THE ASTROPHYSICAL JOURNAL, 683:750–766, 2008 August 20
© 2008. The American Astronomical Society. All rights reserved. Printed in U.S.A.

MAXIMUM LIKELIHOOD FITTING OF TIDAL STREAMS WITH APPLICATION TO THE SAGITTARIUS DWARF TIDAL TAILS

NATHAN COLE,[1] HEIDI JO NEWBERG,[1] MALIK MAGDON-ISMAIL,[2] TRAVIS DESELL,[2] KRISTOPHER DAWSEY,[1]
WARREN HAYASHI,[1] XINYANG FRED LIU,[1] JONATHAN PURNELL,[2] BOLESLAW SZYMANSKI,[2]
CARLOS VARELA,[2] BENJAMIN WILLETT,[1] AND JAMES WISNIEWSKI[1]
Received 2007 December 18; accepted 2008 April 22

ABSTRACT

We present a maximum likelihood method for determining the spatial properties of tidal debris and of the Galactic spheroid. With this method we characterize Sagittarius debris using stars with the colors of blue F turnoff stars in SDSS stripe 82. The debris is located at $(\alpha, \delta, R) = (31.37^\circ \pm 0.26^\circ, 0.0^\circ, 29.22 \pm 0.20$ kpc), with a (spatial) direction given by the unit vector $(-0.991 \pm 0.007$ kpc, 0.042 ± 0.033 kpc, 0.127 ± 0.046 kpc), in galactocentric Cartesian coordinates, and with FWHM $= 6.74 \pm 0.06$ kpc. This 2.5° wide stripe contains 0.9% as many F turnoff stars as the current Sagittarius dwarf galaxy. Over small spatial extent, the debris is modeled as a cylinder with a density that falls off as a Gaussian with distance from the axis, while the smooth component of the spheroid is modeled with a Hernquist profile. We assume that the absolute magnitude of F turnoff stars is distributed as a Gaussian, which is an improvement over previous methods which fixed the absolute magnitude at $\bar{M}_{g_0} = 4.2$. The effectiveness and correctness of the algorithm is demonstrated on a simulated set of F turnoff stars created to mimic SDSS stripe 82 data, which shows that we have a much greater accuracy than previous studies. Our algorithm can be applied to divide the stellar data into two catalogs: one which fits the stream density profile and one with the characteristics of the spheroid. This allows us to effectively separate tidal debris from the spheroid population, both facilitating the study of the tidal stream dynamics and providing a test of whether a smooth spheroidal population exists.

Subject headings: Galaxy: halo — Galaxy: structure — methods: data analysis

1. INTRODUCTION

1.1. *The Milky Way's Spheroid*

In the late 1980s, the Milky Way's spheroid population was described as slowly rotating, with density distribution given by $\rho \propto r^{-3.5}$ (Freeman 1987). There was a controversy concerning the flattening of the spheroid, since kinematic studies of the spheroid stars suggested a flatter spheroid (Gilmore et al. 1989) and star counts often suggested a more spherical spheroid (Bahcall 1986). By early in the 21st century, the spheroid was still thought of as a smooth power-law distribution, but studies were starting to show that the shape of the spheroid depended on the type of star being observed, and it was noted that at least some of the spheroid, if not all, was composed of debris from hierarchical structure formation (Freeman & Bland-Hawthorne 2002). The Sagittarius (Sgr) dwarf galaxy (Ibata et al. 1994) and its associated tidal stream (Yanny et al. 2000; Ibata et al. 2001a, 2001b) was the one known example of merging in the present day. In the last 10 years, the discovery of substructure has dominated the discussion of the Galactic spheroid. The discovery of substructure has been driven primarily by the Sloan Digital Sky Survey (SDSS) and the related Sloan Extension for Galactic Understanding and Exploration (SEGUE), but other surveys such as the Quasar Equatorial Survey Team and the Two Micron All Sky Survey (2MASS) have also been influential. Discoveries include new tidal debris streams, dwarf galaxies, and globular clusters.

The positions, velocities, and metallicities of Sgr dwarf spheroidal tidal debris stars have been measured all across the sky (Newberg et al. 2002, 2003; Bellazzini et al. 2003; Majewski et al. 2003, 2004; Martínez-Delgado et al. 2004; Fellhauer et al. 2006; Chou et al. 2007). These measurements have then been compared with models of tidal disruption (Johnston et al. 1995, 2002; Ibata et al. 1997; Ibata & Lewis 1998; Gómez-Flechoso et al. 1999; Helmi & White 2001; Law et al. 2004, 2005). Given that the data spans three spatial dimensions plus radial velocities, but both the models and data are presented in papers primarily as two-dimensional plots, it has been difficult to make detailed comparisons in multidimensional space. In this paper we develop a technique that would allow us to create a catalog of stars that has the same spatial distribution as the stars in the Sgr tidal stream. That catalog could then be made available to the modeling community, who would be able to transform the coordinates into whatever system they find most natural to make comparisons.

At least three newly discovered tidal debris streams are thought to be associated with dwarf galaxies, including the Monoceros stream in the Galactic plane (Newberg et al. 2002; Yanny et al. 2003), the Virgo stellar stream (Vivas et al. 2001; Duffau et al. 2006; Newberg et al. 2007), and the Orphan stream (Grillmair 2006) in the Field of Streams (Belokurov et al. 2006b). An additional piece of tidal debris was found in Triangulum-Andromeda by Rocha-Pinto et al. (2004) which might be part of the Monoceros stream (Peñarrubia et al. 2005). Tidal tails spanning many tens of degrees across the sky have been found around Pal 5 (Odenkirchen et al. 2001; Rockosi et al. 2002; Grillmair & Dionatos 2006a) and NGC 5466 (Grillmair & Johnson 2006; Belokurov et al. 2006a). Grillmair & Dionatos (2006b) find a similar-looking tidal stream for which there is no known progenitor globular cluster. Eight new low–surface brightness dwarf galaxy satellites of the Milky Way have been discovered (Willman et al. 2005; Zucker et al. 2006a, 2006b; Belokurov et al. 2006c, 2007a; Grillmair 2006; Irwin et al. 2007), nearly doubling the number of known Milky Way dwarfs. Belokurov et al. (2007b) suggest that there is a new structural component of the spheroid called the Hercules-Aquila cloud.

[1] Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY 12180; astro@cs.rpi.edu.
[2] Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180; astro@cs.rpi.edu.

With all of this substructure detected in the densities of spheroid stars, one wonders how much of the spheroid is smooth and what the shape of the smooth spheroid might be. It is no wonder that standard Galactic models, such as the Besançon model (Robin et al. 2003), which uses a power-law density profile, do not match the SDSS star counts. The spheroid shape is a current controversy as Helmi (2004) has found prolate models favorable, while Johnston et al. (2005) has found oblate models best. The large number of very significant overdensities of stars in the spheroid show that pencil-beam studies and studies that extrapolate from the solar neighborhood could be contaminated by local structure and may not be indicative of the global shape of the Milky Way spheroid. Recent results from F turnoff stars from the SDSS suggest that the smooth component of the spheroid might be asymmetric about the Galactic center, so that no axially symmetric spheroid model can be fit (Newberg & Yanny 2005, 2006; Savage et al. 2006; Xu et al. 2006). However, accurate measurements of the smooth component require that spatial substructure be identified and quantified—otherwise, it is difficult to know which stars to fit to smooth spheroid models.

We need to develop a model that is as complex as the data. In this paper we start this process by developing a maximum likelihood method that can be used to fit a smooth stellar spheroid containing one debris stream, using one 2.5° stripe of SDSS data, and show that our algorithm produces the correct results on a simulated data set. In the future, we will extend the algorithm to include larger data sets with multiple pieces of debris. The long-term goal is to develop a model of the spheroid that fits the observed, lumpy density distribution of stars in the spheroid.

1.2. *The Sloan Digital Sky Survey*

The SDSS is a large, international collaboration that was originally established to find the largest structures of galaxies in the universe from an imaging survey of 10,000 deg^2 of sky and spectra of 1,000,000 galaxies selected from the photometry (from images). The SDSS can simultaneously obtain 640 spectra per pointing, using two 320-fiber double (blue/red) spectrographs. The angular position of a galaxy in the sky is easy to measure from the imaging survey, and for distant galaxies, the distance is well estimated from a spectroscopic measurement of the Hubble redshift. By combining the imaging with the spectroscopic survey, the SDSS has built up a three-dimensional picture of the distribution of galaxies in the universe using a dedicated 2.5 m telescope at Apache Point Observatory in New Mexico (York et al. 2000).

Images are obtained with an array of 30 2048×2048 pixel CCD cameras operated in a drift-scan (time delay and integrate) mode, which produces six "scan lines" of imaging data that are each 13.6′ wide and grow longer at a rate of 15° hr^{-1} for the length of that "run." In each of the six "scan lines," the sky is imaged in five optical filters: u, g, r, i, and z; the time at which each astronomical object is imaged is a few minutes different for each passband. When a second set of six scan lines is observed that fill in the gaps between the first set of six scan lines, they produce a contiguous "stripe" of data of width 2.5° and length that depends on the length of time the sky was observed in the "runs." Each 2.5° wide stripe follows a defined great circle on the sky and is built up of two or more "runs" of data, as many runs as are needed to traverse the SDSS survey area.

The entire sky is divided up into 144 numbered stripes that start and end at the survey poles, $(l, b) = (209.33°, -7°)$ and $(29.33°, 7°)$. Stripes 10 and 82 are centered on the celestial equator, with stripe 10 in the north Galactic cap and stripe 82 in the south Galactic cap. The other stripes are sequentially numbered with inclinations 2.5° apart. If the entire length (180°) of every stripe were imaged, we would have 64,800 deg^2 of imaging data for a sky that is 41,253 deg^2. The overlaps between stripes increase toward the survey poles. Since the SDSS generally observes only parts of the sky that are more than 30° from the survey poles, there is about 50% overlap at the ends of the stripes and very little overlap on the survey equator, which is at R.A. = 185°. With the release of SDSS Data Release 6 (DR6; Adelman-McCarthy et al. 2008), a contiguous 8500 deg^2 area of the north Galactic cap is now publicly available. It is pieced together with 31 adjacent SDSS stripes. Three stripes of data are available in the south Galactic cap, including stripe 82 on the celestial equator. Further information about the SDSS, including survey geometry, can be found in Stoughton et al. (2002), Abazajian et al. (2003), Fukugita et al. (1996), Gunn et al. (1998), Hogg et al. (2001), Pier et al. (2003), and Smith et al. (2002).

Because the imaging survey is large and well calibrated and because the spectroscopic survey included many Galactic stars, in some cases by design and in some cases accidentally, the SDSS has made a significant contribution to our knowledge about the Milky Way and, in particular, the discovery of substructure in the spheroid component, as we discussed above. Because of this, SDSS has been expanded to include a new project, SEGUE. This extension will eventually include 3500 deg^2 of new imaging data which is collected in 2.5° wide great circles on the sky, but these great circles do not in general match those laid out on the sky for the SDSS, and they are not adjacent to each other so they do not fill contiguous areas of the sky. The positions of these stripes were chosen to sparsely sample all directions in the sky that are visible from Apache Point Observatory and include scans at constant Galactic longitude that pass through the Galactic plane. SEGUE will also obtain ~250,000 spectra of Galactic stars to study stellar populations and kinematics.

In this paper we wish to use the photometry derived from imaging of Galactic stars to trace the three-dimensional shape of the Galactic spheroid and the substructure, including tidal streams, contained within it. In this demonstration of the algorithm and the results from running it, we focus on only one well-studied stripe of SDSS data: stripe 82 on the celestial equator. We use only F turnoff stars in 300 deg^2 of this stripe, roughly centered on the place where the Sgr stream crosses it. Unlike galaxies, the distances of stars cannot be inferred from their radial velocities. By selecting only stars with colors consistent with spheroid F turnoff stars, we select a stellar population that is not a very good standard candle. But because we observe a large number of F turnoff stars we can find their underlying density distribution through statistical methods. We assume the absolute magnitudes of the F turnoff stars in the population have a mean of $\bar{M}_{g_0} = 4.2$ and a dispersion of $\sigma_{M_{g_0}} = 0.6$. These are reasonable estimates for the Sgr dwarf tidal stream (Newberg & Yanny 2006).

1.3. *The Technique of Maximum Likelihood*

Given a parameterized model and some data generated according to the model for some "true" values of the model parameters, the task of model estimation is to determine the set of parameters used in generating the data. Within a Bayesian setting, the model estimation problem can be reformulated as determining the a posteriori most likely parameters given the data and the model. In our case we have spatial positions for a set of stars in the spheroid and a proposed model for the spheroid and a tidal stream which passes through it, and we would like to find the most likely values of the parameters in that model.

The likelihood of a set of parameters is the probability of obtaining a particular data set for a given set of parameters. Via

Bayes' theorem, one can decompose the a posteriori probability of a particular set of parameters given the data and the model into the product of two terms, the first being the likelihood of the parameters and the second being the prior probability of the parameters. When the prior probability distribution over the parameters is uniform (as is typically assumed), the a posteriori probability is proportional to the likelihood, and so the maximum likelihood technique may be used to find the most likely model parameters (see Fletcher 1987 for more details).

There are two main tasks to implementing the maximum likelihood technique. The first is to develop the likelihood function, which measures the probability of observing a particular data set given a parameterized model. The second is to maximize the likelihood with respect to the parameters.

In developing the likelihood function, one typically assumes that each data point is independently generated; hence, the likelihood of the data set is the product of the likelihoods of each individual data point. Within our setting, we develop a parameterized model for the background Milky Way and Sgr stellar distributions from which we compute the likelihood. We prefer to maximize the logarithm of the likelihood rather than the likelihood, because the computation of the log-likelihood is numerically more stable. We then use the standard technique of optimization via conjugate gradients (Fletcher 1987) to maximize the likelihood and, hence, determine the most likely model parameters fitting the data.

One advantage of the maximum likelihood framework is that the Hessian of the log-likelihood (at the maximum) gives the shape of the probability distribution over the parameters and, hence, allows us to determine the *statistical* error in the estimated model parameters. When the model parameters are physical quantities of interest, this statistical error translates to an error bar on the measured (i.e., estimated) physical quantity.

In the remainder of this paper we present a new automated technique that can detect tidal debris in a spatial input catalog of stars from one stripe of SDSS data. Since the first stripe in which the Sgr dwarf tidal stream was detected in SDSS data was stripe 82 (Newberg et al. 2002), it has therefore been the most well studied; this is the section of data on which we test our maximum likelihood. Building on this, we describe a technique to extract a catalog of stars that fits the density profile of the debris; this catalog can then be used to constrain the dynamical models.

2. ALGORITHM FOR FITTING SPHEROID SUBSTRUCTURE

2.1. *Goals of the Algorithm*

Starting with a parameterized density model for the spheroid, a parameterized density model for the Sgr dwarf tidal stream, the absolute magnitude distribution for F turnoff stars, and a data set of observed (l, b) angular coordinates and apparent magnitudes for color-selected F turnoff stars in a 2.5° stripe through which a tidal stream passes, we find the parameters for a tidal stream and stellar spheroid concurrently that give us the highest likelihood of observing the data. This will allow us to tabulate summary statistics for the Sgr dwarf tidal stream, including position, width, and density. It will also allow us to test models for the spheroid component of the Milky Way, and ultimately, it will allow us to construct a catalog of stars that fit the density profile of the Sgr tidal stream by probabilistically separating the stream from the spheroid.

The parameters in our stellar density model describe the position, direction, and width of the tidal stream, as well as the shape of the smooth stellar spheroid. We start with an initial set of parameters and iterate using a conjugate gradient search technique until the optimum values for each parameter given the data set are determined. From the parameters which produce the maximum likelihood, we estimate the errors in each parameter.

Once we have found the best parameters, we are able to use a separation algorithm to divide the data into two sets of stars, one of which has the density profile of the stream and the other of which has the density profile of the spheroid. While we cannot determine which of the individual stars in the data set belong to the stream and which belong to the spheroid, we can separate the input catalog into two catalogs that will have the density properties of the stream and the spheroid, respectively. These catalogs will allow for a close comparison of the stream density profile with that of simulations of tidal disruption and constrain the dynamical models used in simulations of tidal disruption.

Section 2.2 gives a brief overview of the algorithm. Section 2.3 describes the construction of the probability density function. Section 2.4 discusses the optimization method we use. Section 2.5 describes the method by which we estimate the errors in the parameters. Section 2.6 discusses the application of a separation algorithm to distinguish stream and spheroid stars. Section 2.7 describes the computational time required to run the algorithm and the application of parallel processing as a way to improve efficiency. Section 2.8 discusses some limitations and possible enhancements that could be made in the future. A thorough discussion of the conjugate gradient and line search optimization method can be found in Appendix A.

2.2. *Overview of the Algorithm*

The algorithm starts with initial guesses for the model parameters and an input catalog of observed F turnoff star positions as (l, b, g), which represent the Galactic longitude, in degrees; Galactic latitude, in degrees; and reddening-corrected g magnitude for each star. The apparent magnitudes are corrected for reddening using the Schlegel et al. (1998) reddening maps, as implemented in SDSS DR6.

The essence of the maximum likelihood technique is to create a parameterized probability density function (PDF) and then find parameters in this function which maximize the likelihood of observing the data. We first select a parameterized stellar density function for the spheroid and its substructure. Since the parameterized models describe the density as a function of space (X, Y, Z) and the observations include only Galactic coordinates (l, b) and apparent magnitude g, we need to relate spatial position to (l, b, g). The angular position in the sky is very well known for each star, but it is difficult to determine the distance to each star. We can only estimate the distance from its apparent magnitude, g. If all of the stars in the sample have the same absolute magnitude, then the conversion of apparent magnitude to distance is trivial. Unfortunately, this is not a good approximation for F turnoff stars, whose luminosity can deviate from the mean by a factor of 5 or more, even within the same age and metallicity population. Instead, we approximate the distribution of absolute magnitudes of F turnoff stars as Gaussian.

The likelihood of observing the data is the product of the likelihoods of observing each of the stars within the data set, given the model,

$$\mathcal{L}(\boldsymbol{Q}) = \prod_{i=1}^{N} \mathrm{PDF}(l_i, b_i, g_i|\boldsymbol{Q}), \quad (1)$$

where the index i runs over the N stars, $\boldsymbol{Q}$ is a vector representing the parameters in the model, and the PDF is a normalized version

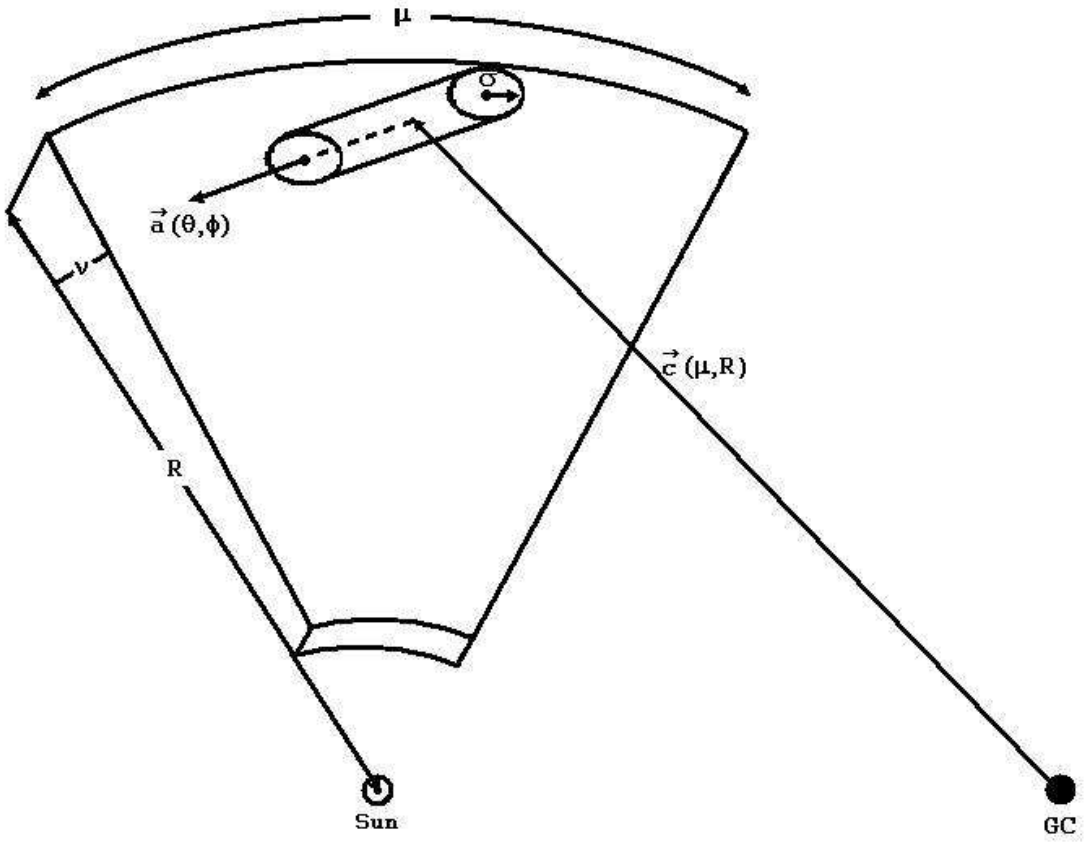

FIG. 1.—Stripe and stream parameter definitions. The segment of a stream which passes through an SDSS stripe is cylindrical within an individual stripe, with density that falls off as a Gaussian with distance from its axis. The coordinates μ, ν, and r are used to define SDSS stripes. We adopt these coordinates to define a vector, $\boldsymbol{c}(\mu, R)$ ($\nu = 0$), which points to the center of the stream from the Galactic center. We then define a stream directional vector $\hat{\boldsymbol{a}}(\theta, \phi)$ of unit length. Finally, we define the stream width as σ, which is the standard deviation of the Gaussian that defines the density falloff of the stream.

of the stellar density function that is derived in § 2.3. Because the individual probabilities are small, we avoid numerical underflow by maximizing the average logarithm of the likelihood,

$$\frac{1}{N} \ln \mathcal{L}(\boldsymbol{Q}) = \frac{1}{N} \sum_{i=1}^{N} \ln \mathrm{PDF}(l_i, b_i, g_i | \boldsymbol{Q}), \qquad (2)$$

which is maximized for the same parameters that the likelihood is maximized.

We have chosen as input to this first version of the maximum likelihood algorithm a volume of the Galaxy sampled in a section of stripe 82. This volume is a piece of a wedge, 2.5° wide, with the point at the Sun, limited in the near and far ends by the apparent magnitude range $16 < g_0 < 22.5$. For F turnoff stars with absolute magnitude $M_g = 4.2$, the approximate distance range is 2.3 kpc $< R <$ 45.7 kpc from the Sun. The near and far edges of the volume are fuzzy due to the inexact relationship between apparent magnitude and distance, and the variable detection efficiency at the faint end. We chose this volume of a single stripe as the first application of the algorithm, for we do not have a global model for a general tidal stream; however, if we choose to look at a small enough volume of the stream (as is the intersection of the stream with stripe 82), we are able to approximate its shape as a cylindrical density with cross-sectional Gaussian falloff from its axis. We model the entire tidal stream as a set of these cylindrical stream segments. This model will not work if the stream is in the plane of the wedge, but we show below that this cylindrical approximation works for this stream in this wedge.

The algorithm begins by calculating the average log-likelihood for an initial set of parameters. A conjugate gradient with the line search method is used to find the next set of parameters and the process iterates. Using this method, each subsequent set of parameters is guaranteed to produce an equal or higher likelihood in each iteration of the algorithm. The algorithm continues to generate new sets of parameters until convergence, which is defined by one of a number of conditions: either the gradient becomes so small that it would not provide foreseeable improvement, the line search returns negligible movement, or a maximum number of iterations is reached.

2.3. *The Probability Density Function*

2.3.1. *Tidal Stream Model*

In general, tidal streams follow a complex path through the sky. The stars in the structures may bunch up at apogalacticon and may have a complex cross-sectional density that varies with position along the stream. However, within a single SDSS stripe through which the stream passes it is reasonable to approximate the path of the stream as linear (see § 3.3).

We model the stream in a piecewise linear fashion with a separate set of parameters for each stripe of data in the SDSS, although in principle we could run this algorithm on any 2.5° wide great circle or set of 2.5° wide great circles on the sky. The length of the cylinder is limited by the edges of the data in one stripe. The cross section is circularly symmetric with a density that falls off as a Gaussian with distance from the stream axis. Figure 1 depicts the data volume and the relationship between the stream

parameters and the segment of a cylinder that describes the stream.

The SDSS great circle coordinate system (μ, ν) is used to measure the angular position on the sky. The coordinate ν is bounded by $-1.25^\circ < \nu < 1.25^\circ$ for all stripes and measures the angular position across the narrow dimension of the stripe; ν is by definition zero along the center of each stripe of data. The coordinate μ measures the angular distance along the great circle swept out by each stripe. The inclination of each stripe is the maximum angle between that stripe and the celestial equator. Thus, μ, ν, and the stripe inclination uniquely specify the angular sky coordinates.

The vector $\boldsymbol{c}$ points from the Galactic center to the axis of the cylinder. Normally, a vector requires three parameters, one for each dimension. However, since we have the freedom to require that the position along the cylinder axis to which it points be in the plane that splits the volume in half along the narrow direction, we reduce that to two parameters. Enforcing the condition that $\nu = 0$, we are thus able to parameterize the stream center with only a radial distance from the Sun, R (in kpc), and the angular position along the stripe, μ (in degrees). Therefore, $\boldsymbol{c}(\mu, R)$ fixes the center point of a piece of tidal debris within an SDSS stripe and lies along the axis of our cylinder.

The unit vector $\hat{\boldsymbol{a}}$ describes the orientation of the axis of the cylinder. Again, we need only two parameters because this vector is constrained to be of unit length. We parameterize this vector using two angles θ and ϕ, both in radians, where θ is the angle between $\hat{\boldsymbol{a}}$ and the Galactic Z-axis and Z is perpendicular to the Galactic plane and points toward the north Galactic cap. The azimuthal angle ϕ is measured counterclockwise around the Z-axis, as viewed looking down on the Galaxy from the north Galactic pole, starting from the X-axis, which points in the direction from the Sun to the Galactic center.

The last stream parameter, σ (in kpc), specifies the stream width and is the standard deviation of the Gaussian distribution used to describe the density fall off with distance from the cylinder axis. For a star with spatial coordinates given by $\boldsymbol{p}$, the distance, d, from the cylinder axis is given by

$$d = \left| (\boldsymbol{p} - \boldsymbol{c}) - \hat{\boldsymbol{a}} * [\hat{\boldsymbol{a}} \cdot (\boldsymbol{p} - \boldsymbol{c})] \right| . \quad (3)$$

In practice, this calculation is performed by first converting each vector to a galactocentric Cartesian coordinate system.

In summary, we use five parameters to define our cylindrical stream: μ, R, θ, ϕ, and σ. Figure 1 depicts how these parameters are defined with respect to the stripe volume. Using these parameters, the stellar density of the stream at point $\boldsymbol{p}$ is

$$\rho_{\rm str}(\boldsymbol{p}) \propto e^{-\left(d^2/2\sigma^2\right)}. \quad (4)$$

Normalization of the stellar density will be considered once we have assembled the entire PDF.

2.3.2. *Spheroid*

We model the stellar spheroid with a standard Hernquist profile (Hernquist 1990),

$$\rho_{\rm sph}(\boldsymbol{p}) \propto \frac{1}{r(r + r_0)^3}, \text{ where}$$
$$r = \sqrt{X^2 + Y^2 + \frac{Z^2}{q^2}}, \quad (5)$$

and X, Y, and Z are Cartesian coordinates centered on the Galactic center. The coordinates X and Y are in the Galactic plane, with X directed from the Sun to the Galactic center and Y in the direction of the Solar motion. The coordinate Z is perpendicular to the Galactic plane and points in the direction of the north Galactic cap. The spheroid is described by two parameters: r_0 which is a core radius, in kpc, and the dimensionless quantity, q, which is the scaling factor in the Z-coordinate direction. For $q < 1$ the spheroid is oblate, for $q = 1$ the spheroid is spherically symmetric, and for $q > 1$ the spheroid is prolate.

There are many Galactic components (thick disk, bulge, etc.) other than the spheroid that we could in principle add to our full model. However, we do not expect many disk stars in our sample because we are fitting data that is too far from the plane of the Milky Way and the color-selected F turnoff stars are bluer than the turnoff of the thick disk stellar population.

2.3.3. *Absolute Magnitude Distribution*

In this section we address the fact that stars within our selected color range, which are primarily F turnoff stars, do not all have the same absolute magnitude. If we assume that all of the color-selected stars have the same absolute magnitude (equal to the mean absolute magnitude of the population) when we estimate their distances from the Sun, any substructure in the spheroid will appear to be elongated along our line of sight. To account for this, we calculate an "observed" spheroid spatial density that is elongated along our line of sight by convolution of the density model with the absolute magnitude distribution function along our line of sight.

We model the distribution of absolute magnitudes as a Gaussian with a center of $\bar{M}_{g_0} = 4.2$ and dispersion $\sigma_{g_0} = 0.6$, which is a simplification of the F turnoff star absolute magnitude distribution found for globular clusters by Newberg & Yanny (2006). Thus, letting M_{g_0} be the absolute magnitude,

$$M_{g_0} = \bar{M}_{g_0} + \Delta M_{g_0}, \quad (6)$$

where $\bar{M}_{g_0} = 4.2$ and ΔM_{g_0} has a Gaussian distribution with zero mean and standard deviation 0.6. To account for this distribution over the absolute magnitude, we first derive the probability of observing a star per unit apparent magnitude per unit solid angle, assuming all stars are of absolute magnitude $\bar{M}_{g_0} = 4.2$, and then convolve in apparent magnitude with a Gaussian of dispersion 0.6 centered at zero. The result is the probability of observing a star per unit apparent magnitude per unit solid angle, with the absolute magnitude distribution taken into account.

We will need to refer to several density functions in different spaces, and so we first lay down some definitions to make this discussion clear. We refer to a density as $\rho_A(x)$, where A is the name of the density which refers to the coordinate space in which it is defined and x refers to a generic variable in this space. Thus, we define the six densities,

$$\rho_X(\boldsymbol{x}), \quad \rho_R(R, \Omega), \quad \rho_{g_{4.2}}(g_{4.2}, \Omega),$$
$$\rho_{g_0}(g_0, \Omega), \quad \rho_{\mathcal{R}}(\mathcal{R}(g_0), \Omega), \quad \rho_{X_c}(\boldsymbol{x}), \quad (7)$$

where the first three densities are the model in Cartesian, spherical, and apparent magnitude coordinate systems;

$$\rho_X(\boldsymbol{x}) = \frac{dV}{dx\, dy\, dz}, \quad \rho_R(R, \Omega) = \frac{dV}{dR\, d\Omega},$$
$$\rho_{g_{4.2}}(g_{4.2}, \Omega) = \frac{dV}{dg_{4.2}\, d\Omega}. \quad (8)$$

The densities which take into account the distribution on the absolute magnitude space are denoted by $\rho_{g_0}(g_0,\Omega)$, $\rho_{\mathcal{R}}(\mathcal{R}(g_0),\Omega)$, and $\rho_{X_c}(\boldsymbol{x})$, where the subscript c stands for convolved. All coordinate systems are centered at the Sun, the solar position is assumed to be 8.5 kpc from the Galactic center, and the direction from the Sun to the Galactic center is in the positive X-direction. Here we have made a distinction between R and $\mathcal{R}$; R denotes the actual distance each star is from the Sun, while $\mathcal{R}(g_0)$ denotes the distance a star of apparent magnitude g_0 would be from the Sun if it had an absolute magnitude of $\bar{M}_{g_0} = 4.2$. We have also made the distinction between g_0 and $g_{4.2}$; g_0 denotes the actual reddening-corrected apparent magnitude of a star, while $g_{4.2}$ denotes the apparent magnitude a star at distance R would be calculated to have if it had an absolute magnitude of $\bar{M}_{g_0} = 4.2$. We adopt these definitions for the remainder of this paper.

The galactocentric Cartesian density, ρ_X, is the actual spatial density of stars as described in equations (4) and (5) for the stream and spheroid, respectively. We need ρ_{X_c}, the observed galactocentric Cartesian density that is elongated along our line of sight, to account for the Gaussian distribution of absolute magnitudes with dispersion ΔM_{g_0}. We obtain ρ_{X_c} through the sequence of transformations

$$\begin{aligned}&\rho_X(\boldsymbol{x}) \rightarrow \rho_R(R,\Omega) \rightarrow \rho_{g_{4.2}}(g_{4.2},\Omega)\\ &\quad\rightarrow \rho_{g_0}(g_0,\Omega) \rightarrow \rho_{\mathcal{R}}(\mathcal{R}(g_0),\Omega) \rightarrow \rho_{X_c}(\boldsymbol{x}).\end{aligned} \tag{9}$$

The relationship between these densities is determined by the transformations which take one coordinate space to the other. The $X \rightarrow R$ mapping is the well-known spherical coordinate transform,

$$\rho_R(R,\Omega) = R^2 \rho_X(\boldsymbol{x}). \tag{10}$$

If all of the stars had an absolute magnitude $\bar{M}_{g_0} = 4.2$, then one would measure an apparent magnitude

$$\begin{aligned}&g_{4.2} = 4.2 + 5\log_{10}\left(\frac{R}{10\ \mathrm{pc}}\right), \text{ therefore}\\ &R = \mathcal{R}(g_{4.2}) = 10^{0.2(g_{4.2}-4.2-10)}\ (\mathrm{kpc}),\\ &dR = \frac{\ln 10}{5} R\, dg_{4.2}.\end{aligned} \tag{11}$$

Thus, the relationship between ρ_R and ρ_g is given by

$$\begin{aligned}\rho_{g_{4.2}}(g_{4.2},\Omega) &= \frac{dg_{4.2}}{dR}\rho_R(R,\Omega) = \frac{\ln 10}{5} R^3 \rho_X(\boldsymbol{x})\\ &= \frac{\ln 10}{5}\mathcal{R}^3(g_{4.2})\rho_X(\boldsymbol{x}).\end{aligned} \tag{12}$$

The measured g_0 is given by

$$g_0 = g_{4.2} + \Delta M_{g_0}. \tag{13}$$

Since g_0 is the sum of independent random variables, its density is the convolution of the two densities, i.e., we have that $\rho_{g_0}(g_0,\Omega) = \rho_g * \rho_{\Delta M_{g_0}}(g_0,\Omega)$, where the convolution is in the g-dimension. Thus,

$$\rho_{g_0}(g_0,\Omega) = \int_{-\infty}^{\infty} dg\, \rho_{g_{4.2}}(g,\Omega)\mathcal{N}(g_0 - g;u), \tag{14}$$

where $\mathcal{N}$ is the Gaussian density function given by

$$\mathcal{N}(x;u) = \frac{1}{u\sqrt{2\pi}} e^{-\left(x^2/2u^2\right)}, \tag{15}$$

with $u = 0.6$. Switching back from apparent magnitude to spherical coordinates,

$$\rho_{\mathcal{R}}(\mathcal{R}(g_0),\Omega) = \frac{5}{\ln 10}\frac{\rho_{g_0}(g_0,\Omega)}{\mathcal{R}(g_0)}. \tag{16}$$

Since the coordinate spaces X_c and $\mathcal{R}$ are related by the spherical coordinate transformation, we may now collect our results and write the convolved density ρ_{X_c} in terms of ρ_X as follows,

$$\begin{aligned}\rho_{X_c}(\boldsymbol{x}) &= \frac{1}{\mathcal{R}^2(\mathcal{G}_0)}\rho_{\mathcal{R}}(\mathcal{R}(g_0),\Omega)\\ &= \frac{5}{\mathcal{R}^3(\mathcal{G}_0)\ln 10}\rho_{g_0}(g_0,\Omega)\\ &= \frac{5}{\mathcal{R}^3(\mathcal{G}_0)\ln 10}\int_{-\infty}^{\infty} dg\, \rho_g(g,\Omega)\mathcal{N}(g_0 - g;u)\\ &= \frac{1}{\mathcal{R}^3(\mathcal{G}_0)}\int_{-\infty}^{\infty} dg\, \mathcal{R}^3(g)\rho_X(\boldsymbol{x}(\mathcal{R}(g),\Omega))\mathcal{N}(g_0 - g;u),\end{aligned} \tag{17}$$

where $(\mathcal{R}(g_0),\Omega)$ are the angular coordinates of $\boldsymbol{x}$.

We have derived the convolved stellar density function ρ_{X_c}, for a generic stellar density ρ_X, which could represent either the stream or the spheroid densities in our present context. This convolution is performed separately on the stream and spheroid densities, to compute the functions $\rho_{\mathrm{str}}^{\mathrm{con}}(l,b,\mathcal{R}(g_0))$ and $\rho_{\mathrm{sph}}^{\mathrm{con}}(l,b,\mathcal{R}(g_0))$. We use the numerical integration technique of Gaussian quadrature (Heath 2002) to do the convolution integral.

2.3.4. *Combined Probability Density Function*

We are now ready to compute the PDF for the combination of the stream and spheroid densities. To do this we need the stellar densities of the stream and spheroid as derived in § 2.3.3, the detection efficiency for finding stars as a function of apparent magnitude, the volume over which the density is defined, and one more parameter, ϵ, which describes the fraction of stars in each of the two components (spheroid and stream).

The detection efficiency function, $\mathcal{E}$, was derived by fitting a sigmoid curve to the detection efficiency measurement in Figure 2 of Newberg et al. (2002). The efficiency function accounts for the decrease in detection efficiency at faint magnitudes,

$$\begin{aligned}&\mathcal{E}(g_0) = \frac{s_0}{e^{s_1(g_0 - s_2)} + 1}, \text{ where}\\ &\boldsymbol{s} = (0.9402, 1.6171, 23.5877).\end{aligned} \tag{18}$$

The dimensionless parameter that defines the fraction of the input stars that are in the stream and the fraction that are in the spheroid is ϵ. The parameter ϵ is modeled, as with the stream parameters themselves, separately for each stripe of data that is analyzed. Thus, the value of ϵ for a given stripe of data gives only the relative number of stars that comprise the stream as compared to the spheroid for that stripe of data and does not measure the fraction of stars in the stream as a function of position within the Galaxy.

By definition, the fraction of stars in the stream is

$$P_{\rm str} = \frac{e^{\epsilon}}{1+e^{\epsilon}}. \tag{19}$$

Similarly for the spheroid,

$$P_{\rm sph} = \frac{1}{1+e^{\epsilon}}. \tag{20}$$

We introduce here the concept of constrained and unconstrained variables. Instead of ϵ, we could have defined a variable f which is the fraction in the stream, and then, the fraction in the spheroid would be $1-f$. However, we would then need to maximize the likelihood subject to the constraint that this parameter must be between zero and one. To avoid that constraint, we instead introduced ϵ, which is not constrained. If ϵ is ∞, then all of the stars are in the stream; if $\epsilon = 0$, the stars are evenly split between the components; and if $\epsilon = -\infty$, then all of the stars are in the spheroid component. The other parameters are naturally unconstrained and do not need a conversion.

In total, then, we fit eight model parameters: five from the stream fragment in the current stripe $(\mu, R, \theta, \phi, \sigma)$, two from the spheroid (r_0, q), and one that specifies the fraction of stars in each component (ϵ). Together, we refer to these parameters as the eight components of the parameter vector $\boldsymbol{Q}$.

The PDF, then, is given by

$$\begin{aligned}&{\rm PDF}(l,b,\mathcal{R}(g_0)|\boldsymbol{Q})\\ &= \frac{e^{\epsilon}}{1+e^{\epsilon}}\frac{\mathcal{E}(\mathcal{R}(g_0))\rho_{\rm str}^{\rm con}(l,b,\mathcal{R}(g_0)|\boldsymbol{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho_{\rm str}^{\rm con}(l,b,\mathcal{R}(g_0)|\boldsymbol{Q})dV}\\ &+ \frac{1}{1+e^{\epsilon}}\frac{\mathcal{E}(\mathcal{R}(g_0))\rho_{\rm sph}^{\rm con}(l,b,\mathcal{R}(g_0)|\boldsymbol{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho_{\rm sph}^{\rm con}(l,b,\mathcal{R}(g_0)|\boldsymbol{Q})dV},\end{aligned} \tag{21}$$

where the integral is over the entire volume probed by the input data.

Calculation of the volume integrals dominates the run time of the algorithm. Since these integrals cannot be solved analytically for our functions, we define an integration mesh which divides the total volume into many volume elements. The edges of the volume elements are along constant g, μ, or ν. We then calculate the spheroid and stream probabilities at the center of each volume element and multiply this by the volume of that element. The values for the spheroid and stream probabilities in all volume elements are then summed and are hereafter referred to as the stream and spheroid integrals, respectively.

2.4. *Optimization*

Now that we have defined a PDF and a likelihood measure, we need to find the parameters that maximize that likelihood. At the beginning of the first iteration of the program, the likelihood is calculated for the input parameters. Thereafter, a new set of parameters and likelihood is calculated using a conjugate gradient search coupled with a line search technique. For a detailed description of the conjugate gradient and line search methods employed, see the Appendices A1 and A2, respectively.

2.5. *Errors in the Parameter Estimates*

The accuracy of the parameter estimates depends on the shape of the likelihood surface at its maximum, which is governed by the number of stars in our input catalog. The lower the number of stars, the wider the peak at the maximum and the larger the statistical error. It is also limited by the accuracy with which we are able to numerically determine the maximum. We address this latter point when we discuss the test data in § 3. Here, we explain how we calculate the accuracy with which we expect to be able to determine each parameter.

TABLE 1
PERTURBATION VALUES USED FOR THE GRADIENT AND HESSIAN

Parameter	h
μ (deg)	3×10^{-5}
R (kpc)	4×10^{-5}
θ (rad)	6×10^{-5}
ϕ (rad)	4×10^{-5}
σ (kpc)	4×10^{-6}
ϵ	1×10^{-6}
q	4×10^{-6}
r_0 (kpc)	8×10^{-4}

We assume that the likelihood surface for the parameters can be reasonably estimated by a Gaussian near the maximum. Then, an estimate for the variance of each parameter can be found from the square matrix of second-order partial derivatives of the log-likelihood function, evaluated at the maximum. This matrix is called a Hessian matrix, $\mathbf{H}$. The variance matrix $\mathbf{V}$ is the inverse of the Hessian matrix, normalized by the number of stars in the data set,

$$\mathbf{V} = \frac{1}{N}\mathbf{H}^{-1}. \tag{22}$$

The square roots of the diagonal elements of the matrix provide us with the statistical error in the measurement of each parameter. We compute the Hessian numerically using a finite difference method (as we did the gradient) as follows,

$$\begin{aligned}H_{ij} &= \frac{H_{ij}^1 - H_{ij}^2 - H_{ij}^3 + H_{ij}^4}{4h_ih_j}, \text{ where}\\ H_{ij}^1 &= \mathcal{L}(Q_j+h_j, Q_i+h_i)\big|_{\text{all other } Q_k \text{ fixed}},\\ H_{ij}^2 &= \mathcal{L}(Q_j-h_j, Q_i+h_i)\big|_{\text{all other } Q_k \text{ fixed}},\\ H_{ij}^3 &= \mathcal{L}(Q_j+h_j, Q_i-h_i)\big|_{\text{all other } Q_k \text{ fixed}},\\ H_{ij}^4 &= \mathcal{L}(Q_j-h_j, Q_i-h_i)\big|_{\text{all other } Q_k \text{ fixed}}.\end{aligned} \tag{23}$$

Here, Q_j is the jth component of parameter vector $\boldsymbol{Q}$, similarly for Q_i; and h_j is the perturbation value for Q_j, similarly for h_i. In practice we use the same h_k to calculate the gradient as the Hessian. These h_k can be seen in Table 1.

2.6. *Separation*

Once we have found the values of the parameters in our model, we can use the model to separate the input catalog of stars into two catalogs, one having the density profile of the stream and the other having the density profile of the spheroid. It is possible to determine the probability that a star is in the stream or in the spheroid; however, it is not possible to determine which stars are in the stream and which are in the spheroid. For example, suppose a star selected from the data set is computed to be a stream star with probability 0.6 and that it is a spheroid star with probability 0.4. We would put that star in the stream catalog with probability 0.6 and in the spheroid catalog with probability 0.4, but there is a 48% chance that the star was put in the wrong file. The separation has proved very useful for evaluating the effectiveness of the model fits to the data and hopefully will be useful

for fitting tidal stripping models to the stream density profiles. However, it may not be the best method for selecting spectroscopic follow-up targets to measure the composition and velocities of stream stars, for example. Spectroscopic follow-up targets might be better selected based on the probability that it is a stream star, rather than on the catalog to which it was randomly assigned.

To perform this separation, we calculate for each star the probability, T, that it was drawn from the stream population given the parameters, where T is defined by

$$T(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q}) = \frac{S(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q})}{S(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q}) + B(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q})}, \tag{24}$$

where $S(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q})$ is the stream portion of equation (21), and $B(l, b, \mathcal{R}(g_0) \mid \boldsymbol{Q})$ is the spheroid portion of equation (21).

A uniform random number between 0 and 1 is then generated and tested against this probability. If the probability of being drawn from the stream population is greater than that of the random number, then the star is placed into the stream star catalog; otherwise, it is placed within the spheroid catalog. We repeat this process, generating a new random number for each star until all stars have been placed into their appropriate catalog. Thus, we get two distinct populations: one that is a smooth spheroid and one that has all the density properties of the stream.

We use this nondeterministic approach of testing the star probability against a random number for extracting the stream from the spheroid stars in order to get an accurate representation of the density profiles. Since the stream probability definition is based solely on the star's distance from the stream axis, assigning all stars with a probability greater than a set value to the stream would result in assignment of all stars within a certain distance to the stream. This would not be an accurate representation of the stream as we would be carving out a cylinder of stars from the data rather than a set of stars that fit a specific distribution with Gaussian cross section in density.

2.7. *Run Time and Distribution*

In determining the likelihood, a trade-off needs to be made in the accuracy of the numerical integral calculation in equation (21). The smaller the step size, ζ, in each dimension the more accurate the integral calculation is, which can lead to faster convergence of the search method being used. However, increasing the precision of the integral calculation leads to polynomially longer calculation times (scaling like ζ^3). Thus, the precision to which the integral is calculated is the dominant factor in the run time and accuracy of our astronomical model.

Testing has shown that in order to generate a meaningful likelihood value in stripe 82, a minimum of 1.4 million points within the stream and spheroid integrals is required. On a single processor, this takes approximately 15 minutes, assuming there is no magnitude distribution and the nonconvolved probabilities are used. As performing a single conjugate gradient descent can take thousands (or more) of integral calculations, these long execution times presented an obstacle to this research.

To generate a meaningful value for the magnitude convolution integral, each integral point requires the calculation of at least 30 surrounding points. This increases the calculation time of a single integral to 7.5 hr. Again, assuming a bare minimum of 1000 likelihood evaluations for a single conjugate gradient descent (typically the range is closer to 5000), evaluating one astronomical model would take almost a year on a single processor.

To alleviate these massive computational requirements, a generic maximum likelihood evaluator (GMLE[3]) has been developed and used (Desell et al. 2008). GMLE allows for a likelihood calculation to use various distributed computing environments, such as multiple processors in a cluster, a heterogeneous grid of clusters, or even a supercomputer.

Using this distributed framework with 88 processors on the Rensselaer Grid resulted in a 65 times speedup over a single processor, and using 512 nodes of an IBM BlueGene/L system resulted in a 148 times speedup. It now takes only a few days to compute the maximum likelihood parameters for a particular model and stripe of data. In both the grid and supercomputing environments used, the overhead of communication was very small compared to the calculation time. Communication time consists of approximately 1%–10% of the time to calculate the likelihood and does not noticeably increase as more processors are used. With these observations, the calculation should scale to at least 1000 processors on a grid and 10,000 processors on the IBM BlueGene/L. More precise calculations of the integral and magnitude convolution will allow the distributed calculation to scale to even larger numbers of processors, due to the increased calculation time assuming relatively fixed communication times.

GMLE has also been extended to allow maximum likelihood evaluation over the Berkeley Open Infrastructure for Network Computing (BOINC) Internet computing framework (Anderson et al. 2005). This allows users to volunteer computing resources by downloading a BOINC client and attaching to the MilkyWay@home project.[4] Currently, over 500 computers from around the world are being volunteered for the project and performance is comparable to 512 nodes of the IBM BlueGene/L.

2.8. *Limitations and Enhancements*

Generically, all enhancements to this algorithm either improve the Galaxy model, increase the quantity of data to which the model is fit, or improve the speed and accuracy of the convergence. We plan to improve the algorithm in all of these aspects.

We plan to build up a model of the Milky Way stellar density, starting with the spheroid component. Over time, we will be able to include additional streams and additional pieces of the Sgr dwarf tidal stream, a more sophisticated model of the smooth component of the spheroid, and disk components. The spheroid substructure is best fit in a piecewise manner as we have done in this paper, but the smooth components require that more directions be fit simultaneously. To be able to fit multiple stripes during a single optimization it must be possible to fit multiple debris pieces during a single run as well. To add additional tidal debris, we need only to add a term for each additional stream to the PDF,

$$\begin{aligned}\mathrm{PDF} = \sum_{i=1}^{k} \left[\frac{e^{\epsilon_i}}{(1 + \sum_{j=1}^{k} e^{\epsilon_j})} \frac{\rho_{\mathrm{str}}^{\mathrm{con}}(l, b, \mathcal{R}(g_0) | \boldsymbol{Q}_{\mathrm{str}_i})}{\int \rho_{\mathrm{str}}^{\mathrm{con}}(l, b, \mathcal{R}(g_0) | \boldsymbol{Q}_{\mathrm{str}_i}) dV} \right] \\ + \frac{1}{(1 + \sum_{i=1}^{k} \epsilon_i)} \frac{\rho_{\mathrm{sph}}^{\mathrm{con}}(l, b, \mathcal{R}(g_0) | \boldsymbol{Q}_{\mathrm{sph}})}{\int \rho_{\mathrm{sph}}^{\mathrm{con}}(l, b, \mathcal{R}(g_0) | \boldsymbol{Q}_{\mathrm{sph}}) dV},\end{aligned} \tag{25}$$

where i and j denote the ith and jth stream of k total streams, respectively. We add five new parameters, $\boldsymbol{Q}_{\mathrm{str}_i}$, and a sixth parameter, ϵ_i, for each new stream segment. This change to the PDF would allow k pieces of tidal debris to be fit within a single or multiple stripes. We are considering whether we should enforce continuity conditions between adjacent sections of the tidal stream.

[3] GMLE is available for download along with more information at http://wcl.cs.rpi.edu/gmle.

[4] See http://milkyway.cs.rpi.edu/milkyway/.

TABLE 2
Results for the Simulated Data Set of 205,708 Stars

Parameter	True Value	Expected Deviation	Numerical Error	Optimized Value	Actual Deviation	Std. Deviation of Optimizations
μ (deg)	31.361	0.233	0.05	31.443	0.082	0.064
R (kpc)	29.228	0.167	0.04	29.217	0.011	0.010
θ (rad)	1.445	0.032	0.003	1.421	0.024	0.0005
ϕ (rad)	3.186	0.049	0.001	3.182	0.004	0.002
σ (kpc)	2.854	0.033	0.015	2.858	0.004	0.009
ϵ	−1.828[1]	0.005	0.002	−1.833	0.005	0.005
q	0.670	0.013	0.000	0.671	0.001	0.0004
r_0 (kpc)	13.500	0.276	0.15	13.917	0.417	0.016

[1] Corresponds to a stream star ratio of 13.85%: 28,498 in stream and 177,210 in spheroid.

Since it has been suggested that the smooth portion of the spheroid may not be symmetric, we would also like to increase the number of parameters in the smooth spheroid component, following Newberg & Yanny (2006).

We are currently fitting only F turnoff stars and assuming that the absolute magnitude distribution of these stars is a standard Gaussian for all populations. Ideally, we would use stars in a wider color range and develop a probability distribution in magnitude and color for each population of stars. This would increase the number of stars available to the algorithm and would allow us to use additional information about real structures in the Milky Way—insisting that each population with a turnoff also has a main sequence and giant branch, for example. Implementing this requires significant development not only in the computation of the likelihood, but also in characterizing Galactic stellar distributions. Small steps in improving this aspect of the algorithm are currently underway.

All of the preceding improvements to the algorithm increase the number of fit parameters and the complexity of calculating the likelihood function. In parallel with scientific development of the algorithm, we will be improving the speed of convergence. We are experimenting with other methods of finding the maximum and increasing the number of architectures on which our software can be run in parallel. Future architectures that we will use include supercomputers and BOINC, which among other products supports the SETI@home application.

3. SIMULATED DATA

3.1. *Generation*

To test our algorithm we generated a simulated version of SDSS stripe 82. The "True Value" column of Table 2 shows the parameter values used for the data generation, while Figure 2 shows a plot of the simulated data. The data set contained a total of 205,708 simulated stars, 28,498 (13.85%) of which are stream stars. The stream model parameters and star ratio were chosen to be the same as the real data for stripe 82. The techniques of generation are described below.

3.1.1. *Stream*

We create the simulated stream using an active generation technique, which means that all points that are generated are valid stream stars. First, we define a set of stream parameters to generate over. From these parameters we then calculate the $\boldsymbol{c}$ and $\hat{\boldsymbol{a}}$ vectors defining our stream position and direction. We then generate three random numbers: a uniform random number that determines where along the stream axis the point is, and two Gaussian random numbers (with standard deviation σ as defined in the parameters) to define the stream cross-sectional coordinates. The point is then converted to Galactic longitude, latitude, and apparent magnitude. Thus far, we have assumed a constant absolute magnitude for each star in the stream. To account for the fact that there is an absolute magnitude distribution, we add or subtract a small amount from the apparent magnitudes of each of the stream stars in the data set. A Gaussian random variable with standard deviation 0.6 is generated (one for each point); this value is added to the apparent magnitude of the generated point. Finally, we keep points with a probability given by the efficiency function to produce a data set for a simulated stream. Stars near the magnitude limit of the data are tossed out of the sample with a probability equal to the efficiency of our object detection as a function of apparent magnitude in the real data.

3.1.2. *Spheroid*

The simulated spheroid is generated using a rejection sampling technique. Here we cannot apply the same active generation technique we used for the stream because the density distribution in the smooth component of the spheroid is more complex than a Gaussian. Instead, we define a rejection technique that allows us

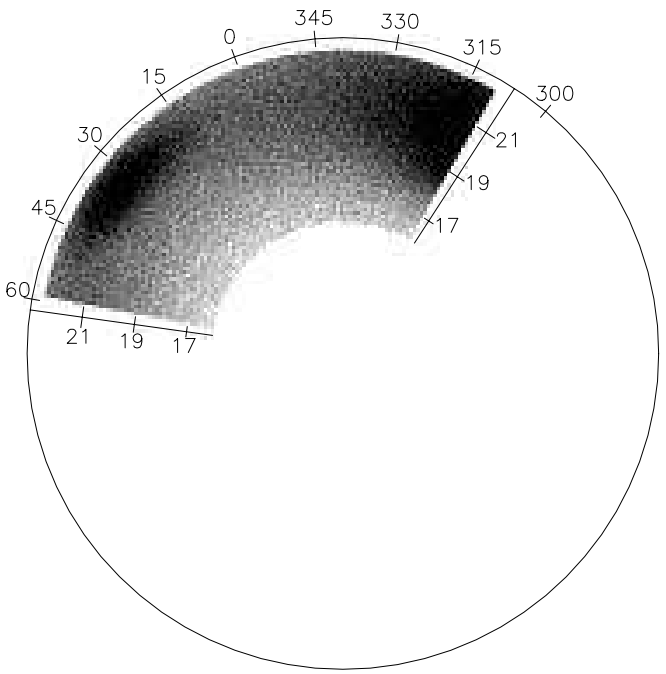

Fig. 2.—Simulated data wedge plot. We show a Sun-centered density plot of the 205,708 stars generated to mimic SDSS stripe 82. The quantity g_0 is labeled along the radial spokes; the circle indicates where $g_0 = 23$. The angle μ in degrees is marked along this circle. For stripe 82 $\mu = \alpha$ and $\nu = \delta$. The simulated stream is easily discernible at $(\alpha, g_0) = (33.4, 21.4)$.

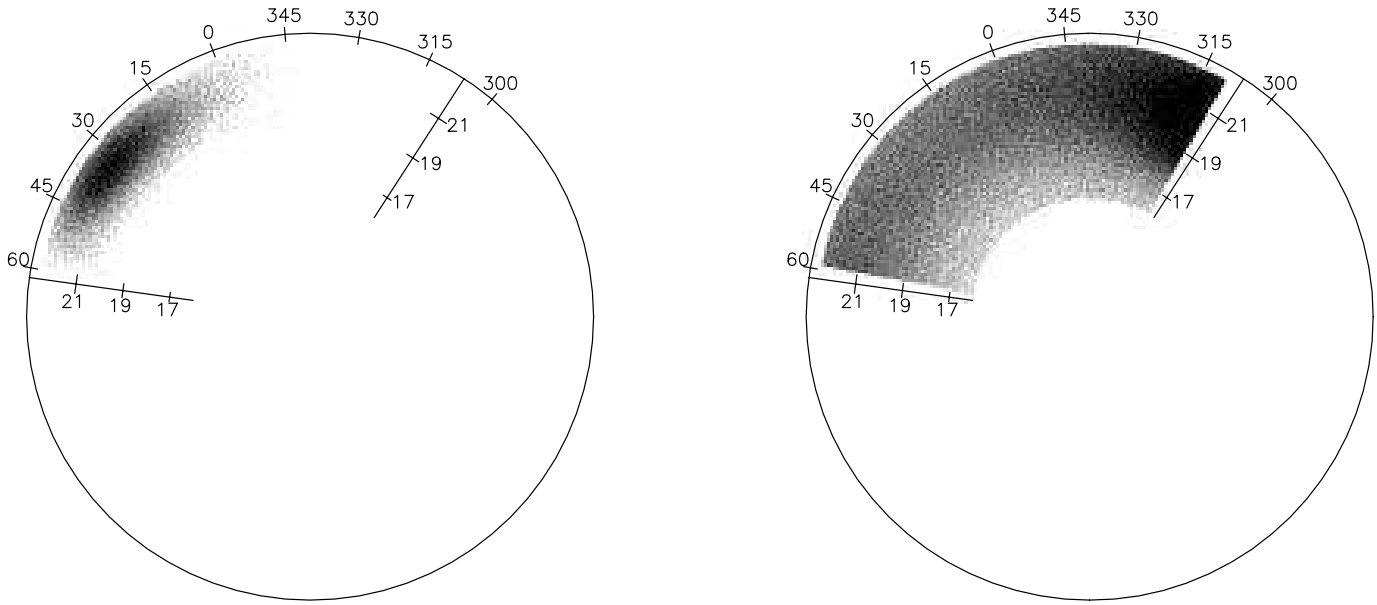

FIG. 3.—Simulated stream and spheroid stars after separation. Here we have plotted the same 205,708 star simulated data set similarly to Fig. 2. However, we have separately plotted the stream (*left*) and spheroid (*right*) catalogs returned from running our separation algorithm. The spheroid is recovered as a smooth Hernquist function after the removal of the stream component.

to generate over a given stripe. We cannot simply generate values for μ, ν, and R, however, because the stars are not uniformly distributed in these variables. The distribution in ν varies according to cos (ν), and the distribution in R must take into account the growing volume at increasing distance. The quantity μ may be generated uniformly over the stripe, as it corresponds to a longitude, and the volume of space in each longitude bin is the same. The volume element in a stripe is given by

$$dV = R^2 \cos(\nu) dr\, d\mu\, d\nu, \tag{26}$$

and the corresponding stripe volume is

$$V = \frac{R_{\max}^3}{3}(\mu^+ - \mu^-)[\sin(\nu^+) - \sin(\nu^-)], \tag{27}$$

where $R_{\max}$ is the maximum radial distance, and the positive and negative superscripts refer to the maximum and minimum values of that coordinate for the stripe. These equations are used to define the functions which allow us to generate in the other two coordinates,

$$\nu = \sin^{-1}\{u * [\sin(\nu^+) - \sin(\nu^-)] + \sin(\nu^-)\},$$
$$R = R_{\max} w^{1/3}, \tag{28}$$

where u and w are uniform random numbers between zero and one. The quantity μ is generated by a uniform random number between the stripe limits. Once a point is generated, its probability is calculated using the likelihood function developed in § 2.3. The probability is then divided by the total probability possible for a star. The total probability possible for a star is simply the maximum value that can be returned by the likelihood function given the current volume and parameter set. The value of this ratio is then tested against a uniform random number between zero and one; the point is put into the data set if the value is greater than the value of the random number.

3.1.3. *Accuracy of Parameter Determination*

Using the method described in § 2.5 we calculated the expected accuracy for each parameter. Since we created the simulated data with a known set of parameters, we therefore know the true parameter values of the data set. We can calculate the Hessian matrix at the optimum values and, subsequently, get a set of error bars at the optimum values. These error bars therefore give us the accuracy with which we expect to find the values of the parameters. The values of these errors can be found in the

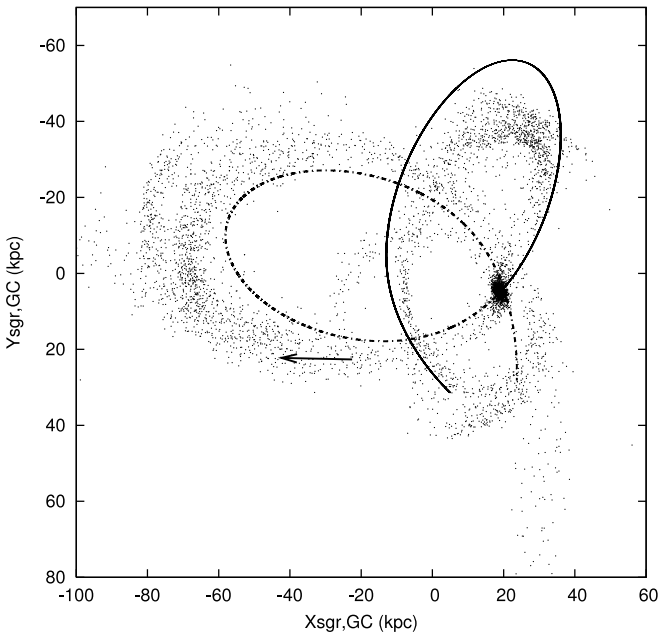

FIG. 4.—Plot of the disruption of the one million particle simulated Sgr dwarf with a spherical dark matter halo ($q = 1$) and a velocity dispersion of $v_{\rm halo} = 114$ km s^{-1} in the Sgr X-Y orbital plane. To reduce crowding, the stream was sampled 1 in 100. The thin line is the future orbit of the core, while the dotted line is the past orbit. The past orbit is not closed, but in fact follows the lower trailing stream. The arrow depicts the center and direction returned by our algorithm optimization.

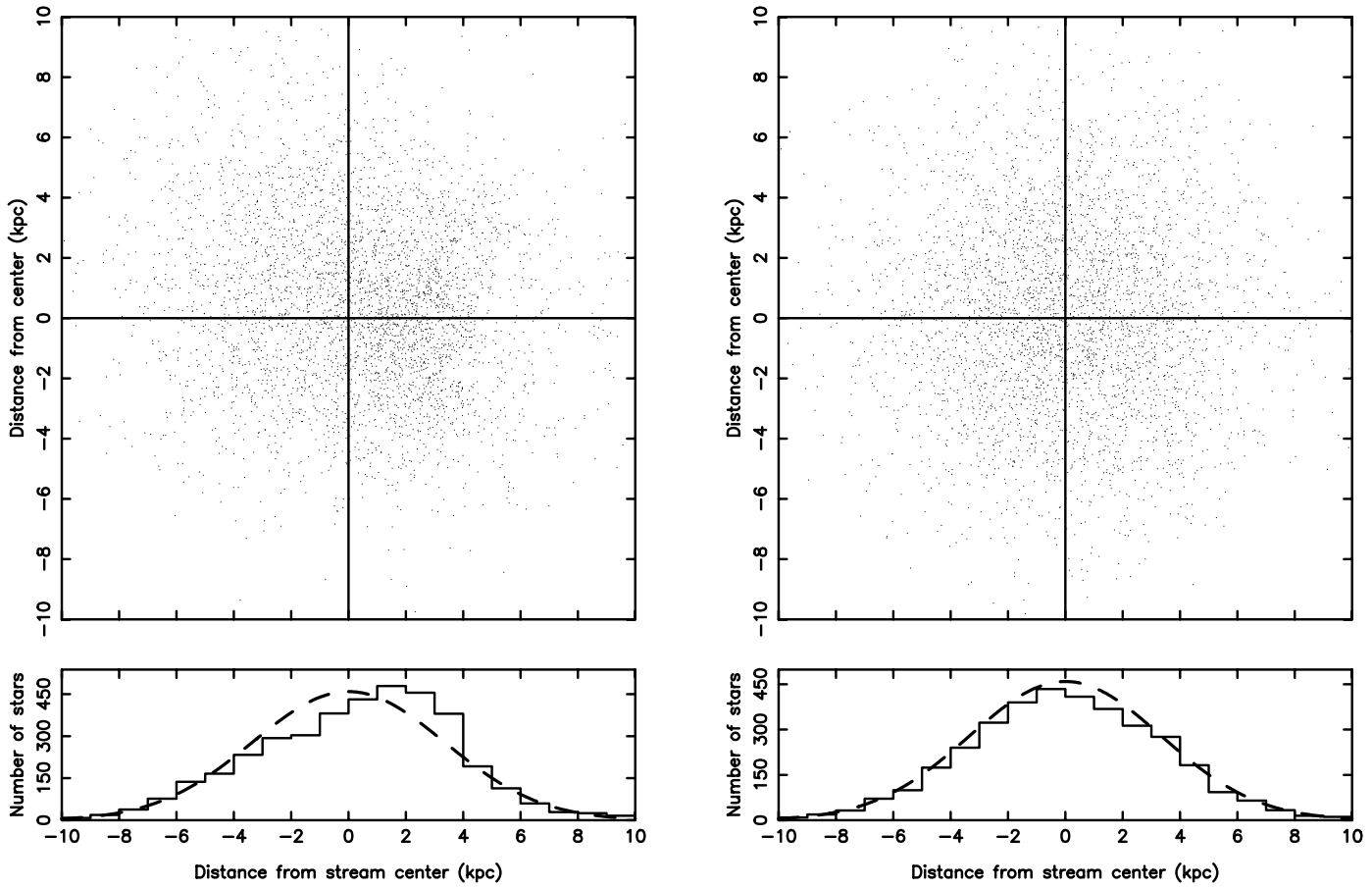

FIG. 5.—Cross sections of Sgr *N*-body simulation stream (*left*) and simulated stream (*right*). The simulated stream matches our stream model and was generated from the parameters fit to the *N*-body simulation. The top figure in both panels shows the 1 kpc thick cross section of the respective data set. The cross sections are centered at the best-fit value of the optimization of the center in the *N*-body simulation from Fig. 4. Here the axes are perpendicular to the stream direction. The *X*-axis is $0.053X + 0.055Y + 0.997Z$ and the *Y*-axis is $0.055X + 0.997Y - 0.058Z$, where X, Y, Z are galactocentric Cartesian coordinates with the Sun at $X = -8.5$ kpc and moving in the direction of positive *Y*. The bottom figure of both panels is a histogram of those stars within the cross section binned along the *X*-axis. The thick dashed line shows a Gaussian distribution with standard deviation given by σ from the fit to the *N*-body simulation. Note that the cross section of the *N*-body simulation is non-Gaussian and somewhat asymmetric. Also note the model simulated stream is well fit by a Gaussian. The density distributions in the left and right panels are not the same; however, the algorithm still fits a reasonable center for the non-Gaussian *N*-body data set.

column "Expected Deviation" (the statistical error bar) in Table 2 for the 205,708 star simulated data set.

In addition to the statistical errors, there can be numerical errors driven primarily by the accuracy with which we determine the integrals. These numerical errors were determined by calculating the accuracy of the likelihood for a fixed number of points in the numerical integration and then calculating the "error" in each parameter by perturbing a single parameter, holding all others fixed, until a change in the total likelihood is observed that is greater than the minimum accuracy threshold of the numerical integration. It should be noted that the values quoted as "Numerical Error" are heuristic errors, meaning they are estimated and not a true error bar.

3.2. *Testing*

We tested our algorithm by letting it optimize to convergence for eight randomized sets of input parameters. Randomized here means a random perturbation of the actual values by 75% of the parameter's value. Although parameters outside of this range are in principle allowed, the likelihood surface for parameter values very far from the correct values is so flat, and the gradient so small, that numerical errors dominate our measurement of the gradient. In these eight optimizations, five optimized to the true values, two optimized to a local maximum which has a lower likelihood than that of the true values, and one did not converge because it was far enough from the correct parameter values that the gradient could not be accurately determined. The local maximum appears to be related to the fact that when the stream crosses the stripe at low inclination, the algorithm sometimes fits the data with a much higher inclination angle and larger width to compensate. All other parameters are still optimized to their true values at this local maximum.

The average values of the returned parameters from the five optimizations to the true values are presented in the "Optimized Value" column of Table 2, and the "Actual Deviation" is the

TABLE 3
STREAM FITTING RESULTS FOR TESTS WITH INCORRECT SPHEROID MODEL

			Optimization		
Parameter	True Value	Expected Deviation	$\alpha = 3.5$	$\alpha = 2.5$	$\alpha = 2.0$
μ (deg)	31.361	0.233	31.449	31.526	31.457
R (kpc).....	29.228	0.167	29.094	29.326	29.108
θ (rad).....	1.445	0.032	1.426	1.452	1.437
ϕ (rad).....	3.186	0.049	3.172	3.168	3.160
σ (kpc).....	2.854	0.033	2.869	2.848	2.865

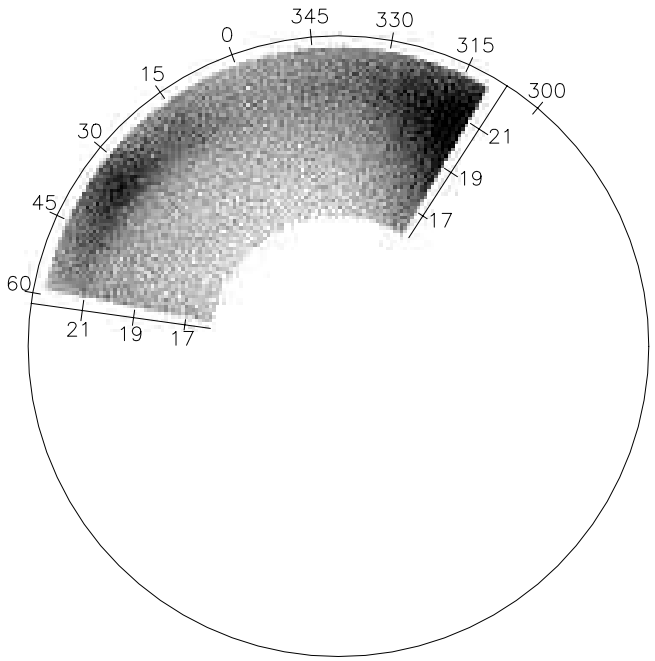

FIG. 6.—F turnoff stars within SDSS stripe 82. Here we plot, as in Fig. 2, F turnoff stars with color cut $0.1 < (g-r)_0 < 0.3$ and $(u-g)_0 > 0.4$ that are within the volume defined by $310 < \alpha < 419$, $\pm 1.25 = \delta$, and $16 < g_0 < 22.5$ that were detected in SDSS stripe 82. Sgr tidal debris is visible around $(\alpha, g_0) = (31, 21.5)$.

difference between the "True Value" and the "Optimized Value." The returned parameters have very little deviation. The standard deviation of the five optimizations were calculated to be around an order of magnitude lower than the statistical error calculated from the Hessian. These can be seen in the "Std. Deviations of Optimizations" column of Table 2. Note that if the theoretical deviation is large for a particular parameter, then near the maximum, the likelihood changes little with the variation of that parameter. This leads to a larger "Actual Deviation" and can also make it more difficult for the algorithm to numerically find the maximum, since the gradient will be very small. The errors calculated from the Hessian assume that the maximum of the PDF is found exactly.

As can be seen in Table 2, all of the parameters have "Actual Deviations" which are smaller than their "Theoretical Deviations" with the exception of r_0. It has been discovered that the likelihood changes so little compared to perturbations in these parameters when close to the maximum that this parameter is not being calculated accurately enough to reach the "True Value" seen in Table 2. The true error bar should be taken as the sum, in quadrature, of the "Expected Deviation" and "Numerical Error" columns. Once this is done, the "Optimized Value" of all parameters are comparable, within the errors, to the "True Value."

Finally, we ran our separation algorithm on the average returned parameter values to create two catalogs of stars given those parameters, and we plotted these separately in Figure 3. Clearly visible is the stream with a Gaussian density fall off in the left plot, while the right plot depicts a smooth Hernquist profile.

We also tested our algorithm on this same simulated data set of 205,708 stars without using an absolute magnitude correction, thereby simply assuming that all F turnoff stars have an absolute magnitude, $M_{g_0} = 4.2$. This was done to see how important this correction was and to see whether the algorithm could be effectively used without it. The results showed that the parameters of μ and r_0 were the only values not affected by this correction; the rest of the parameters deviated wildly from their actual values with the stream parameters R, θ, and ϕ being the worst of the eight parameters reaching upward of 34 σ in error.

3.3. *Robustness of Models*

Section 3.2 showed that the algorithm produces the correct answer given that the data is drawn from our model. In this section we test the algorithm on data that was not generated according to the assumed spatial model. We constructed a reproduction of the Sgr dwarf galaxy tidal disruption produced by Law et al. (2005) using a semianalytic *N*-body approach. Using their parameters for the Galactic potential and kinematic values, we constructed an orbit for Sgr using the NEMO Stellar Dynamics Toolbox (Teuben 1995). Specifically, we used the case of a spherical dark matter halo, with a velocity dispersion of 114 km s^{-1}. We used a Plummer sphere with one million particles as our model for Sgr. The sphere was evolved along the orbit for 3.18 Gyr until it reached the present position of the Sgr dwarf spheroidal. The resulting plot of the disruption of Sgr is shown in Figure 4. It is consistent with the results obtained by Law et al. (2005).

We then selected the volume corresponding to SDSS stripe 82 from the *N*-body simulation and fit it using our model Hernquist profile spheroid and maximum likelihood technique. Using the output model parameters found by our algorithm, we created a new simulated stream as described in § 3.1.1. Finally, we compared the cross sections of this simulated stream with those of a similar stream generated with our density model.

A cross section of these two streams 1 kpc thick and centered at the returned stream center are plotted, along with a histogram of the stars within the cross section, in Figure 5. Figure 5 (*left*) depicts the *N*-body simulation stream, while Figure 5 (*right*) is that of our simulated stream. It can be seen that the *N*-body simulation is indeed not Gaussian. The drawback to testing with an *N*-body model is that there are no correct model parameters with which to compare; but the center, as shown in Figure 5 (*left*), and

TABLE 4
RESULTS FOR THE 115,907 F TURNOFF STARS IN SDSS STRIPE 82

Parameter	Optimized Value	Statistical Error	Numerical Error	Std. Deviation of Optimizations
μ (deg)	31.373	0.244	0.08	0.008
R (kpc)	29.218	0.184	0.07	0.012
θ (rad)	1.444	0.044	0.01	0.001
ϕ (rad)	3.184	0.034	0.008	0.002
σ (kpc)	2.862	0.025	0.008	0.009
ϵ	−1.827[1]	0.005	0.001	0.000
q	0.458	0.023	0.005	0.001
r_0 (kpc)	19.404	0.581	0.09	0.051

[1] Corresponds to a stream star ratio of 13.86% of the 115,907 F turnoff stars, about 16,000 stream stars.

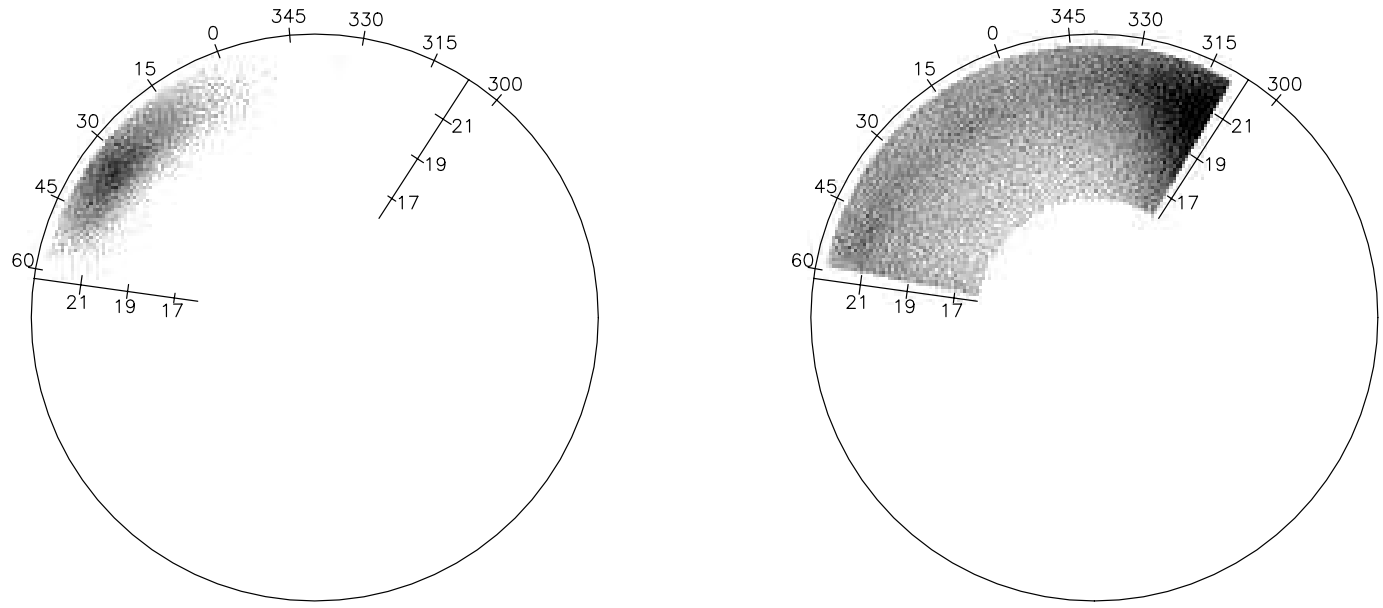

FIG. 7.—F turnoff stars within SDSS stripe 82 after separation. Here we have plotted the data from Fig. 6 after being separated into two distinct catalogs: one for stream and one for spheroid. The stream (*left*) has clearly been extracted from the spheroid (*right*).

the direction, as shown in Figure 4, are reasonable. In order to assess the error in the determination of the stream center, one would have to define what is meant by the center of an asymmetric distribution, but the algorithm's choice seems reasonable.

Next, we estimate the curvature of the Sgr tidal stream in stripe 82. Taking the distance to the center of the stream in stripe 82 to be 29 kpc (Newberg et al. 2003), we calculate that the 2.5° wide SDSS stripes would be 1.3 kpc thick at this distance. The length of the stream in stripe 82, taking an inclination of 30° from the stripe (Freese et al. 2005), is then 2.5 kpc. We estimate the radius of curvature of the Sgr tidal tail in the orbital plane at this point to be 18 kpc by fitting a circle to the two southern stripes with detections in Newberg et al. (2003). The deviation from linear can then be found to be $d = 0.06$ kpc at the edge and is very small when compared to the 6 kpc (Freese et al. 2005) width of the stream or the 6.74 kpc width found in this paper. The linear approximation is quite reasonable for the Sgr stream in stripe 82.

Finally, we tested our algorithm with differing spheroid models. Because it was easier to change the spheroid model we fit than to regenerate many sets of spheroid stars, we modified the exponential component of the model Hernquist profile the algorithm fits to the data. Three tests were performed with the exponent of the Hernquist profile set to 3.5, 2.5, and 2.0 as compared to the value of 3.0 used to generate the data. The results from these tests can be found in Table 3. In all cases, the stream parameters and the epsilon parameter saw little to no change (all optimizations were within the errors calculated above). As expected, the spheroid parameter values were incorrect, but represent the best parameters for the model profile fit to the simulated data within our generated volume.

4. RESULTS FROM SDSS STRIPE 82

After validating the algorithm on the simulated data, we then applied the algorithm to the stripe 82 data as described in § 2.2. Figure 6 shows a plot of this data where a piece of the Sgr tidal stream is clearly visible as a dense structure on the left of the image at $(\alpha, g_0) \approx (31, 21.5)$.

We selected from the SDSS DR6 all of the sources in stripe 82 that had magnitudes $16 < g_0 < 22.5$, colors of $0.1 < (g-r)_0 < 0.3$ and $(u-g)_0 > 0.4$, were identified as point sources, and were not EDGE or SATURATED. Since the selected stars are far from the Galactic plane, the reddening can reasonably be estimated from the total amount of dust in that direction in the sky. The subscript "0" indicates that we selected reddening-corrected magnitudes from the SDSS databases. The cut in $(u-g)$ is made to remove low-redshift quasi-stellar objects. We limited the angular length along stripe 82 to 109°, with right ascension in the range $310° < \alpha < 360°$, $0 < \alpha < 59°$. Since stripe 82 is centered on the celestial equator, the edges of the stripe are ±1.25° in declination and the angle along the stripe is given exactly by right ascension. A small section of data with right ascension $323.2° < \alpha < 323.6°$ and declination $-1.0° < \delta < -0.7°$ was removed in order to avoid a globular cluster. The final data set includes 115,907 stars.

As was done for the test data, a number of randomized sets of initial parameters were used (having the same 75% perturbation), and the average of the parameters of those optimizations that converged to a minima was taken. In this case, 10 total optimizations were performed, five optimized to a minima, four to a local minima with lower likelihood than the true minima, and one was again a poor starting point allowing for no optimization. The Hessian method was then used to calculate the error bars for these parameters at the true minima. The errors, as well as the average parameter values, are tabulated in Table 4. There are additional, unknown systematic errors for real data that were not present for the model data. The quoted errors assume that the model is an accurate representation of the stellar density function, including our formula for the smooth spheroid, tidal stream, and absolute magnitude distribution.

The average parameter values were used as input for our separation algorithm to create stream and spheroid catalogs for stripe 82. These catalogs were then plotted in Figure 7. As can be seen, the Sgr tidal stream has been clearly extracted from the spheroid in the left panel, while the spheroid remains in the right panel.

We compared the average stream center and direction with Figure 4 from Newberg et al. (2007) in Figure 8 of this paper. As can be seen in the figure, the stream direction is roughly tangent to

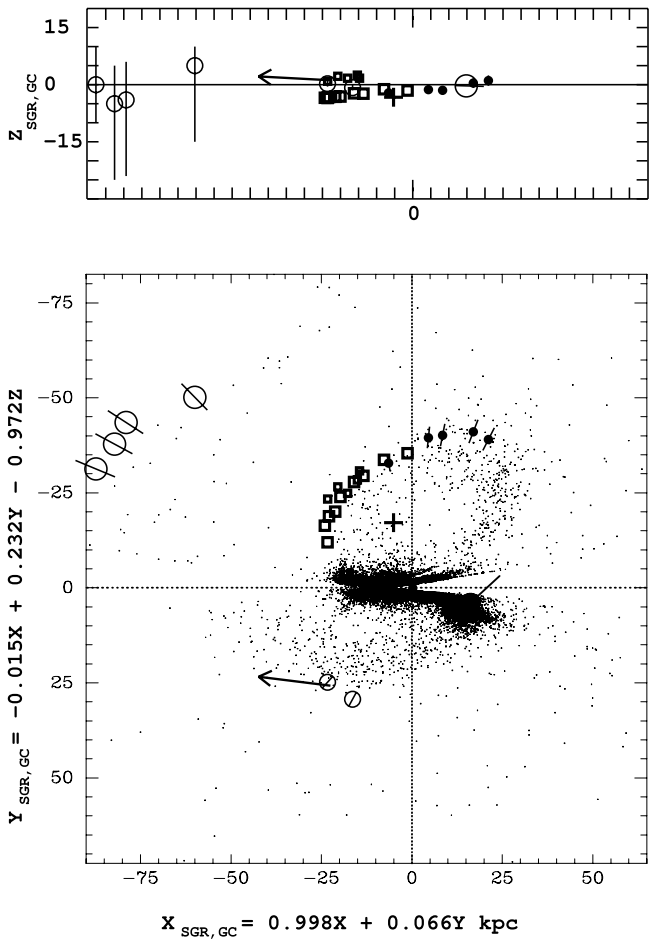

FIG. 8.—Plot of the stream center and direction in the Sgr dwarf plane compared to other Sgr detections. Here we reproduce Fig. 4 of Newberg et al. (2007), which displays detections of the Sgr stream in A-colored stars in eleven stripes, where the filled circles and larger squares represent leading debris, the open circles trailing debris, and the smaller squares Sgr debris on the opposing side of the Sgr orbital plane. These are plotted along with the positions of 2MASS M giants from Fig. 11 of Majewski et al. (2003). The arrow shows our improved measurement of the position and direction in one place on the Sgr tidal stream. The length of the directional vector is arbitrary, representing only the projection of an elongated unit vector onto the respective planes. Note that the direction is tangent to the 2MASS M star data and plausibly on a smooth path from stripe 86 (the open circle just to the right of our detection for stripe 82) to stripe 27 (the open circle farthest to the left in the bottom panel).

the stream of 2MASS M giants (although at a different distance—there is a difference in scale between distances measured with the M giants and distances measured with F turnoff stars and RR Lyraes). Notice also that the stream direction does not lie exactly in the plane of the Sgr dwarf, but is tilted slightly with respect to the $X_{\rm Sgr}$-$Y_{\rm Sgr}$ plane ($3^\circ \pm 1^\circ$). We also compare our direction to the directional results of Freese et al. (2005), who estimated the angle between the observational plane and the Sgr dwarf stream to be 30° and the angle between the normal to the line of sight toward a point and the tangent of the stream at that point to be 10°. We calculated the former angle from our parameters by taking the angle between the observational plane of stripe 82 and the directional vector. Doing this, we get a value of $30^\circ \pm 3^\circ$. Our value agrees perfectly. The second angle was calculated by first finding the line-of-sight vector to the center of the stream, calculating a normal to that, and then finding the angle between the stream direction and that normal. We found a value of $22.5^\circ \pm 0.2^\circ$. This difference is explained by the drastically improved accuracy of the maximum likelihood algorithm over the estimate by eye for this more difficult angle. A tabulation of the stream center position and the stream direction in various coordinate systems can be found in Table 5. Note that the Cartesian center and Cartesian direction correspond to the vectors $\boldsymbol{c}$ and $\hat{\boldsymbol{a}}$, respectively, that are used in § 2.3.1.

TABLE 5
STREAM CENTER AND DIRECTION AS CALCULATED FOR DIFFERENT COORDINATE SYSTEMS

Physical Quantity	Coordinate System	Value 1	Value 2	Value 3
Center	Equatorial (α, δ, R)	31.37	0.0	29.22[1]
	Equatorial (α, δ, g_0)	31.36	0.0	21.53
	Galactic (l, b, R)	159.223	−57.558	29.22[1]
	Galactic (l, b, g_0)	159.223	−57.558	21.53
	Cartesian (X, Y, Z)	−23.154	5.560	−24.658
Direction	Cartesian (X, Y, Z)	−0.991	−0.042	0.127

NOTE.—All distances are in kpc and angles are in degrees.
[1] Assuming an absolute magnitude of $M_{g_0} = 4.2$.

Our results are in reasonable agreement with those of Newberg et al. (2003), but with much tighter error bars. The center of the stream is shifted from $\alpha = 33.99^\circ \pm 1^\circ$ to $31.37^\circ \pm 0.25^\circ$. This shift is slightly larger than expected, but not overly so. Note also that the Newberg et al. (2002) estimate of the center in F turnoff stars is $\alpha = 33^\circ$, which is closer to our value. Our detection of radial distance, however, corresponds exactly to that seen in the Newberg et al. (2003). Our stream width of $\sigma = 2.86$ is equivalent to a FWHM of 6.7 kpc, which is in good agreement with the value of 6 kpc estimated in Freese et al. (2005).

The stream ratio ϵ provides a very good separation as shown in Figure 7 and corresponds to a detection of approximately 16,000 stream stars out of the total 115,907 F turnoff stars in the stripe. By performing a calculation paralleling equation (10) of Freese et al. (2005), we are able to get an estimate of the total stellar mass in stripe 82 as a fraction of the Sgr dwarf itself. We first calculate the number of stars in Figure 6 of Newberg et al. (2002) with $0.1 < (g-r)_0 < 0.3$ and $16.0 < g_0 < 22.5$. This was calculated to be 2194 stars. We then use this in place of the value calculated for F/G-type stars in equation (10) of Freese et al. (2005) to get the total number of F turnoff stars currently in the Sgr dwarf itself. This results in a total of 1,798,700 F turnoff stars in the Sgr dwarf. Dividing this by the 16,050 stars found to be in the stream in stripe 82 yields that stripe 82 contains 0.9% of the Sgr dwarf's current F turnoff stars.

Our spheroid parameters of $q = 0.46$ and $r_0 = 19.4$ are slightly unexpected given the previous work in Bell et al. (2008), which finds that the flattening parameter for the stellar halo is closer to $q = 0.6$. However, this does not discount that the best likelihood for the stellar halo is obtained from these results, only that they were not the expected ones. Further results from other data sets and other stripes will be needed to determine the consistency of these results among other stripes and confirm such a strong flattening of the stellar halo.

5. CONCLUSIONS

We have measured the position, direction, distribution, width, and number of F turnoff stars of the Sgr tidal debris in SDSS stripe 82 as well as the stellar spheroid parameters corresponding to flattening and core radius. Our work has resulted in the improved measurement of the Sgr stream width (FWHM $= 6.74 \pm 0.06$ kpc), position $[(\alpha, \delta, R) = (31.37^\circ \pm 0.26, 0.0^\circ, 29.22 \pm 0.20$ kpc)], and direction $[(X, Y, Z)_{\rm gal} = (-0.991 \pm 0.007$ kpc, -0.042 ± 0.033 kpc, 0.127 ± 0.046 kpc)]. We measured the

number of F turnoff stars in the stream in SDSS stripe 82 (13.86% ± 0.06% of 115,907). There are 16,050 Sgr F turnoff stars in stripe 82 compared to 1.8×10^6 turnoff stars in the Sgr dwarf itself (0.9%). We also calculated values for the stellar spheroid parameters: the flattening parameter ($q = 0.46 \pm 0.024$) and the core radius ($r_0 = 19.40 \pm 0.59$). While these spheroid parameters are not the expected values, they are intriguing and deserve further study. Finally, we were able to generate separate catalogs of stream stars and spheroid stars for the F turnoff stars in stripe 82. These catalogs, while not a true representation of the actual stars in the stream or spheroid, provide the correct density characteristics of these structures and can be compared with the density profiles of *N*-body simulations.

The results were calculated using a new method to study tidal debris that is more robust than previous methods. This method will prove valuable to the study of tidal disruption and Galactic structure, as it provides a means to accurately and efficiently study many variables used to model the spheroid and tidal debris. Also, we are able to extract a catalog of stream stars which can be used to better constrain models of the Sgr dwarf disruption by providing actual stream stellar positions to compare with the data.

The algorithm itself is quite flexible. The stream and spheroid models can be easily changed should more accurate ones be discovered. The new model definitions would simply replace the existing probability code with no other changes. Also, with minor changes to the volume definitions it would be possible to use data from surveys other than the SDSS as an input catalog. With some revisions, the code could be adapted to search for multiple debris and use multiple volumes to provide more accurate results.

This publication is based on work supported by the National Science Foundation under grant SEI(AST) 06-12213. Additional support was provided by the NSF grants AST 06-07618 and CNS 0323324, NASA New York Space Grant, and John A. Huberty. We thank Brian Yanny for his help in the extraction of data and overall troubleshooting he did to aid us in this work.

We use data from the Sloan Digital Sky Survey. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

APPENDIX A

OPTIMIZATION TECHNIQUE

A1. CONJUGATE GRADIENT

The conjugate gradient method is similar to the gradient method in that we perturb the current parameter value, Q_i, by a small amount, h_i, and get a direction for the change in the parameter space. However, unlike the gradient method, the conjugate gradient method enforces the condition that each direction is conjugate to the previous direction. This means that as we move along the new search direction, the component of the gradient parallel to the previous search direction must remain zero. This condition speeds up the optimization by using search directions that are noninterfering. A further discussion may be found in Fletcher (1987).

The ith component of the gradient vector, $\boldsymbol{G}$, is calculated numerically by perturbing Q_i, holding all other Q_j fixed,

$$G_i = \left.\frac{\mathcal{L}(Q_i + h_i) - \mathcal{L}(Q_i - h_i)}{2h_i}\right|_{\text{all other } Q_j \text{ fixed}}, \tag{A1}$$

where G_i is the ith component of the gradient vector, Q_i is the ith parameter, h_i is the perturbation amount for that parameter, and $\mathcal{L}$ is the likelihood function from § 2.2. This gradient is used to calculate the direction, $\boldsymbol{D}$, for the first iteration,

$$\boldsymbol{D}_1 = \boldsymbol{G}_1. \tag{A2}$$

We take the positive value of the gradient, for we want to maximize the likelihood. Should we want to perform a simple gradient ascent we would do this same calculation for every iteration and use solely the direction calculated above; however, we wish to apply the conjugate gradient technique which always calculates a direction conjugate to the previous directions. This is done by calculating a multiplier, B_i, based on the current gradient and the previous gradient, as

$$B_i = \frac{\boldsymbol{G}_i \cdot (\boldsymbol{G}_i - \boldsymbol{G}_{i-1})}{\boldsymbol{G}_{i-1}^2}, \tag{A3}$$

where $\boldsymbol{G}_i$ denotes the gradient vector for the ith iteration. This value is then used to calculate the new conjugate gradient direction,

$$\boldsymbol{D}_i = \boldsymbol{G}_i + B_i * \boldsymbol{D}_{i-1}, \tag{A4}$$

where $\boldsymbol{D}_i$ denotes the direction vector of the ith iteration (the direction to move the current parameters) and $\boldsymbol{G}_i$ is the gradient of the current parameters, Q_i, at the ith iteration.

Because of the large variety of parameters that we utilize, all on different scales, we do not use one value for our perturbation value, h, for it could potentially produce poor estimates for the gradient if the perturbation is too large or small for a given parameter. If h is too large, then the computed gradient is not an accurate approximation of the true gradient at the current point; if h is too small, then we may encounter computational errors. We therefore choose an appropriate value for each parameter independently. Table 1 shows the perturbation values for each parameter. These numbers were chosen in two ways. The values of the gradient were calculated at a number of differing perturbation values in order to determine how much the gradient changed. Also, plots of all parameter spaces were generated in order to determine the distribution of each parameter. Combining these two techniques, we were able to determine a value for the perturbation for each parameter that would be small enough to provide a very accurate gradient calculation, but also large enough to overcome any anomalous behavior within the parameter's distribution.

A2. LINE SEARCH

The line search technique seeks to determine the value, α, that minimizes the function

$$\Phi(\alpha) = \mathcal{L}(\boldsymbol{Q}_k + \alpha \boldsymbol{D}_k), \tag{A5}$$

where $\boldsymbol{D}_k$ is the direction and $\boldsymbol{Q}_k$ is the current set of parameters. After finding α^*, which minimizes $\Phi(\alpha)$, we then update

$$\boldsymbol{Q}_{k+1} \leftarrow\!\!- \boldsymbol{Q}_k + \alpha^* \boldsymbol{D}_k. \tag{A6}$$

A bracketing method must be employed first to ensure that the minimum of the function is within the range we are searching. This is done by calculating the likelihood at three points along the current search direction: the current parameter values plus 0, 1, and 2 times $\boldsymbol{D}_k$. This corresponds to the current position, the current position plus the full direction, and the current position plus twice the full direction. If the middle point does not have a likelihood greater than the two endpoints, the endpoint with the highest probability becomes the new center point, the center point the end, and a new third point is calculated by expanding to 2 times the current factor times the direction and the new likelihood is calculated. Iteration continues until the midpoint has a greater likelihood than both endpoints.

These three points are then passed to the line search algorithm which will use them to find the peak of the parabolic function defined by moving along the current directional vector and calculating the likelihood at points along this path. The line search iterates as follows. First, a guess is made for the value of α by fitting a parabola via the calculation

$$\begin{aligned} u &= a_1^2 * (L_2 - L_3) + a_2^2 * (L_3 - L_1) + a_3^2 * (L_1 - L_2), \\ b &= a_1 * (L_2 - L_3) + a_2 * (L_3 - L_1) + a_3 * (L_1 - L_2), \\ \alpha &= 0.5 * (u/b), \end{aligned} \tag{A7}$$

where a_1, a_2, and a_3 are the factors multiplied by the current direction that produce the parameters with likelihood L_1, L_2, and L_3, respectively. For the first iteration, a_1, a_2, and a_3 are the values returned from the bracketing method. The quantity α is then the current guess for the value to minimize equation (A5). The likelihood L_α is then calculated using the parameters generated using α. The new order of the three points is

$$\begin{aligned} &\text{if } \alpha > a_2 \text{ and } L_\alpha > L_2, \text{ then: } 1, 2, \alpha, \\ &\text{if } \alpha > a_2 \text{ and } L_\alpha < L_2, \text{ then: } 2, \alpha, 1, \\ &\text{if } \alpha < a_2 \text{ and } L_\alpha > L_2, \text{ then: } \alpha, 2, 3, \\ &\text{if } \alpha < a_2 \text{ and } L_\alpha < L_2, \text{ then: } 1, \alpha, 2. \end{aligned} \tag{A8}$$

In short, the distance along the search direction is reduced based on the value of α and its corresponding likelihood. These new points are then used to calculate a new guess for α using equations (A6)–(A8). Iteration continues for a set number of iterations until returning the optimum value of α; we currently use three.

Once we have calculated the change in the parameters, utilizing the conjugate gradient and line search methods, we then update the value of the probability for the fit using the same method as discussed in § 2.3 and continue to maximize our probability through the use of the conjugate gradient and line search methods. The algorithm stops iterating and returns the current values when the value of the largest component of the conjugate gradient drop below a set threshold, the line search returns a value that would cause negligible movement, or the algorithm has completed a maximum number of iterations.

REFERENCES

Abazajian, K., et al. 2003, AJ, 126, 2081
Adelman-McCarthy, J. K., et al. 2008, ApJS, 175, 297
Anderson, D. P., Korpela, E., & Walton, R. 2005, in First Int. Conf. on e-Science and Grid Computing, ed. H. Stockinger, R. Buyya, & R. H. Perrott (Los Alamitos: IEEE Computer Society), 196
Bahcall, J. 1986, ARA&A, 24, 577
Bell, E. F., et al. 2008, ApJ, 680, 295
Bellazzini, M., Ibata, R., Ferraro, F. R., & Testa, V. 2003, A&A, 405, 577
Belokurov, V., Evans, N. W., Irwin, M. J., Hewett, P. C., & Wilkinson, M. I. 2006a, ApJ, 637, L29
Belokurov, V., et al. 2006b, ApJ, 642, L137
———. 2006c, ApJ, 647, L111

Belokurov, V., et al. 2007a, ApJ, 654, 897
———. 2007b, ApJ, 657, L89
Chou, M.-Y., et al. 2007, ApJ, 670, 346
Desell, T., Cole, N., Magdon-Ismail, M., Newberg, H., Szymanski, B., Varela, C. 2008, in Third Int. Conf. on e-Science and Grid Computing, in press
Duffau, S., et al. 2006, ApJ, 636, L97
Fellhauer, M., et al. 2006, ApJ, 651, 167
Fletcher, R. 1987, Practical Methods of Optimization (2nd. ed.; New York: Wiley-Interscience)
Freeman, K. C. 1987, ARA&A, 25, 603
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Freese, K., Gondolo, P., & Newberg, H. J. 2005, Phys. Rev. D, 71, 043516
Fukugita, M., Ichikawa,T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
Gilmore, G., Wyse, R. F. G., & Kuijken, K. 1989, ARA&A, 27, 555
Gómez-Flechoso, M. A., Fux, R., & Martinet, L. 1999, A&A, 347, 77
Grillmair, C. J. 2006, ApJ, 645, L37
Grillmair, C. J., & Dionatos, O. 2006a, ApJ, 641, L37
———. 2006b, ApJ, 643, L17
Grillmair, C. J., & Johnson, R. 2006, ApJ, 639, L17
Gunn, J. E., et al. 1998, AJ, 116, 3040
Heath, M. T. 2002, Scientific Computing (2nd. ed.; New York: McGraw-Hill)
Helmi, A. 2004, ApJ, 610, L97
Helmi, A., & White, S. D. M. 2001, MNRAS, 323, 529
Hernquist, L. 1990, ApJ, 356, 359
Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, AJ, 122, 2129
Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, Nature, 370, 194
Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001a, ApJ, 547, L133
Ibata, R. A., & Lewis, G. F. 1998, ApJ, 500, 575
Ibata, R. A., et al. 1997, AJ, 113, 634
———. 2001b, ApJ, 551, 294
Irwin, M. J., et al. 2007, ApJ, 656, L13
Johnston, K. V., Law, D. R., & Majewski, S. R. 2005, ApJ, 619, 800
Johnston, K. V., Spergel, D. N., & Hadyn, C. 2002, ApJ, 570, 656
Johnston, K. V., Spergel, D. N., & Hernquist, L. 1995, ApJ, 451, 598
Law, D. R., Johnston, K. V., & Majewkski, S. R. 2005, ApJ, 619, 807
Law, D. R., et al. 2004, in ASP Conf. Ser. 327, Satellites and Tidal Streams, ed. F. Prada, D. Martinez Delgado, & T. J. Mahoney (San Francisco: ASP), 239
Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, ApJ, 599, 1082
Majewski, S. R., et al. 2004, AJ, 128, 245
Martínez-Delgado, D., Gómez-Flechoso, M. A., Aparicio, A., & Carrera, R. 2004, ApJ, 601, 242
Newberg, H. J., & Yanny, B. 2005, in ASP Conf. Ser. 338, Astrometry in the Age of the Next Generation of Large Telescopes, ed. P. K. Seidelmann & A. K. B. Monet (San Francisco: ASP), 210
———. 2006, J. Phys.: Conf. Ser., 47, 195
Newberg, H., et al. 2002, ApJ, 569, 245
———. 2003, ApJ, 596, L191
———. 2007, ApJ, 668, 221
Odenkirchen, M., et al. 2001, ApJ, 548, L165
Peñarrubia, J., et al. 2005, ApJ, 626, 128
Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Z. 2003, AJ, 125, 1559
Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523
Rocha-Pinto, H. J., Majewski, S. R., Skrutskie, M. F., Crane, J. D., & Patterson, R. J. 2004, ApJ, 615, 732
Rockosi, C. M., et al. 2002, AJ, 124, 349
Savage, C., Newberg, H. J., Freese, K., & Gondolo, P. 2006, J. Cosmol. Astropart. Phys., 7, 3
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Smith, J. A., et al. 2002, AJ, 123, 2121
Stoughton, C., et al. 2002, AJ, 123, 485
Teuben, P. J. 1995, in ASP Conf. Ser. 77, in Astronomical Data Analysis Software and Systems IV, ed. R. Shaw, H. E. Payne & J. J. E. Hayes (San Francisco: ASP), 398
Vivas, A. K., et al. 2001, ApJ, 554, L33
Willman, B., et al. 2005, ApJ, 626, L85
Xu, Y., Deng, L. C., & Hu, J. Y. 2006, MNRAS, 368, 1811
Yanny, B., et al. 2000, ApJ, 540, 825
———. 2003, ApJ, 588, 824
York, D. G., et al. 2000, AJ, 120, 1579
Zucker, D. B., et al. 2006a, ApJ, 643, L103
———. 2006b, ApJ, 650, L41