A Fully Self-Consistent Constraint on the Mass of M31 & the Local Group

by

Daniel Thomas Foreman-Mackey

A thesis submitted to the

Department of Physics, Engineering Physics & Astronomy

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

August 2010

Abstract

We present the first fully self-consistent, axisymmetric, dynamical model of the Andromeda galaxy (M31). We constrain the physical parameters of the model with datasets on all radial scales: the bulge projected velocity dispersion, rotation curve, surface brightness profile, and the kinematics of globular clusters and satellite galaxies. Combining these highly heterogeneous datasets into a single self-consistent analysis is natural in the framework of Bayesian inference. Using a geometric argument, we also infer the three-dimensional velocity of M31 relative to the Milky Way. From this orbit, we constrain the total mass of the Local Group by the "timing argument". We find that the virial mass of M31 is $M_{\mathrm{M31,vir}} = 5.0^{+2.2}_{-1.7} \times 10^{12} \, M_{\odot}$ and the mass of the Local Group is $M_{\mathrm{LG}} = 8.8^{+8.0}_{-4.2} \times 10^{12} \, M_{\odot}$. We conclude that the large uncertainties in our results are due primarily to the small sample size at large radii and that either a significantly larger sample or unjustifiably informative priors are necessary to improve the constraint.

Acknowledgments

First, I would like to thank my supervisors Dave Hanes and Larry Widrow for their patience, encouragement and support during the past two years. I would specifically like to thank Dave for tolerating — and hopefully even accepting — my move into the world of theoretical astrophysics and Larry for sharing his enthusiasm with me. I am also thankful for conversations with Stéphane Courteau and Terry Bridges about science and life in general. If it wasn't for Stéphane's encouragement and his (largely unfounded) faith in my abilities, I probably wouldn't even be writing this. For that, I am forever grateful. I would also like to thank my colleagues for tea time, lake lunches and Friday grad club trips.

I acknowledge the use of the Canada Foundation for Innovation funded CITA Sunnyvale cluster and John Dubinski for providing access to this resource.

Finally, I am extremely grateful to Elaine, Clarke, Annie and Laura for being the most wonderful family in the world and providing me with love and support that I can always count on. I would especially like to thank Laura for letting me into her life, for cooking me dinner when I've been at school too long and for always being my best friend.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Andromeda galaxy (hereafter M31) is the closest disk galaxy to the Milky Way at a distance of 785 kpc or $\sim 2.6 \times 10^6$ light-years (McConnachie & Irwin 2006). The disk of M31 spans about 2° on the sky but the full Andromeda system (including the associated satellite galaxies and star clusters) stretches nearly 15°. M31 forms a gravitationally bound system with the Milky Way in an association called the Local Group. The Milky Way and M31 are the most luminous and most massive members of the Local Group by a factor of ten, so they are assumed to dominate the gravitational potential of the system. This assumption is supported by the fact that in the Milky Way's rest frame, M31 is approaching at ~ 130 km s^{-1}. By studying the physical properties of M31, we gain important insight into the formation and future evolution of the Local Group and the distribution of dark matter. Furthermore, on the larger scale, studies of galaxy dynamics can improve our understanding of distant systems that are impossible to observe with the resolution we can obtain in M31.

The literature associated with the mass modelling of M31 is extensive but as the sample of dynamical tracers continues to grow, more sophisticated approaches

are possible and necessary. Bayesian inference in the context of galaxy dynamics was first introduced by Little & Tremaine (1987) to model the Milky Way using the kinematics of satellite galaxies and an analytic spherical mass estimator. Evans & Wilkinson (2000) performed the first Bayesian analysis of M31 based on the kinematics of satellite galaxies, globular clusters and planetary nebulae.

In each case, these authors also approximated the galaxy as a spherical, isolated dark matter halo. By ignoring the disk and bulge, Evans & Wilkinson (2000) cannot produce an astrophysically relevant rotation curve or any surface brightness profile, the two best established and most powerful galactic scale mass constraints that we have. Accepting a large computational cost, we employ a state-of-the-art semi-analytic, multi-component galaxy model to fit all the available data sets. This is the first self-consistent constraint of M31's mass distribution using the full suite of kinematic observations and this method will continue to be applicable as the datasets continue to grow.

Evans & Wilkinson (2000) made the additional assumption that the orbit of M31 with respect to the Milky Way is purely radial. A complementary analysis by van der Marel & Guhathakurta (2008) showed that this assumption is unnecessary. Using only the line-of-sight velocities of satellites at large angular separation, van der Marel & Guhathakurta (2008) inferred the systematic transverse velocity of M31 on the plane of the sky. These authors modelled the peculiar velocities of the satellites as isotropic and isothermal which implicitly requires that the satellites feel no gravitational effect from the galaxy whatsoever. van der Marel & Guhathakurta (2008) claimed that their results were insensitive to this assumption but the same geometric argument can be applied to our model to produce a simultaneous, self-consistent constraint on

both the mass of M31 and its transverse velocity.

By inferring the full three-dimensional velocity of M31 relative to the Milky Way, we also constrain the total mass of the Local Group by the so-called "timing argument" (e.g. Kahn & Woltjer 1959; Einasto & Lynden-Bell 1982; Kroeker & Carlberg 1991; Kochanek 1996; Li & White 2008; van der Marel & Guhathakurta 2008). Using this constraint, we can infer the mass ratio of the Milky Way to M31.

Our galaxy model is a generalization of the model described by Widrow et al. (2008). It is constructed from a multi-component, semi-analytic phase space distribution function that is tuned to produce a generalized cuspy double power-law (Dubinski & Carlberg 1991; Navarro et al. 1997) dark matter halo distribution, a de-projected Sérsic bulge and an exponential thin disk. We use Bayes' theorem to constrain the model parameters of interest based on five datasets: satellite galaxy kinematics, outer globular cluster kinematics, the rotation curve, the central velocity dispersion and the surface brightness profile.

After summarizing the literature that has led to this work in chapter 2, we describe our sample of observational constraints from the literature in chapter 3. In chapter 4, we review several basic results from the field of galaxy dynamics and outline our galaxy model in this context. Next, we describe the Markov chain Monte Carlo algorithm that we employ to simultaneously calculate the probabilistic distribution of physical parameters that satisfy the data and draw samples from that distribution. Finally, in chapters 6 and 7, we present our results and discuss their implications in a cosmological context.

Chapter 2

Background & Previous Work

It has been known for some time that either there is more mass in galaxies than can be accounted for by the luminous matter (e.g. Zwicky 1937; Rubin & Ford 1970) or our current laws of physics are flawed on the scales of galaxies or galaxy clusters (e.g. Milgrom 1983; Sanders & McGaugh 2002). A discussion of alternate theories of gravity is beyond the scope of this work so in what follows, we will accept the currently accepted paradigm of cold dark matter (e.g. Blumenthal et al. 1984) as given.

In this paradigm, after the big bang, the Universe was nearly isotropic but small energy density fluctuations induced gravitational collapse and clumping of dark matter particles. These clumps continued to grow into a filamentary structure until galaxies formed in the densest regions. The overdensity of dark matter where a galaxy formed is known as the dark matter halo and understanding the shape of these halos is still a very active area of research. For our purposes, however, it is sufficient to approximate them as roughly spherically symmetric. The baryonic matter in these halos flattened into a thin disk and centrally concentrated bulge as a

result of gravitationally induced angular momentum. The high density of gas and dust in the disk and bulge induced star formation in these regions. Similar processes produce dwarf galaxies in smaller halos embedded in the galactic halo to produce satellite galaxies that are gravitationally bound to the main galaxy. In the galactic halo, there is also a collection of star clusters (called globular clusters) that form at early times — although the formation mechanism is not fully understood (Brodie & Strader 2006).

At this point, there is no method of detecting dark matter directly so we must rely on the observable motions of stars and gas to inform our knowledge of dark matter. The rotational speed in the disk and the bulge constrain the mass distribution towards the centre of the galaxy using only simple Newtonian dynamics

$$v^2 = \frac{GM}{r} \tag{2.1}$$

where v is the circular speed of the stars in the disk, r is radius, M is the mass contained within r and G is Newton's gravitational constant. At larger radii, we can study the motions of satellite galaxies and globular clusters to constrain the mass distribution in a similar way but the dataset is extremely sparse. Plagued by the limited phase space information, heterogeneous datasets with large uncertainties and the unknown physics associated with dark matter itself, conclusive modelling of the phase space distribution of dark halos at large radii is impossible without significant simplifying assumptions. As fewer simplifications are made, inferring this distribution becomes quickly intractable.

There are two complementary methods of studying dark matter and galaxy formation theoretically: ab initio numerical simulations and isolated equilibrium modelling.

High-resolution galaxy simulations are, however, extremely computationally expensive. Therefore, it is not possible to fine-tune the parameters of a simulation to match observational constraints in a statistically rigorous manner. In contrast, equilibrium galaxy models based on smoothed density distributions can be fit to observations efficiently and self-consistently. In particular, the distribution function based analysis that we develop in this thesis generates equilibrium models that can then be used as the initial conditions for numerical simulations of galaxy evolution.

2.1   Classical Mass Estimators

In many extragalactic systems, the simplest method of constraining the dark mass within a given radius comes from the Jeans equation in the form of a projected mass estimator (PME)

$$M \propto \frac{\langle v^2 R \rangle}{G} \tag{2.2}$$

(Bahcall & Tremaine 1981) where the brackets indicate an ensemble average over a system of dynamical tracers, v has units of speed – being either velocity or velocity dispersion (Wolf et al. 2010) – and R is the distance from the centre of the galaxy in three dimensions or in projection (Watkins et al. 2010). The PME can be applied to systems of globular clusters (GCs), satellite galaxies, planetary nebulae or any other homogeneous population that traces the DM gravitational potential.

Little & Tremaine (1987) improved on the standard PME by using Bayesian inference to constrain the mass distribution. The motivation for these authors was the fact that with a small sample size, the propagation of uncertainties from measurements to confidence intervals on inferred quantities is extremely important, yet there

is no rigorous method of doing this in a standard frequentist statistical framework.
Little & Tremaine (1987) considered an analytic distribution function (based on the
PME) for a system of satellite galaxies orbiting a point mass with isotropic velocities.
With a sample of 15 satellite galaxies, Little & Tremaine (1987) estimated the mass
of the Milky Way as $2.4^{+1.3}_{-0.7} \times 10^{11} \, M_\odot$ and concluded that the dark halo extended no
further than ~ 50 kpc from the centre of the Galaxy. Both of these results seem to
be significant under-estimates but this was nonetheless a pivotal paper in the history
of Local Group dynamics because it emphasized the need for rigorous probabilistic
inference.

Bayesian inference provides a natural framework for the self-consistent combined
analysis of heterogenous datasets. In the second Bayesian study of the Milky Way,
Kochanek (1996) exploited this characteristic and inferred the mass of the Galaxy
using satellite kinematics, the local escape velocity, the rotation curve and the Local
Group timing. Kochanek (1996) constructed a dark halo model based on the analytic
but astrophysically inspired Jaffe (1983) density distribution

$$\rho_{\mathrm{J}} = \frac{v_c^2}{4\pi G} \frac{r_j^2}{r^2(r_j + r)^2} \tag{2.3}$$

where r_j is the so-called Jaffe radius. This model produces a flat rotation curve for
$r \lesssim r_j$ with rotational speed v_c. Using this model, Kochanek (1996) estimated the
mass of the Milky Way within 50 kpc as $(4.9 \pm 1.1) \times 10^{11} \, M_\odot$.

With a larger dataset of satellite galaxies and a different analytic phase space
distribution function, Wilkinson & Evans (1999) estimated the total mass of the Milky
Way as $19^{+36}_{-17} \times 10^{11} \, M_\odot$. The spherically symmetric density distribution considered

by Wilkinson & Evans (1999) is called the "truncated, flat rotation curve" (TF) model

$$\rho_{\text{TF}} = \frac{M}{4\pi} \frac{a^2}{r^2(r^2 + a^2)^{3/2}} \tag{2.4}$$

where M is the total mass and a is the radial scale. The TF model has the advantageous property that the self-consistent[1] gravitational potential generated by equation 2.4 is analytic, namely

$$\psi_{\text{TF}} = \frac{GM}{a} \log\left(\frac{\sqrt{r^2 + a^2} + a}{r}\right). \tag{2.5}$$

Furthermore, equation 2.4 can be rewritten as a function of ψ_{TF} and hence the phase space distribution function of dynamical tracers can be generated analytically.

Evans & Wilkinson (2000) were the first to model the Andromeda galaxy (M31) using dynamical tracers and Bayesian inference using the same TF model as the previously mentioned paper by the same authors. They expanded their analysis to include prescriptions for globular clusters and planetary nebulae and estimated the total mass of M31 as $12.3^{+18}_{-6} \times 10^{11} \, M_\odot$. When compared to the mass of the Milky Way inferred using the same method by Wilkinson & Evans (1999), this result suggested that the Milky Way could be the more massive member of the Local Group in contrast to estimates based on surface brightness and globular cluster specific frequency.

The problem with both equations 2.3 and 2.4 is that cosmological dark matter simulations suggest that the density of DM halos falls off much less steeply than either $\rho_{\text{J}} \propto r^{-4}$ or $\rho_{\text{TF}} \propto r^{-5}$. Also, the density distribution towards the centre of the galaxy is probably much less "cuspy" with $\rho \propto r^{-1}$ instead of $\rho_{\text{J,TF}} \propto r^{-2}$. Both the Jaffe

---

[1]The self-consistent potential ψ generated by a density distribution ρ is one which satisfies Poisson's equation $\nabla^2 \psi = 4\pi G\rho$.

and TF models are much too restrictive and the estimated mass is very sensitive to these effects. A preferred empirical density law is the Navarro, Frenk, & White (1997, NFW) model

$$\rho_{\text{NFW}} = \frac{v_0^2}{4\pi G a^2} \frac{1}{(r/a)(1 + r/a)^2} \tag{2.6}$$

or even more generally,

$$\rho = \frac{\rho_0}{(r/a)^\gamma (1 + r/a)^{\alpha - \gamma}} \tag{2.7}$$

(Binney & Tremaine 2008). Equation 2.7 is a generalization of both the Jaffe and NFW models and it can closely approximate the TF model.

2.2 Multi-component Axisymmetric Galaxy Models

Up to this point, each galaxy model that we have mentioned is a spherically symmetric dark matter halo in isolation. Of course, the true distribution of matter in galaxies is much more complicated than this simple assumption with flattened or triaxial halos, complex streams and other dark matter substructure (e.g. Ibata et al. 2001, 2007; Belokurov et al. 2006). The details of studies considering this substructure are beyond the scope of this work but a necessary improvement on the isolated spherical halo model is to include the visible matter. The kinematic distribution of baryonic matter (stars and gas) in galaxies is much easier to measure than dark matter. As a result, the mass distribution near the galactic centre is quite well constrained by the rotation curve, surface brightness profile and velocity dispersion profile. Unfortunately, no spherical halo can reproduce the observed rotation curves in isolation. In order to fit

rotation profiles, therefore, we must consider a more sophisticated model that breaks spherical symmetry.

An idealized schematic of an axisymmetric disk galaxy is shown in figure 2.1. It is composed of a centrally concentrated bulge, a thin disk and and an extended dark matter halo. To construct the phase space distribution function of a tracer particle in the potential of a multi-component model galaxy of this type, Kuijken & Dubinski (1995) extended the disk distribution function from Shu (1969) to three dimensions and combined it with a lowered Evans halo model and a King (1966) bulge model. The resulting multi-component distribution function was a function of only three integrals of motion: energy, angular momentum about the symmetry axis and vertical energy of the disk stars which is approximately conserved in cool disks. The advantage of the Kuijken & Dubinski (1995) method is that they develop an analytic model for the higher order terms in the spherical harmonic expansion of the disk density. As a result, they could solve Poisson's equation and generate a fully self-consistent galaxy model for the nearly spherical residual potential without a restrictively large number of spherical harmonics.

Widrow et al. (2003) applied the Kuijken & Dubinski (1995) model to M31 using a maximum likelihood method to simultaneously fit the rotation curve, surface brightness profile and bulge velocity dispersion profile. From this analysis, they estimated the mass of M31 within 200 kpc as $\sim 10^{12}\,M_\odot$ without any constraint on the propagated uncertainty. Subsequently, using similar observational constraints and a χ^2-statistic, Widrow & Dubinski (2005) modelled both M31 and the Milky Way. Widrow & Dubinski (2005) also adapted the galaxy model to be composed of more realistic density distributions. For the bulge, they adopted a Hernquist (1990) model
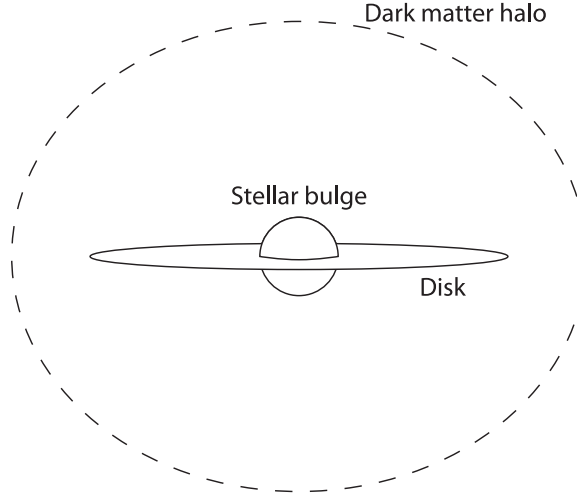
Figure 2.1 Schematic of a typical disk galaxy. Not drawn to scale.

and for the halo, they chose the NFW profile (equation 2.6).

In order to obtain an objective, probabilistic constraint on the model parameters, Widrow et al. (2008) further modified the Kuijken & Dubinski (1995) model and modelled the Milky Way using Bayesian inference and Markov chain Monte Carlo to evaluate the marginalized posterior probability distribution of parameters satisfying a suite of observational constraints. Widrow et al. (2008) generalized the halo model to be equation 2.7 with $\alpha = 3$ and introduced a deprojected Sérsic law (Prugniel & Simien 1997; Terzić & Graham 2005) for the bulge. With this 15 parameter model, Widrow et al. (2008) estimated the mass of Milky Way within 100 kpc as $4.0^{+2.2}_{-1.9} \times 10^{11} \, M_\odot$.

The Widrow et al. (2008) model is especially useful for modelling galaxies using Bayesian inference because it is so general. Recently, Puglielli et al. (2010) used Markov chain Monte Carlo to constrain the mass distribution in NGC 6503 using this model and in this work we present a statistically robust study of M31 and we rely heavily on the generality of this galaxy model.

2.3   The Mass of Andromeda

Besides the aforementioned studies of the DM distribution in M31, there is a plethora of other estimates based on various permutations of observational constraints. Table 2.1 lists a selection of these estimates and figure 2.2 shows the same values plotted as a function of radius. Many of the studies listed in table 2.1 assumed different distances to M31 so we must correct for this effect in order to compare the results. Following Evans & Wilkinson (2000), we correct the results to our mean assumed distance (785 kpc) using

$$\Delta M = \frac{\Delta D}{D} M \tag{2.8}$$

where ΔM and ΔD are the changes in mass and distance respectively and M and D are the mass and distance from the literature.

Figure 2.2 also shows the uncertainty on the mass profile from Evans & Wilkinson (2000). The TF profile (equation 2.4) generates the mass profile

$$M(< r) = M \frac{r}{\sqrt{r^2 + a^2}} \tag{2.9}$$

when integrated in spherical coordinates. The preferred model from Evans & Wilkinson (2000) has $M \sim 12.5 \times 10^{11} \, M_\odot$ and $a \sim 93$ kpc.

Klypin et al. (2002) proposed two theoretical multi-component mass models of M31 fit to the data using "χ-by-eye". The Klypin et al. (2002) model was composed of an NFW halo, a central black hole and a baryonic component — roughly an exponential disk and triaxial bulge. The mass distributions of these two models are shown in figure 2.2 as the dotted and dashed lines.

Table 2.1. Mass estimates for M31 from the literature

Source	Tracer	r (kpc)	$M(<r)$ $(10^{11}\,M_\odot)$
Rubin & Ford (1970)	Rotation curve	27	2.1 ± 0.1
Gottesman & Davies (1970)	Rotation curve	34	~ 2.5
Hartwick & Sargent (1974)	Globular clusters	20	4.0 ± 1.6
Deharveng & Pellet (1975)	Rotation curve	23	1.9 ± 0.2
Rood (1979)	Satellite galaxies	To IC 1613	2.9 ± 0.4
van den Bergh (1981)	Satellite galaxies	To LGS3	8.8 ± 4.6
	Globular clusters	7	1.1 ± 0.2
	Globular clusters	23	2.8 ± 1.4
Braun (1991)	Rotation curve	32	2.3 ± 0.1
Courteau & van den Bergh (1999)	Satellite galaxies	To LGS3	13.7 ± 1.9
Evans et al. (2000)	Satellite galaxies	To IC 1613	8.7 ± 1.5
Evans et al. (2003)	Globular clusters	102	~ 12.2
Lee et al. (2008)	Globular clusters	55	$5.5^{+0.4}_{-0.3}$
	Globular clusters	101	$19.3^{+1.4}_{-1.3}$
Chemin et al. (2009)	Rotation curve	38	4.7 ± 0.5
Watkins et al. (2010)	Satellite galaxies	100	2.1 ± 1
	Satellite galaxies	200	12.4 ± 3.8
	Satellite galaxies	300	14.0 ± 4.0
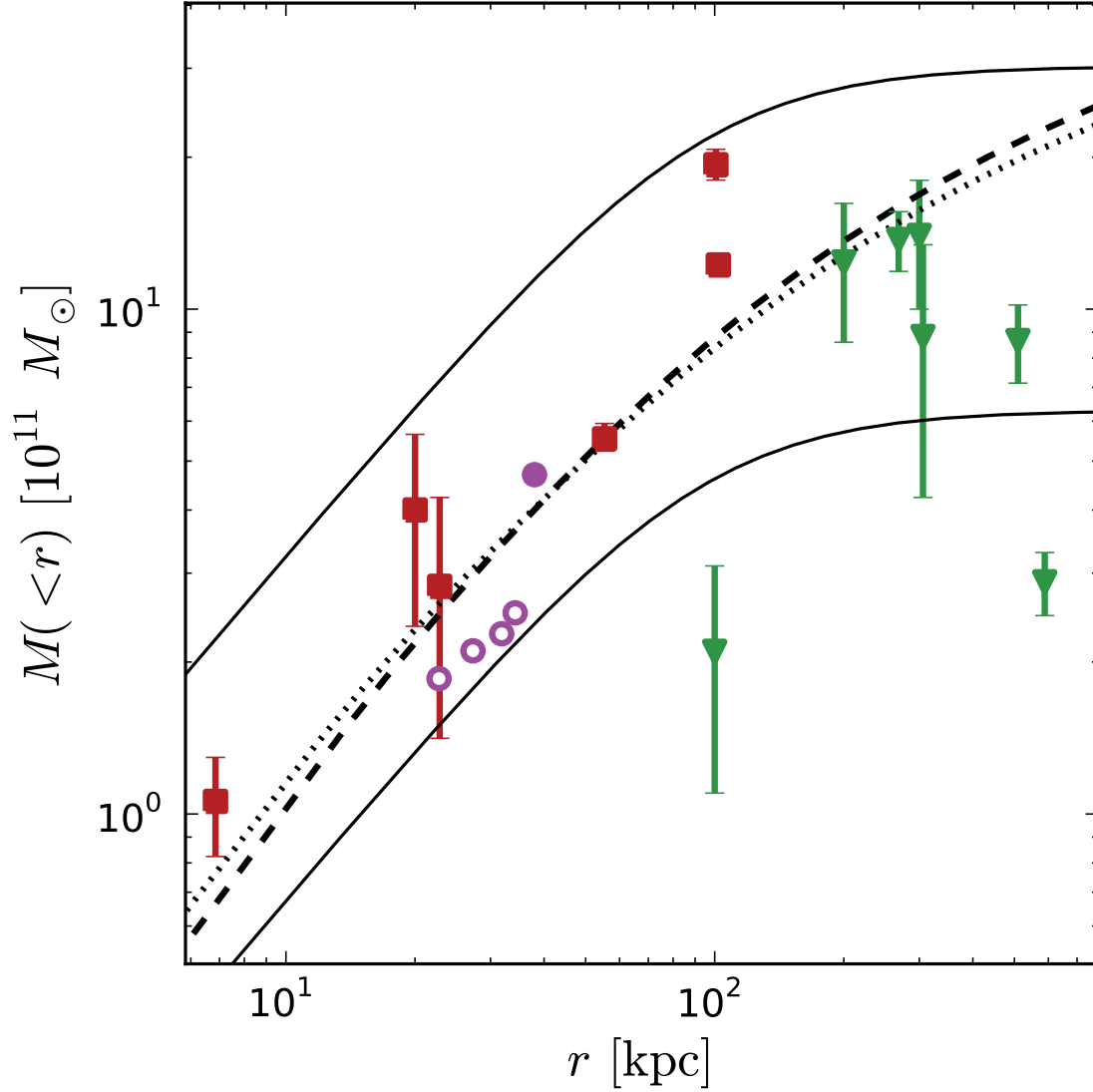
Figure 2.2 Mass estimates for M31 from the literature constrained using the rotation curve (violet points), globular clusters (red squares) and satellite galaxies (green triangles). See table 2.1 for the list of references. The filled violet point is from Chemin et al. (2009) using their measurement of the rotation curve and the open violet points previous estimates using older, low-resolution data. The black curves indicate the 1σ uncertainty on the mass profile from Evans & Wilkinson (2000). The dotted and dashed curves are the two theoretical models from Klypin et al. (2002).

2.4   The Orbit of the Local Group

Kahn & Woltjer (1959) were the first to argue that since M31 is approaching the Milky Way, the Local Group (LG) is a physical unit that decoupled from the Hubble flow at very early times. Under this assumption, the evolution of the LG has been governed by Newtonian dynamics since the big bang. With a constraint on the age of the universe, this argument (called the "timing argument") yields an estimate of the mass of the LG. Approximating the LG as two point masses with no angular momentum, Kahn & Woltjer (1959) set a lower limit of $M_{\mathrm{LG}} \geq 1.8 \times 10^{12}\,M_{\odot}$ on the LG mass which was much larger than the mass estimates of either galaxy at that time. Kahn & Woltjer (1959) suggested that there must be significant quantities of intergalactic matter in the LG and argued that ionized hydrogen could account for this discrepancy.

More recently, other authors (including Lynden-Bell 1981, 1999; Einasto & Lynden-Bell 1982; Raychaudhury & Lynden-Bell 1989; Kochanek 1996) applied the timing argument assuming a purely radial orbit. Since the MW and M31 are — by at least an order of magnitude — the most luminous galaxies in the LG, there are no objects massive enough to induce significant torques in the LG system. Therefore, it is reasonable to postulate a purely radial orbit. This is a powerful assumption because, while the line-of-sight velocity of extragalactic systems is easy to measure, the motion in the plane of the sky is currently unmeasurable for most objects outside of the solar neighbourhood.

To quantify the bias caused by the assumption of a purely radial orbit, Kroeker & Carlberg (1991) selected galaxy pairs with properties similar to the LG from an N-body cold dark matter (CDM) simulation and compared the true mass to the mass

calculated from the radial timing argument. Kroeker & Carlberg (1991) found that the radial timing argument underestimated the mass by, on average, a factor of ~ 2. If the full three-dimensional velocity information was available, however, Kroeker & Carlberg (1991) found that the timing argument was an unbiased estimator of the total mass, despite the significant simplifications involved.

In contrast, Li & White (2008) selected LG-like galaxy pairs from the Millennium Run simulation (Springel et al. 2005) and did not confirm the findings of Kroeker & Carlberg (1991). They found that the radial timing argument was an unbiased estimator of the total mass. The scatter on the estimated mass increased with transverse velocity but the bias did not.

While the cause of the discrepancy between the conclusions of Kroeker & Carlberg (1991) and Li & White (2008) is not clear, van der Marel & Guhathakurta (2008) presented an argument that made this assumption unnecessary. Making the weaker assumption that the system of satellite galaxies around M31 moves, on average, with the systematic motion of the galaxy, van der Marel & Guhathakurta (2008) demonstrated that a component of the systematic transverse velocity projects onto the line-of-sight velocity of the satellites with the amplitude $\sim v_\text{t} \sin \Phi_\text{S}$ where Φ_S is the angular separation of the satellite. We outline the geometry of this effect in detail is section 3.6 but figure 2.3 shows an intuitive sketch of the system. From this effect, they inferred the transverse velocity of M31. Using this constraint and the timing argument, they estimated the total mass of the LG as $5.58^{+0.85}_{-0.72} \times 10^{12} \, M_\odot$.
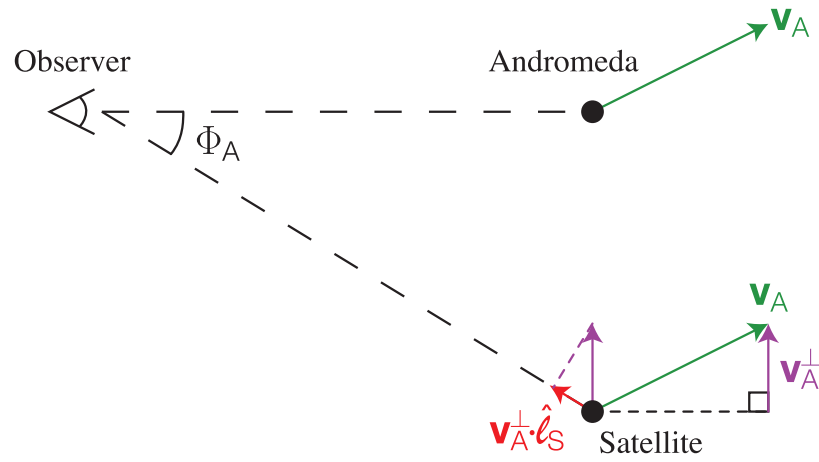
Figure 2.3 The systematic transverse velocity of M31 (purple vector) projects onto the line-of-sight velocity vector of the satellite galaxies (red vector) under the assumption that the satellite system moves, on average, with the systematic motion (green vectors). The effect is strongest for satellites at large radii.

Chapter 3

Observational Constraints

In this thesis, we improve on the literature presented in the previous chapter by developing a self-consistent inference of M31's mass and transverse velocity. Our constraint is based on five datasets from the literature spanning the full range of radial scales. At the centre of the galaxy, we constrain the projected line-of-sight velocity dispersion of stars in the bulge along a single line-of-sight (at a projected radius of ~ 0.25 kpc). We simultaneously fit the rotation curve and surface brightness profile to a radius of ~ 30 kpc. At larger radii, we constrain the mass using dynamical tracers: globular clusters ($R \lesssim 100$ kpc) and satellite galaxies ($R \lesssim 500$ kpc).

3.1   Satellite Galaxies

Table 3.1 lists the relevant information for the 23 satellite galaxies in our sample. This sample was constructed by supplementing the catalogues from Evans et al. (2000) and McConnachie & Irwin (2006) with more recently discovered dwarf galaxies and updated distance and velocity measurements. We restrict our sample to galaxies with

measurements of both line-of-sight velocity and distance.

Andromeda IX and X were discovered by Zucker et al. (2004) and Zucker et al. (2007) respectively in the Sloan Digital Sky Survey (SDSS; York et al. 2000) and subsequently observed spectroscopically by Collins et al. (2010) and Kalirai et al. (2010). Kalirai et al. (2010) also refined the distance and line-of-sight velocity measurements for Andromeda I, II, III, VII and XIV. Andromeda XI, XII and XIII were discovered by Martin et al. (2006) in a wide-field photometric survey of M31's outer halo their distances and line-of-sight velocities were measured by Collins et al. (2010). Andromeda XIV was discovered by Majewski et al. (2007) and followed up spectroscopically by Kalirai et al. (2010). The two most recently discovered satellites with both line-of-sight velocity and distance measurements – Andromeda XV and XVI – were both discovered by Ibata et al. (2007) in the first-year data from the Pan-Andromeda Archeological Survey (PAndAS; McConnachie et al. 2008) and observed spectroscopically by Letarte et al. (2009).

The observed velocities of Andromeda XII and XIV seem qualitatively to be inconsistent with those of the other satellites (figure 3.1). Quantitatively, the observed velocities of Andromeda XII and XIV are 3.4σ and 2.5σ from the mean of the remainder of the sample. It has been suggested that these two satellites form a system that is falling into M31 for the first time (Chapman et al. 2007; Majewski et al. 2007). If this hypothesis is correct then including these galaxies in our sample could bias our result. To test this, we perform our analysis with and without these two satellites and find that the effect on our final result is negligible.

An interesting inference problem presented by this dataset is that the distance
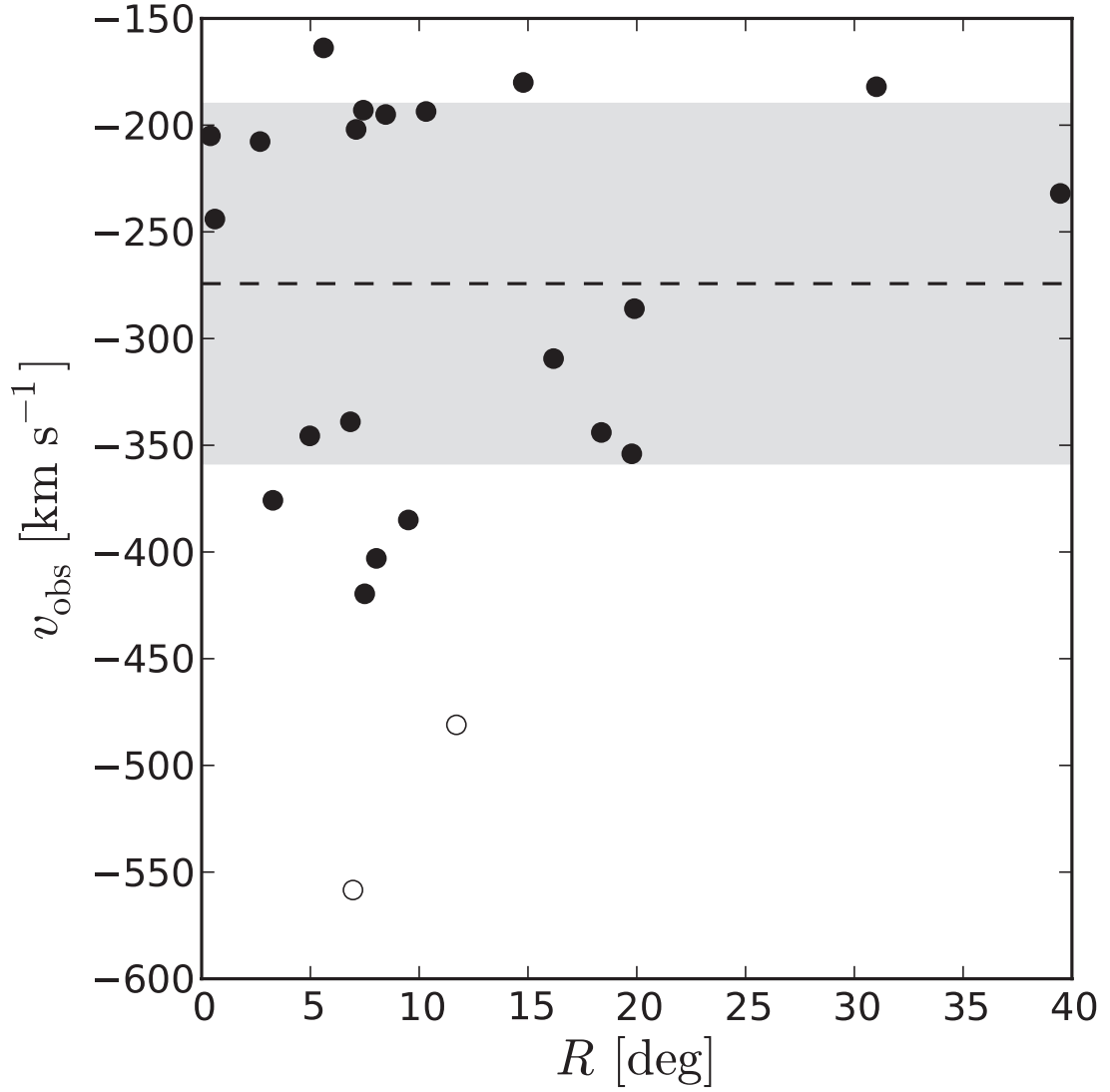
Figure 3.1 The observed satellite line-of-sight velocities as a function of projected radius. The mean and 1σ uncertainty for the main sample (without Andromeda XII and XIV, open circles) are indicated by the dashed line and shaded region.

measurement uncertainties for Andromeda IX, XI, XII and XIII are extremely asymmetric about the mode. In the limit of symmetric uncertainties, we can follow the standard procedure of modelling the measurement probability distribution function (PDF) as a Gaussian random variable. This model breaks down, however, when the uncertainties are significantly asymmetric. Without samples from the PDF of the original distance measurement, we are free to assume a parametric model for the measurement PDF. In this thesis, we choose a simple three-parameter generalization of the normal (Gaussian) distribution called the skew-normal distribution (appendix A). As the shape parameter $\alpha \to 0$, the distribution simplifies to a Gaussian and the skew-normal PDF can be efficiently sampled pseudo-randomly for any value of α. Table 3.2 lists the skew-normal parameterizations of the asymmetric distance measurement PDFs.

The positions and velocities of the satellite galaxies in our sample are all measured in the heliocentric rest frame but we are interested in modelling their distribution with respect to M31. Our procedure for calculating the M31-centric positions is summarized in appendix B and section 3.6 outlines the velocity correction. For both of these corrections, we marginalize over the uncertainties on all the relevant measurements.

Figure 3.2 shows the equal-area projection of our sample of satellite galaxies and globular clusters as seen from the centre of M31. The uncertainties in the distances to the satellite and M31 have both been propagated by Monte Carlo.
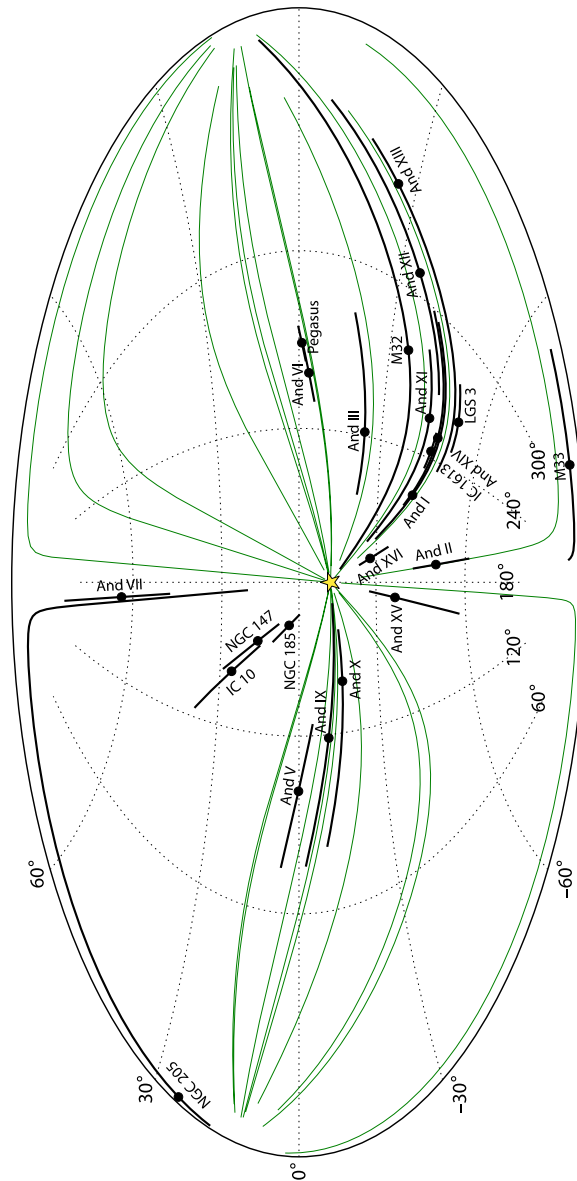
Figure 3.2 An Aitoff projection of the positions of satellite galaxies (black) and globular clusters (green) in M31-centric coordinates. The solid black lines show the uncertainty in the positions of the satellite galaxies taking into account the distance uncertainty of both the satellite and M31. The green lines show the positions of the globular clusters marginalized over the distance to the cluster along the line-of-sight. The yellow star is the position of the Sun and the equator ($b = 0$) is the plane of M31's disk.

Table 3.1.   Our sample of 23 satellite galaxies.

Galaxy	R.A. (J2000.0)	Dec. (J2000.0)	D (kpc)	v_{los} (km s^{-1})	Sources
M33	$1^h33^m50.9^s$	$+30°39'36.8''$	809 ± 24	-180 ± 1	1,2
M32	$0^h42^m41.8^s$	$+40°51'54.6''$	785 ± 25	-205 ± 3	1,2
IC 10	$0^h20^m17.3^s$	$+59°18'13.6''$	660 ± 65	-344 ± 5	1
NGC 205	$0^h40^m22.1^s$	$+41°41'7.1''$	824 ± 27	-244 ± 3	1,2
NGC 185	$0^h38^m58.0^s$	$+48°20'14.6''$	616 ± 26	-202 ± 7	1,2
IC 1613	$1^h4^m47.8^s$	$+2°7'4.0''$	715 ± 35	-232 ± 5	1
NGC 147	$0^h33^m12.1^s$	$+48°30'31.5''$	675 ± 27	-193 ± 3	1,2
Pegasus	$23^h28^m36.2^s$	$+14°44'34.5''$	919 ± 30	-182 ± 2	1,2
LGS 3	$1^h3^m55.0^s$	$+21°53'6.0''$	769 ± 23	-286 ± 4	1,2
And I	$0^h45^m39.8^s$	$+38°2'28.0''$	745 ± 24	-375.8 ± 1.4	1,2,3
And II	$1^h16^m29.8^s$	$+33°25'9.0''$	652 ± 18	-193.6 ± 1.0	1,2,3
And III	$0^h35^m33.8^s$	$+36°29'52.0''$	749 ± 24	-345.6 ± 1.8	1,2,3
And V	$1^h10^m17.1^s$	$+47°37'41.0''$	774 ± 28	-403 ± 4	1,2
And VI	$23^h51^m46.3^s$	$+24°34'57.0''$	783 ± 25	-354 ± 3	1,2
And VII	$23^h26^m31.0^s$	$+50°41'31.0''$	763 ± 35	-309.4 ± 2.3	2,3
And IX	$0^h52^m51.1^s$	$+43°11'48.6''$	765^{+5}_{-150}	-207.7 ± 2.7	4,5
And X	$1^h6^m33.7^s$	$+44°48'15.8''$	703 ± 70^a	-163.8 ± 1.2	3,6
And XI	$0^h46^m21.0^s$	$+33°48'22.0''$	760^{+10}_{-150}	-419.6 ± 4.4	5,7
And XII	$0^h47^m27.0^s$	$+34°22'29.0''$	830^{+170}_{-30}	-558.4 ± 3.2	5,7
And XIII	$0^h51^m51.0^s$	$+33°0'16.0''$	910^{+30}_{-160}	-195.0 ± 8.4	5,7
And XIV	$0^h51^m35.0^s$	$+29°41'49.0''$	740 ± 74^a	-481.0 ± 2.0	3,8
And XV	$1^h14^m18.7^s$	$+38°7'3.0''$	630 ± 60	-339 ± 7	9,10
And XVI	$0^h59^m29.8^s$	$+32°22'36.0''$	525 ± 50	-385 ± 6	9,10

Note. — Sources: 1 - Evans et al. (2000), 2 - McConnachie & Irwin (2006), 3 - Kalirai et al. (2010), 4 - Zucker et al. (2004), 5 - Collins et al. (2010), 6 - Zucker et al. (2007), 7 - Martin et al. (2006), 8 - Majewski et al. (2007), 9 - Ibata et al. (2007), 10 - Letarte et al. (2009)

[a]Assuming distance uncertainty of 10%.

Table 3.2. Skew-normal parameterization of distance measurements

Galaxy	Location (ξ)	Scale (ω)	Shape (α)
And IX	730	115	-7
And XI	717	107	-4
And XII	889	109	2
And XIII	854	103	-2

3.2 Globular Clusters

Our sample of globular clusters (hereafter GCs) was selected from the Revised Bologna Catalogue[1] (RBC; Galleti et al. 2004) by choosing the confirmed clusters (class 1 objects) with measured line-of-sight velocities and projected radius $\gtrsim 0.5°$ (20 kpc for $D_{\mathrm{M31}} = 785$ kpc; McConnachie & Irwin 2006). The final sample of 22 clusters is listed in table 3.3. The line-of-sight velocity distribution is shown in figure 3.3. We choose to omit cluster G339-BA30 since it appears to be an outlier in velocity space with a line-of-sight velocity $\sim 3\sigma$ from the mean of the remainder of the sample.

We must also transform the GC positions and velocities into the M31 centric coordinate system. Since there is no distance measurement to any of the clusters in our sample, we must marginalize over all reasonable distances as described in section 5.2.2.

---

[1]v.4, December 2009 – http://www.bo.astro.it/M31/

Table 3.3. Our sample of 22 outer globular clusters.

RBC Id.	Name	R.A. (J2000.0)	Dec. (J2000.0)	v_{los} (km s^{-1})	Φ[a] (arcmin)
0283	B289	$0^h34^m20.9^s$	$+41°47'51.1''$	-181 ± 30	99.4
0284	B290	$0^h34^m20.9^s$	$+41°28'18.2''$	-381 ± 26	95.2
0285	B291-G009	$0^h36^m5.0^s$	$+42°2'9.3''$	-215 ± 30	87.6
0289	B295-G014	$0^h36^m46.7^s$	$+40°19'42.1''$	-408 ± 11	88.1
0295	B301-G022	$0^h38^m21.6^s$	$+40°3'37.2''$	-382 ± 11	88.0
0392	B400-G343	$0^h48^m1.4^s$	$+42°25'33.0''$	-258 ± 11	91.1
0460	B468	$0^h43^m12.5^s$	$+39°47'57.2''$	-278 ± 22	88.4
0601	G001-MII	$0^h32^m46.5^s$	$+39°34'40.7''$	-332 ± 3	152.4
0602	G002-MIII	$0^h33^m33.8^s$	$+39°31'18.9''$	-313 ± 17	148.2
0927	G327-MVI	$0^h46^m49.5^s$	$+42°44'46.7''$	-251 ± 11	99.7
0939	G339-BA30	$0^h47^m50.2^s$	$+43°9'16.4''$	33 ± 30	126.5
0953	G353-BA13	$0^h50^m18.2^s$	$+42°35'44.1''$	-295 ± 26	116.0
0961	EXT8	$0^h53^m14.5^s$	$+41°33'24.5''$	-154 ± 30	119.4
3011	BA11	$0^h48^m45.6^s$	$+42°23'37.7''$	-97 ± 12	95.3
4001	B514-MCGC4	$0^h31^m9.8^s$	$+37°54'0.1''$	-458 ± 23	242.4
4004	B517	$0^h59^m59.9^s$	$+41°54'6.8''$	-272 ± 54	197.3
4007	MCGC1-B520	$0^h26^m47.7^s$	$+39°44'46.3''$	-219 ± 15	203.5
4008	MGC1	$0^h50^m42.5^s$	$+32°54'58.8''$	-355 ± 2	510.1
6005	MCGC5-H10	$0^h35^m59.7^s$	$+35°41'3.6''$	-358 ± 2	344.3
6008	MCGC8-H23	$0^h54^m25.0^s$	$+39°42'55.7''$	-381 ± 15	162.6
6009	MCGC9-H24	$0^h55^m44.0^s$	$+42°46'16.4''$	-147 ± 20	170.5
6010	MCGC10-H27	$1^h7^m26.3^s$	$+35°46'48.4''$	-291 ± 2	438.5

[a]Φ is the angular separation of the cluster from the galaxy centre.
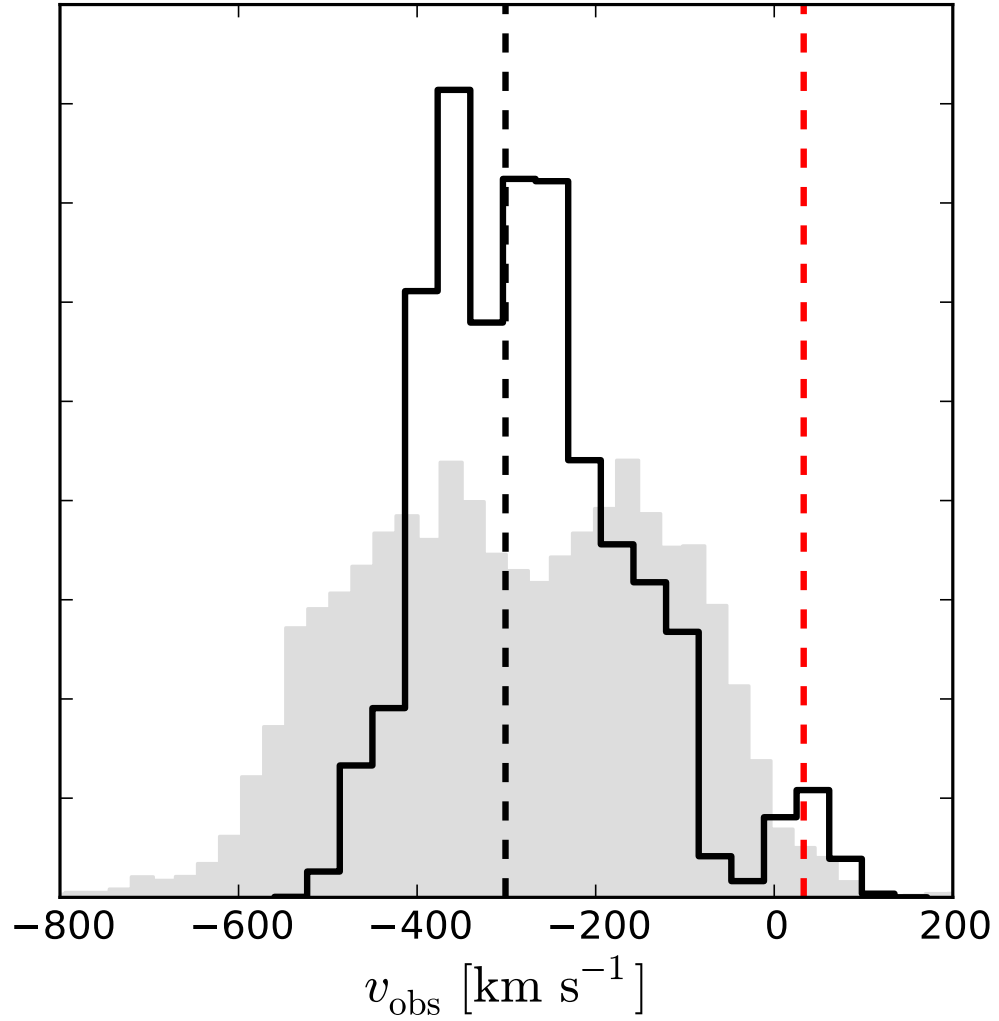
Figure 3.3 A histogram of the observed line-of-sight velocities for all confirmed globular clusters from the RBC (shaded region) and our sample (black histogram). All velocity uncertainties have been propagated using Monte Carlo sampling. The black dashed line indicates the systematic velocity of -301 km s^{-1} and the red dashed line is at the mean velocity of the outlying cluster G339-BA30 ($+33$ km s^{-1}).
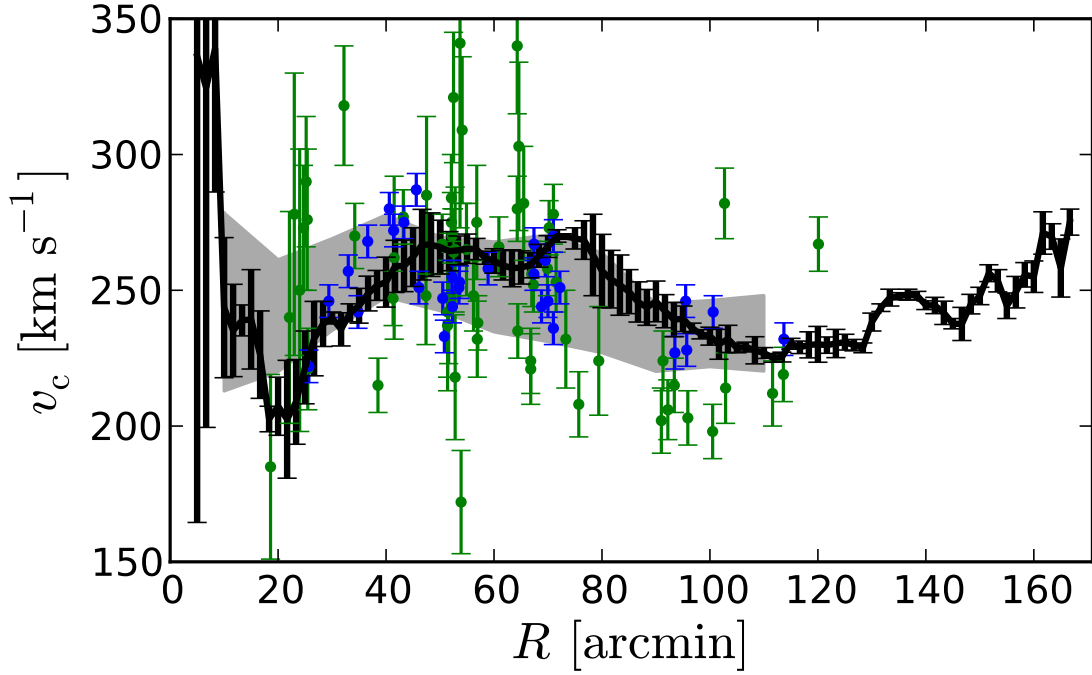
Figure 3.4 The rotation curve from Chemin et al. (2009) (black error bars). The measurements from Rubin & Ford (1970, green points) and Kent (1989, blue points) and the smoothed dataset used by Widrow & Dubinski (2005, shaded region) are shown for illustration.

3.3 Rotation curve

The rotation curve of M31 has been used extensively for mass modelling and it is credited as one of the first decisive arguments for dark matter (Rubin & Ford 1970). Most measurements of gas kinematics are more than 20 years old and they are all of quite low resolution. Chemin et al. (2009) recently published a high resolution study of the HI kinematics in M31 (figure 3.4). Such a high resolution measurement is extremely powerful for the study of phase space substructure in the disk and other dynamical experiments. For our purposes, however, we only need to model the general shape so we introduce a smoothing parameter in our fit to the data (see section 5.2.3).

3.4   Bulge velocity dispersion

To constrain the inner parts of our galaxy model, we include a simple constraint based on the bulge velocity dispersion profile from Saglia et al. (2010). Figure 3.5 shows this profile and there is an obvious peak at projected radius $R \sim 60''$. We choose to constrain our model along this line-of-sight.

To construct our constraint, we select all the data in the radial range $50'' \leq R \leq 70'' \sim 200$ pc (the final result is insensitive to the chosen range) and plot the dispersion as a function of position angle (figure 3.5). In this figure, there is a clear oscillatory signature that is possibly the signature of a bar or triaxial bulge. We model this oscillation as

$$\sigma_p(R) = \langle \sigma_p(R) \rangle + \mathcal{O}(\Phi) \tag{3.1}$$

where

$$\mathcal{O}(\Phi) = A \cos \left[B \cdot (\Phi + \Phi_0) \right] \, \mathrm{km \, s^{-1}}. \tag{3.2}$$

Assuming Gaussian uncertainties, we fit the maximum likelihood solution of equation 3.1 and find the mean $\langle \sigma_p(R) \rangle = 171.6 \, \mathrm{km \, s^{-1}}$ and higher order terms

$$\mathcal{O}(\Phi) = 4.719 \cos \left[2.227 \cdot (\Phi + 27.9°) \right] \, \mathrm{km \, s^{-1}}. \tag{3.3}$$

The amplitude of the oscillation and the scatter in velocities are of nearly the same magnitude so we set a lower limit of $\sim 5 \, \mathrm{km \, s^{-1}}$ on the uncertainty in the mean velocity dispersion. As mentioned, this uncertainty should only be considered a lower limit given the crude method used to estimate it.
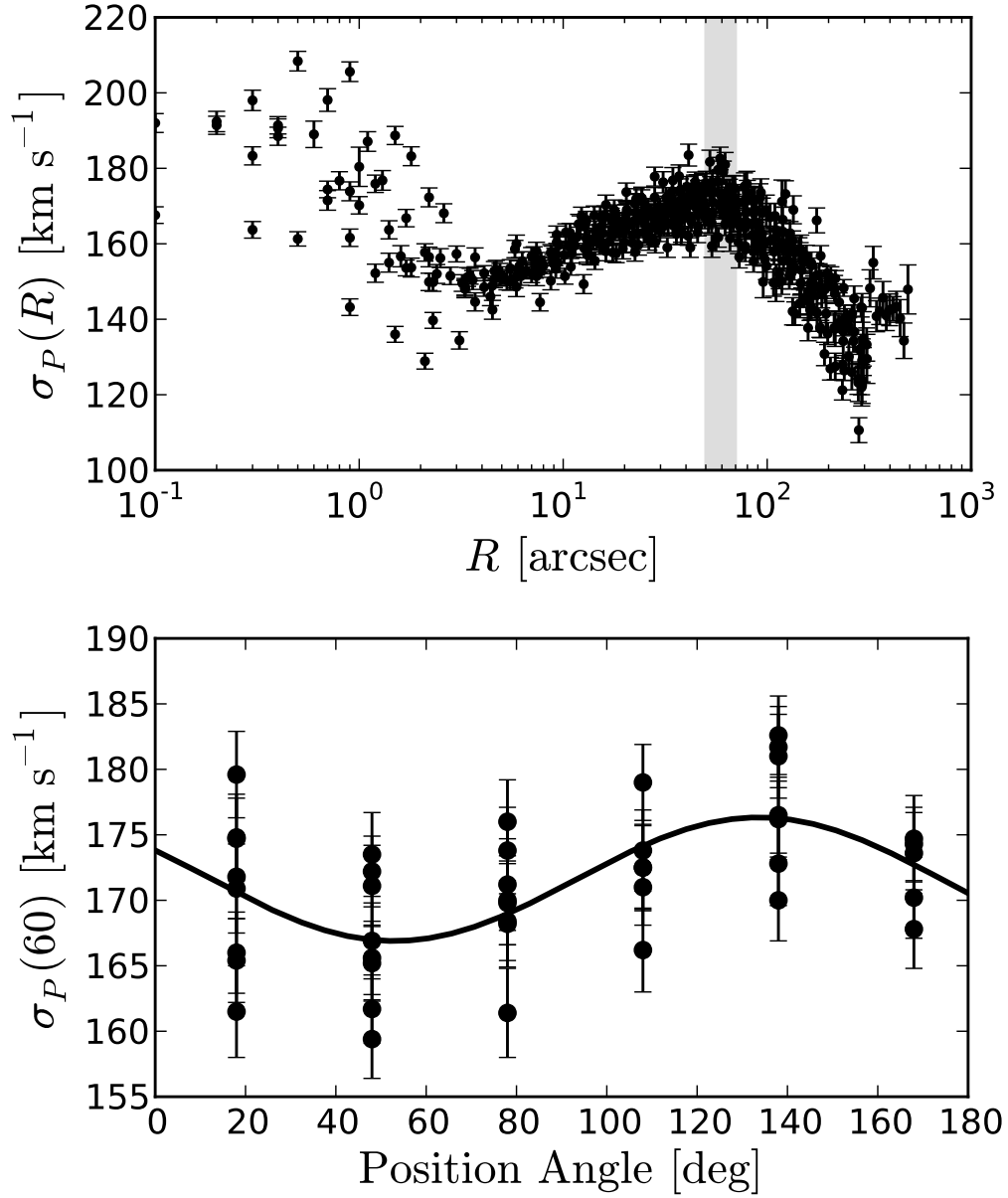
Figure 3.5 *Top*: The projected line-of-sight velocity dispersion profile from Saglia et al. (2010) plotted for all position angles. *Bottom*: The same data from the shaded region ($50'' \leq R \leq 70''$). The black line is the maximum likelihood estimate of the oscillatory behaviour with a mean 172 ± 5 km s^{-1}.

3.5 Surface Brightness Profile

The kinematic data discussed above provides a strong constraint on the mass distribution of the galaxy but we should also recover the surface brightness profile with reasonable assumptions about the mass-to-light ratios of the disk and bulge. For our observational constraint, we consider the classic measurement of the R-band surface brightness profile from Walterbos & Kennicutt (1987, 1988). Instead of a bluer waveband, we choose to use the R-band since it is less affected by dust and more sensitive to the low-mass stars that dominate the mass distribution (e.g. Binney & Merrifield 1998).

Since we are uninterested in the details of the profile and only wish to estimate the overall normalization, we fit the profile along the disk's major axis and we mask the radial region between projected radius $R = 20'$ and $70'$ where there is substructure that we would be unable to constrain with our smooth model. Figure 3.6 shows the data and a representative decomposition for a simple exponential disk and Sérsic bulge model.

3.6 Satellite Velocity Corrections

The tracer kinematics that we measure are always with respect to the Sun and various corrections are necessary to recover the quantity that we are interested in for the purpose of dynamical modelling — the peculiar velocity of an object in M31's rest frame. The concept behind this correction is very simple but van der Marel & Guhathakurta (2008) showed that the result can be quite powerful.

We begin by writing the full three-dimensional velocity of a given satellite galaxy
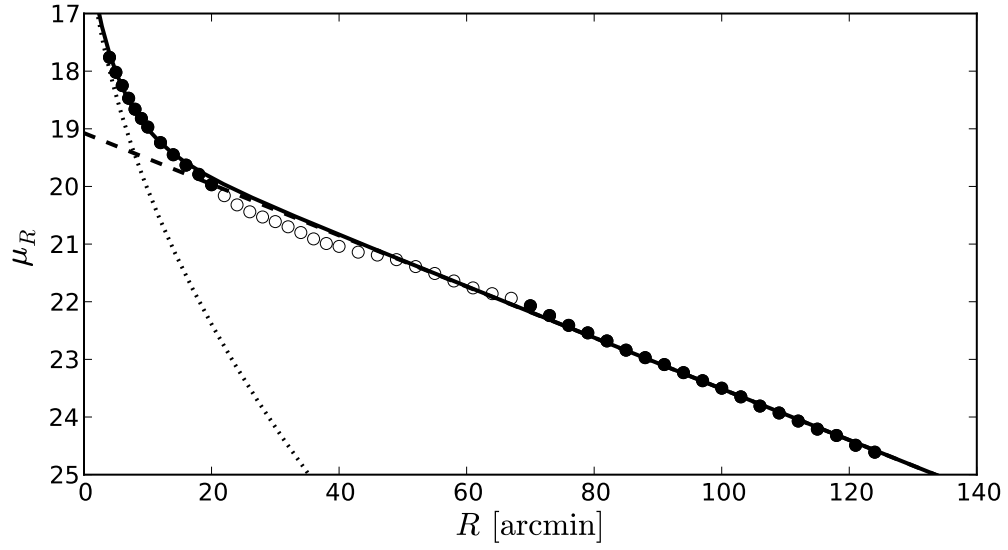
Figure 3.6 R-band surface brightness profile along the major axis of M31 from Walterbos & Kennicutt (1987). The solid line is a least-squares estimate of the disk (dashed line, $R_d = 5.6$ kpc) and bulge (dotted line, $R_e = 1.1$ kpc) decomposition for constant mass-to-light ratios. The filled points are included in the fit and the open points are not.

S with respect to M31 (labelled A) as

$$\mathbf{v}_{S,A} = \mathbf{v}_{S,\odot} + \mathbf{v}_{\odot,A} \tag{3.4}$$

where $\mathbf{v}_{X,Y} = \mathbf{v}_X - \mathbf{v}_Y$ is the velocity of X with respect to Y. This can be rewritten

$$\mathbf{v}_{S,A} = \mathbf{v}_{S,\odot} + \mathbf{v}_{\odot,G} + \mathbf{v}_{G,\odot} + \mathbf{v}_{\odot,A} = \mathbf{v}_{S,G} + \mathbf{v}_{G,A} \tag{3.5}$$

where G indicates the Milky Way. Then, we decompose $\mathbf{v}_{G,A}$ into a component along our line-of-sight to M31 $\hat{l}_A$ and a component in the plane of the sky $\hat{l}_A^{\perp}$

$$\mathbf{v}_{S,A} = \mathbf{v}_{S,G} + \left(\mathbf{v}_{G,A} \cdot \hat{l}_A\right) \hat{l}_A + \left(\mathbf{v}_{G,A} \cdot \hat{l}_A^{\perp}\right) \hat{l}_A^{\perp}. \tag{3.6}$$

$(\mathbf{v}_{G,A} \cdot \hat{l}_A)$ has been tightly constrained by Courteau & van den Bergh (1999) but $\mathbf{v}_{\odot,A} \cdot \hat{l}_A^{\perp}$ is not directly measurable. Therefore, all measurements of satellite line-of-sight velocities are contaminated by the transverse velocity term. Evans & Wilkinson (2000) argue that $\mathbf{v}_{G,A} \cdot \hat{l}_A^{\perp}$ should be negligible since the Milky Way-M31 orbit should be nearly radial and the vector between the Galactic centre and M31 is approximately equal to our line-of-sight $\hat{l}_A$. As discussed in section 2.4, this is not an unreasonable assumption since there are no large galaxies in the LG to generate significant angular momentum. It is, however, an unnecessary assumption and in making it, the previous authors are losing significant information. Instead, we can actually exploit this contamination to obtain a probabilistic constraint on the transverse motion of the M31 system.

Bahcall & Tremaine (1981) – and subsequently Evans & Wilkinson (2000) and van

der Marel & Guhathakurta (2008) – pointed out that the line-of-sight velocities of satellites at large angular distances will be most significantly affected by this unmeasured systematic transverse motion and dynamical modelling is highly sensitive to this. van der Marel & Guhathakurta (2008) realized, however, that this effect could actually be used to *infer* the systematic motion. Specifically, if we project equation 3.4 onto the line-of-sight of the satellite $\hat{l}_S$ we find that the peculiar velocity of a satellite along the line-of-sight is

$$
\begin{aligned}
v_{S,A}^{los} \;=\; & \mathbf{v}_S \cdot \hat{l}_S - (\mathbf{v}_A \cdot \hat{l}_A)\,\cos\Phi_S \\
& -(\mathbf{v}_A \cdot \hat{l}_A^{\perp})\,\sin\Phi_S\,\cos(\Theta_S - \Theta_{\perp})\,.
\end{aligned}
\tag{3.7}
$$

where $\mathbf{v}_S$ and $\mathbf{v}_A$ are heliocentric velocities and Φ_S is the angular separation between $\hat{l}_A$ and $\hat{l}_S$. In equation 3.7, Θ_S and $\Theta_{\perp}$ are the position angles of the satellite with respect to M31 and $\hat{l}_A^{\perp}$ respectively measured from north through east (van der Marel & Guhathakurta 2008).

It is clear from equation 3.7 that the line-of-sight velocities of the satellites will exhibit an oscillation as a function of position angle that will depend on the transverse velocity of M31. This effect will be strongest for satellites at large angular separations. This effect, though not immediately visible in the dataset (figure 3.7), can be used to constrain the transverse velocity as demonstrated by van der Marel & Guhathakurta (2008). van der Marel & Guhathakurta (2008) modelled $v_{S,A}^{los}$ as a Gaussian velocity deviate with constant dispersion σ along the line-of-sight and inferred $\mathbf{v}_A \cdot \hat{l}_A^{\perp}$ by fitting equation 3.7 to the data using least-squares (assuming uniform priors in $\mathbf{v}_A \cdot \hat{l}_A^{\perp}$ and $\Theta_{\perp}$).

Instead of assuming that the transverse motion of M31 is negligible or that we
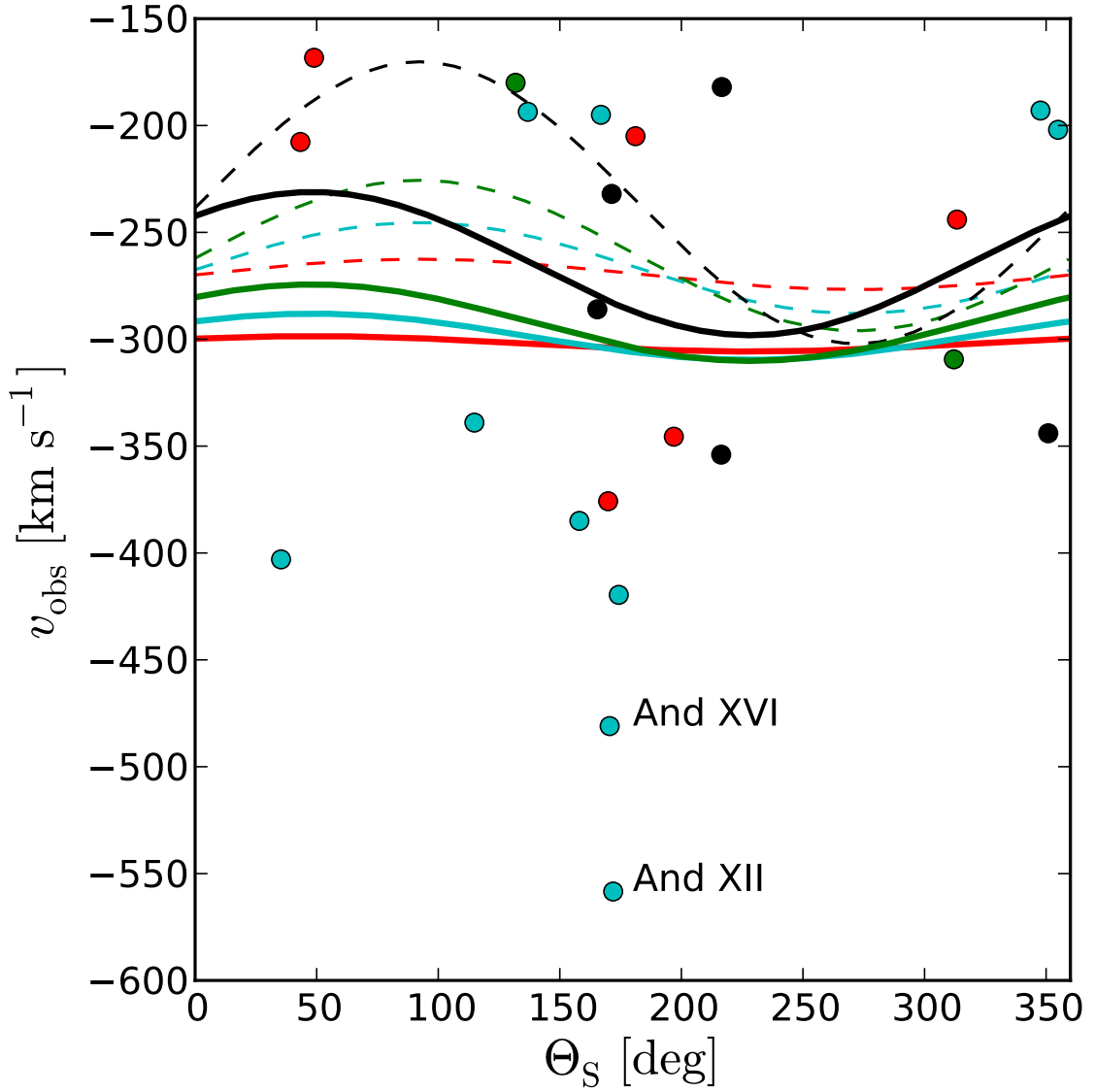
Figure 3.7 The projected line-of-sight velocities of the satellite galaxies separated into Φ_S bins: red – $(0°, 6°)$, cyan – $(6°, 12°)$, green – $(12°, 18°)$, black – $(18°, 40°)$. The solid lines show the least-squares fit of equation 3.7 plotted for $\Phi_S = 3°$ (red), $9°$ (cyan), $15°$ (green) and $29°$ (black). The dashed lines show the equivalent results from van der Marel & Guhathakurta (2008) for reference.

have no information regarding the peculiar velocities, we choose to model $v_{\mathrm{S,A}}^{\mathrm{los}}$ with a realistic, cosmologically motivated distribution function and include the transverse velocity of M31 as a model parameter. In doing so, we infer both the mass profile and the transverse motion simultaneously and self-consistently.

In order to compare results with van der Marel & Guhathakurta (2008), we decompose the systematic transverse velocity into northern and western components

$$v_{\mathrm{W}} \equiv (\mathbf{v}_{\mathrm{A}} \cdot \hat{l}_{\mathrm{A}}^{\perp}) \cos{(\Theta_{\perp} + 90°)}$$
$$v_{\mathrm{N}} \equiv (\mathbf{v}_{\mathrm{A}} \cdot \hat{l}_{\mathrm{A}}^{\perp}) \sin{(\Theta_{\perp} + 90°)} \quad . \tag{3.8}$$

Repeating the fitting procedure from van der Marel & Guhathakurta (2008) with our full sample, we find $(v_{\mathrm{M31,los}}, v_{\mathrm{W}}, v_{\mathrm{N}}) = (-303, -52, 46) \pm (24, 197, 107)$ km s^{-1} where $v_{\mathrm{M31,los}} \equiv \mathbf{v}_{\mathrm{A}} \cdot \hat{l}_{\mathrm{A}}$. This is in reasonably good agreement with the result from van der Marel & Guhathakurta (2008) using only satellite line-of-sight velocities: $(v_{\mathrm{M31,los}}, v_{\mathrm{W}}, v_{\mathrm{N}}) = (-270, -136, 5) \pm (19, 148, 75)$ km s^{-1}. The median and 1σ confidence interval on the total mass of the LG predicted from the timing argument and the transverse velocity inferred from our full sample is $7.9_{-2.3}^{+4.8} \times 10^{12} \, M_{\odot}$. This result is consistent with the value quoted by van der Marel & Guhathakurta (2008) but it is significantly larger than the $(2.3 \pm 0.6) \times 10^{12} \, M_{\odot}$ calculated by Courteau & van den Bergh (1999) based on the projected mass estimator.

Chapter 4

Theory

The quantity and quality of dynamical information about Andromeda is orders of magnitude more substantial than for other individual galactic systems. It is therefore reasonable to imagine that we could infer statistically significant probabilistic information about the dark matter distribution using very general models. For the purpose of mass modelling, disk galaxies are generally either approximated as a spherically symmetric dark matter halo in isolation or the dynamics of the system are ignored. While this assumption can yield excellent first-order approximations to galaxy parameters, the true systems are much more complex. In fact, disk galaxies are composed of several dynamically distinct components including a stellar disk, a halo and often a centrally concentrated bulge. Beyond this, each of these components displays significant substructure in their phase space distributions for which modelling is currently intractable.

For this work, we employ an *axisymmetric, multi-component, dynamical* galaxy model from Widrow et al. (2008) based on Kuijken & Dubinski (1995). We construct semi-analytic collisionless, equilibrium phase-space distribution functions (DFs) for a

thin stellar disk, centrally concentrated bulge and a dark matter halo. Then, using an iterative scheme, we adjust the DF parameters to yield a fully self-consistent density distribution that closely approximates cosmologically inspired empirical parameterizations.

4.1 Galaxy Dynamics

Even though galaxies are composed of many discrete stars and dark matter particles, the density distribution can be well described as samples from a phase space DF $f(\mathbf{x}, \mathbf{v}, t)\, d^3\mathbf{x}\, d^3\mathbf{v}$ if the relevant time scale is less than the relaxation time of the system (Binney & Tremaine 2008). $f(\mathbf{x}, \mathbf{v}, t)\, d^3\mathbf{x}\, d^3\mathbf{v}$ is proportional to the probability of finding a randomly chosen star in the phase space volume $d^3\mathbf{x}\, d^3\mathbf{v}$. In practice, we adopt the convention when $f(\mathbf{x}, \mathbf{v}, t)$ is scaled by the total mass of the system M_{tot} so that

$$\int d^3\mathbf{x}\, d^3\mathbf{v}\, f(\mathbf{x}, \mathbf{v}, t) = M_{\text{tot}}. \tag{4.1}$$

The dynamics of a system described by f are governed by the collisionless Boltzmann equation (CBE)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{x}} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \tag{4.2}$$

where $\Phi \equiv \Phi(\mathbf{x})$ is the gravitational potential. This potential is considered *self-consistent*, if Φ and ρ solve the Poisson equation

$$\nabla^2 \Phi = 4\pi G \rho. \tag{4.3}$$

To solve equation 4.2, we employ the *Jeans theorem* (Jeans 1915) which states that

any function of only a system's integrals of motion will be a steady-state, equilibrium solution of the CBE. An integral of motion is any quantity that is constant for a particle's full orbit.

Isolated, spherically symmetric, isotropic systems can be fully specified by a DF depending only on energy — a so-called "ergodic" DF — that satisfies the Eddington equation

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2} \left[\int_0^{\mathcal{E}} \frac{d\Psi}{\sqrt{\mathcal{E} - \Psi}} \frac{d^2\rho}{d\Psi^2} + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho}{d\Psi} \right)_{\Psi=0} \right] \qquad (4.4)$$

(Binney & Tremaine 2008) where $\Psi \equiv -\Phi$ and $\mathcal{E} \equiv \Psi - v^2/2$ are the relative potential and energy respectively. The second term on the right hand side of equation 4.4 is zero for the distributions that we will consider.

For a self-consistent, isolated system the density distribution generated by integrating equation 4.4 over all velocities must be spherically symmetric. In the presence of an external potential, however, this is no longer true. We can rewrite equation 4.4 as

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2} \int_0^{\mathcal{E}} \frac{d\Psi_{\text{tot}}}{\sqrt{\mathcal{E} - \Psi_{\text{tot}}}} \frac{d^2\rho}{d\Psi_{\text{tot}}{}^2} \qquad (4.5)$$

where $\Psi_{\text{tot}} \equiv \Psi + \Psi_{\text{ext}}$ is the total potential. If Ψ_{tot} is not spherically symmetric then neither is the density distribution generated by integrating equation 4.5 because the isodensity surfaces lie at the surfaces of constant potential.

4.1.1 Multipole Expansion

The solution of the Poisson equation (equation 4.3) in spherical coordinates is found by expanding the potential in a series of spherical shells and solving Laplace's equation

inside and outside each shell. The solution is then

$$\Psi(r, \boldsymbol{\Omega}) = 4\pi \sum_{l=0}^{\infty} \frac{P_l(\boldsymbol{\Omega})}{2l+1} \left(\frac{1}{r^{l+1}} \int_0^r d\tilde{r} \, \tilde{r}^{l+2} \rho_l(\tilde{r}) + r^l \int_r^{\infty} \frac{d\tilde{r}}{\tilde{r}^{l-1}} \rho_l(\tilde{r}) \right) \qquad (4.6)$$

(Binney & Tremaine 2008) where P_l are the Legendre polynomials and $\boldsymbol{\Omega}$ is the solid angle. This method can be applied to any density distribution and the potential can be calculated to arbitrarily high precision. However, calculating the potential using equation 4.6 is not efficient for highly flattened distributions, since a restrictively large number of poles would be required. Since our model does contain a disk, Kuijken & Dubinski (1995) proposed an analytic potential/density ansatz for the disk and solved equation 4.6 for the residual potential and density.

4.2 Our Semi-Analytic, Multi-component, Axisymmetric Generative Model for Galaxy Dynamics

Analytically tractable density-potential pairs have always been an attractive tool for dynamical modelling in astrophysics. The phase-space DFs for particles in these models can be easily or trivially evaluated and the results can yield excellent first order constraints on physically interesting parameters. However, these models are generated exclusively by spherically symmetric isolated density distributions and more sophisticated generative models are needed if we wish to use all the available datasets effectively. For example, it is unphysical to model a disk galaxy's rotation curve using only a smooth spherically symmetric dark matter halo in isolation. Instead,

the model must include prescriptions for a disk and bulge.

Our galaxy model is generated from a semi-analytic composite equilibrium DF of the form

$$f(\mathcal{E}, L_Z, E_Z) = f_{\mathrm{h}}(\mathcal{E}) + f_{\mathrm{b}}(\mathcal{E}) + f_{\mathrm{d}}(\mathcal{E}, L_Z, E_Z) \tag{4.7}$$

where $\mathcal{E} \equiv -\Phi - v^2/2$ is the relative energy, L_Z is the angular momentum about the symmetry axis and E_Z is the vertical kinetic energy of the stars in the disk. The first two variables ($\mathcal{E}$ and L_Z) are exact integrals of motion and E_Z is an approximate integral of motion for cold disks (see Kuijken & Dubinski 1995). By the Jeans theorem, any DF generated by equation 4.7 will be an equilibrium solution to the CBE. In isolation, f_{h} and f_{b} would generate spherically-symmetric distributions of particles with isotropic velocities since they are functions of only $\mathcal{E}$. However, the composite distribution breaks spherical symmetry since it is also a function of L_Z and E_Z.

Integrating equation 4.7 over all velocities yields the composite axisymmetric density distribution

$$\rho(s, z, \Psi) = \rho_{\mathrm{h}}(\Psi) + \rho_{\mathrm{b}}(\Psi) + \rho_{\mathrm{d}}(s, z, \Psi) \tag{4.8}$$

in terms of the relative potential $\Psi \equiv -\Phi$ and cylindrical coordinates (s, z).

Using an iterative procedure (Kuijken & Dubinski 1995; Widrow & Dubinski 2005; Widrow et al. 2008), we fine-tune equation 4.7 so that equation 4.8 closely approximates the sum of cosmologically relevant, highly general target density profiles. This procedure produces a fully self-consistent, axisymmetric, semi-analytic density-potential-DF triplet that can be directly compared to observations.

4.2.1 Target density profiles

For the halo density profile, we choose a two-parameter generalization of the Navarro, Frenk, & White (1997, NFW) profile

$$\rho_{\rm h}(r) = \frac{2^{1-\gamma_{\rm h}}v_{\rm h}{}^2}{4\pi a_{\rm h}{}^2}\left(\frac{r}{a_{\rm h}}\right)^{-\gamma_{\rm h}}\left(1+\frac{r}{a_{\rm h}}\right)^{\gamma_{\rm h}-\alpha_{\rm h}} \tag{4.9}$$

where r is the spherical radius, $\gamma_{\rm h}$ is the cusp strength, $\alpha_{\rm h}$ is the outer slope, $a_{\rm h}$ is the halo scale length and $v_{\rm h}$ is the characteristic velocity scale. In practice, we must multiply equation 4.9 by a truncation function that drops from one to zero over a finite range so that the mass does not diverge at large radii. We chose a function of the form

$$C\left(r; r_{\rm h}, \delta r_{\rm h}\right) \equiv \frac{1}{2}{\rm erfc}\left(\frac{r-r_{\rm h}}{\sqrt{2}\,\delta r_{\rm h}}\right) \tag{4.10}$$

where $r_{\rm h}$ is well beyond the radius of our most distant data point.

For the bulge, we chose a de-projected Sérsic profile with three free parameters: scale length R_e, velocity scale $v_{\rm b}$ and Sérsic index n (Prugniel & Simien 1997; Terzić & Graham 2005). This profile is parameterized as

$$\rho_{\rm b}(r) = \rho_0\left(\frac{r}{R_e}\right)^{-p}\exp\left[-b\left(\frac{r}{R_e}\right)^{1/n}\right] \tag{4.11}$$

where R_e is the half-mass radius, n is the Sérsic index and

$$\rho_0 \equiv \frac{v_{\rm b}{}^2}{4\pi n b^{n(p-2)}\Gamma\left[n(2-p)\right]R_e{}^2}. \tag{4.12}$$

In equation 4.12, $\Gamma(x)$ is the standard gamma function. Neither b or p are free parameters; b is adjusted numerically to ensure that R_e contains half the total mass

and $p = 1 - 0.6097/n + 0.05563/n^2$ (Prugniel & Simien 1997) so that integrating equation 4.11 along the line-of-sight produces the surface density

$$\Sigma_{\mathrm{b}}(R) = \Sigma_0 \exp\left[-b\left(\frac{R}{R_e}\right)^{1/n}\right] \tag{4.13}$$

where R is projected radius and

$$\Sigma_0 \equiv \frac{v_{\mathrm{b}}{}^2\, b^n\, \Gamma\left[n(3-p)\right]}{2\pi\, R_e\, n\, \Gamma\left[n(2-p)\right]\Gamma(2n)} \tag{4.14}$$

(Terzić & Graham 2005).

The self-consistent potential generated by equation 4.11 is analytically

$$
\begin{aligned}
\Psi(r) \;&=\; \frac{v_{\mathrm{b}}}{\Gamma\left[n(2-p)\right]} \\
&\times \left\{ \Gamma\left[n(2-p), b\left(\frac{r}{R_e}\right)^{1/n}\right] + \frac{R_e\, b^{-n}}{r}\gamma\left[n(3-p), b\left(\frac{r}{R_e}\right)^{1/n}\right]\right\}
\end{aligned}
\tag{4.15}
$$

(Terzić & Graham 2005) where $\Gamma(a, x)$ and $\gamma(a, x)$ are the incomplete gamma functions

$$\Gamma(a, x) = \int_x^\infty t^{a-1}e^{-t}\, dt \text{ and } \gamma(a, x) = \int_0^x t^{a-1}e^{-t}\, dt. \tag{4.16}$$

Since none the kinematic datasets that we consider place strong constraints on the value of n and because there is a strong degeneracy between n and R_e, we choose to fix $n = 2$ with minimal effect on our final conclusions.

We parameterize the disk as a thin exponential disk with mass M_{d}, scale length

R_d and scale height z_d. In cylindrical coordinates this distribution is

$$\rho_\mathrm{d}(s, z) = \frac{M_\mathrm{d}}{4\pi R_\mathrm{d}{}^2 z_\mathrm{d}} \exp\left(-\frac{s}{R_\mathrm{d}}\right) \operatorname{sech}^2\left(\frac{z}{z_\mathrm{d}}\right).$$ (4.17)

In practice, we must also include a truncation function in analogy to equation 4.10. Since our data provide no constraint on the disk scale height and since the value will have negligible effects, we fix $z_\mathrm{d} = 0.5$ kpc in all that follows. Integrating equation 4.17 over all z, we find that the surface mass density of this disk is

$$\Sigma_\mathrm{d}(R) = \frac{M_\mathrm{d}}{2\pi R_\mathrm{d}{}^2} \exp\left(-\frac{R}{R_\mathrm{d}}\right).$$ (4.18)

As mentioned previously, we model the higher order disk potential harmonics analytically in order to solve Poisson's equation more efficiently. Following Kuijken & Dubinski (1995), we start with an ansatz for the disk potential

$$\Psi^\dagger_\mathrm{disk}(r, z) = \frac{G\mathcal{M}_\mathrm{d} z_\mathrm{d}}{2 R_\mathrm{d}{}^2} \ln\left[\cosh\left(\frac{z}{z_\mathrm{d}}\right)\right] \exp\left(-\frac{r}{R_\mathrm{d}}\right) \times C(r; r_\mathrm{d,out}, \delta r_\mathrm{d,out}).$$ (4.19)

This potential is generated self-consistently by the density profile

$$
\begin{aligned}
\rho^\dagger_\mathrm{disk}(R, z) &= \frac{1}{4\pi G}\nabla^2 \Psi^\dagger_\mathrm{disk}(r, z) \\
&= \frac{\mathcal{M}_\mathrm{d} z_\mathrm{d}}{8\pi R_\mathrm{d}{}^2}\left[\psi_r{}'' \ln\cosh z + 2\frac{\psi_r{}'}{r}\left(\frac{z}{z_\mathrm{d}}\tanh z + \ln\cosh z\right) + \frac{1}{z_\mathrm{d}{}^2}\psi_r \operatorname{sech}^2 z\right]
\end{aligned}
$$
                    (4.20)

where primes denote differentiation in spherical radius r and $\psi_r \equiv \psi_r(r)$ is the radial

component of the potential

$$\psi_r(r) \equiv \exp\left(-\frac{r}{R_\mathrm{d}}\right) \times C(r; r_\mathrm{d,out}, \delta r_\mathrm{d,out}). \tag{4.21}$$

We can then calculate the potential for the galaxy density distribution — after subtracting our approximate disk density — to high precision without calculating high order spherical harmonics. Kuijken & Dubinski (1995) show that equation 4.19 matches the higher order disk harmonics well. This method provides a significant boost of efficiency over calculating the potential without the disk ansatz.

4.2.2 Galaxy Light

The relationship between a galaxy's mass distribution and its surface brightness profile is governed by a quantity called the mass-to-light ratio. In this thesis we will discuss stellar mass-to-light ratios that relate the baryonic mass in the disk and bulge of a galaxy to the observed light. Stellar mass-to-light ratios are non-trivially correlated with type, age and metallicity of the stars in the composite stellar populations of each component. The detailed physics of this relationship is an active area of research and rigorous comparisons to the literature are difficult but it is still of interest to determine — at least roughly — the mass-to-light ratios predicted by our modelling procedure.

For simplicity, we will follow common practice and assume that the disk and bulge have constant but different mass-to-light ratios. Given the significant colour gradient in M31's disk (e.g. Walterbos & Kennicutt 1988), it is likely that the disk mass-to-light ratio of the disk depends on radius but for our purposes the first order approximation will be sufficient.

Since we only wish to estimate the magnitude of mass-to-light ratios needed to reproduce the surface brightness profile, we will also assume that the density profiles of the disk and the bulge are given by the target densities (equations 4.11 and 4.17). In this case, we can calculate the surface mass density of the disk by integrating equation 4.17 over z (equation 4.18). By definition, the surface mass density of the bulge is a Sérsic law (equation 4.13). The major-axis luminosity profile at wavelength λ is therefore

$$\mathcal{L}_\lambda(R) = \frac{\Sigma_{\mathrm{d}}(R)}{\Upsilon_{\mathrm{d},\lambda} \cos i} + \frac{\Sigma_{\mathrm{b}}(R)}{\Upsilon_{\mathrm{b},\lambda}} \tag{4.22}$$

where $\Upsilon_{\mathrm{d/b},\lambda}$ is the mass-to-light ratio and i is the inclination of the galactic disk with respect to the line-of-sight. No inclination correction is necessary for the bulge because the model is spherically symmetric. The surface brightness profile (in units of mag arcsec^{-2}) is

$$\mu_\lambda(R) = \mathcal{M}_{\odot,\lambda} + 36.572 - 2.5 \log_{10} \mathcal{L}_\lambda(R) \tag{4.23}$$

where $\mathcal{M}_{\odot,\lambda}$ is the solar magnitude in the wavelength band of interest (Binney & Merrifield 1998).

4.3   The Mass of the Local Group

From our constraint on M31's three-dimensional velocity (section 3.6), we can infer the orbit of the Local Group (LG). The parameters of this orbit yield an estimate of the total mass of the LG by the timing argument. It is interesting for us to consider this procedure in the context of this work because we can simultaneously estimate the mass of both M31 and the LG.

4.3.1   An Orbital Generative Model

To infer the orbital parameters of the LG, we must first construct a model that could generate the "data" (M31's inferred three-dimensional velocity). To do this, we approximate the LG as two point masses of mass M_{M31} and M_{MW}. Under the assumption that the LG decoupled from the Hubble flow at early times, the first pericentre passage must have occurred at the time of the big bang and the evolution of the system has since been governed by Newtonian dynamics. We can parameterize the kinematics of an obit of eccentricity e and semi-major axis length a as

$$r = a\left(1 - e\cos\eta\right)$$
$$t = \left(\frac{a^3}{M}\right)^{1/2}\left(\eta - e\sin\eta\right) \tag{4.24}$$

where $M \equiv M_{\text{MW}} + M_{\text{M31}}$ is the total mass and η is called the eccentric anomaly (Binney & Tremaine 2008). The LG has $\pi < \eta < 2\pi$ since the orbit is between the first apocentre and second pericentre and t is the age of the universe 13.75 ± 0.13 Gyr (Larson et al. 2010). Although $\eta > 2\pi$ is not theoretically excluded, the masses predicted are unphysically large. The components of the velocity vector of this orbital parameterization are

$$V_{\text{rad}} = \left(\frac{M}{a}\right)^{1/2}\frac{e\sin\eta}{1 - e\cos\eta}$$
$$V_{\text{tan}} = \left(\frac{M}{a}\right)^{1/2}\frac{\sqrt{1 - e^2}}{1 - e\cos\eta} \tag{4.25}$$

(Kochanek 1996; van der Marel & Guhathakurta 2008) where V_{rad} is the radial velocity along the vector from the Galactic centre to M31 and V_{tan} is the tangential velocity

in the orbital plane.

To convert V_{rad} and V_{tan} to a vector in the observed heliocentric velocity frame, we require several more parameters: the position angle of V_{tan} (Θ_{tan}), the distance to M31, the distance between the Sun to the Galactic centre and the solar motion relative to the Galaxy.

Chapter 5

Method

Probabilistic parameter estimation is the practice of inferring constraints on the parameters of a chosen model by minimizing a physically justified objective function of the data. This procedure is the only rigorous method of constraining physical quantities of interest within the limitations of the available data. To construct the objective function for a given dataset, we must postulate a physically motivated generative model for the data. A generative model is a parameterized description of a statistical process from which the data could be sampled, such as a phase-space distribution function. The generative model then motivates an objective function which is the probability of the data being drawn from a given particular set of model parameters.

In chapter 4, we outlined our general disk galaxy model. From this model we can generate all five of the datasets described in chapter 3 and hence calculate a combined objective function. This leaves the engineering challenge of optimizing this objective function to determine the distribution of physical parameters that satisfy the datasets. In this chapter, we will outline the method that we use to sample the distribution of model parameters that satisfy the available data to develop rigorous

probabilistic constraints on the physical quantities of interest.

To constrain our model parameters using all the available datasets, we are interested in evaluating and sampling the distribution called the posterior PDF or the probability of a model given the data $p(\boldsymbol{\Theta}|\mathbf{X})$. Here and throughout $\mathbf{X}$ is the set of all the data and $\boldsymbol{\Theta}$ is the vector of model parameters. The posterior PDF can be calculated from the likelihood of the data given a model $p(\mathbf{X}|\boldsymbol{\Theta})$ using Bayes' theorem

$$p(\boldsymbol{\Theta}|\mathbf{X}) = \frac{p(\boldsymbol{\Theta}) \times p(\mathbf{X}|\boldsymbol{\Theta})}{p(\mathbf{X})} \tag{5.1}$$

where $p(\boldsymbol{\Theta})$ is our prior knowledge and $p(\mathbf{X})$ is a normalization constant. We can then integrate this over any number, N, of nuisance parameters $\boldsymbol{\Theta}_{\mathrm{nuis}}$ to find the marginalized posterior PDF for the model parameters of interest $\tilde{\boldsymbol{\Theta}}$

$$p(\tilde{\boldsymbol{\Theta}}|\mathbf{X}) = \int p(\boldsymbol{\Theta}|\mathbf{X}) \, d^N \boldsymbol{\Theta}_{\mathrm{nuis}} \,. \tag{5.2}$$

The integration in equation 5.2 is generally intractable analytically but it can be evaluated efficiently by Markov chain Monte Carlo (MCMC). In particular, we employ a Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970) with simulated annealing (Gregory 2005; MacKay 2003).

5.1 Markov Chain Monte Carlo

For our purposes, MCMC is an extension of simple Monte Carlo numerical integration that uses a random walk to sample the integrand. After a sufficient number of steps in the chain, the distribution of samples in the chain converges to the PDF of interest.

5.1.1 The Standard Metropolis-Hastings Method

The most common method for evaluating a posterior PDF is called the Metropolis-Hastings method and it was initially proposed by Metropolis et al. (1953) and adapted to the current algorithm by Hastings (1970).

To solve the integral

$$I \equiv \int_V f(\boldsymbol{\Theta}) \, d^M \boldsymbol{\Theta} \tag{5.3}$$

using basic Monte Carlo integration, we sample N points $\boldsymbol{\Theta}_i$ at random from the M-dimensional volume V. Then, the solution of the integral is

$$I \approx \frac{V}{N} \sum_{i=1}^{N} f(\boldsymbol{\Theta}_i) \tag{5.4}$$

with error on I decreasing as $1/\sqrt{N}$. The main shortcoming of this method is that if $f(\boldsymbol{\Theta})$ is sharply peaked anywhere in V, too many samples will be drawn from regions with little contribution to the final integral. To improve the efficiency of this method, MCMC samples are correlated.

Given a position $\boldsymbol{\Theta}_i$ in V, we propose a new sample $\boldsymbol{\Theta}_{i+1}$ drawn from the easily calculable proposal distribution $q(\boldsymbol{\Theta}_{i+1}|\boldsymbol{\Theta}_i)$ that depends only on $\boldsymbol{\Theta}_i$. This sample is accepted with probability

$$r \equiv \min \left(1, \frac{p(\boldsymbol{\Theta}_{i+1}|\mathbf{X})}{p(\boldsymbol{\Theta}_i|\mathbf{X})} \frac{q(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{i+1})}{q(\boldsymbol{\Theta}_{i+1}|\boldsymbol{\Theta}_i)} \right). \tag{5.5}$$

Otherwise, $\boldsymbol{\Theta}_i$ is *sampled again*. In what follows, we will only consider symmetric proposal distributions where $q(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{i+1}) = q(\boldsymbol{\Theta}_{i+1}|\boldsymbol{\Theta}_i)$. For these proposal distributions,

equation 5.5 simplifies to

$$r \equiv \min\left(1, \frac{p(\boldsymbol{\Theta}_{i+1}|\mathbf{X})}{p(\boldsymbol{\Theta}_i|\mathbf{X})}\right). \tag{5.6}$$

The major shortcoming of the standard Metropolis-Hastings method as discussed here is that it is only efficient if the proposal distribution is a good approximation of the posterior PDF. If the proposed step size is too large then too few samples will be accepted and it will take a restrictively large number of steps for the chain to converge to its equilibrium distribution. On the other hand, if the step size is too small, the chain will not explore the full parameter space. In this case, there is no guarantee that the resulting chain is not "stuck" in a secondary extremum of the PDF. To choose the proposal distribution in a data-driven manner and to ensure that the chain settles to the highest probability state, we employ simulated annealing, reevaluating the proposal distribution at each temperature.

5.1.2 Simulated Annealing

Inspired by statistical mechanics, simulated annealing involves "heating" the motion of the Markov chain in order to explore the full parameter space. The chain is subsequently "cooled" into a lower energy — higher probability — state than would have been otherwise possible. In practice, we start by heating the chain to the temperature T_{max} with

$$\log_2 T_{\mathrm{max}} \equiv \left\lfloor \log_2\left(\frac{\sqrt{\langle\sigma^2\rangle - \langle\sigma\rangle^2}}{2}\right) \right\rfloor \tag{5.7}$$

where σ is the variance of an initial warm-up chain. Equation 5.5 is then replaced by

$$r \equiv \min \left[1, \left(\frac{p(\mathbf{\Theta}_{i+1}|\mathbf{X})}{p(\mathbf{\Theta}_i|\mathbf{X})} \right)^{1/T} \right] \tag{5.8}$$

and we run a chain with this acceptance probability. The temperature is then incrementally decreased by a factor of two until it reaches unity. At higher temperatures, the chain can freely explore the full parameter space and "find" the regions of high probability. As the chain cools, the chain is driven to higher probability states and eventually it settles to the maximum.

Each time the temperature is reduced, the proposal distribution is reevaluated to yield an ideal acceptance rate of $\sim 23\%$ at $T = 1$ (Roberts et al. 1997). At each step, the proposal distribution is re-defined as the M-dimensional Gaussian

$$q(\mathbf{\Theta}_{i+1}|\mathbf{\Theta}_i) \equiv \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{\Theta}_{i+1} - \mathbf{\Theta}_i)^{\mathrm{T}} \Sigma^{-1} (\mathbf{\Theta}_{i+1} - \mathbf{\Theta}_i) \right] \tag{5.9}$$

where the covariance Σ is chosen empirically as

$$\Sigma \equiv \frac{2.4^2}{4M} \mathbf{C}, \tag{5.10}$$

M is the dimensionality of the parameter space and $\mathbf{C}$ is the covariance matrix of the previous, hotter chain.

5.2 Calculating the likelihood

5.2.1 Satellite kinematics

Sales et al. (2007) analyzed the distribution of satellite galaxies formed in N-body/gas-dynamical simulations including dark matter, gas and stars and found that within the statistical errors, the satellites follow the same density distribution as the dark matter. Following this result, we assume that the satellite galaxies are sampled directly from the dark matter distribution. The true distribution of satellites is probably more complex than this simple assumption (e.g. Ibata et al. 2007), however, given the sparseness of the dataset, this is a necessary assumption.

Under this assumption, the phase space distribution function of satellite galaxies will be identically $f_{\mathrm{h}}(\mathcal{E})$ (the halo DF from equation 4.7) with $\mathcal{E} = \Psi(s, z) - \left(v_{\mathrm{los}}^2 + v_{\mathrm{t}}^2\right)/2$ where (s, z) are the M31-centric cylindrical coordinates of the satellite (appendix B), $v_{\mathrm{los}} \equiv v_{\mathrm{S,A}}^{\mathrm{los}}$ is given by equation 3.7 and v_{t} is the unknown satellite velocity in the plane of the sky. Therefore, the likelihood of observing satellite i at its position in phase space data given a model Θ is

$$p_{\mathrm{s}}(\mathbf{X}_i | \Theta, I) = \int d^6 \mathbf{X} \, S(\mathbf{X} | \mathbf{X}_i) f_{\mathrm{h}}(\mathcal{E}) \bigg/ \int d^6 \mathbf{X} \, f_{\mathrm{h}}(\mathcal{E}) \qquad (5.11)$$

where $S(\mathbf{X} | \mathbf{X}_i)$ is the distribution of the measured quantities (including the uncertainties). For our entire sample of satellite galaxies, 3 components of position space

(right ascension, declination and distance) and one component of velocity space (line-of-sight) have been measured while the two transverse velocity components are unknown. Therefore, the first integral in equation 5.11 can be rewritten

$$\int d^6\mathbf{X}\, S(\mathbf{X}|\mathbf{X}_i) f_{\mathrm{h}}(\mathcal{E}) \;=\; 2\pi \int_{-\infty}^{\infty} dD \int_{-\infty}^{\infty} dv_{\mathrm{obs}} \int_{0}^{v_{\mathrm{max}}} v_{\mathrm{t}}\, dv_{\mathrm{t}}\, g(D, v_{\mathrm{obs}}, v_{\mathrm{t}}) \quad (5.12)$$

where

$$g(D, v_{\mathrm{obs}}, v_{\mathrm{t}}) = \tilde{\mathcal{N}}(D|\xi_i, \omega_i, \alpha_i)\, \mathcal{N}(v_{\mathrm{obs}}|v_{\mathrm{obs},i}, \delta v_{\mathrm{obs},i})\, f_{\mathrm{h}}(\mathcal{E}), \qquad (5.13)$$

$v_{\mathrm{max}} \equiv \sqrt{2\Psi - v_{\mathrm{los}}^2}$ is the maximum transverse component of the satellite's velocity; $v_{\mathrm{obs}} \equiv \mathbf{v}_{\mathrm{S}} \cdot \hat{l}_{\mathrm{S}}$ is the observed line-of-sight velocity; $\mathcal{N}(v_{\mathrm{los}}; v_{\mathrm{los},i}, \delta v_{\mathrm{los},i})$ is the Gaussian PDF; and $\tilde{\mathcal{N}}(D; \xi_i, \omega_i, \alpha_i)$ is the skew-normal parameterization of the distance measurement (see appendix A). The angular position integrals are simply delta functions (because the measurement uncertainty is negligible) so they have been left out of equation 5.12.

Using the coordinate transformations from appendix B and section 3.6, the integration in equation 5.12 can be evaluated numerically unsing Monte Carlo integration with importance sampling for D and v_{obs} and Simpson's rule for v_{t}. We calculate the full log-likelihood of the satellite data given the model by taking the logarithm of equation 5.12 and summing over the set of satellites

$$\ln p_{\mathrm{s}}(\mathbf{X}|\Theta, I) = \sum_i \ln p_{\mathrm{s}}(\mathbf{X}_i|\Theta, I). \qquad (5.14)$$

Similarly, the normalization due to the second integral in equation 5.11 is

$$N_{\mathrm{s}} \times \ln\left[\int d^6\mathbf{X}\, f_{\mathrm{h}}(\mathcal{E})\right] = N_{\mathrm{s}} \ln M_{\mathrm{h}} \qquad (5.15)$$

where N_{s} is the number of satellites in the sample and the integral is the total halo mass, by definition. In practice, we assume spherical symmetry and integrate equation 4.9 from $r = 0$ to r_{max} where r_{max} is approximately the distance of the furthest satellite from M31 (the assumed edge of the survey). We choose a constant limiting radius $r_{\mathrm{max}} = 550$ kpc but the results are insensitive to this choice.

5.2.2 Globular cluster distribution

Like the satellite galaxies, we expect the GCs to trace the potential of the galaxy. It is not realistic, however, to assume that the clusters *sample* the dark matter halo since even GCs in the outer stellar halo are much more centrally concentrated than the satellites. To this end, we parameterize the GC density distribution as a single power law

$$\rho_{\mathrm{g}}(r) = \frac{\sigma_{\mathrm{g}}^2}{4\pi a_{\mathrm{g}}^2} \left[1 + \left(\frac{r}{a_{\mathrm{g}}} \right)^2 \right]^{-\alpha_{\mathrm{g}}}. \tag{5.16}$$

It has been argued that the three-dimensional M31 GC density distribution falls off as $r^{-3.5}$ at large radii (e.g. Crampton et al. 1985) which would imply that $\alpha_{\mathrm{g}} = 1.75$. We test this assumption by projecting equation 5.16 (with α_{g} fixed at 1.75) along the line-of-sight and scaling it to match the cumulative number density of all 40 confirmed GCs from the RBC with projected radius $\gtrsim 20$ (figure 5.1). The fit is qualitatively very good and given the small sample size, we fix $\alpha_{\mathrm{g}} = 1.75$ in all that follows. We also fix $a_{\mathrm{g}} = 10$ kpc since, as long as $a_{\mathrm{g}} \lesssim 20$ kpc, the outer GCs will be unaffected by its value.

From the parameterization in equation 5.16, we calculate the phase space DF of
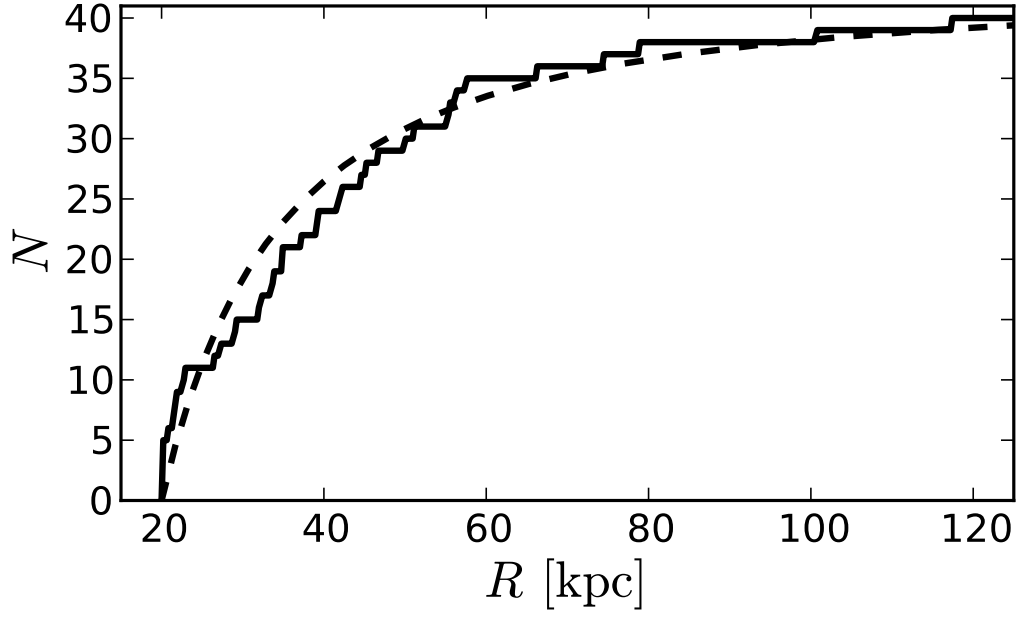
Figure 5.1 The cumulative number of confirmed GCs with $R \gtrsim 20$ kpc from the RBC (solid line) and the scaled number density produced by projecting equation 5.16 (with $\alpha_{\mathrm{g}} = 1.75$) along the line-of-sight (dashed curve).

the GCs using the Eddington inversion

$$f_{\mathrm{g}}(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2} \int_0^{\mathcal{E}} \frac{d\Psi_{\mathrm{tot}}}{\sqrt{\mathcal{E} - \Psi_{\mathrm{tot}}}} \frac{d^2\rho_{\mathrm{g}}}{d\Psi_{\mathrm{tot}}^2} \tag{5.17}$$

where Ψ_{tot} is the full multicomponent axisymmetric galaxy potential (Binney & Tremaine 2008).

From this DF, we construct the expression analogous to equation 5.12 for the GCs. The distance integral is different, however, because no distance measurements have been made for any of the clusters in our sample. Instead, the likelihood of observing

$$\int d^6\mathbf{X}\, S(\mathbf{X}|\mathbf{X}_i) f_{\mathrm{g}}(\mathcal{E}) = 2\pi \int_{D_{-,i}}^{D_{+,i}} dD \int_{-\infty}^{\infty} dv_{\mathrm{los}}\, \mathcal{N}(v_{\mathrm{los}}|v_{\mathrm{los},i}, \delta v_{\mathrm{los},i})$$
$$\int_0^{v_{\mathrm{max}}} v_{\mathrm{t},i}\, dv_{\mathrm{t},i}\, f_{\mathrm{g}}(\mathcal{E}) \tag{5.18}$$

where $D_{\pm,i}$ are the maximum and minimum distance at which the cluster could be be located and still be closer to M31 than $r_{\mathrm{max,g}}$, the limiting radius of the GC survey. For the limiting radius, we choose 200 kpc but the results are again insensitive to this choice.

The GC normalization constant is similar to equation 5.15 but the region of integration is no longer spherical because we have restricted our sample to contain only clusters beyond a projected radius of $\sim 1.5°$. Therefore, we must integrate equation 5.16 over a sphere of radius $r_{\mathrm{max,g}}$ minus a cone along the line-of-sight. This integral is easily evaluated using Monte Carlo integration.

5.2.3 Rotation curve

The circular velocity in the disk of our galaxy model is simply

$$v_c^2(s) = -sF_s(s) \tag{5.19}$$

where s is cylindrical radius and F_s is the force in the equatorial plane. The force is calculated by taking the gradient of equation 4.6 and summing the spherical harmonics. If we assume Gaussian uncertainties on the rotation curve measurements then the log-likelihood of the data is simply

$$\ln \mathcal{L}_{\mathrm{RC}} = K - \frac{1}{2} \sum_i \frac{[v_i - v_c(R_i)]^2}{\delta_i^2} \tag{5.20}$$

where K is a constant and we sum over each data point with rotational speed $v_i \pm \delta_i$ at projected radius R_i.

Equation 5.20 is, however, much too prescriptive. It is clear from figure 3.4 that there is significant velocity substructure in the rotation curve measurements. This substructure could be due to non-circular motions or spiral arms and we have no chance of modelling these details with our smooth model. To quantify this effect, we introduce a "noise" parameter $\sqrt{\epsilon_{\mathrm{RC}}}$ added to δ_i in quadrature. Introducing this parameter allows us to constrain the model without being overly restrictive and the posterior PDF of ϵ_{RC} yields a measure of the velocity substructure in the rotation curve. With this new parameter, the log-likelihood of the data becomes

$$\ln \mathcal{L}_{\mathrm{RC}} = K - \frac{1}{2} \sum_i \left[\ln \left(\delta_i^2 + \epsilon_{\mathrm{RC}} \right) + \frac{(v_i - v_c)^2}{\delta_i^2 + \epsilon_{\mathrm{RC}}} \right]. \tag{5.21}$$

The first term in the sum on the right-hand side of equation 5.21 penalizes overly general values of ϵ_{RC}.

5.2.4 Bulge velocity dispersion

Assuming isotropic velocities, the radial velocity dispersion of a spherical system with the density profile ρ is

$$\sigma_r^2(r) = \frac{1}{\rho(r)} \int_r^\infty \frac{\rho(\tilde{r})\, M(\tilde{r})}{\tilde{r}^2} dr \tag{5.22}$$

(Binney 1980) where $M(r)$ is the mass within radius r. For our composite system, we can use this equation to estimate the velocity dispersion near the centre of the galaxy by replacing $M(r)$ with $M_{\mathrm{tot}}(r)$, the total mass within radius r. The projected line-of-sight velocity dispersion is therefore

$$\sigma_{\mathrm{P}}^2(R) = \frac{2}{\Sigma_{\mathrm{b}}(R)} \int_R^\infty \frac{\sqrt{r^2 - R^2}}{r^2} \rho_{\mathrm{b}}(r)\, M_{\mathrm{tot}}(r)\, dr \tag{5.23}$$

where R is the projected radius of the line-of-sight (Prugniel & Simien 1997; Terzić & Graham 2005). As discussed above, we choose to constrain the dispersion at $R \sim 60''$ to be $\sigma_{\mathrm{obs}} \pm \delta\sigma = 172 \pm 5$ km s^{-1}. In analogy to equation 5.21, we add a fudge parameter ϵ_{b} to the observed uncertainty since $\delta\sigma$ is a lower limit. Assuming Gaussian uncertainty, the log-likelihood of the data is

$$\ln \mathcal{L}_{\mathrm{b}} = K - \frac{1}{2} \ln \left(\delta\sigma^2 + \epsilon_{\mathrm{b}}\right) - \frac{1}{2} \frac{\left(\sigma_{\mathrm{obs}} - \sigma_{\mathrm{P}}\right)^2}{\delta\sigma^2 + \epsilon_{\mathrm{b}}} \tag{5.24}$$

where K is a constant and σ_{P} is the model dispersion at $R = 60''$. Again, the second term on the right hand side of equation 5.24 penalizes large values of ϵ_{b} for being

overly general.

5.2.5 Surface brightness profile

We calculate the surface brightness profile of our model $\mu_R(R)$ in the R-band using equation 4.23 and compare it to the data assuming a Gaussian uncertainty with variance $\sqrt{\epsilon_\mu}$. The likelihood of the data is therefore

$$\ln \mathcal{L}_\mu = K - \frac{1}{2} \sum_i \left[\ln \epsilon_\mu + \frac{[\mu_i - \mu_R(R_i)]^2}{\epsilon_\mu} \right] \tag{5.25}$$

where the observed surface brightness is μ_i at projected radius R_i.

5.3 Assumed priors

Table 5.1 lists all the model parameters and their assumed priors.

For any dimensioned parameter that can vary from zero to infinity, it is best to assume a Jeffreys prior, especially if there is large uncertainty in the physically allowed range (Gregory 2005). The Jeffreys prior implies scale invariance and it is equivalent to sampling in the logarithm of the parameter of interest. To this end, we adopt uniform priors on the logarithm of the scale parameters in our model unless a more informative prior is available. For the cusp strength of our halo model and the bulge Sérsic index, we also assume uniform priors.

For the R-band mass-to-light ratios of the disk and bulge, we set a conservative hard upper limit of $\Upsilon_R \leq 10 \, M_\odot/\mathcal{L}_{\odot,R}$ to ensure that the models don't underestimate the contribution of the baryonic component. For the components of the systematic transverse velocity, we assume a uniform prior on v_W and v_N but set a hard upper

limit on the total transverse velocity of $\sqrt{v_{\mathrm{W}}^2 + v_{\mathrm{N}}^2} < 400$ km s^{-1} for computational purposes.

Throughout, we adopt a Gaussian prior on the distance with mean 785 kpc and variance 25 kpc (McConnachie et al. 2005; McConnachie & Irwin 2006). Similarly, for the systematic heliocentric line-of-sight velocity, we adopt the value -301 ± 1 km s^{-1} (Courteau & van den Bergh 1999).

In order to calculate the mass of the LG group with the timing argument (section 4.3) we also require the distance to the Galactic centre and the solar motion with respect to the Milky Way. For the distance to the Galactic centre, we combine two recent estimates $D_{\mathrm{GC}} = 8.0 \pm 0.6$ kpc (Ghez et al. 2008) and $D_{\mathrm{GC}} = 8.33 \pm 0.35$ (Gillessen et al. 2009) with equal weights. For the circular velocity of the Milky Way at the radius of the Sun, we adopt the value $v_{\mathrm{circ}} = 236 \pm 11$ km s^{-1} (Bovy et al. 2009a) based on the combined constraints from the kinematics of Galactic masers, the proper motion of Sgr A* and orbital fitting of stellar streams. For the solar peculiar velocity, we adopt the components: 10.1 ± 0.5 km s^{-1} towards the Galactic centre, 4.0 ± 0.8 km s^{-1} in the direction of Galactic rotation and 6.7 ± 0.2 km s^{-1} towards the north Galactic pole (Hogg et al. 2005). The quoted uncertainties on the peculiar motion are probably underestimates due to the effect of moving groups in the Hipparcos data (Bovy et al. 2009b; Bovy & Hogg 2010) but this uncertainty is negligible compared to the other uncertainties in the problem and our final results are insensitive to this effect.

Table 5.1. Model parameters and their prior distributions

Parameter	Description	Prior
α_{M31}	Right ascension (J2000.0)[a]	$00^h42^m44.4^s$
δ_{M31}	Declination (J2000.0)[a]	$+41°16'08''$
θ	Position Angle[a]	$37.7°$
i	Inclination[a]	$77.5°$
D_{M31}	Distance to M31[b]	785 ± 25
D_{GC}	Distance to the Galactic centre[c]	8.2 ± 0.5
v_{circ}	Galactic circular velocity[d]	236 ± 11
$u_\odot$		10.1 ± 0.5
$v_\odot$	Solar peculiar velocity[e]	4.0 ± 0.8
$w_\odot$		6.7 ± 0.2
$v_{M31,los}$	Line-of-sight velocity[f]	301 ± 1
(v_W, v_N)	Systemic transverse velocity	$\sqrt{v_W^2 + v_N^2} < 400$
$\log_{10} a_h$	Halo scale length	$[0, 3]$
$\log_{10} v_h$	Halo characteristic velocity	$[0.5, 1.5]$
γ_h	Halo cusp	$[0, 2]$
α_h	Halo outer slope	$[2, 6]$
$\log_{10} M_d$	Disk mass scale	$[-1.6, -0.6]$
$\log_{10} R_d$	Disk scale length	$[0, 1]$
Υ_d	Disk R-band mass-to-light ratio	$[0, 10]$
$\log_{10} v_b$	Bulge characteristic velocity	$[0, 1]$
$\log_{10} R_e$	Bulge scale length	$[-2, 0.7]$
Υ_b	Bulge R-band mass-to-light ratio	$[0, 10]$
$\log_{10} \epsilon_{RC}$	Rotation curve noise parameter	$[-1.5, 2]$
$\log_{10} \epsilon_b$	Bulge velocity dispersion noise parameter	$[-8, 3]$
$\log_{10} \epsilon_\mu$	Surface brightness profile noise parameter	$[-4, 3]$
$\log_{10} v_g$	GC characteristic velocity	$[0, 3]$
t_0	The age of the Universe[g]	13.75 ± 0.13 Gyr
$\ln M_{LG}$	Total mass of the Local Group (LG)	$[0, 5]$
e	Ellipticity of the LG orbit	$[0, 1]$
η	Eccentric anomaly of the LG orbit	$[\pi, 2\pi]$
Θ_{LG}	Position angle of the LG orbital plane	$[0, 2\pi]$

Note. — Units: distance – kpc, speed – km s^{-1}, mass – $10^{12} M_\odot$ and mass-to-light ratios – $(M_\odot/L_\odot)_R$. Priors: $[a, b]$ – uniform in the quantity listed in column 1 between a and b, $x \pm \delta x$ – Gaussian with mean x and variance δx.

[a]de Vaucouleurs (1958); de Vaucouleurs et al. (1991)

[b]McConnachie et al. (2005)

[c]Ghez et al. (2008); Gillessen et al. (2009)

[d]Bovy et al. (2009a)

[e]Hogg et al. (2005)

[f]Courteau & van den Bergh (1999)

[g]Larson et al. (2010)

Chapter 6

Results

We now describe the results of the dynamical modelling procedure outlined in the previous chapters. We have analyzed three different subsets of the data to investigate the robustness of our conclusions to the inclusion and exclusion of data. Chain A was constrained by the bulge dispersion, rotation curve, surface brightness profile and satellite galaxy kinematics. Chain B used all the data from chain A except the two outlying satellite galaxies Andromeda XVI and XII. Chain C was run to fit the entire dataset, including the GCs. The decision to exclude the GCs from chains A and B was both scientific and practical. The density distribution of GCs is poorly constrained, none of the clusters have distance measurements and their velocity uncertainties are quite large. Therefore, we expected that the GC dataset would have little effect on the conclusions. Also, the likelihood calculation for the GC dataset is restrictively computationally expensive since it requires two extra numerical integrations for each cluster in the sample and a further integration to evaluate the normalization for the likelihood. The statistics obtained from chain C are not currently sufficient to be conclusive, so in what follows we will only discuss the results from chains A and B.

The results from chain C are still in preparation for publication. See section 7.2.1 for a discussion of some preliminary results of this analysis.

There is a strong covariance between many of the parameters in our model, making it impossible to ensure convergence to the absolute maximum probability in each model parameter. To study this effect, we ran five independent subchains for each of the chains discussed above. Each subchain was initialized with unique, random initial conditions and allowed to anneal to the highest probability state completely independently. After sufficient statistics were obtained, we combined these chains and performed our final analysis on the composite chain. It is still not clear that the full parameter space has been sampled for each parameter — especially the halo parameters — despite this precaution. Therefore, the results presented here should still be considered somewhat preliminary. The final conclusions should be unaffected, however.

Each subchain was parallelized and run on a dedicated node (8 cpus @ 1.60GHz) of the CITA Sunnyvale[1] cluster for > 1000 cpu-hours to improve the statistics.

6.1   Convergence Diagnostic

Quantifying convergence for Markov chains is difficult and all methods are currently empirical and somewhat ad-hoc (e.g. Cowles & Carlin 1996). It is unreasonable to expect that any diagnostic will correctly detect all convergence problems but recently, Dunkley et al. (2005) proposed a convergence diagnostic that is easy to implement and we will adopt this test for our work. The diagnostic was designed to test models of measurements of the cosmic microwave background power spectra and it has proven

---

[1]http://www.cita.utoronto.ca/mediawiki/index.php/Sunnyvale

effective for various applications including cosmological parameter estimation (Dunkley et al. 2009), high-redshift supernovae detection (Riess et al. 2007) and stellar population synthesis (Conroy et al. 2009).

Dunkley et al. (2005) proposed a parametric model of the power spectrum $\hat{P}$ of a single parameter Markov chain. For wavenumber j, the empirical model is

$$\ln \hat{P}(j) = \ln P_0 + \ln \left[\frac{(j^*/j)^\alpha}{1 + (j^*/j)^\alpha} \right] - 0.577216 \tag{6.1}$$

where P_0 quantifies the amplitude of white noise in the spectrum and j^* parameterizes the turnover from the white noise regime to a power-law behaviour. From this parameterization, Dunkley et al. (2005) estimated the "convergence ratio" $r = P_0/N$ where N is the length of the chain. For an ideally converged chain, $j^* \to \infty$ and $r \to 0$. Empirically, Dunkley et al. (2005) defined the convergence criteria $j^* > 20$ and $r < 0.01$.

For each chain in our sample, we fit equation 6.1 to the power spectrum using the iterative least-squares method from Dunkley et al. (2005) and confirmed that each parameter satisfied the above criteria.

6.2 Galaxy Properties

Most of the model parameters listed in table 5.1 can be considered nuisance parameters since the value of interest is the mass of M31. It is interesting, however, to quote the inferred distributions of these parameters and discuss various results. Table 6.1 summarizes the median and 1σ uncertainty ranges on a selection of parameters from table 5.1. The parameters with informative priors have been omitted since there is

no further information to be gained from those results and the parameters associated with the timing argument are discussed in section 6.6.

6.2.1 Dark Matter Halo

The top panel of figure 6.1 shows the marginalized posterior PDF for the halo's velocity scale and scale length. Interestingly, there is a positive correlation between these two parameters. This correlation stems from the fact that as the halo model becomes more extended, it must also be more massive if all of the satellite galaxies in the sample are to remain bound.

The halo model for chains A and B are nearly equivalent given the large statistical uncertainties. Both models are weakly cusped with a very steep outer slope. The final halo models are also both extremely massive and extended. Surprisingly, the halo profile is quite similar to a TF model (equation 2.4) but with a larger mass than predicted by Evans & Wilkinson (2000). This discrepancy is due to the strong covariance between disk mass and halo mass (bottom panel of figure 6.1) — or equivalently, scale length. It is likely (see section 7.2.2) that we have over-estimated the disk mass by a factor of ~ 2. If this is true, then the value of v_{h} is also too large.

6.2.2 Disk & Bulge

Figure 6.2 shows the marginalized posterior PDF for the disk scale parameters. Since the scale length is well constrained by the surface brightness profile, there is no significant covariance between the disk mass and scale length. Both chain A and chain B estimate the disk scale length $R_{\mathrm{d}} = 5.3 \pm 0.1$ kpc which is consistent with the Walterbos & Kennicutt (1988) measurement and more recent values (Widrow,

Table 6.1. Median and 1σ uncertainties for the model parameters

Parameter	Chain A	Chain B
$v_{\mathrm{W}}\,[\mathrm{km\,s}^{-1}]$	-19.9 ± 132.0	-72.2 ± 143.0
$v_{\mathrm{N}}\,[\mathrm{km\,s}^{-1}]$	78.7 ± 217.8	39.5 ± 181.7
ρ_{WN}	-0.526	0.550
$a_{\mathrm{h}}\,[\mathrm{kpc}]$	$79.4^{+42.3}_{-68.8}$	$119.5^{+47.0}_{-102.2}$
$v_{\mathrm{h}}\,[\mathrm{km\,s}^{-1}]$	$1258.6^{+343.1}_{-747.5}$	$958.1^{+607.8}_{-490.5}$
γ_{h}	$0.2^{+0.2}_{-0.1}$	$0.8^{+0.3}_{-0.5}$
α_{h}	$5.1^{+0.7}_{-2.2}$	$5.2^{+0.6}_{-2.4}$
$M_{\mathrm{d}}\,[10^{10}\,M_{\odot}]$	$11.6^{+1.0}_{-2.2}$	$11.4^{+1.3}_{-1.8}$
$R_{\mathrm{d}}\,[\mathrm{kpc}]$	5.2 ± 0.2	5.3 ± 0.1
$\Upsilon_{\mathrm{d}}\,[M_{\odot}/\mathcal{L}_{\odot,R}]$	$4.8^{+0.4}_{-0.7}$	$4.6^{+0.7}_{-0.6}$
$v_{\mathrm{b}}\,[\mathrm{km\,s}^{-1}]$	$464.9^{+15.7}_{-29.2}$	$438.0^{+29.3}_{-42.6}$
$M_{\mathrm{b}}\,[10^{10}\,M_{\odot}]$	2.8 ± 0.3	2.7 ± 0.4
$R_{\mathrm{e}}\,[\mathrm{kpc}]$	0.8 ± 0.1	0.9 ± 0.2
$\Upsilon_{\mathrm{b}}\,[M_{\odot}/\mathcal{L}_{\odot,R}]$	$1.4^{+0.1}_{-0.2}$	1.3 ± 0.2
$\sqrt{\epsilon_{\mathrm{RC}}}\,[\mathrm{km\,s}^{-1}]$	$34.3^{+1.9}_{-1.6}$	$34.9^{+1.7}_{-1.6}$
$\sqrt{\epsilon_{\mathrm{b}}}\,[\mathrm{km\,s}^{-1}]$	$6.1^{+78.8}_{-6.0}$	$16.7^{+66.3}_{-16.6}$
$\sqrt{\epsilon_{\mu}}\,[\mathrm{mag\,arcsec}^{-2}]$	0.24 ± 0.02	$0.23^{+0.03}_{-0.02}$
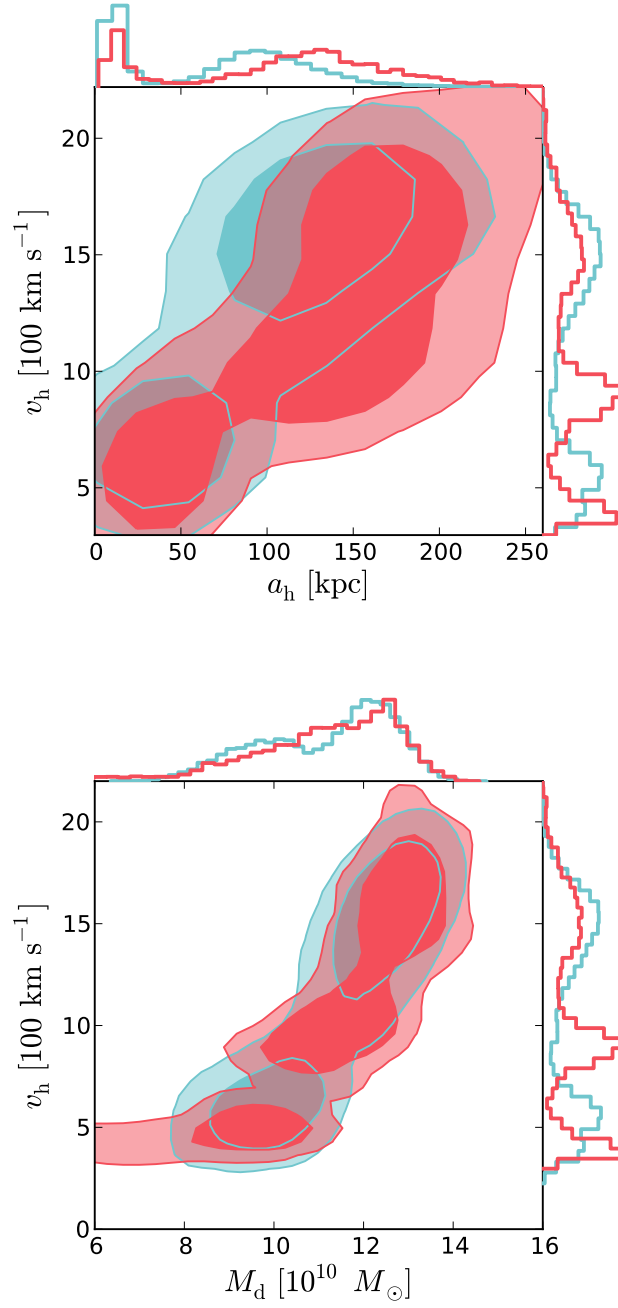
Figure 6.1 The marginalized posterior PDF of the halo scale parameters from chains A (blue) and B (red). The contour levels correspond to the 1- and 2-σ confidence intervals and the histograms show the projected PDF in each dimension.

private communication). In contrast, the estimated disk mass is somewhat larger than previous measurements (Klypin et al. 2002; Widrow & Dubinski 2005). It has been shown (Debattista & Sellwood 2000; Widrow et al. 2008) that if the disk dominates the gravitational force at the radius $R \sim 2.2 R_\mathrm{d}$, the galaxy is highly unstable and it will quickly form a strong bar that significantly disrupts the disk density distribution. In a very short time $\lesssim 2.5$ Gyr, the density profile no longer reproduces any observations whatsoever. This effect is important for any model with a large disk mass (see section 7.2.2).

The distribution of mass-to-light ratios fit by our modelling procedure is shown in figure 6.3. The covariances of these ratios with the disk and bulge parameters are shown in figure 6.4. Assuming the standard corrections for internal and foreground extinction (Binney & Merrifield 1998; Widrow & Dubinski 2005), our corrected mass-to-light ratios for the disk and bulge are $\sim 2.2 \, M_\odot/\mathcal{L}_{\odot,R}$ and $\sim 1.2 \, M_\odot/\mathcal{L}_{\odot,R}$ respectively. Following the prescription of Bell et al. (2003) and using the colours calculated by Walterbos & Kennicutt (1988), the mass-to-light ratios predicted by stellar population modelling are $\Upsilon_\mathrm{d} \sim 1.7 \, M_\odot/\mathcal{L}_{\odot,R}$ and $\Upsilon_\mathrm{b} \sim 2.2 \, M_\odot/\mathcal{L}_{\odot,R}$ with a predicted scatter of about 25% (Chemin et al. 2009). A rigorous comparison to this results is beyond the scope of this work but these mass-to-light ratios and those predicted by Widrow & Dubinski (2005) are consistent with our results within all the associated uncertainties.

6.3 Convergence to Observational Constraints

Both the rotation curve and surface brightness profile are extremely informative datasets. As a result, our modelling procedure constrains these values very tightly
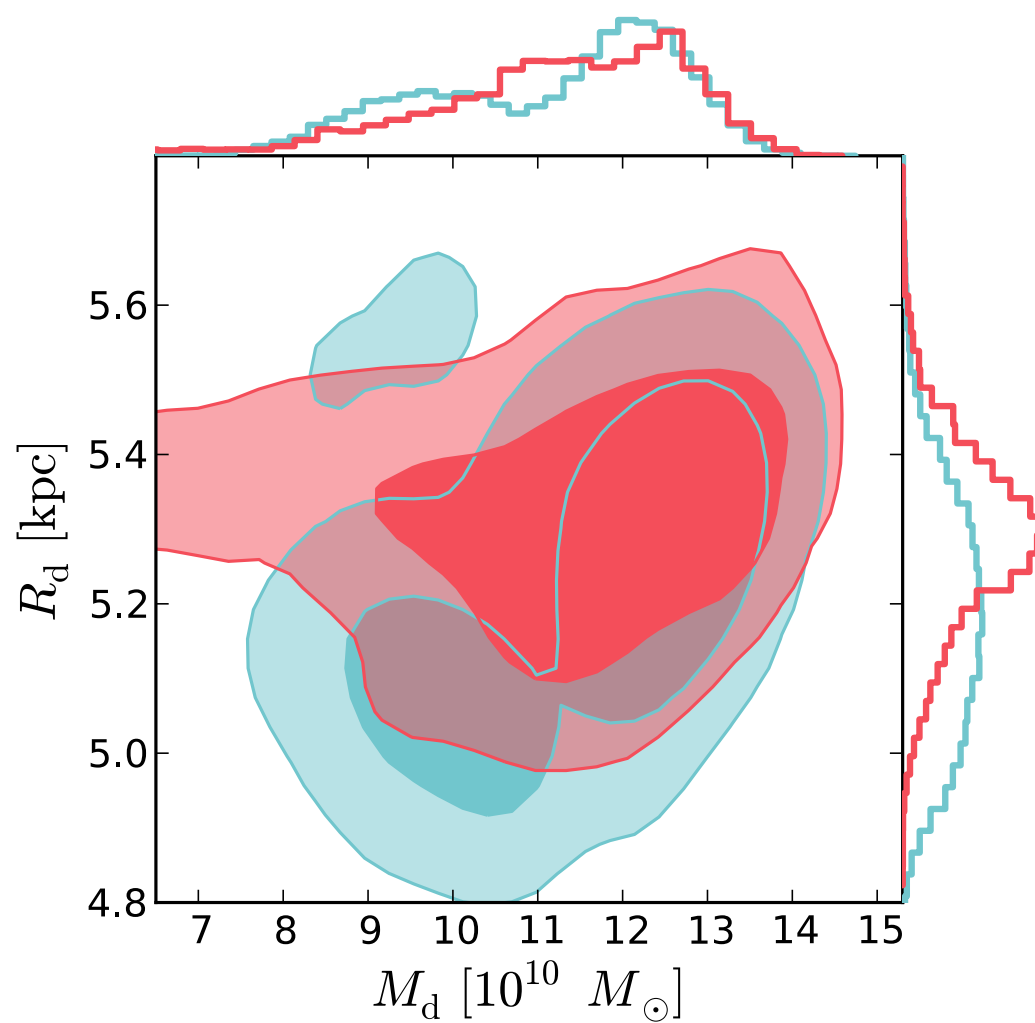
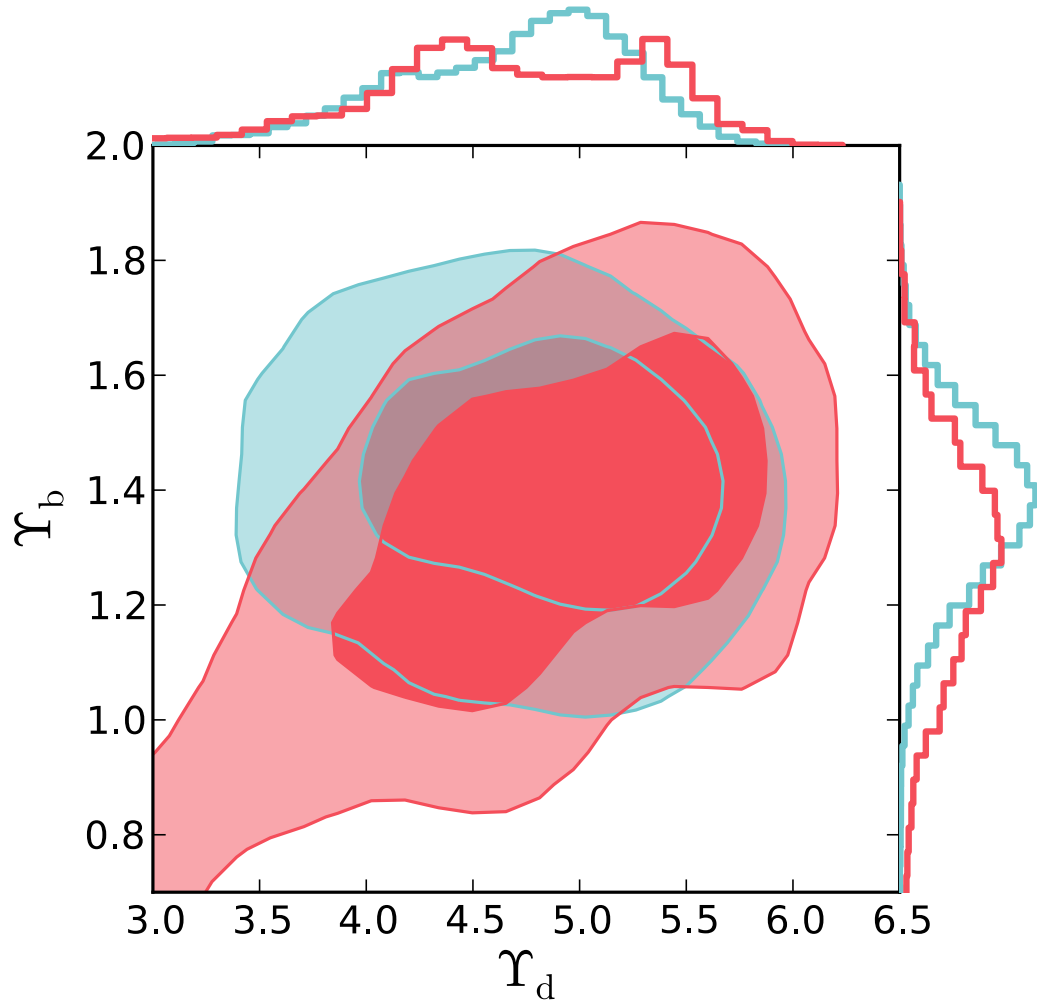Figure 6.2 Same as figure 6.1 for the disk scale parameters.

Figure 6.3 Same as figure 6.1 for the R-band mass-to-light ratios in units of $M_\odot/\mathcal{L}_{\odot,R}$.
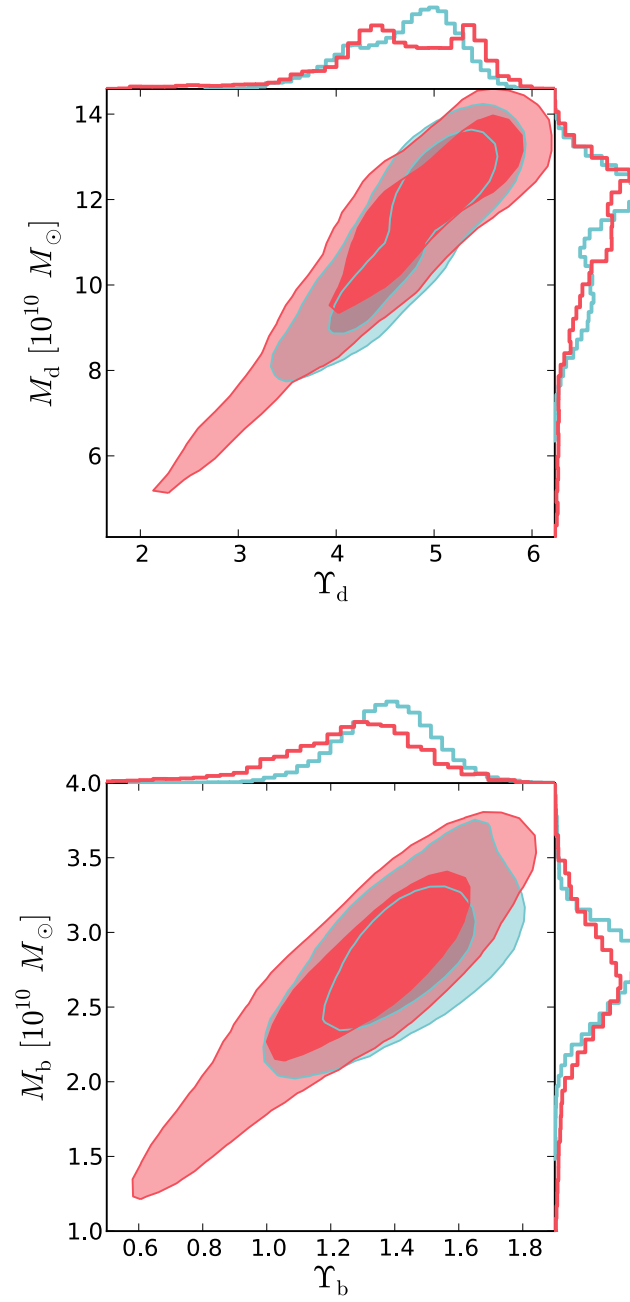
Figure 6.4 Same as figure 6.1 but showing the strong covariance between mass and mass-to-light ratio for the disk and bulge. The mass-to-light unit is $M_\odot/\mathcal{L}_{\odot,R}$.

and it fits the data well.

6.3.1 Rotation curve

Figure 6.5 shows the fully marginalized median and 1σ contour of the rotation curve for chains A and B. As expected, the statistical uncertainties are quite small (darkly shaded region in figure 6.5) and the fit is nearly identical for both chains. The distribution of $\sqrt{\epsilon_{RC}}$ (figure 6.6) quantifies the velocity substructure in the rotation curve. For both chain A and B, $\sqrt{\epsilon_{RC}} \sim 35$ km s^{-1}. This median value of $\sqrt{\epsilon_{RC}}$ is also shown in figure 6.5 added in quadrature to the statistical uncertainty in the rotation curve fit (lightly shaded region). This region perfectly encloses all of the structure in the data, implying that this parameter is fulfilling its function.

Figure 6.7 shows the relative contribution of each dynamical component to a representative rotation curve. The thin solid line in this figure is the contribution from the disk and it is clear that this component dominates at the radius $R \sim 2.2R_{d} \sim 50'$. As discussed previously, this result indicates that the model is highly unstable and susceptible to strong bar formation.

6.3.2 Surface brightness profile

Figure 6.8 shows the fully marginalized fits of equation 4.23 to the surface brightness profile data. The median disk and bulge contributions to the profile are also indicated as dashed and dotted lines respectively. As in figure 6.5, the median value of $\sqrt{\epsilon_{\mu}}$ is shown as the light grey region when added in quadrature to the statistical uncertainty on the fit. The fit to the profile of the disk is excellent at large radii but it is also clear that the quality of the fit in the inner regions is not quite as good. This implies that
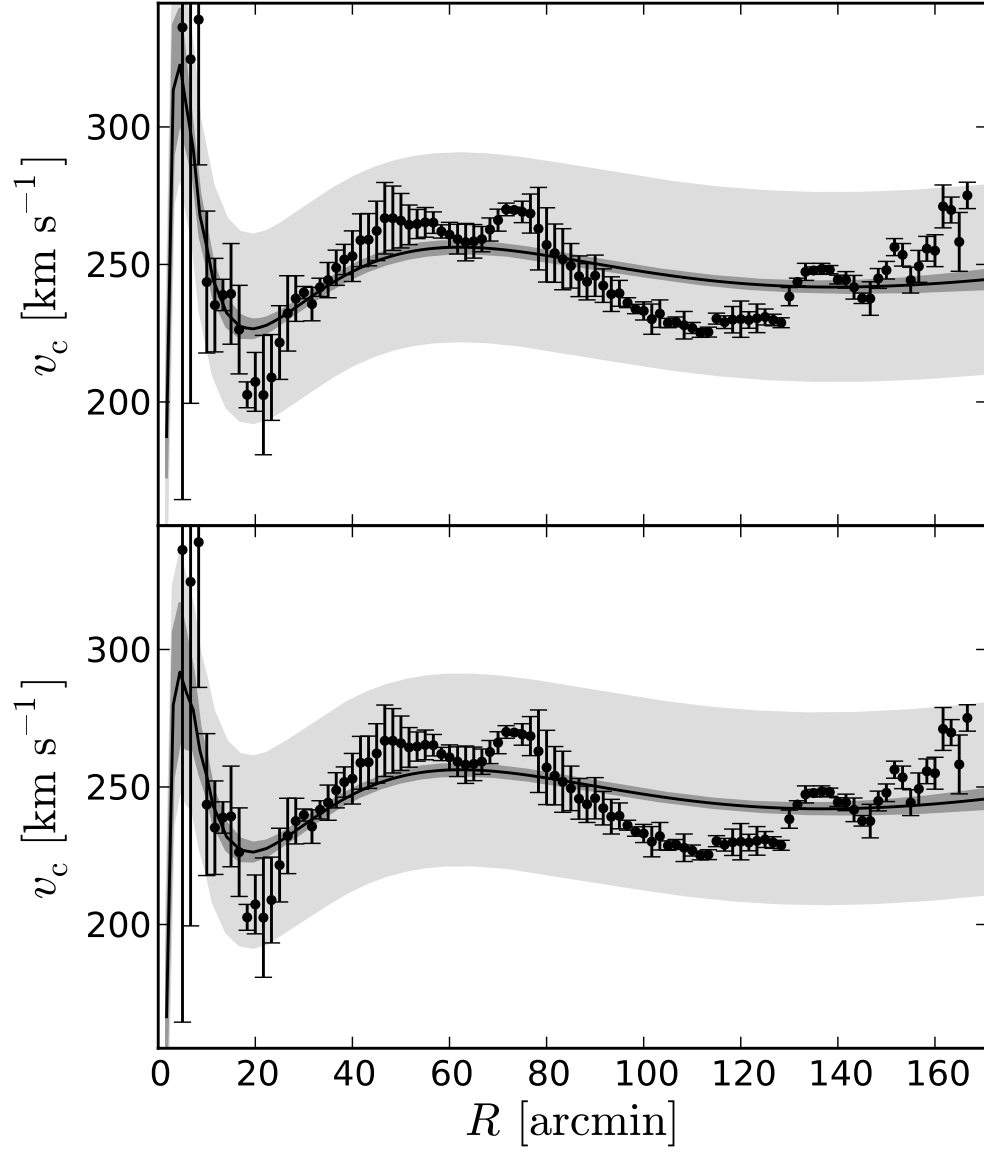
Figure 6.5 The error bars show the data from Chemin et al. (2009). The marginalized fit constrained by chains A (top) and B (bottom) are shown as the mean (black curve) and 1σ uncertainty (dark grey region). The light grey region shows the median value of $\sqrt{\epsilon_{\mathrm{RC}}}$ added to the statistical uncertainty in quadrature.
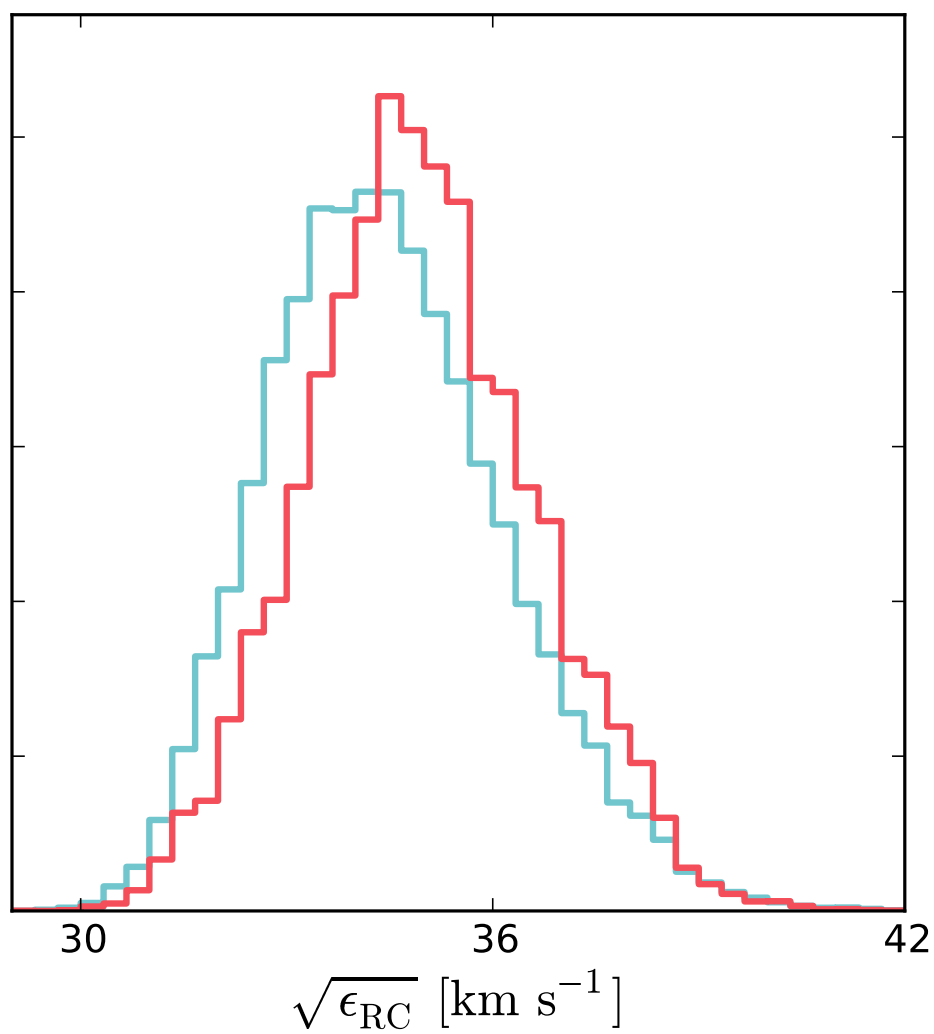
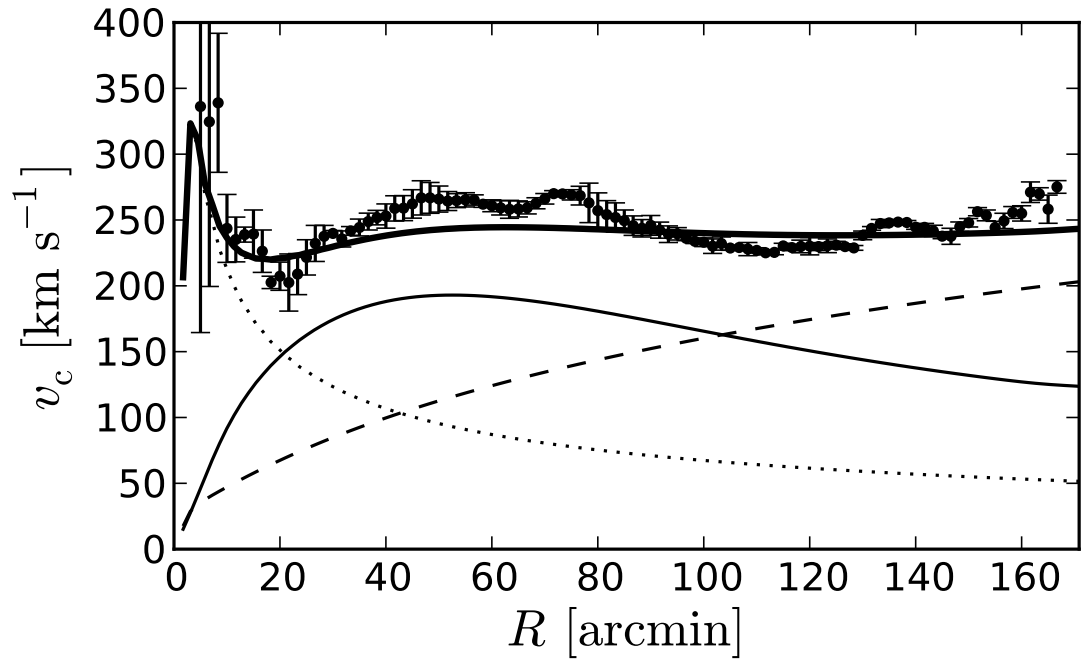Figure 6.6 The marginalized PDF of $\sqrt{\epsilon_{\mathrm{RC}}}$. The colours are the same as figure 6.1

Figure 6.7 The same data as figure 6.5 with a representative rotation curve plotted with the thick black line. The individual contributions from the disk (solid curve), bulge (dotted curve) and halo (dashed curve) are also shown.

our assumption that the Sérsic index of the bulge $n \equiv 2$ is not completely consistent with the data. However, as mentioned previously, the mass distribution at large radii is quite insensitive to the value of n so this does not affect our final conclusions.

6.4 The Mass of M31

For every model accepted in each chain, we calculate the composite mass profile as a function of spherical radius. The mass of a spherical density distribution within a given radius r is given by

$$M(r) = -r^2 \times \hat{\mathbf{e}}_r \cdot \mathbf{F}(r) \tag{6.2}$$

where $\hat{\mathbf{e}}_r$ is the radial unit vector and $\mathbf{F}$ is the gravitational force vector (Binney & Tremaine 2008). Since our galaxy is not spherical, we can't use equation 6.2 directly. Instead, we approximate the mass profile as

$$M(s, z = 0) = -s^2 \times \hat{\mathbf{e}}_s \cdot \mathbf{F}(s, z = 0) \tag{6.3}$$

where s is the cylindrical radius in the plane of the disk.

Figures 6.9 and 6.10 show the mass profile over nearly three orders of magnitude in radius as constrained by chains A and B respectively. Overplotted on this constraint are the literature values from figure 2.2. The results from both chains look nearly identical and the residual mass difference between these two models is on average $\lesssim 10\%$ although the scatter is nearly 50% at large radii (figure 6.11).

One feature that is immediately obvious is that the uncertainty on the mass in the range $\sim 4 - 40$ kpc is negligible compared to the uncertainty from Evans & Wilkinson (2000). This is the exact radial range where the likelihood is dominated
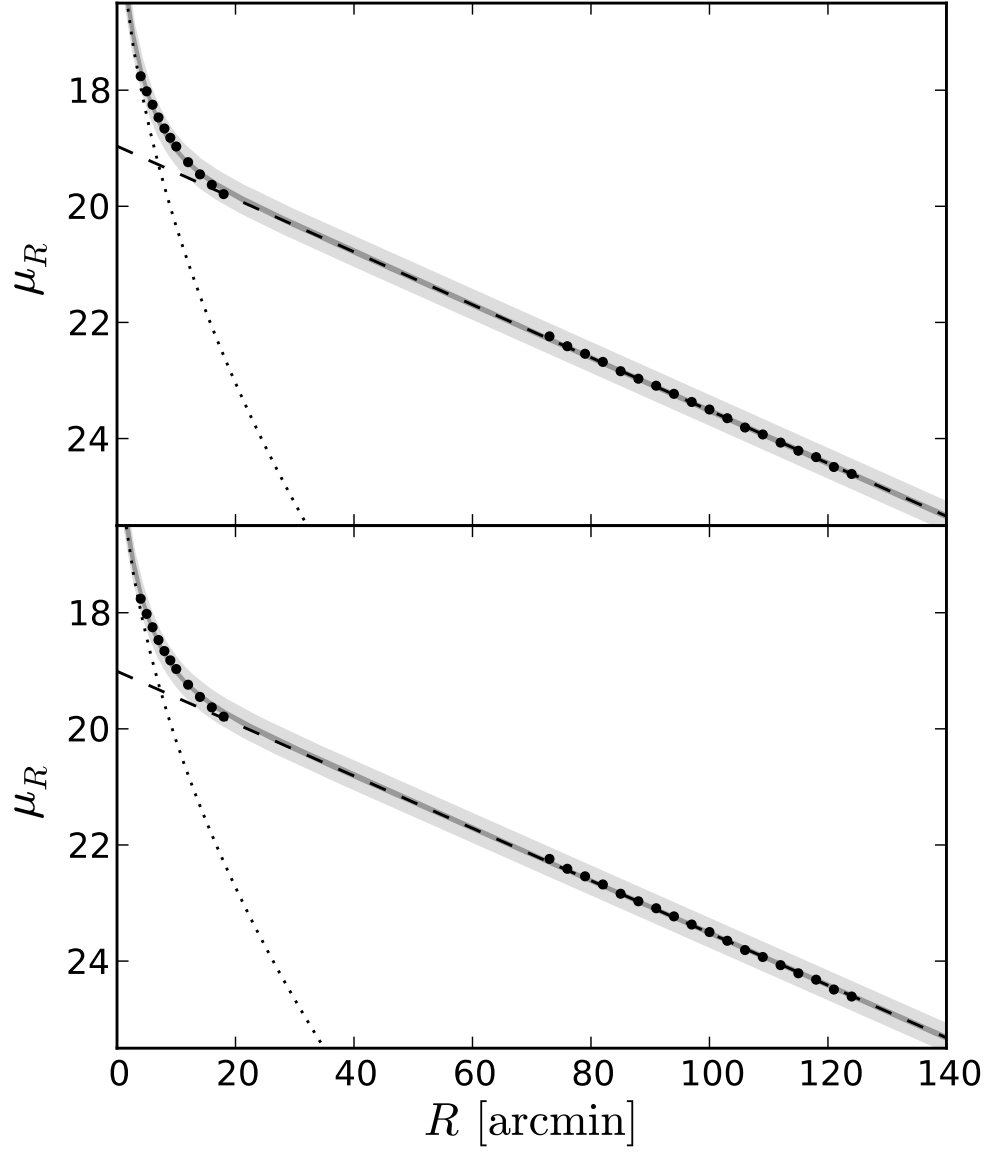
Figure 6.8 Like figure 6.5, the dark grey region shows the 1σ statistical uncertainty on the marginalized surface brightness profile constrained by chains A (top) and B (bottom). The light grey region has the median value of $\sqrt{\epsilon_\mu}$ added in quadrature to the statistical uncertainty. The black points are the data from figure 3.6.

by the rotation curve and surface brightness profile and as expected, the constraint is very tight. The mass contained within ~ 40 kpc agrees with the value estimated by Chemin et al. (2009) using the same rotation curve and surface brightness profile.

In the outer regions of the galaxy, the uncertainties on the mass profile blow up since the dataset is so small at large radii. Despite these large uncertainties, our mass estimate is larger than most literature values at the 1σ level and the results are consistent with the maximum mass estimated by Evans & Wilkinson (2000) at the 2σ level. This result shows that when the highly informative rotation curve and surface brightness profile are considered, an extremely massive and extended halo is preferred and this is not inconsistent with the satellite kinematics.

6.4.1 Connection to Cosmology

The mass of most of our galaxy models diverges at large radii but a good proxy for the total mass is the mass within the virial radius. There are various definitions of the virial radius in the literature and here we will consider two different definitions in order to compare to previous work.

The classical definition of virial radius is called r_{200} and it is defined as the radius within which the mean density is 200 times the critical density $\rho_{\mathrm{crit}} \equiv 3H^2/8\pi$ of the Universe (Navarro et al. 1996). In the definition of the critical density, H is the Hubble constant for which we adopt a recent value 71.0 km s^{-1} Mpc (Larson et al. 2010).

The definition of r_{200} is independent of the cosmological model, however, so Bullock et al. (2001) proposed a more general definition. These authors defined r_{vir} as the radius within which the mean density is Δ_{vir} times the mean matter density of
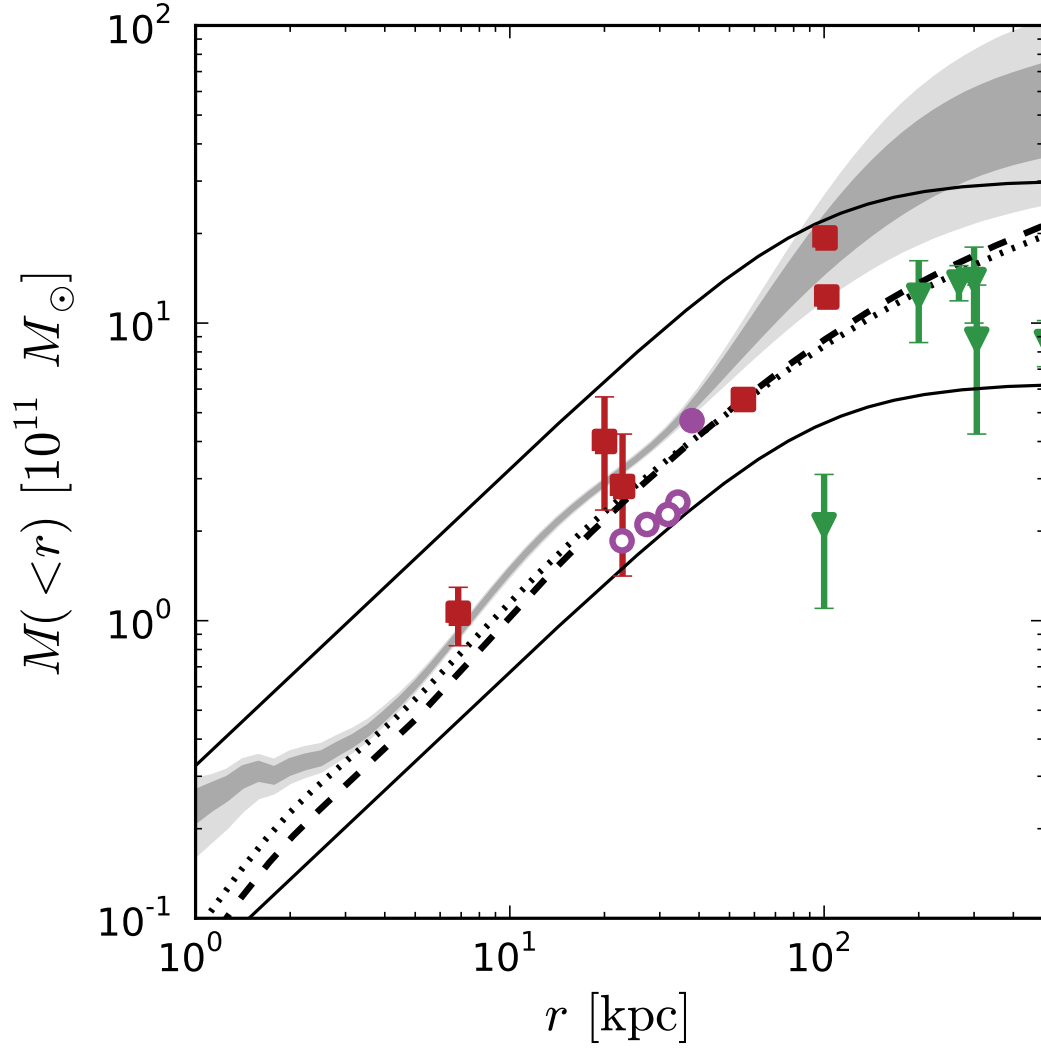
Figure 6.9 The 1σ uncertainty on the marginalized mass profile constrained by chain A (shaded region). The curves and points are the same as figure 2.2.
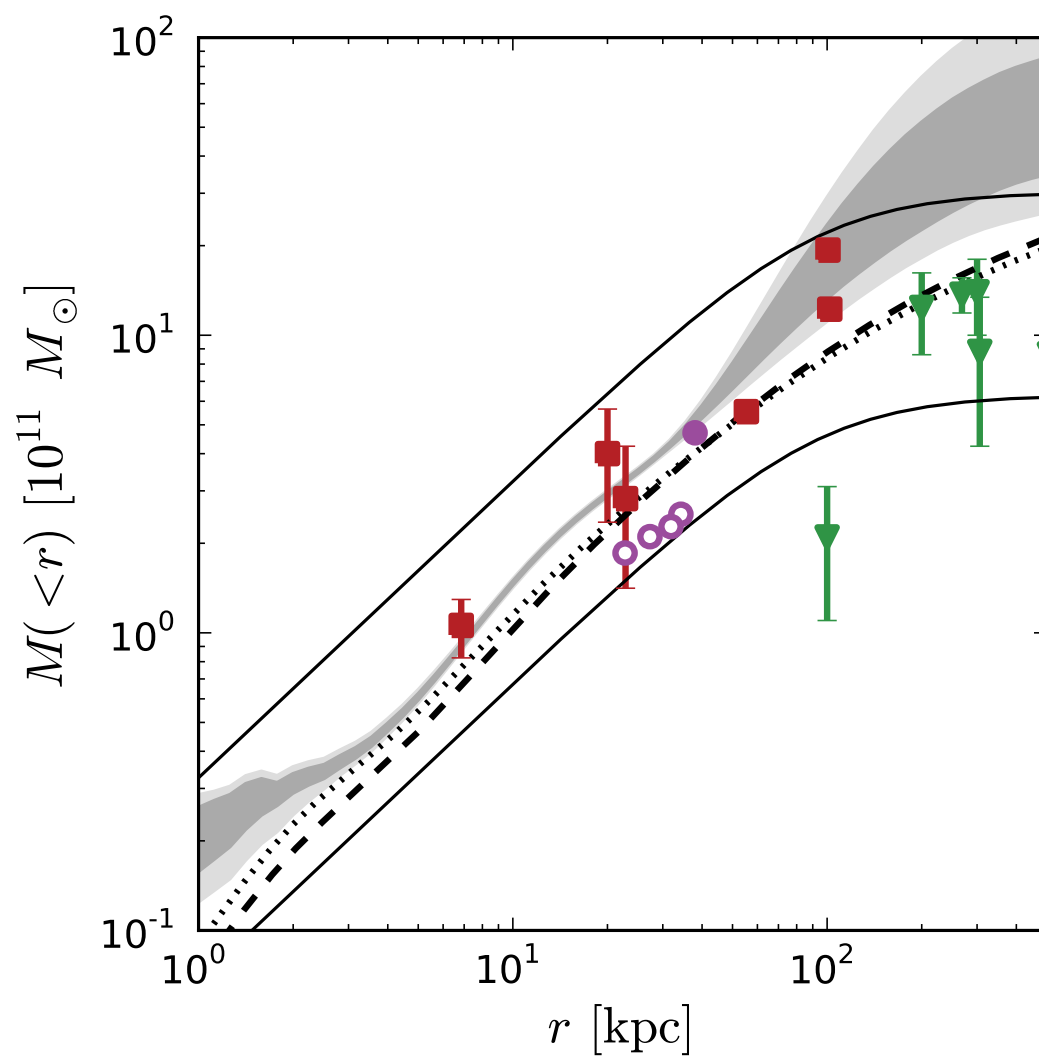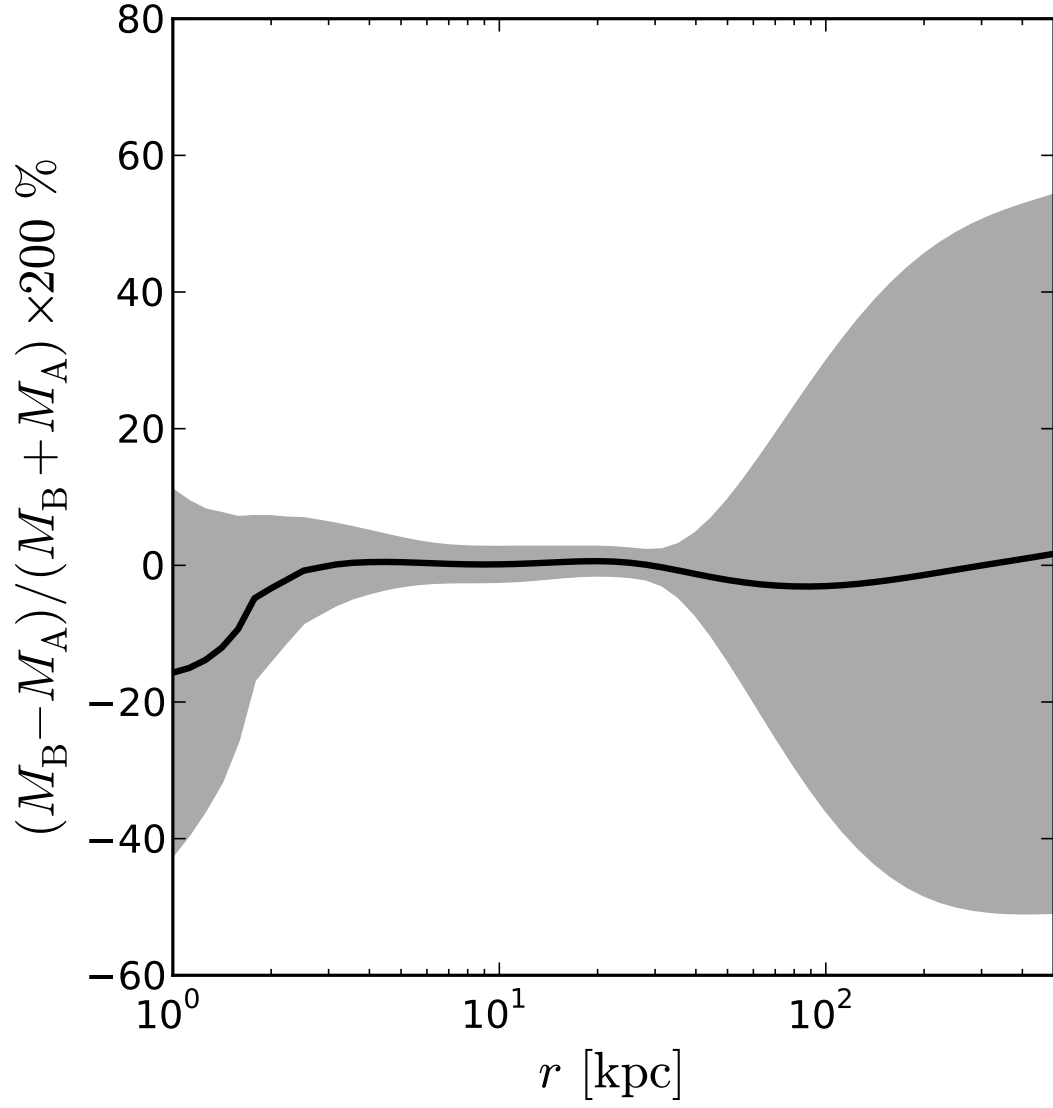
Figure 6.10 The same as figure 6.9 for chain B.

Figure 6.11 The residual mass difference between chains A and B as a function of radius. The mean is shown in black and the 1σ uncertainty on the variance is shown as grey. The uncertainties have been propagated by Monte Carlo.

the Universe $\rho_{\mathrm{m}} \equiv \Omega_{\mathrm{m}} \rho_{\mathrm{crit}}$ and it can be calculated by solving the equation

$$M_{\mathrm{vir}} \equiv M(< r_{\mathrm{vir}}) = \frac{4\pi}{3} \Delta_{\mathrm{vir}} \rho_{\mathrm{m}} r_{\mathrm{vir}}^3. \tag{6.4}$$

We will assume a flat cosmology with $\Omega_{\mathrm{m}} = 0.3$, therefore, $\Delta_{\mathrm{vir}} \sim 337$.

We solve equation 6.4 numerically for both r_{200} and r_{vir} and calculate M_{vir} and $M(< r_{200})$. We also calculate the concentration

$$c_x \equiv \frac{r_x}{r_{-2}} \tag{6.5}$$

where x is either '200' or 'vir' and $r_{-2} \equiv a_{\mathrm{h}}(2 - \gamma_{\mathrm{h}})/(\alpha_{\mathrm{h}} - 2)$ is the radius where the logarithmic slope of the halo density profile is -2 (Bullock et al. 2001).

Figure 6.12 shows the marginalized posterior PDF of M_x, r_x and c_x for chains A and B. The empirical relationship between c_{vir} and M_{vir} calculated by Bullock et al. (2001) for a ΛCDM cosmology is also shown and it is consistent with our results at the 1σ level. The points in figure 6.12 are the two theoretical mass models from Klypin et al. (2002). Our models are generally more massive and extended than these values.

6.5 Andromeda's Systematic Transverse Velocity

Figure 6.13 shows our constraint on the systematic transverse velocity of M31. Compared to the results from van der Marel & Guhathakurta (2008, shown as error bars on the figure), our constraint is very weak. This is unsurprising since while we marginalized over a very general dynamical model, van der Marel & Guhathakurta (2008)
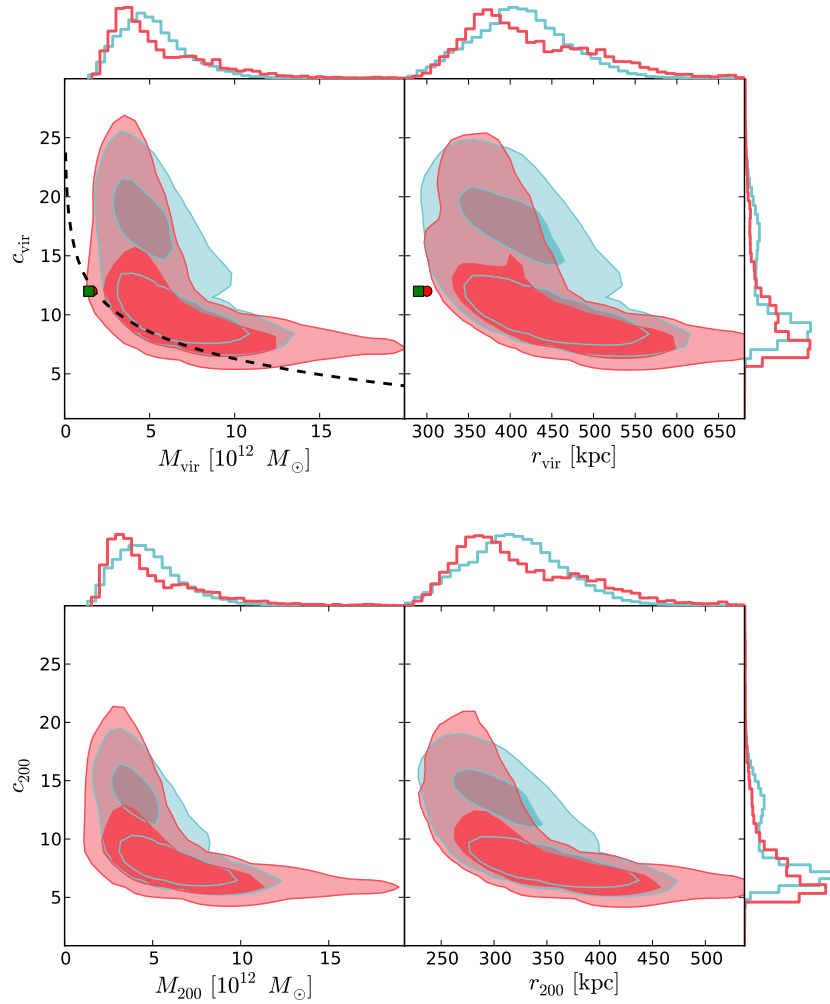
Figure 6.12 Same as figure 6.1 for the virial mass, radius and concentration. The top panels have been computed using the Bullock et al. (2001) definition and the bottom panels are for the Navarro et al. (1996) definition. The green and red points are the theoretical models from Klypin et al. (2002) and the black dashed curve is the empirical relationship calculated by Bullock et al. (2001) assuming a ΛCDM cosmology.

assumed an extremely restrictive galaxy model of only one parameter. This shows that the simplifying assumptions made by van der Marel & Guhathakurta (2008) did have a significant effect on their conclusions. Without these assumptions, it is not clear that such a tight constraint on the transverse motion is justified.

In order to estimate the mass of the LG based on our weak constraint on the orbit, we parameterize the posterior PDF of v_W and v_N as a two-dimensional Gaussian with mean

$$\langle \mathbf{v}_t \rangle \equiv \begin{pmatrix} \langle v_W \rangle \\ \langle v_N \rangle \end{pmatrix} \tag{6.6}$$

and covariance

$$\Sigma = \begin{pmatrix} \sigma_W^2 & \rho_{WN} \sigma_W \sigma_N \\ \rho_{WN} \sigma_W \sigma_N & \sigma_N^2 \end{pmatrix}. \tag{6.7}$$

In equation 6.6, the angled brackets indicate the ensemble average of the chain of $v_{W/N}$ values produced by the Markov chain. The values of $\langle v_{W/N} \rangle$, $\sigma_{W/N}$ and ρ_{WN} are listed in table 6.1.

6.6   The Mass of the Local Group

Given the two-dimensional parameterization of the inferred transverse velocity from equations 6.6 and 6.7, the likelihood of the true transverse velocity being

$$\mathbf{v}_t \equiv \begin{pmatrix} v_W \\ v_N \end{pmatrix} \tag{6.8}$$
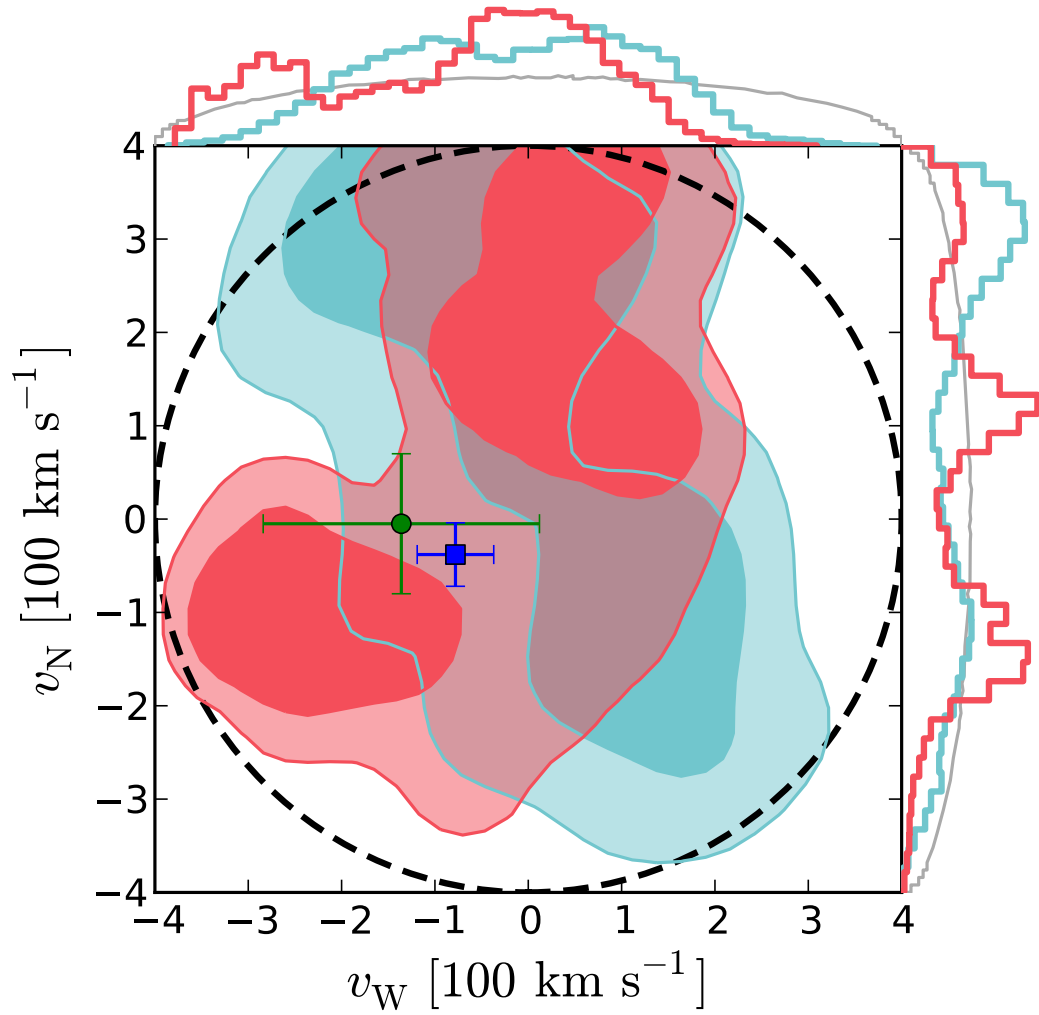
Figure 6.13 Same as figure 6.1 but showing the PDF of the components of M31's systematic transverse velocity. The green error bars show the value measured by van der Marel & Guhathakurta (2008) using only the satellite line-of-sight velocities and the blue error bars show the weighted average of their various estimates from different methods. The black dashed circle and the grey histograms show the prior on $|\mathbf{v_t}|$

is

$$p(\mathbf{v}_t| \langle \mathbf{v}_t \rangle, \Sigma) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(\mathbf{v}_t - \langle \mathbf{v}_t \rangle)^{\mathrm{T}} \Sigma^{-1} (\mathbf{v}_t - \langle \mathbf{v}_t \rangle)\right]. \qquad (6.9)$$

The LG orbital model described in section 4.3.1 generates the velocity vector of M31 relative to the Milky Way $\mathbf{v}_{\mathrm{A,G}} \equiv (V_{\mathrm{rad}}, V_{\mathrm{tan}}\cos\Theta_{\mathrm{LG}}, V_{\mathrm{tan}}\sin\Theta_{\mathrm{LG}})$ where V_{rad} is along the vector from the Milky Way to M31 and V_{tan} is in the perpendicular plane. To transform this vector into the $(v_{\mathrm{M31,los}}, v_{\mathrm{W}}, v_{\mathrm{N}})$ coordinate system, we require the motion of the Sun in the rest frame of the Milky Way and the heliocentric distances to M31 and the Galactic centre. Therefore, the full set of model parameters is $\{\ln M_{\mathrm{LG}}, e, \eta, \Theta_{\mathrm{LG}}, D_{\mathrm{M31}}, D_{\mathrm{GC}}, v_{\mathrm{circ}}, u_{\odot}, v_{\odot}, w_{\odot}\}$. The prior assumptions on these parameters are listed in table 5.1 and the informative prior on the orbital time is $t \equiv t_0 = 13.75 \pm 0.13$ Gyr: the age of the Universe. With these priors, we can invert the likelihood (equation 6.9) using Bayes' theorem to evaluate the marginalized posterior PDF for the mass of the LG.

Using the MCMC algorithm described in the previous chapter, we constrain the mass of the LG based on the inferred three-dimensional velocity. Figure 6.14 shows the marginalized posterior PDFs for the LG mass M_{LG} and the magnitude of the tangential component of M31's orbital velocity v_{tan} (equation 4.25) inferred by the timing argument. For chain A, the median and 1σ uncertainties are $M_{\mathrm{LG}} = 9^{+7}_{-3} \times 10^{12}\,M_{\odot}$ and $V_{\mathrm{tan}} = 161^{+106}_{-82}\,\mathrm{km\,s}^{-1}$. Similarly, for chain B, $M_{\mathrm{LG}} = 7^{+5}_{-2} \times 10^{12}\,M_{\odot}$ and $V_{\mathrm{tan}} = 132^{+91}_{-69}\,\mathrm{km\,s}^{-1}$.

6.7   The Implied Mass of the Milky Way

Li & White (2008) studied LG-like galaxy pairs selected from the Millennium Run

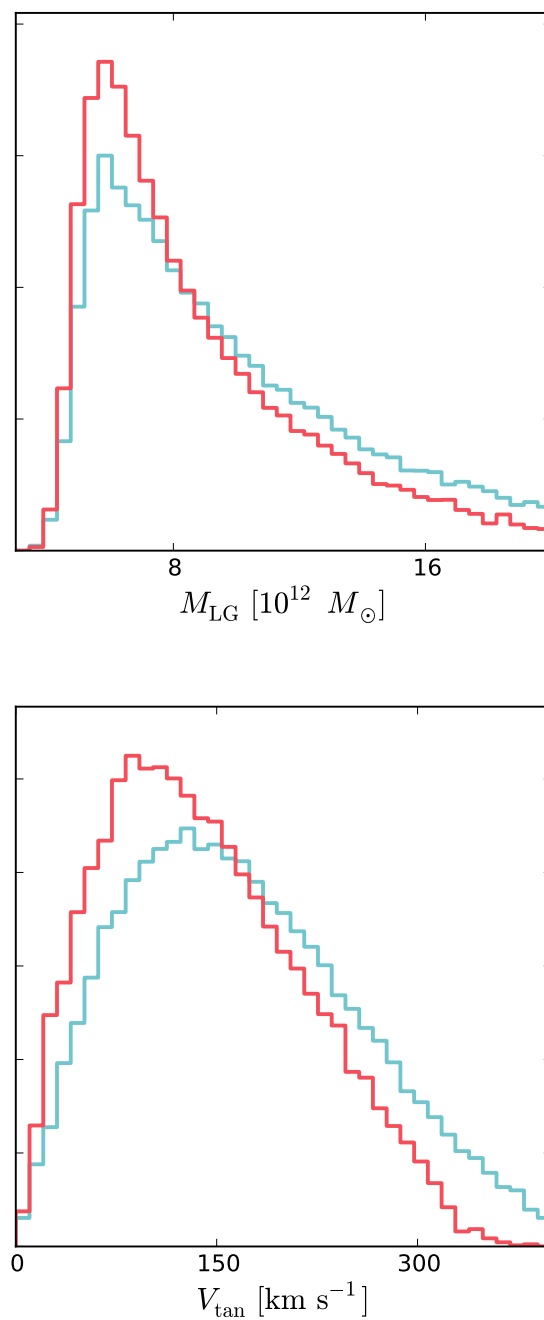Figure 6.14 The mass of the Local Group (top) and the tangental component of the orbital velocity (bottom) constrained by chains A (blue) and B (red).

numerical simulation (Springel et al. 2005) and found that the timing argument was an unbiased estimator of the sum of the two galaxy masses. Using this result, we can infer the mass of the Milky Way indirectly based on our mass estimate of M31 and our results from the timing argument.

To estimate the true masses of the galaxies in their sample, Li & White (2008) sum the values of $M(< r_{200})$ for each galaxy. Therefore, we estimate the Milky Way's mass as simply

$$M_{\mathrm{MW}}(< r_{200,\mathrm{MW}}) \equiv M_{\mathrm{LG}} - M_{\mathrm{M31}}(< r_{200,\mathrm{M31}}). \tag{6.10}$$

Figure 6.15 shows the inferred mass of the Milky Way propagated using this argument. Also included in this figure is is a histogram showing the effect of the theoretical uncertainty in the timing argument. Li & White (2008) found empirically that the ratio of true mass to the timing argument estimate is $A = 0.99 \pm 0.55$ (interquartile range). Converting the uncertainty to a standard Gaussian variance gives $A = 0.99 \pm 0.41$. Propagating the combined statistical and theoretical uncertainties via Monte Carlo, we find that the implied mass ratio $M_{\mathrm{M31}}/M_{\mathrm{MW}} \sim 0.7$ with $M_{\mathrm{M31}}/M_{\mathrm{MW}} = 1$ consistent at the 1σ level.

6.8 Inferred Quantities

Table 6.2 lists the inferred median and 1σ constraints on several physical quantities of interest. These constraints are also plotted in the $M_{\mathrm{M31}} - M_{\mathrm{MW}}$ plane in figure 6.16.
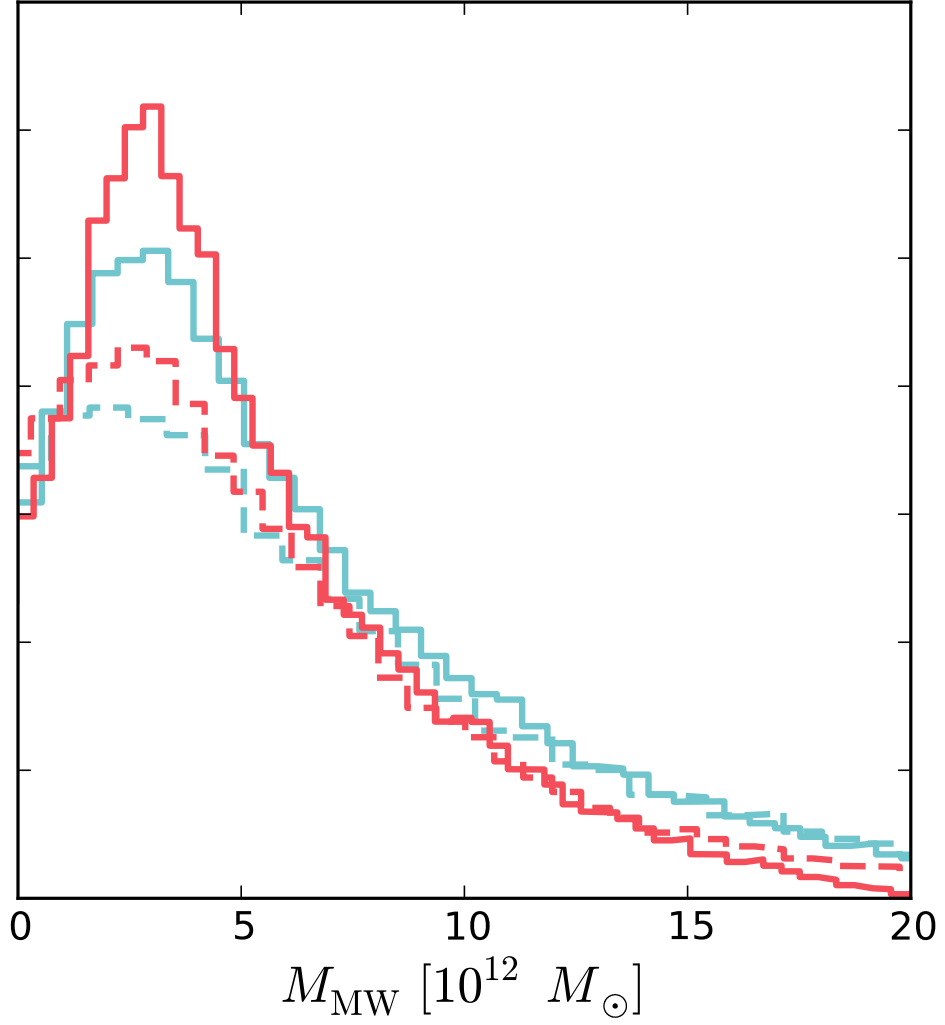
Figure 6.15 The inferred mass of the Milky Way constrained by chain A (blue) and chain B (red). The solid histograms measure only the statistical uncertainty and the dashed histograms combine the theoretical uncertainties as determined by Li & White (2008).
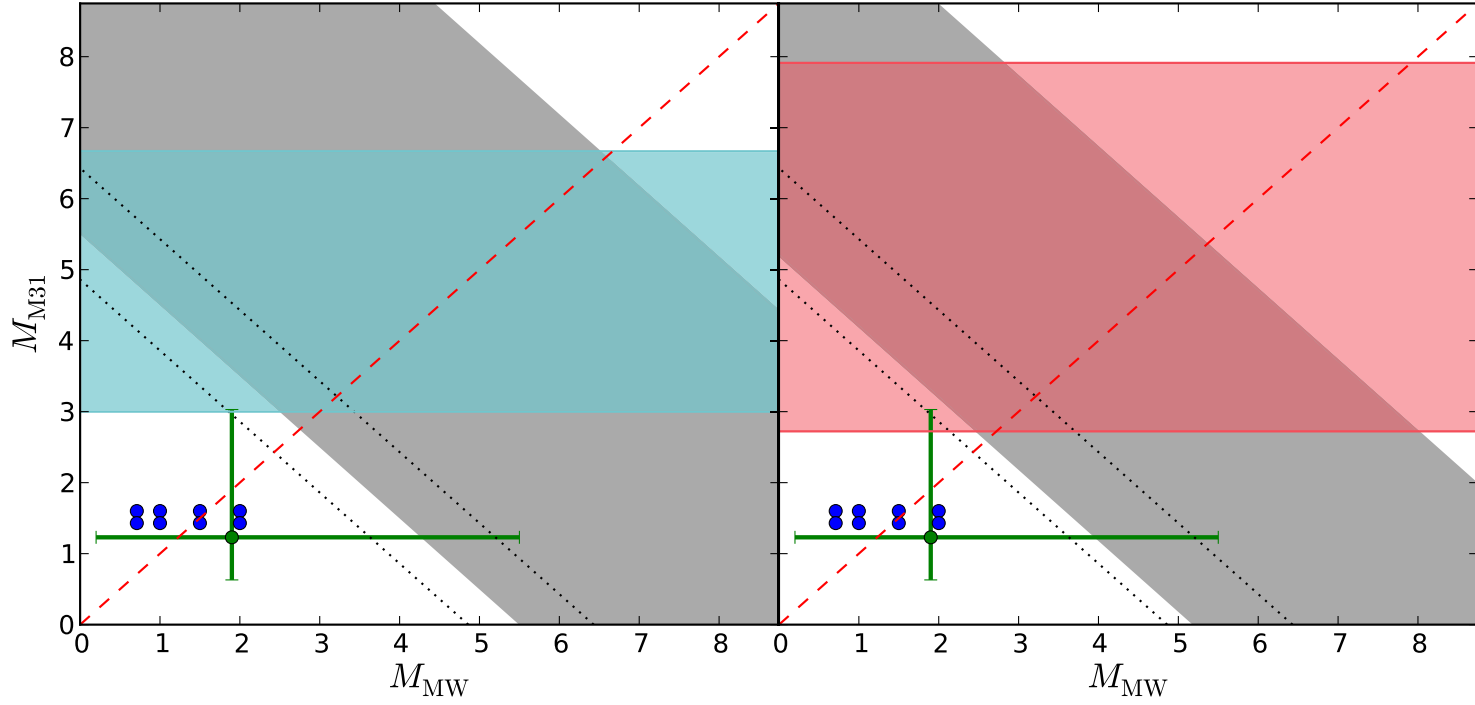
Figure 6.16 Constraints in the $M_{M31} - M_{MW}$ plane from chain A (left) and chain B (right). The mass of M31 is shaded in blue and red respectively. The dark grey region is the constraint on the total LG mass from the timing argument. The black dotted lines are the constraint on the LG mass from van der Marel & Guhathakurta (2008). The red dashed line is the one-to-one line. The green error bar is the combined estimate Wilkinson & Evans (1999) and Evans & Wilkinson (2000) and the blue points are from Klypin et al. (2002).

Table 6.2. Median and 1σ uncertainties for the inferred parameters

Parameter	Chain A	Chain B
$M_{\mathrm{vir}}\,[10^{12}\,M_\odot]$	$5.0^{+2.2}_{-1.7}$	$4.6^{+3.8}_{-1.5}$
$r_{\mathrm{vir}}\,[\mathrm{kpc}]$	$412.0^{+58.9}_{-57.1}$	$403.7^{+97.4}_{-51.8}$
c_{vir}	$9.8^{+7.6}_{-1.6}$	$9.2^{+5.5}_{-2.2}$
$M_{200}\,[10^{12}\,M_\odot]$	$4.5^{+2.1}_{-1.6}$	$4.1^{+3.6}_{-1.3}$
$r_{200}\,[\mathrm{kpc}]$	$319.4^{+47.1}_{-45.1}$	$310.5^{+79.4}_{-40.5}$
c_{200}	$7.7^{+5.6}_{-1.3}$	$7.2^{+4.1}_{-1.7}$
$M_{\mathrm{LG}}\,[10^{12}\,M_\odot]$	$9.1^{+6.8}_{-3.2}$	$7.8^{+5.0}_{-2.2}$
$M_{\mathrm{LG,theor}}\,[10^{12}\,M_\odot]$	$8.8^{+8.0}_{-4.2}$	$7.7^{+6.0}_{-3.5}$
$M_{\mathrm{MW}}\,[10^{12}\,M_\odot]$	< 12.7	< 9.9
$M_{\mathrm{MW,theor}}\,[10^{12}\,M_\odot]$	< 12.9	< 9.9
$\log_{10}\,M_{\mathrm{M31}}/M_{\mathrm{MW}}$	$-0.2^{+0.6}_{-0.4}$	$-0.1^{+0.6}_{-0.4}$

Chapter 7

Conclusions

7.1 Summary

In this thesis, we have presented the first fully self-consistent, axisymmetric dynamical model of M31. We constrained the model using datasets on all radial scales and fit the data using Markov chain Monte Carlo. We also allowed the systematic transverse velocity to vary, unlike the standard assumption that the Local Group is on a purely radial orbit. From this constraint on the orbit, we estimated the mass of the Local Group by the timing argument. The model that we used is extremely general and most of our assumed priors were highly uninformative.

At small radii, when fitting the rotation curve and surface brightness profile, we achieved an excellent, tight constraint that is consistent with previous findings. At larger radii, however, the uncertainty in the mass profile diverges due to the sparse sample and large observational uncertainties. When compared to previous results (e.g. Evans & Wilkinson 2000; van der Marel & Guhathakurta 2008), our constraints on the total mass of M31 and its transverse velocity are very poor. At first, this seems

to be a disappointing result given the sophistication of the modelling procedure and the large computational expense invested. However, by being statistically rigorous and cautious about the assumptions that we make, we have shown that the tight conclusive constraints by others on the large-scale physical parameters of M31 are unjustified.

The main sources of uncertainty for this procedure are the small sample size and large distance uncertainties for the satellite galaxies at large radii. The first year of data from the Pan-Andromeda Archeological Survey (PanDAS) revealed many previously unknown satellite galaxies and globular clusters. The second year of data is now available to the collaboration and targeted spectroscopic observations are in progress to measure the kinematics of many globular clusters at large radii (Huxor, private communication). It is also expected that many more satellite galaxies will be detected in the second year of data from PanDAS and Pan-STARRS (e.g. Irwin et al. 2008). Our modelling procedure will continue to be applicable as more data become available.

The framework of Bayesian inference is essentially a method of propagating information. As scientists using this framework as a tool, we must make a choice about what prior information can be justifiably included. In our modelling process, we were extremely conservative about the adoption of any informative priors to allow our conclusions to be exclusively data-driven. We conclude that there is insufficient information at large radii to constrain such a general model. Without a large increase in the size of the dataset, more informative priors are necessary to obtain a stronger constraint on the mass at large radii.

7.2 Future Work

The results from this project naturally lead to several future studies, some of which are currently underway. A very interesting project would be to systematically test the effects of various astrophysical assumptions on our modelling method. This would require generating and modelling a Monte Carlo sample of synthetic datasets that mimic the observational properties. This process would be tremendously computationally expensive and it is not feasible at this time but it would yield significant insight into the systematic effects of this procedure and the dynamical modelling of disk galaxies in general. It would also be beneficial to perform a similar independent analysis to infer the mass of the Milky Way for the sake of comparison.

Next, we outline two projects that we are currently investigating.

7.2.1 Globular Cluster Distribution

Given the huge computational cost required to calculate the likelihood for the globular cluster model, chain C did not have sufficient statistics in time for the preparation of this report. The preliminary results look quite promising, however and the model mass profile (figure 7.1) seems more consistent with previous estimates. Further interpretation is not yet possible due to the lack of statistics.

7.2.2 Disk Stability

The two most popular diagnostics of disk stability are the Toomre (1964) Q parameter and the Goldreich & Tremaine (1978, 1979) X parameter. It has been shown
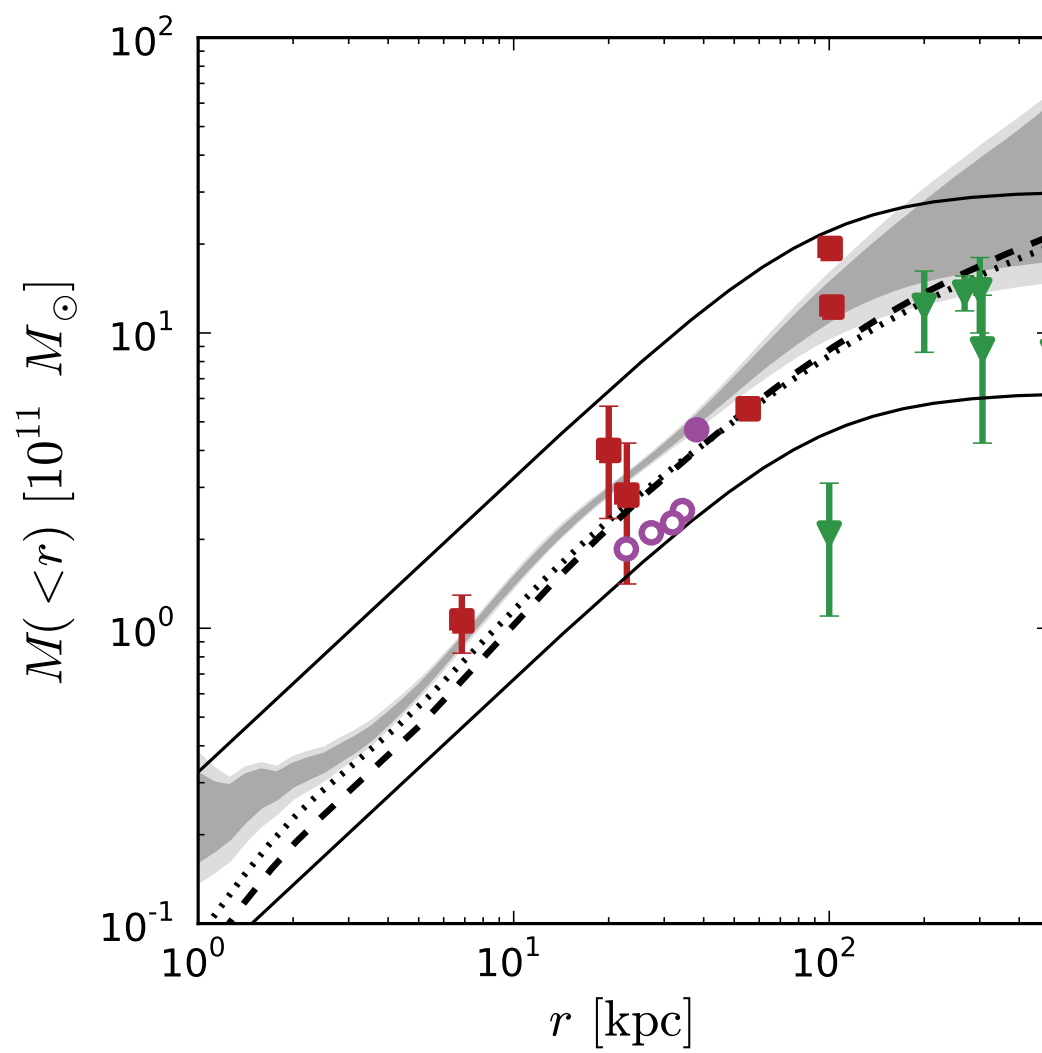
Figure 7.1 The same as figure 6.9 for chain C.

(Debattista & Sellwood 2000; Widrow et al. 2008) that the parameter

$$X' = \frac{v_{\text{tot}}^2(2.2\,R_{\text{d}})}{v_{\text{d}}^2(2.2\,R_{\text{d}})} \tag{7.1}$$

is a good proxy for X where v_{tot} is the total circular velocity and v_{d} is the contribution from the disk. Widrow et al. (2008) found that when models with $X' \lesssim 2$ were evolved numerically, they immediately became unstable. The density profile of the galaxy quickly became extremely distorted[1] and no longer matched the observational constraints. Unfortunately, the distribution of X' for our chain A is shown in figure 7.2 to have a peak well below ~ 2. To counteract this, we propose the prior

$$p(X') = \begin{cases} 0 & \text{for } X' < x_0 \\ 3 \cdot \left(\dfrac{X' - x_0}{x_1 - x_0} \right)^2 - 2 \cdot \left(\dfrac{X' - x_0}{x_1 - x_0} \right)^3 & \text{for } x_0 < X' < x_1 \\ 1 & \text{for } X' > x_1 \end{cases} \tag{7.2}$$

which varies smoothly from zero to one between x_0 and x_1. This prior is shown in figure 7.2 for $x_0 = \sqrt{2}$ and $x_1 = 2$. Admittedly, this is somewhat arbitrarily defined but we hope that the result will be insensitive to the details of the parameterization.

---

[1]See http://www.cita.utoronto.ca/ dubinski/DynamicalBlueprints/ for a visualization.
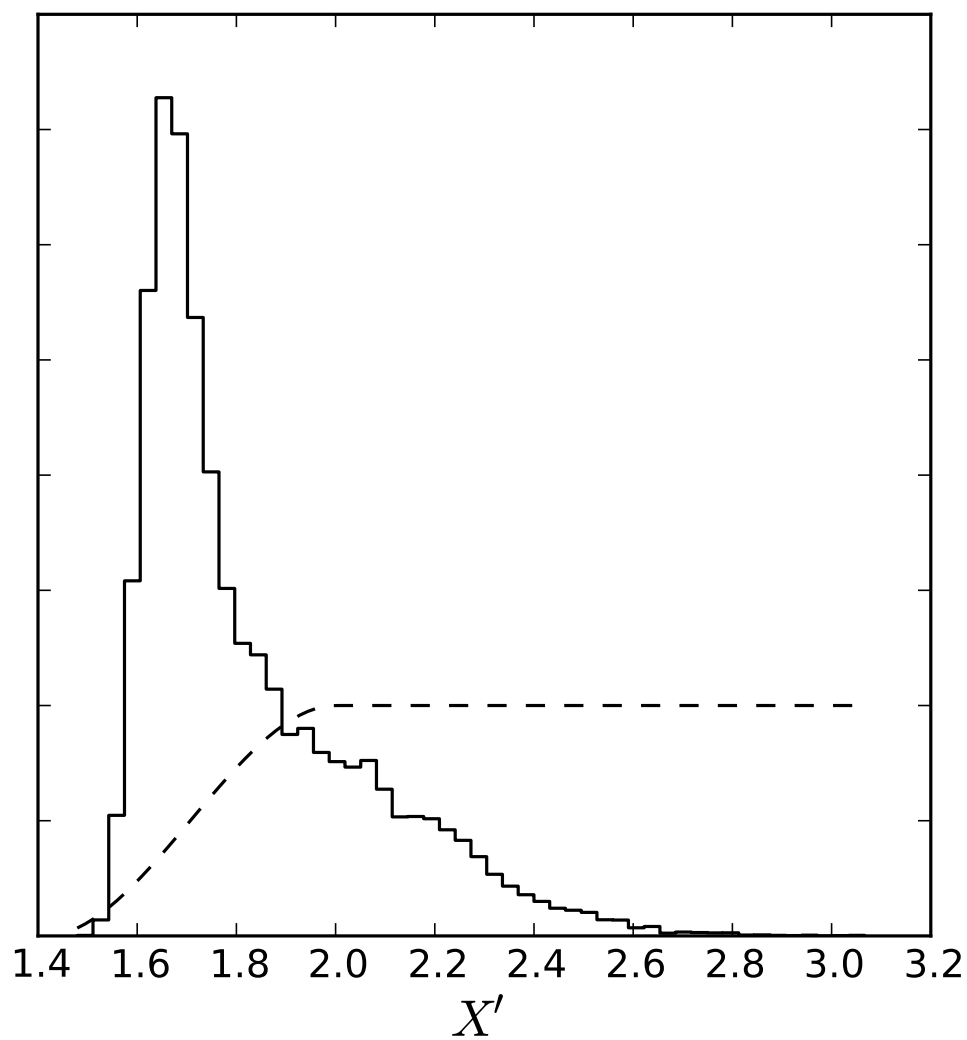
Figure 7.2 The distribution of X' for chain A. Our proposed prior is overplotted as a dashed curve.

Bibliography

Azzalini, A., & Capitanio, A. 1999, J. Royal Statistical Society, B, 579

Azzalini, A., & Dalla Valle, A. 1996, Biometrika, 83, 715

Bahcall, J. N., & Tremaine, S. 1981, ApJ, 244, 805

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Belokurov, V. et al. 2006, ApJ, 642, L137

Binney, J. 1980, MNRAS, 190, 873

Binney, J., & Merrifield, M. 1998, Galactic astronomy, ed. M. Binney, J. Merrifield

Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bovy, J., & Hogg, D. W. 2010, ApJ, 717, 617

Bovy, J., Hogg, D. W., & Rix, H. 2009a, ApJ, 704, 1704

Bovy, J., Hogg, D. W., & Roweis, S. T. 2009b, ApJ, 700, 1794

Braun, R. 1991, ApJ, 372, 54

Brodie, J. P., & Strader, J. 2006, ARA&A, 44, 193

Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R., & Dekel, A. 2001, MNRAS, 321, 559

Chapman, S. C. et al. 2007, ApJ, 662, L79

Chemin, L., Carignan, C., & Foster, T. 2009, ApJ, 705, 1395

Collins, M. L. M. et al. 2010, MNRAS, 1119

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Courteau, S., & van den Bergh, S. 1999, AJ, 118, 337

Cowles, M. K., & Carlin, B. P. 1996, Journal of the American Statistical Association, 91, 883

Crampton, D., Cowley, A. P., Schade, D., & Chayer, P. 1985, ApJ, 288, 494

de Vaucouleurs, G. 1958, ApJ, 128, 465

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Jr., H. G., Buta, R. J., Paturel, G., & Fouque, P. 1991, Third Reference Catalogue of Bright Galaxies, ed. de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., Buta, R. J., Paturel, G., & Fouque, P.

Debattista, V. P., & Sellwood, J. A. 2000, ApJ, 543, 704

Deharveng, J. M., & Pellet, A. 1975, A&A, 38, 15

Dubinski, J., & Carlberg, R. G. 1991, ApJ, 378, 496

Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, MNRAS, 356, 925

Dunkley, J. et al. 2009, ApJ, 701, 1804

Einasto, J., & Lynden-Bell, D. 1982, MNRAS, 199, 67

Evans, N. W., & Wilkinson, M. I. 2000, MNRAS, 316, 929

Evans, N. W., Wilkinson, M. I., Guhathakurta, P., Grebel, E. K., & Vogt, S. S. 2000, ApJ, 540, L9

Evans, N. W., Wilkinson, M. I., Perrett, K. M., & Bridges, T. J. 2003, ApJ, 583, 752

Galleti, S., Federici, L., Bellazzini, M., Fusi Pecci, F., & Macrina, S. 2004, A&A, 416, 917

Ghez, A. M. et al. 2008, ApJ, 689, 1044

Gillessen, S., Eisenhauer, F., Trippe, S., Alexander, T., Genzel, R., Martins, F., & Ott, T. 2009, ApJ, 692, 1075

Goldreich, P., & Tremaine, S. 1978, ApJ, 222, 850

——. 1979, ApJ, 233, 857

Gottesman, S. T., & Davies, R. D. 1970, MNRAS, 149, 263

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support, ed. Gregory, P. C. (Cambridge University Press)

Hartwick, F. D. A., & Sargent, W. L. W. 1974, ApJ, 190, 283

Hastings, W. K. 1970, Biometrika, 57, 97

Hernquist, L. 1990, ApJ, 356, 359

Hogg, D. W., Blanton, M. R., Roweis, S. T., & Johnston, K. V. 2005, ApJ, 629, 268

Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001, ApJ, 547, L133

Ibata, R., Martin, N. F., Irwin, M., Chapman, S., Ferguson, A. M. N., Lewis, G. F., & McConnachie, A. W. 2007, ApJ, 671, 1591

Irwin, M. J., Ferguson, A. M. N., Huxor, A. P., Tanvir, N. R., Ibata, R. A., & Lewis, G. F. 2008, ApJ, 676, L17

Jaffe, W. 1983, MNRAS, 202, 995

Jeans, J. H. 1915, MNRAS, 76, 70

Kahn, F. D., & Woltjer, L. 1959, ApJ, 130, 705

Kalirai, J. S. et al. 2010, ApJ, 711, 671

Kent, S. M. 1989, PASP, 101, 489

King, I. R. 1966, AJ, 71, 64

Klypin, A., Zhao, H., & Somerville, R. S. 2002, ApJ, 573, 597

Knop, R. 1969, Commun. ACM, 12, 281

Kochanek, C. S. 1996, ApJ, 457, 228

Kroeker, T. L., & Carlberg, R. G. 1991, ApJ, 376, 1

Kuijken, K., & Dubinski, J. 1995, MNRAS, 277, 1341

Larson, D. et al. 2010, arXiv:1001.4635

Lee, M. G., Hwang, H. S., Kim, S. C., Park, H. S., Geisler, D., Sarajedini, A., & Harris, W. E. 2008, ApJ, 674, 886

Letarte, B. et al. 2009, MNRAS, 400, 1472

Li, Y., & White, S. D. M. 2008, MNRAS, 384, 1459

Little, B., & Tremaine, S. 1987, ApJ, 320, 493

Lynden-Bell, D. 1981, The Observatory, 101, 111

Lynden-Bell, D. 1999, in IAU Symposium, Vol. 192, The Stellar Content of Local Group Galaxies, ed. P. Whitelock & R. Cannon, 39–+

MacKay, D. J. C. 2003, Information Theory, Inference, and Learning Algorithms (Cambridge University Press)

Majewski, S. R. et al. 2007, ApJ, 670, L9

Martin, N. F., Ibata, R. A., Irwin, M. J., Chapman, S., Lewis, G. F., Ferguson, A. M. N., Tanvir, N., & McConnachie, A. W. 2006, MNRAS, 371, 1983

McConnachie, A. W. et al. 2008, ApJ, 688, 1009

McConnachie, A. W., & Irwin, M. J. 2006, MNRAS, 365, 902

McConnachie, A. W., Irwin, M. J., Ferguson, A. M. N., Ibata, R. A., Lewis, G. F., & Tanvir, N. 2005, MNRAS, 356, 979

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Metz, M., Kroupa, P., & Jerjen, H. 2007, MNRAS, 374, 1125

Milgrom, M. 1983, ApJ, 270, 365

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

——. 1997, ApJ, 490, 493

O'Hagan, A., & Leonard, T. 1976, Biometrika, 63, 201

Press, W. H., Flannery, B. P., & Teukolsky, S. A. 1986, Numerical recipes. The art of scientific computing

Prugniel, P., & Simien, F. 1997, A&A, 321, 111

Puglielli, D., Widrow, L. M., & Courteau, S. 2010, ApJ, 715, 1152

Raychaudhury, S., & Lynden-Bell, D. 1989, MNRAS, 240, 195

Riess, A. G. et al. 2007, ApJ, 659, 98

Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, Ann. Appl. Probab., 7, 110

Rood, H. J. 1979, ApJ, 232, 699

Rubin, V. C., & Ford, Jr., W. K. 1970, ApJ, 159, 379

Saglia, R. P. et al. 2010, A&A, 509, A260000+

Sales, L. V., Navarro, J. F., Abadi, M. G., & Steinmetz, M. 2007, MNRAS, 379, 1464

Sanders, R. H., & McGaugh, S. S. 2002, ARA&A, 40, 263

Shu, F. H. 1969, ApJ, 158, 505

Springel, V. et al. 2005, Nature, 435, 629

Terzić, B., & Graham, A. W. 2005, MNRAS, 362, 197

Toomre, A. 1964, ApJ, 139, 1217

van den Bergh, S. 1981, PASP, 93, 428

van der Marel, R. P., & Guhathakurta, P. 2008, ApJ, 678, 187

Walterbos, R. A. M., & Kennicutt, Jr., R. C. 1987, A&AS, 69, 311

——. 1988, A&A, 198, 61

Watkins, L. L., Evans, N. W., & An, J. H. 2010, MNRAS, 406, 264

Widrow, L. M., & Dubinski, J. 2005, ApJ, 631, 838

Widrow, L. M., Perrett, K. M., & Suyu, S. H. 2003, ApJ, 588, 311

Widrow, L. M., Pym, B., & Dubinski, J. 2008, ApJ, 679, 1239

Wilkinson, M. I., & Evans, N. W. 1999, MNRAS, 310, 645

Wolf, J., Martinez, G. D., Bullock, J. S., Kaplinghat, M., Geha, M., Muñoz, R. R., Simon, J. D., & Avedo, F. F. 2010, MNRAS, 778

York, D. G. et al. 2000, AJ, 120, 1579

Zucker, D. B. et al. 2004, ApJ, 612, L121

——. 2007, ApJ, 659, L21

Zwicky, F. 1937, ApJ, 86, 217

Appendix A

The Skew-Normal Distribution

When a measurement μ is published, it is generally reported as $\mu \pm \sigma$, where σ is the uncertainty range to some confidence level (generally 68%). In practice, the probability density function for this measurement is often modelled as a Gaussian with mean μ and variance σ^2:

$$\mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{A.1}$$

When the uncertainty on the published value is asymmetric (e.g. $x = \mu_{-\sigma_-}^{+\sigma_+}$), a generalization of this assumption is necessary. A one-parameter (shape $\alpha \in \mathbb{R}$) generalization of equation A.1 is the skew-normal (SN; O'Hagan & Leonard 1976; Azzalini & Dalla Valle 1996) distribution

$$\tilde{\mathcal{N}}(x|\xi, \omega, \alpha) = \left(\frac{2}{\omega}\right) \phi\left(\frac{x-\xi}{\omega}\right) \Phi\left[\alpha\left(\frac{x-\xi}{\omega}\right)\right], \tag{A.2}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{A.3}$$

is the standard normal distribution and

$$\Phi(x) = \int_{-\infty}^{x} \phi(x') \, dx' = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \tag{A.4}$$

is the cumulative distribution of $\phi(x)$.

The parameters $\xi \in \mathbb{R}$ and $\omega > 0$ in equation A.2 are called the location and scale, respectively. This notation is used to emphasize that ξ is the distribution mean and ω^2 is the variance *only* when $\alpha = 0$. In fact, it is clear that when $\alpha = 0$, equation A.2 simplifies to equation A.1. For any SN distribution with $\alpha \neq 0$, the parameterization of equation A.2 satisfying a given observation must be modelled numerically since there is no analytic relationship between the SN parameters and the mode and confidence intervals of the distribution.

To calculate the SN parameterization, we first calculate the value of α required to produce the given σ-ratio (σ_+/σ_-) then, we calculate the maximum-likelihood of ξ and ω given α. We find the empirical relationship

$$\alpha(\sigma_+/\sigma_-) \approx \begin{cases} -|\alpha|, & \text{if } \sigma_+ < \sigma_- \\ 0, & \text{if } \sigma_+ = \sigma_- \\ |\alpha|, & \text{if } \sigma_+ > \sigma_- \end{cases} \tag{A.5}$$

where

$$|\alpha| \approx 46.26\Delta - 341.35\Delta^2$$
$$+1642.86\Delta^3 - 3434.54\Delta^4 + 3702.57\Delta^5 - 1397.62\Delta^6 \qquad (\text{A.6})$$

and

$$2 \times \Delta = \left| \log_{10} \left(\frac{\sigma_+}{\sigma_-} \right) \right|. \qquad (\text{A.7})$$

Figure A.2 shows the error on alpha resulting from the approximation in equation A.5. It is clear from this figure that for the range of interest $0 \ll \Delta \lesssim 1$, the error on α is $\lesssim 1\%$.

A major advantage of the SN distribution is that it can be easily sampled pseudo-randomly (Azzalini & Capitanio 1999). To sample the SN distribution with location ξ and scale ω, we generate two normally distributed random variables u and v (Knop 1969; Press et al. 1986) and return $\xi + \omega z$ where

$$z \equiv \begin{cases} \delta u + v\sqrt{1 - \delta^2}, & \text{if } u \geq 0 \\ -\delta u - v\sqrt{1 - \delta^2}, & \text{otherwise} \end{cases} \qquad (\text{A.8})$$

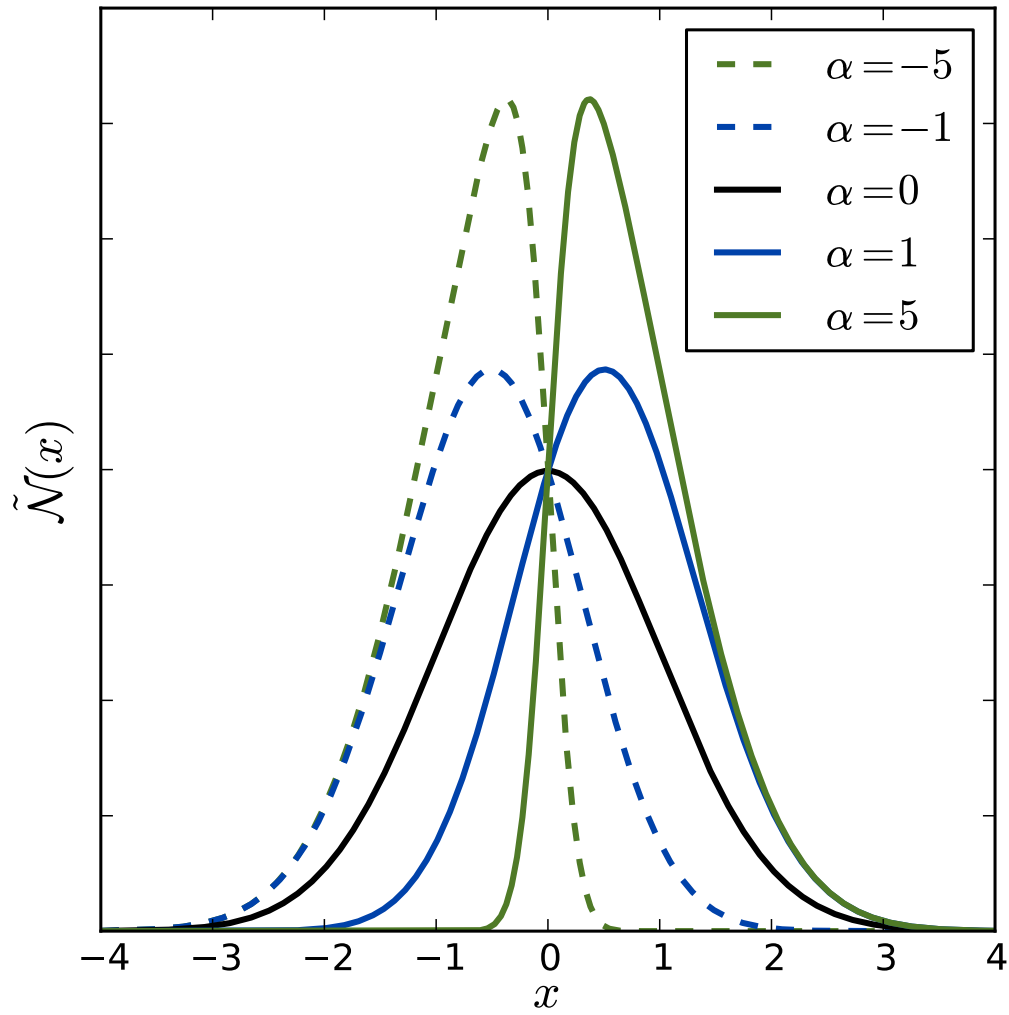and $\delta^2 = \alpha^2/(1 + \alpha^2)$.

Figure A.1 Examples of the skew normal distribution with $\xi = 0$ and $\omega = 1$ for different values of α.
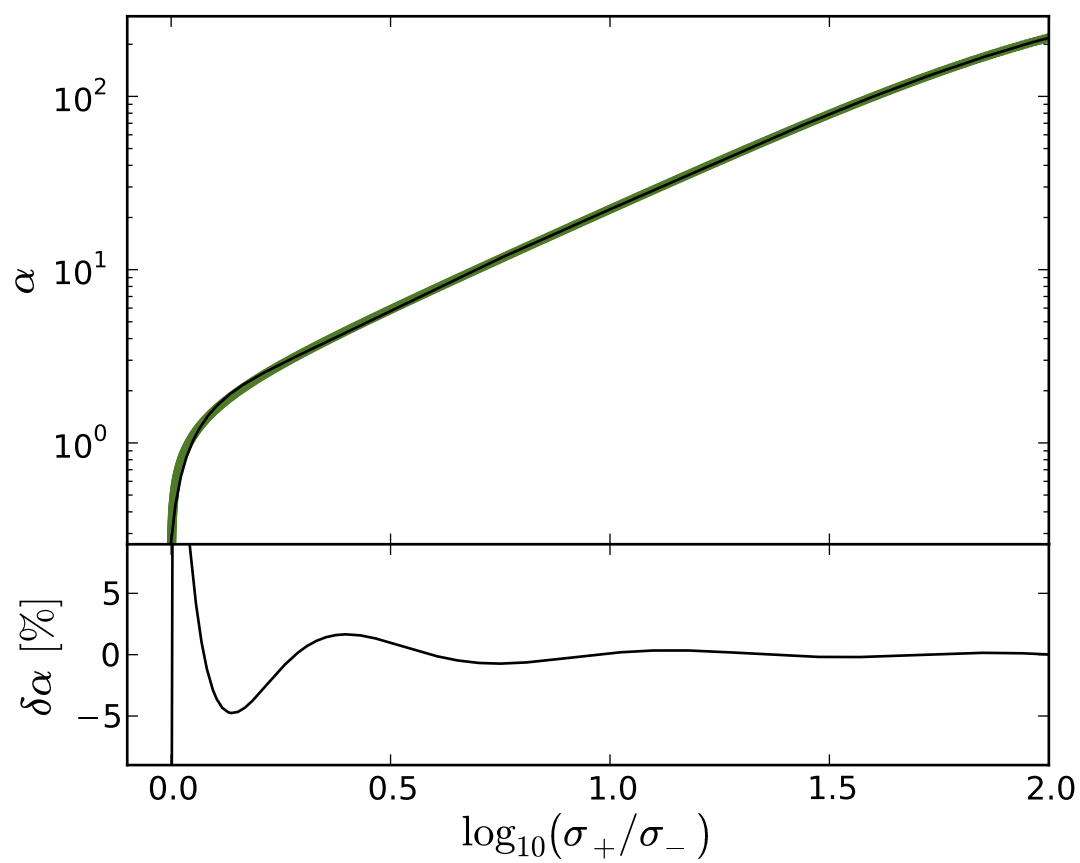
Figure A.2 *Top:* The true values of α (green) compared to those calculated using equation A.5 (black). *Bottom:* The residuals of the above fit.

Appendix B

M31-centric Coordinate System

To convert heliocentric positions from observed right ascension α, declination δ and distance r to M31-centric coordinates, we follow a derivation similar to Metz et al. (2007, their appendix B) and include it here for clarity. We start with a position $\mathbf{r}$ measured in the equatorial basis

$$
S \equiv \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix} = \begin{pmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\delta \\ \sin\delta \end{pmatrix} \tag{B.1}
$$

centred on the Sun and we wish to transform to a basis $S' \equiv (\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}'_z)^{\mathrm{T}}$ centred on M31 (with equatorial coordinates $\mathbf{r}_{\mathrm{M31}}$). To simplify calculations in this primed frame, we will rotate the coordinate system so that $\mathbf{e}'_x$ and $\mathbf{e}'_y$ lie in M31's disk and $\mathbf{e}'_z$ is the normal. $\mathbf{e}'_x$ will be the projection of our line-of-sight onto the disk of M31. This rotation must correct for (a) the orientation of the 'normal triad' at the position of M31, (b) the position angle of the stellar disk and (c) the inclination of the disk

with respect to the line-of-sight.

The final transformation from $\mathbf{r} \in S$ to $\mathbf{r}' \in S'$ takes the form

$$\mathbf{r}' = \mathbf{R}(\mathbf{r} - \mathbf{r}_{\mathrm{M31}}) \tag{B.2}$$

where $\mathbf{R}$ is a rotation matrix that corrects for the effects mentioned previously. To correct for the orientation of the triad at M31's position, we multiply S by $\mathbf{R}_{rpq}(\alpha_{\mathrm{M31}}, \delta_{\mathrm{M31}})$ where

$$\mathbf{R}_{rpq}(\alpha, \delta) \equiv \begin{pmatrix} \cos\delta\cos\alpha & \cos\delta\sin\alpha & \sin\delta \\ -\sin\alpha & \cos\alpha & 0 \\ -\sin\delta\cos\alpha & -\sin\delta\sin\alpha & \cos\delta \end{pmatrix}. \tag{B.3}$$

With this transformation and the corrections for inclination and position angle, the full rotation becomes

$$\mathbf{R} = \mathbf{R}_y(90° - i)\,\mathbf{R}_x(\theta - 90°)\,\mathbf{R}_{rpq}(\alpha_{\mathrm{M31}}, \delta_{\mathrm{M31}}) \tag{B.4}$$

where $\mathbf{R}_{x/y/z}$ is the *right-handed* Cartesian rotation matrix about the given axis, $i = 77.5°$ is M31's inclination and $\theta = 37.7°$ is the position angle of the stellar disk measured from north over east (de Vaucouleurs 1958). Numerically, this matrix is the constant

$$\mathbf{R} \equiv \begin{pmatrix} 0.7752 & 0.3204 & 0.5445 \\ -0.6261 & 0.5042 & 0.5947 \\ -0.0840 & -0.8019 & 0.5915 \end{pmatrix} \tag{B.5}$$

which differs from the transformation from Metz et al. (2007) by a z-rotation of $\sim 0.4°$

since they define $\mathbf{e}'_z$ in the direction of the Galactic centre instead of the Sun. For our axisymmetric purposes, this rotation is unnecessary.