mass density falls to half its central value—is $r_c \approx 375 \pm 100$ kpc (90% confidence). This is somewhat greater than the core radius as determined by the distribution of bright galaxies (e.g., Kent & Gunn 1982), though the two values are consistent given the respective uncertainties.

One undesirable feature of the potential just considered is that it becomes unphysical when the central scale length a_2 is too small, since equation (14) then implies a divergent mass at small radii. This fact led us to consider the second potential:

$$\rho_2(r) = \frac{a_1}{4\pi a_2^3}\left(1 + \frac{r^2}{a_2^2}\right)^{-1}, \quad r \le r_0\,,$$
$$= 0\,, \qquad r > r_0\,,$$
$$\Phi_2(r) = -\frac{Ga_1}{a_2}\left[1 - \frac{a_2}{r}\tan^{-1}\left(\frac{r}{a_2}\right) + \frac{1}{2}\log\left(\frac{1 + r_0^2/a_2^2}{1 + r^2/a_2^2}\right)\right], \quad r \le r_0\,;$$
$$= -\frac{Ga_1}{a_2}\left[\frac{r_0}{r} - \frac{a_2}{r}\tan^{-1}\left(\frac{r_0}{a_2}\right)\right], \quad r > r_0\,. \tag{15}$$

Here r_0 is a radius beyond which the mass density is set to zero. We chose $r_0 = 4$ Mpc, well outside of the region in which most of the galaxies in our sample are likely to lie. According to Figure 9*b*, the most likely potential from this family has essentially a zero core radius, $a_2 \lesssim 0.05 \approx 50$ kpc, with $a_1 \approx 4.0 a_2$. In other words, the most likely potential is close to that of a singular isothermal sphere:

$$\rho(r) = \frac{a_1}{4\pi a_2}\frac{1}{r^2} \approx \frac{0.3}{r^2}\,. \tag{16}$$

The likelihood of this potential is nearly the same as that of the most likely potential from Φ_1. Thus, although very small dark-matter core radii are consistent with the data, we are unable to determine whether they are more or less likely than a core radius similar to that of the bright galaxies.

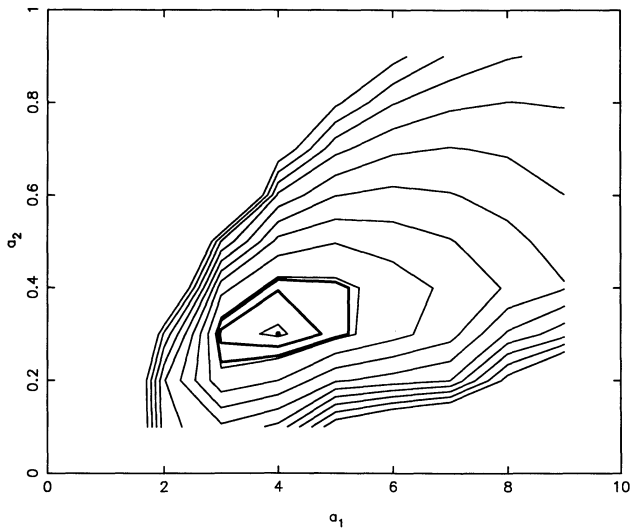


FIG. 9*a*

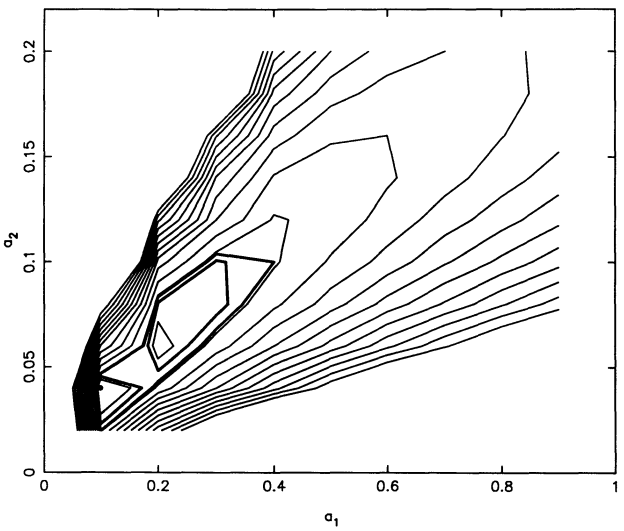


FIG. 9*b*

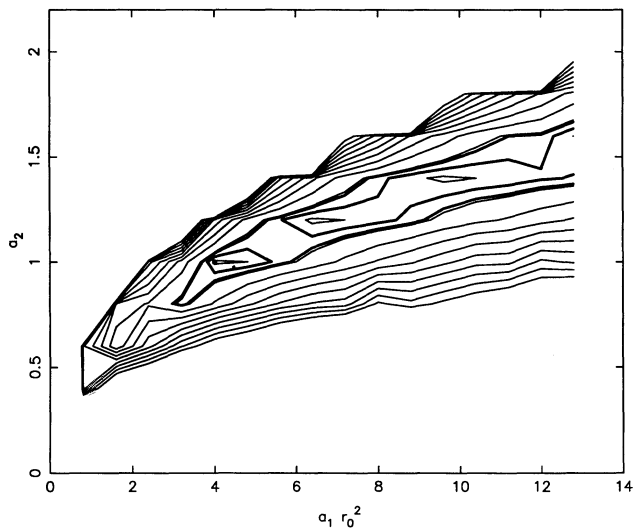


FIG. 9*c*

FIG. 9.—Contours of constant likelihood in the space of parameters (a_1, a_2) that define the potentials fit to the Coma Cluster data. Heavy curves denote the approximate 50% and 90% confidence regions. (*a*) Potential 1; (*b*) potential 2; (*c*) potential 3.

Our third family of potentials contains elements of the first two:

$$\rho_3(r) = \frac{a_1}{4\pi a_2^3}\left(1+\frac{r^2}{a_2^2}\right)^{-1}\left(1+\frac{r^2}{r_0^2}\right)^{-1},$$

$$\Phi_3(r) = -\frac{a_1 r_0^2}{a_2(a_2^2-r_0^2)}\left[\frac{a_2}{r}\tan^{-1}\left(\frac{r}{a_2}\right) - \frac{r_0}{r}\tan^{-1}\left(\frac{r}{r_0}\right) - \frac{1}{2}\log\left(\frac{r_0^2+r^2}{a_2^2+r^2}\right)\right]. \quad (17)$$

Here $r_0 \ll a_2$ is the central core radius, set to 20 kpc. At small radii, ρ_2 approximates a singular isothermal sphere; at large radii, the density falls as r^{-4}. The most likely potential from this family (Fig. 9*c*) appears, once again, to be the singular isothermal sphere: the likelihood tends slowly toward a maximum as a_1 and a_2 increase, such that

$$\frac{a_1 r_0^2}{a_2^3} \approx 4.0,$$

corresponding to a density law

$$\rho(r) = \frac{a_1 r_0^2}{4\pi a_2^3}\frac{1}{r^2} \approx \frac{0.3}{r^2}, \quad (18)$$

the same result obtained from family 2. However in this case, mass models with a_2 as small as ~ 0.7 fall within the 90% confidence region, corresponding to a density that falls as r^{-2} within about 700 kpc ($H_0 = 100$), and as r^{-4} outside.

These two distinct, "most likely" models for the potential of the Coma Cluster have rather different associated distribution functions for the cluster galaxies. The f corresponding to Φ_1 (i.e., large core radius) has a velocity distribution that is radially biased inside of ~ 1 Mpc and approximately isotropic outside. The f corresponding to Φ_2 and Φ_3 (zero core radius) has a nearly isotropic velocity distribution inside of ~ 500 kpc, and a tangentially biased distribution outside.

A number of other attempts have been made to constrain the Coma Cluster dark matter distribution using this kinematical sample or similar ones (e.g., Rood et al. 1972; Bailey 1982; Kent & Gunn 1982; The & White 1987). All of these studies were based on a comparison of models with number density and velocity dispersion profiles of the galaxies, and thus, for the reasons outlined above, cannot in principle distinguish between very different models for the potential. While some of these authors claimed to find "most likely" forms for the potential, in every case the existence of an apparently preferred Φ can be traced to the use of restricted forms for f or its moments. Merritt (1987) used the global velocity distribution of the Coma galaxies as a means of distinguishing between some rather extreme models of the potential; the present study supersedes that work. Hughes (1989) placed limits on the Coma dark matter distribution using X-ray data from a number of satellite observatories. His results were quite similar to ours: while he could place no useful lower limit on the dark

matter core radius, values as large as ~ 20 arcmin ≈ 400 kpc ($H_0 = 100$) were consistent with the existing X-ray data.

Our analysis of the Coma data affirms that a sample of a few hundred positions and velocities is sufficient to place usefully tight constraints on a potential characterized by only two free parameters. When the form of Φ is completely unknown a priori, however, we have found that rather different potentials can be equally consistent with kinematical samples of this size. Our results demonstrate the importance of allowing great freedom when constructing dynamical models of hot stellar systems, and the need for caution when claiming to have found a preferred model. The strongest statement that we feel comfortable making about the distribution of dark matter in the Coma Cluster is that its core radius is unlikely to be greater than about 500 kpc ($H_0 = 100$), or its central density less than about $2 \times 10^{-3}\ M_\odot\ \mathrm{pc}^{-3}$. However, smaller core radii, and higher central densities, are equally consistent with the available data. Larger numbers of radial velocities will be needed to improve these constraints.

6. DISCUSSION

The approach to potential estimation outlined here differs in two ways from previous work. The more fundamental difference is that we view the data as an approximation to the projected distribution function $v_p(r_p, v_p)$, rather than as a set of numbers from which the surface density and velocity dispersion profiles can be estimated. As discussed in Paper I, it is only through using the additional information contained within the line-of-sight velocity distributions that one can hope to eliminate the generally extreme indeterminacy associated with low-order velocity moments. The less fundamental difference is our particular algorithm for finding the most likely Φ and f. The algorithm described here is suited to cases where velocities exist for nearly every object identified as a likely member. Examples of such data sets are galaxies in galaxy clusters, as discussed above; systems of planetary nebulae around individual galaxies (e.g., Ford et al. 1989); or the OH–IR stars in the Galactic bulge (te Linkel Hekkert 1990). Qualitatively different sorts of data would most naturally be analyzed with modified versions of the algorithm discussed here. For high surface brightness objects such as elliptical galaxies, it is possible in principle to calculate $v_p(r_p, v_p)$ directly, by deconvolving integrated spectra (e.g., Bender 1990). Estimation of f in this case would involve a straightforward inversion of the integral equation (1), rather than a maximum likelihood technique. However such data are likely to be available only in the central regions of galaxies, where sky subtraction is not a serious problem. Another class of data would be obtained from a galaxy or star cluster where the velocity data are discrete, but where the density profile of the population from which the kinematical sample was drawn is independently known. One example is the system of globular clusters surrounding the giant elliptical galaxy M87: here, the radial distribution of clusters is very well determined, through background-subtracted counts (Harris 1986), while accurate radial velocities exist for only ~ 45 objects (Mould et al. 1990). Similarly, in individual globular clusters, the sample of stars from which the surface density profile is determined is generally much larger than the sample for which radial velocities are known (e.g., Meylan & Mayor 1986). Clearly, an analysis based on only those stars with measured velocities would be ill-advised; instead, one should accept solutions only if they yield the known density profile as well. The proper approach in these cases is to represent f through a much larger number of basis functions—say, as a histogram—and to require any solution to reproduce the known number density profile at a set of radial points.

Finally, we discuss the possibility of generalizing our approach to nonspherical systems. We note at the outset that there do not yet exist proofs, like the one for spherical systems (Paper I), of the correspondence between projected and intrinsic distribution functions for axisymmetric or triaxial systems. In the absence of such proofs, it is difficult to answer questions about the uniqueness of solutions, even when the potential is completely specified. In a (fully integrable) axisymmetric system, the distribution function generally depends on three orbital integrals, $f = f(E, L_z^2, I_3)$, one more than in the spherical case. However, the projected distribution function is also a function of three variables:

$$v_p = v_p(x_p, y_p, v_p)\ ,$$

where x_p and y_p are the projected coordinates on the plane of the sky. Thus it seems feasible that the axisymmetric problem is "as constrained" by the data as the spherical one, except for uncertainties about the orientation, and the fact that more velocities are required to specify a three-dimensional function than a two-dimensional function. In practice, however, it will not be easy to handle the nonclassical integral I_3; short of a completely numerical approach, one could imagine approximating the integrals via perturbation theory (Gerhard & Saha 1992). A good place to apply an axisymmetric algorithm might be the Galactic bulge, as traced by OH/IR stars (te Linkel Hekkert 1990), or dwarf spheroidal galaxies. Triaxial systems would be rather more difficult due to the need to test different assumptions about the intrinsic axis ratios and orientation.

This work was supported by NSF grant AST 90-16515. We benefited from conversations with H. Dejonghe and J. Sellwood.

REFERENCES

Bailey, M. E. 1982, MNRAS, 201, 271
Bender, R. 1990, A&A, 229, 441
Bertin, G., Saglia, R. P., & Stiavelli, M. 1988, AJ, 87, 945
Casertano, S., & van Albada, T. S. 1990, in Baryonic Dark Matter, ed. D. Lynden-Bell & G. Gilmore (Dordrecht: Kluwer), 159
Dahlquist, G., Björck, A., & Anderson, N. 1974, Numerical Methods (Englewood Cliffs: Prentice-Hall), 117
Dejonghe, H. 1987, MNRAS, 224, 13
———. 1989, ApJ, 343, 113
Dejonghe, H., & Merritt, D. 1992, ApJ, 391, 531 (Paper I)
Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, Statistical Methods in Experimental Physics (Amsterdam: North Holland), 207
Fabian, A. C., Thomas, P. A., Fall, S. M., & White, R. E. 1986, MNRAS, 221, 1049
Ford, H. C., Ciardullo, R., Jacoby, G. H., & Hui, X. 1989, in IAU Symp. 131, Planetary Nebula, ed. S. Torres-Peimbert (Dordrecht: Kluwer), 335
Fricke, W. 1952, Astron. Nach., 280, 193
Gerhard, O. E., & Saha, P. 1991, MNRAS, 251, 251
Gunn, J. E., & Griffin, R. F. 1979, AJ, 84, 752
Harris, W. E. 1986, AJ, 91, 822
Heisler, J., Tremaine, S., & Bahcall, J. N. 1985, ApJ, 298, 8
Hughes, J. P. 1989, ApJ, 337, 21
Kent, S. M. 1991, MNRAS, 247, 702
Kent, S. M., & Gunn, J. E. 1982, AJ, 87, 945
King, I. R. 1966, AJ, 71, 64
Kuijken, K. 1991, ApJ, 372, 125
Kuijken, K., & Gilmore, G. 1989, MNRAS, 239, 651
Little, B., & Tremaine, S. 1987, ApJ, 320, 493
Martin, B. R. 1971, Statistics for Physicists (London: Academic), 167

Merritt, D. 1987, ApJ, 313, 121
———. 1988, AJ, 95, 496
Meylan, G., & Mayor, M. 1986, A&A, 166, 122
Miller, G. F. 1974, in Numerical Solution of Integral Equations, ed. L. M. Delves & J. Walsh (Oxford: Clarendon), 181
Mould, J. R., Oke, J. B., de Zeeuw, P. T., & Nemec, J. M. 1990, AJ, 99, 1823
Newton, A. J., & Binney, J. 1984, MNRAS, 210, 711
Oort, J. H. 1932, Bull. Astron. Inst. Netherlands, 6, 749
Peebles, P. J. E. 1984, ApJ, 277, 470
Pryor, C. 1992, in Morphological and Physical Classification of Galaxies, ed. G. Busarello, M. Capaccioli, & G. Longo (Dordrecht: Kluwer), 163
Pryor, C., & Kormendy, J. 1990, AJ, 100, 127
Pryor, C., McClure, R. D., Fletcher, J. M., & Hesser, J. E. 1988, in IAU Symp. 126, Globular Cluster Systems in Galaxies, ed. J. E. Grindlay & A. G. Davis Phillips (Dordrecht: Kluwer), 661
Rice, J. R. 1964, The Approximation of Functions, Vol. 1 (Reading: Addison-Wesley), 119
Richstone, D. O., & Tremaine, S. 1984, ApJ, 286, 27
———. 1988, ApJ, 327, 82
Rood, H. J. 1981, Rep. Prog. Phys., 44, 1077
Rood, H. J., Page, T. L., Kintner, E. C., & King, I. R. 1972, ApJ, 175, 627
Schittkowski, K. 1985, Annals Operations Res., 5, 485
Schwarzschild, M. 1979, ApJ, 232, 236
Tapia, R. A., & Thompson, J. R. 1978, Nonparametric Probability Density Estimation (Baltimore: Johns Hopkins Univ. Press), 99
te Linkel Hekkert, P. 1990, Ph.D. thesis, Leiden Univ.
The, L. S., & White, S. D. M. 1986, AJ, 92, 1248
Turchin, V. F., Kozlov, V. P., & Malkevich, M. S. 1971, Soviet Phys.—Uspekhi, 13, 681