Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 19 July 2013 (MN LATEX style file v2.2)

A non-parametric method for mass modelling spherical systems

P. Steger[1⋆], D. von Rickenbach[1], J. I. Read[1,2]
[1] *Institute for Astronomy, Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland*
[2] *Department of Physics, University of Surrey, Guildford, GU2 7XH, UK*

19 July 2013

ABSTRACT
We propose a new non-parametric method to determine the mass distribution in spherical systems. A high dimensional parameter space encoding tracer density, line of sight velocity dispersion and total mass density is sampled with an Monte Carlo Markov Chain.

Without assumptions on the functional form of any of these profiles, we can reproduce liably the total mass density of mock dwarf galaxies, and disentangle the degeneracy between dark matter density and tracer velocity anisotropy.

We show early applications to observed dwarf galaxies, and point out what data quality is required to yield a sensible estimate.

Key words: galaxies: dwarf – galaxies: fundamental parameters – galaxies: kinematics and dynamics – cosmology: dark matter

1 INTRODUCTION

Cosmological ΛCDM simulations of representative patches of the Universe predict the dark matter to assemble self-similarly.

Assuming that only dark matter might influence the physics on all scales down to the stellar regime, Navarro et al. (1997) found that the density profiles of the resulting halos are best described by a function diverging as r^{-1} towards the center. Given that only a finite amount of dark matter is available in the halo, there exists a very small lower bound for a turnover radius, where the density approaches a constant value.

Observations of low surface brightness galaxies measure rotation curves and deduce from them a constant density below a rather large scale radius r_S (de Blok et al. 2001), (McGaugh et al. 2001), contrary to the theoretical predictions.

This fact became known as the cusp/core problem.

Many solutions have been proposed to solve it. It was deemed possible that difficulties in data modelling prevent observations from resolving cusps.

From a theoretical point, it was proposed that dark matter is not cold as assumed, with a warm component smearing out density peaks below a certain scale.

Dark matter could also be self-interacting, as e.g. postulated by (Spergel & Steinhardt (2000), Vogelsberger et al. (2012)), and have a large scattering cross-section and low annihilation or dissipation cross-section, and thus prevent formation of overly-dense cusps.

For the simulations in question, it became clear that baryonic physics plays a crucial role for the buildup of the overall density profile on small scales, and thus dark matter.

In particular, modelling of stellar feedback, the introduction of a higher density threshold for star formation and an increase of resolution for treatment of individual star-forming regions in a cosmological context led simulations to reveal dwarf galaxies with shallow central density slopes in dark matter (Governato et al. 2010). This compares well with what is found in THINGS dwarf galaxies (Oh et al. 2011).

Another type of dwarf galaxies, the dwarf spheroidals, lie closer to the Sun and can be resolved much better, often allowing us to observe individual stars. We need a method to model their mass distribution from those observations to investigate, whether they exhibit a cusp or a core, and whether simulations are able to reproduce the dark matter distribution.

An early approach for general triaxial systems in dynamical equilibrium was proposed by Schwarzschild (1979): Based on a density profile and a corresponding gravitational potential, an ensemble of orbits are calculated and superposed to yield the underlying density profile.

Another method makes use of the Jeans equations encompassing tracer density, velocity dispersion and the gravitational potential to solve for the potential, and ultimately get the underlying dark matter density (Binney & Tremaine 2008). Mostly, a functional form with some free parameter(s)

⋆ E-mail: psteger@phys.ethz.ch

is assumed for the density profile, and fitting routines are used to yield the best agreeing form.

Another assumption is required for the velocity anisotropy profile to get a mass density in the case of lowest-order Jeans equations. Since it is not known a priori how much the system is supported by rotational motion, this leads to a degeneracy between mass and velocity anisotropy.

Higher order Jeans equations can help to break this degeneracy (Łokas 2002). Another degeneracy between inner DM slope and concentration shows up, though. With better data available for the local dSphs, distribution function based methods may be used. Breddels et al. (2012) cannot distinguish between cuspy or cored profiles for Sculptor.

Yet another approach is to use the motion of globular clusters inside dwarf galaxies (Goerdt et al. 2006), (Cole et al. 2012). If the density distribution follows a cored profile, globular clusters will not fall in, or will even get pushed out of the core if they formed inside. In Fornax dSph, there is evidence for a core.

Walker & Peñarrubia (2011) split the stellar tracers into two populations with different metallicities and half-mass radii, and make use of the fact that the enclosed mass is found the same at the half-mass radius, independent of the underlying velocity anisotropy profile, to get two points in the center of the dwarf galaxy, from which constraints on the inner DM density slope can be drawn.

We want to get the full density profile, though.

In this paper, we propose a new non-parametric mass-modelling technique based on the Jeans approach, with no assumptions on the functional form of the dark matter density, nor the velocity anisotropy profile.

We assume spherical symmetry in a first step. Beware that dSph galaxies of the local group are known to be slightly non-spherical, with an average ellipticity of 0.3 (Mateo 1998).

Our method has following advantages over other methods presented so far:

(i) no assumptions on the functional form of the underlying dark matter;
(ii) applicable to any gravitational model, since Jeans equation and Poisson equation are solved each on their own;
(iii) robust to noise in the data, as no numerical differentiation is used as soon as the three-dimensional model stands.

2 METHOD

The collisionless Boltzmann equation for a spherical system with gravitational potential Φ,

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \nabla_{\vec{x}} f \cdot \vec{v} - \nabla_{\vec{v}} f \cdot \nabla_{\vec{x}} \Phi = 0, \tag{1}$$

describes the motion of tracer stars with distribution function $f(\vec{x}, \vec{v})$.

In spherical coordinates (r, θ, ϕ), the collisionless Boltzmann equation then reads as

$$\frac{\partial f}{\partial t} + \dot{r}\frac{\partial f}{\partial r} + \dot{\theta}\frac{\partial f}{\partial \theta} + \dot{\phi}\frac{\partial f}{\partial \phi} + \dot{v}_r\frac{\partial f}{\partial v_r} + \dot{v}_\theta\frac{\partial f}{\partial v_\theta} + \dot{v}_\phi\frac{\partial f}{\partial v_\phi} = 0 \tag{2}$$

with velocities

$$\dot{r} = v_r, \tag{3}$$
$$\dot{\theta} = v_\theta / r \tag{4}$$
$$\dot{\phi} = v_\phi / r \sin\theta. \tag{5}$$

The assumption of steady state hydrodynamic equilibrium gives $\partial f/\partial t = 0$ and $\bar{v}_r = 0$, and using spherical symmetry $\bar{v}_\theta = 0$, $\bar{v}_\phi = 0$, with a unique tangential velocity dispersion $\sigma_\phi^2 = \sigma_\theta^2 = \sigma_t^2$ yields

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma_r^2) + 2\frac{\sigma_r^2 - \sigma_t^2}{r} = -\frac{\partial\Phi}{\partial r} = -\frac{GM(<r)}{r^2} \tag{6}$$

with enclosed mass $M(<\ r)$, gravitational constant $G = 6.67398 \cdot 10^{-11}\mathrm{m}^3/\mathrm{kg\,s}^2$. The departure from spherical hydrostatic equilibrium $\sigma_r^2 = \sigma_t^2$ is measured by the anisotropy parameter

$$\beta \equiv 1 - \frac{\sigma_t^2}{\sigma_r^2} \tag{7}$$

with values in the range from $-\infty$ (purely circular orbits) through 0 (hydrostatic equilibrium) to 1 (purely radial orbits).

Integrating both sides of equation 6 gives the main equation of this paper,

$$\sigma_r^2(R) = \frac{1}{\nu(R)} \exp\left(-2\int_{r_{min}}^{r} \frac{\beta(s)}{s}\mathrm{d}s\right) \cdot \tag{8}$$

$$\int_R^\infty \frac{GM(r)\nu(r)}{r^2} \exp\left(2\int_{r_{min}}^{r} \frac{\beta(s)}{s}\mathrm{d}s\right)\mathrm{d}r.$$

For distant spherical systems, only the projected velocity dispersion σ_{LOS} can be measured, which in our case is given by

$$\sigma_{\mathrm{LOS}}^2(R) = \frac{2}{\Sigma(R)}\int_R^\infty \left(1 - \beta\frac{R^2}{r^2}\right)\frac{\nu(r)\sigma_r^2(r)r}{\sqrt{r^2 - R^2}}\mathrm{d}r, \tag{9}$$

where $\Sigma(R)$ denotes the surface mass density at radius R.

In the following, we present a non-parametric method for the solution of equation 9 for the total gravitating mass density $\rho(r)$, given observed $\nu(r)$ and $\sigma_{\mathrm{LOS}}(r)$, where r denotes the projected two-dimensional radius from the center of mass of the spherical system. We get the enclosed mass $M(<r)$ from the density via

$$M(<r) = \int_0^r \rho(r) r^2 dr, \tag{10}$$

which shows up in eq. 8. In principle, the method can be generalized to investigate alternative gravity models, if the acceleration $GM(r)/r^2$ is replaced with the respective form of $-\partial\Phi/\partial r$.

The degeneracy between mass M and velocity anisotropy β is accounted for: For any non-isothermal system, we let vary the anisotropy $\beta(r)$ as well. We checked that in the case of a simple Hernquist profile, $\beta(r) \approx 0$ is retrieved correctly.

The main idea is to let an Monte Carlo Markov Chain

marginalize over the parameter space $[\nu_i, \sigma_{\rm LOS,i}, \rho]$ for distinct populations $i = 1...N$ of stellar or gaseous tracers. There are a large number of parameters from the representation of the radial profiles of each of those in N_{bin} bins, with only very few constraints from physical priors. The functional form of the profiles is not predescribed. This is what we call *non-parametric*.

Two different approaches are taken into account for the sampling of densities:

Tracer densities ν_i are expected to fall with increasing radii. To ensure this, one can explicitly build a monotonic function

$$\nu_i(r) = \int_0^r \tilde{\nu}_i(r')dr' \tag{11}$$

with parameters $\tilde{\nu}_i(r') > 0$ discretized in bins. The other possibility is to let $\tilde{\nu}_i(r')$ vary freely, or just sample $\nu_i(r') = \tilde{\nu}_i|_{r'}$ directly. As it turns out, the MCMC prefers decreasing densities anyhow, so the ν_i parameters were allowed to vary within the priors described below.

Furthermore, the sampling might be done in a linear or logarithmic fashion. Wherever negative components are required, $\beta_i(r)$, linear sampling was chosen, s.t.

$$\beta_i^{(n+1)}(r) = \beta_i^{(n)}(r) + \delta\beta_i(r) \tag{12}$$

with new parameter $\beta_i^{(n+1)}(r)$ in iteration $n+1$ determined from its old value $\beta_i^{(n)}(r)$ at iteration n and the parameter stepsize $\delta\beta_i(r)$ drawn from a random uniform distribution. On the other hand, for any positive definite parameter that needs to span a range in logarithmic space, $\nu_i(r)$ and $\rho(r)$, we sample logarithmically,

$$\nu_i^{(n+1)}(r) = 10^{\tilde{\nu}_i^{(n+1)}}, \qquad \tilde{\nu}_i^{(n+1)}(r) = \tilde{\nu}_i^{(n)}(r) + \delta\tilde{\nu}_i(r) \tag{13}$$

In a next step, $\sigma_{\rm LOS,i}(r)$ is calculated from ν_i, $\rho(r)$, and $\beta_i(r)$ according eq. 9. This is done numerically, involving three integrations, which are performed with polynomial extrapolations of the integrands up to infinity, s.t. contributions from $\rho(r > r_{max})$ hinder an artificial falloff of $\sigma_{\rm LOS}$. The additional parameter of the slope is calculated from the second half of all bins.

The last step involves comparison of the projected $\nu_i(r)$, $\sigma_i(r)$ and $beta_i(r)$, if available, to the data for the tracer populations to get an error function

$$\chi^2 = \sum_{i=1}^{N} \chi^2_{\nu,i} + \chi^2_{\sigma,i} + \chi^2_{\beta,i} \tag{14}$$

$$\chi^2_{\nu,i} = \sum_{j=1}^{N_{bin}} \left(\frac{\nu_{i,\rm data}(r_j) - \nu_{i,\rm model}(r_j)}{\epsilon_\nu(r_j)} \right)^2. \tag{15}$$

and accordingly for $\chi^2_{\sigma,i}$ and $\chi^2_{\beta,i}$. In absence of a measured $\beta_i(r)$, we set $\chi^2_{\beta,i} = 0$.

The model for iteration $n+1$ is accepted if its close or below the previous iteration,

$$\exp(\chi^2_n - \chi^2_{n+1}) < \varepsilon, \qquad \varepsilon \in [0, 1) \tag{16}$$

for a uniform random ε. Otherwise the model is rejected.

2.0.1 *Initialization phase and step sizes*

The stepsize can vary for each bin, and is changed during an initialization phase. If the acceptance rate of models lies between 0.24 and 0.26, it is decreased by factor of 1.01, else, it is increased by the same amount. **(TODO: reference why 0.25 is best acceptance rate)**

To ensure fast convergence in the first few iterations, we employ a polynomial representation of the overall density $\rho(r)$,

$$\rho(r) = \sum_{i=0}^{N-1} a_i \cdot \left(\frac{r_s - r}{r_s} \right)^i, \tag{17}$$

$$a_i = \{a_0, (r_s^i / i^\gamma)_{i=1,N-1}\} \tag{18}$$

with scale radius $r_s = 1. \cdot \max(r)$, an offset of $a_o = -2.5$ and a suppression of higher order terms by a factor $\gamma = 1.1$. This specific setup yields a working approximation of the density. Other combinations of parameters were tried and shown to yield convergence as well. The parameters a_i are changed by the MCMC during the burn-in phase, whereby any higher order polynomial is allowed to be enhanced if necessary.

After a burn-in phase of several 100 accepted models with $\chi^2 < \chi^2_{end} = 70$, the stepsize is frozen and the MCMC starts storing the accepted models for further statistical analysis. A default 10^5 models are taken, where nothing else is indicated.

2.1 Binning characteristics

The number of bins for ν, σ, β, ρ are free parameters. They are set to fulfill

$$n_\nu = n_\sigma = n_\beta = n_\rho \leqslant n_{\rm data} \tag{19}$$

with number of datapoints $n_{\rm data}$. This choice simplifies integration greatly, and prevents invention of information on scales smaller than the frequency of datapoints.

(TODO: check that nbin is set s.t.)

$$\chi^2_{red} = \frac{\chi^2}{n_{\rm data} - n_\nu - n_\sigma - n_\beta - n_\rho - 1} \tag{20}$$

is minimized, and still the whole parameter space is tracked.

The whole parameter space for β_i is sampled with $n_\beta = 12$ if $n_{\rm data} = 30$ in the case of 10000 tracers.

The dark matter density is calculated by subtracting the measured baryon density from the dynamical mass density.

2.2 Priors

Following priors are included in the MCMC, and can help to reject unphysical models from the start:

1) cprior: $M(r = 0) = 0$, the central mass is set to 0;
2) bprior: $\rho(r) \geqslant \rho_b(r) - \epsilon_{\rho,b}(r) \forall r \geqslant 0$, ensures that no models with overall densities below the measured baryon density (reduced by the measurement error) are considered any further;

3) lbprior: $M(r > r_{max}) \leqslant M(< r_{max})/3.$, rejects any model which has more than 33% of the overall mass up to the outermost radius in the extrapolated bins;

4) rising ρ prior: $(\rho(r + \Delta r) - \rho(r))/\rho(r) \leqslant 0.5$, prevents ρ rising more than 50% for the next bin. There is no reason for the overall mass density to rise outwards in dynamically old systems. It might be favourable for convergence, though, if a dip in $\rho(r)$ does not lead to immediate rejection of all models with correct $\rho(r + \Delta r)$;

5) $\beta_i(r + \Delta r) - \beta_i(r) < 0.5$: prevent any sudden jumps in β_i;

We show in the appendix what effects the disabling of some of these priors have.

3 RESULTS

We apply our method to another set of mock data, the spherical models for the Gaia challenge by Walker and Penarrubia. They consist of dynamical tracer populations with density distribution

$$\nu_*(r) = \nu_0 \left(\frac{r}{r_*}\right)^{-\gamma_*} \left[1 + \left(\frac{r}{r_*}\right)^{\alpha_*}\right]^{(\gamma_* - \beta_*)/\alpha_*} \tag{21}$$

inside dark matter halos of the form

$$\rho_{\rm DM} = \rho_0 \left(\frac{r}{r_{\rm DM}}\right)^{-\gamma_{\rm DM}} \left[1 + \left(\frac{r}{r_{\rm DM}}\right)^{\alpha_{\rm DM}}\right]^{(\gamma_{\rm DM} - \beta_{\rm DM})/\alpha_{\rm DM}} \tag{22}$$

with scale radii $r_*, r_{\rm DM}$, inner and outer logarithmic slopes of $\gamma_*, \gamma_{\rm DM}$ and $\beta_*, \beta_{\rm DM}$, with transition parameters $\alpha_*, \alpha_{\rm DM}$.

The anisotropy follows the functional form of **?** and **?**,

$$\beta_{\rm anisotropy}(r) = 1 - \frac{\sigma_\theta^2}{\sigma_r^2} = \frac{r^2}{r^2 + r_a^2}. \tag{23}$$

with scale radius r_a, turning over from nearly isotropic at $r \to 0$ to radially biased at $r_* = r_a$.

Of these distributions, finite samplings are taken and converted to mock observational data including spectral indices, systemic velocities, proper motions, binary motion.

3.1 Cusps and Cores

Applied on a profile with a core in the DM density profile, our method converges fast in the beginning and reproduces the density profile very well after 50000 iterations (fig. 1).

The retrieved models show a distribution with 90 percent certainty levels (light gray) that encompass the underlying theoretical model over all radii. For all but the most central bin, the underlying density profile lies even within 1σ (dark gray) of all models.

(TODO: d ln rho/d ln r, central density step, constant number of particles / bin)

The errorbars around the scale radius of 1000 pc of the dark matter cusp or the scale lengths of 500 pc or 1000 pc for the scale radii of the stellar components are not decreased, thus indicating that the relative values of the enclosed mass are not better constrained than at any other radius.

3.2 Cored Model

For a cored profile, we have a similar result, see fig. 2. **(TODO: rerun)**

Best restrictions are around 500pc again, only this time a little too low.

3.3 Triaxial mock data

(TODO: motivation: triaxiality of observed dwarfs, effects from other mass modelling schemes)

The models were generated with a Made 2 Measure algorithm of **?** and are tailored to follow a similar profile to the profiles specified above for the dwarf galaxies. They show a density profile of

$$\rho(r) = \frac{\rho_S}{\left(\frac{r}{r_S}\right)^\gamma \left(1 + \left(\frac{r}{r_S}\right)^{1/\alpha}\right)^{\alpha(\beta-\gamma)}} \tag{24}$$

with radius r, scale radius $r_S = 1.5$kpc, $\alpha = 1$, $\beta = 4$. For the cusped profiles we have an inner logarithmic slope of $-\gamma = -1$, scale density $\rho_S = 5.522 \cdot 10^7 M_\odot/\text{kpc}^3$, and $M_{\rm tot} = 1.171 \cdot 10^9 M_\odot$, while for the cored one we have $\gamma = 0.23$, $\rho_S = 1.177 \cdot 10^8 M_\odot$, $M_{\rm tot} = 1.802 \cdot 10^9 M_\odot$. The axis ratios are $b/a = 0.8$ and $c/a = 0.6$. The stars have negligible mass and follow the same functional form in the density profile as dark matter, with $\alpha = 0.34, \beta = 5.92, \gamma = 0.23, r_S = 0.81$kpc.

The velocity anisotropy of the stellar part is calculated via

$$\beta(r) = \frac{r_{s,\beta}^\eta \beta_0 + r^\eta \beta_\infty}{r^\eta + r_{s,\beta}^\eta}, \tag{25}$$

with $r_{s,\beta} = 0.81$kpc, $\beta_0 = 0$, $\beta_\infty = 0.5$ and $\eta = 0.5$, going from isotropic to radially anisotropic with increasing radius.

The retrieved density profile (fig. 3) is constantly overestimating the density.

(TODO: other projections) (TODO: reason: projection)

4 CONCLUSIONS

Conclusions.

5 ACKNOWLEDGEMENTS

JIR would like to acknowledge support from SNF grant PP00P2_128540/1.

REFERENCES

Baes M., Dejonghe H., 2002, A&A, 393, 485
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Breddels M. A., Helmi A., van den Bosch R. C. E., van de Ven G., Battaglia G., 2012, ArXiv e-prints

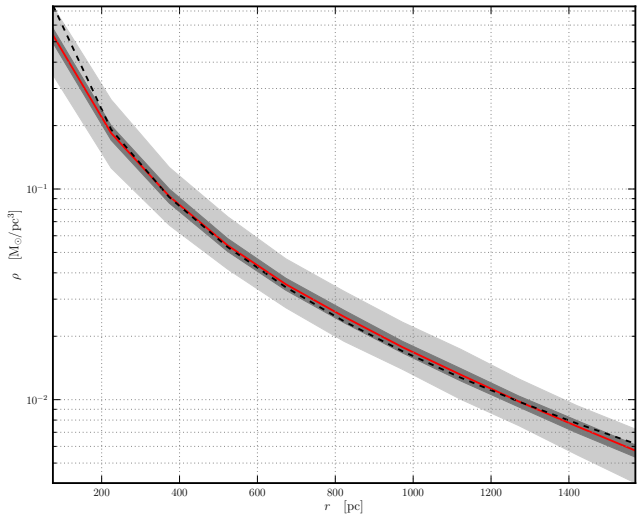

Figure 1. Reconstructed mass of a cusped model (red shows median, shaded areas are 68 and 90 percentiles) for $(2800, 4033)$ tracer particles, after 50000 iterations. The black dashed curve shows the underlying theoretical model.

Figure 2. A cored profile: Reconstructed mass of the MCMC model (red shows median, shaded areas are 68 and 90 percentiles) for 10^4 tracer particles after 30000 iterations. The black dashed curve shows the underlying theoretical model.

Cole D. R., Dehnen W., Read J. I., Wilkinson M. I., 2012, MNRAS, 426, 601
de Blok W. J. G., McGaugh S. S., Rubin V. C., 2001, AJ, 122, 2396
Goerdt T., Moore B., Read J. I., Stadel J., Zemp M., 2006, MNRAS, 368, 1073
Governato F. et al., 2010, Nature, 463, 203
Hernquist L., 1990, ApJ, 356, 359
Łokas E. L., 2002, MNRAS, 333, 697
Mateo M. L., 1998, ARA&A, 36, 435
McGaugh S. S., Rubin V. C., de Blok W. J. G., 2001, AJ, 122, 2381
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Oh S.-H., Brook C., Governato F., Brinks E., Mayer L., de Blok W. J. G., Brooks A., Walter F., 2011, AJ, 142, 24
Read J. I., Pontzen A. P., Viel M., 2006, MNRAS, 371, 885
Schwarzschild M., 1979, ApJ, 232, 236
Spergel D. N., Steinhardt P. J., 2000, Physical Review Letters, 84, 3760
Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740
Walker M. G., Peñarrubia J., 2011, ApJ, 742, 20

6 APPENDIX

6.1 Hernquist Models: Check of Integration routines

To check the correct function of the integration routine, we used the analytic formulas from Hernquist (1990):

$$M(r) = M\frac{r^2}{(r+a)^2} \tag{26}$$

$$\nu(r) = \frac{M}{2\pi}\frac{a}{r}\frac{1}{(r+a)^3} \tag{27}$$

$$\beta(r) = 0 \tag{28}$$

with mass scale M and length scale a as inputs, and calculated $\sigma_{\rm LOS}(r)$ with our numerical routine, including three integrations. Working with extrapolations to the first bin turns out to give the most stable results, corresponding nicely to the analytic value (Baes & Dejonghe 2002) of

$$\begin{aligned}
\sigma_{\rm LOS}(r) &= \frac{1}{I(r)}\frac{1}{24\pi(1-r^2)^3}\times \\
&\quad [3r^2(20-35r^2+28r^4-8r^6)X(r) \\
&\quad +(6-65r^2+68r^4-24r^6)]-\frac{r}{2}, \\
\sigma_r(r) &= r(1+r)^3\ln\left(\frac{1+r}{r}\right) \\
&\quad -\frac{r(25+52r+42r^2+12r^3)}{12(1+r)}, \\
I(r) &= \frac{1}{2\pi}\frac{(2+r^2)X(r)-3}{(1-r^2)^2}, \\
X(r) &= \begin{cases}(1-r^2)^{-1/2}\mathrm{arcsech}\, r, & \text{for } 0\leqslant r\leqslant 1,\\ (r^2-1)^{-1/2}\mathrm{arcsecans}\, r, & \text{for } 1\leqslant r\leqslant\infty\end{cases}.
\end{aligned}$$

A single component model is set up according a simple Dehnen split power-law sphere Read et al. (2006),

$$\rho(r) = \frac{M_\infty(3-\gamma)}{4\pi r_S}\left(\frac{r}{r_S}\right)^{-\gamma}\left(1+\frac{r}{r_S}\right)^{4-\gamma}, \tag{29}$$

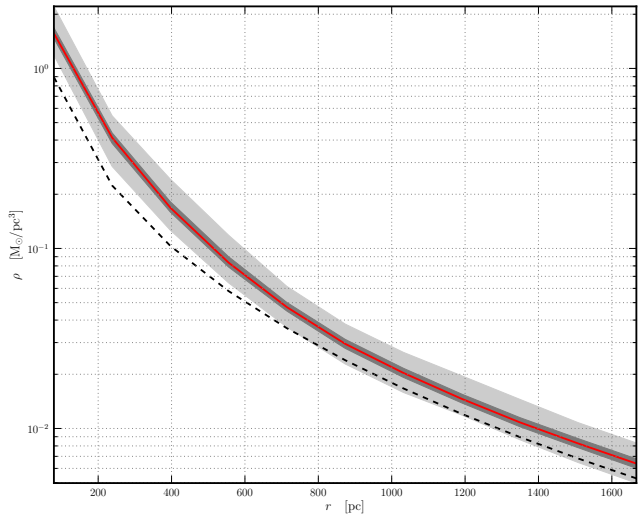

Figure 3. Density profile of a triaxial mock dwarf, for which the line of sight is inclined with 45 degrees with respect to all axes.

where $M_\infty = 1$ denotes the total mass, $\gamma = 1$ the logarithmic central asymptotic slope, and $r_S = 1$ the scale length. We use 10^6 sample points, outof which 10^4 are extracted for further analysis.

We then let ν, ρ, β vary. The MCMC correctly recovers the underlying mass distribution.

6.2 Data quality

How many tracer stars are needed to determine the overall density profile reliably? To address this question, we performed three runs with a restricted set of tracer particles. In the first, 10^3 particles were chosen out of the 10^6 simulated particles. With 10^4 particles, the confidence intervals shrink. These 10^4 particles are split then into two populations of each $5 \cdot 10^3$ particles, with different scalelengths of r_S and $r_S/10$. Most of the second population particles are inside the first two bins, so the overall convergence is not visibly affected above the third bin. However, the models are better constrained around the scalelengths of both tracer populations. This is expected from Walker & Peñarrubia (2011), as any velocity anisotropy sampling yields the same mass constraint there.

6.3 Effects of turning off priors

(TODO: turning off $\delta dens > 0$) (TODO: $\beta < 0$ prior)

6.4 Convergence in MCMC

We check the convergence of the MCMC twofold: first, the range of density profiles swept after 5k, 50k and to the end of our run is increased from 5k to 50k, but stays approximately constant for another 16k runs, thus giving us confidence that we found and fully explored the valid regions of phase space in density. See fig. 5.

Second, the range of δ_i, the only other unrestricted parameters, show a similar behaviour, as shown in 6 for δ_1, with analogous δ_2. Not all the parameter space is sampled, and this is a major cause of worry: any profile found in the first iterations in the init phase will be stuck there, if δ cannot sample a big subset of $]-\infty, 1]$.

(TODO: rerun cusp for proper 10^6 iterations!)

6.5 Convergence in Mass Models

(TODO: constant number of particles per bin)

6.6 Splitting by metallicities

Observations of the abundances of metals and chemical species in the stellar atmospheres show that the ensemble of stars in a dwarf galaxy or globular cluster can be split into populations.

The first approach by **?** showed that if the population of e.g. Fornax is split into two populations, and each of their half-light radius and mass are determined, restrictions on the overall potential can be drawn. Using this approach, they prefer a cored DM profile for Fornax.

In our test suite there are dwarf galaxies with different scale radii and small differences in the mean of the metallicity for the two populations of stars. In order to reproduce the underlying populations we use an inset MCMC with assumptions that

(i) Foreground stars are younger than most of the dSph member stars. Therefore, they show a high metallicity and can be removed from the dataset with a single cut in metallicity;
(ii) the remaining stellar components are divided into two populations;
(iii) the fraction of stars in population 1 is sampled in a uniform way in the range $[0.2, 0.8]$;
(iv) both populations show a normal distribution in metallicity with the same width;
(v) the initial values of the means are set to half and twice the mean metallicity, to allow for a reasonable difference between the means. This difference is then subsequently sampled assuming a normal distribution;

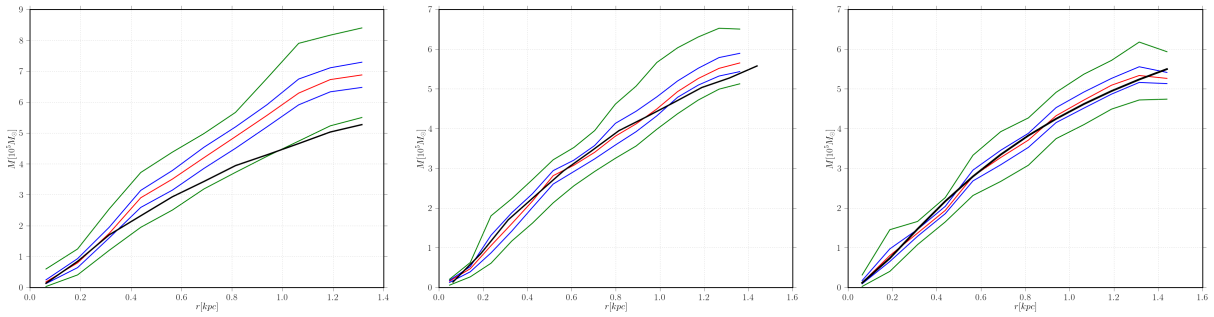

Figure 4. Hernquist profile found by MCMC model (red) for 10^3, 10^4 and two times $5 \cdot 10^3$ tracer particles. The black curve shows the enclosed mass derived from the theoretical model.

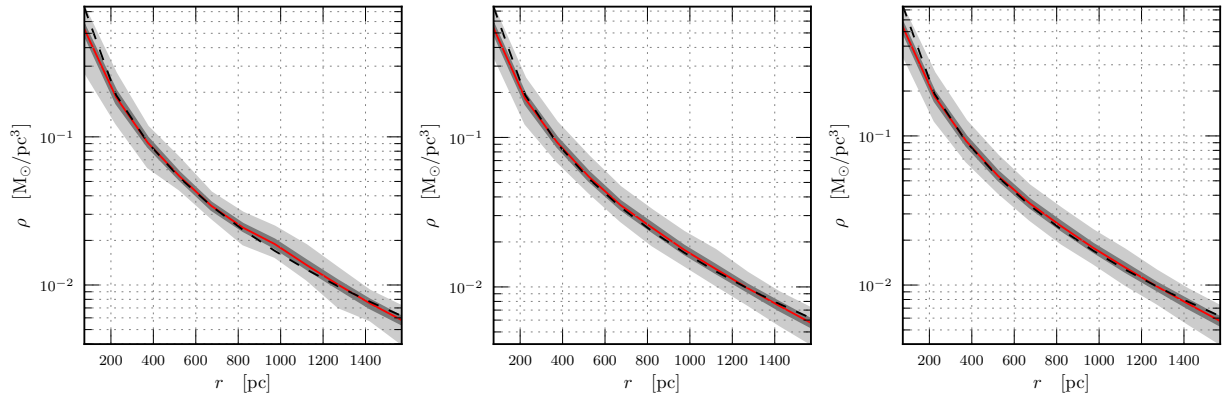

Figure 5. Convergence of the density profile after (5k,50k,67k) iterations (left to right).

(vi) 40000 iterations for burn-in and 10000 iterations for subsequent parameter estimates are used, with a thinning by factor 10 to reduce correlations between subsequent models. This procedure converges in all tested cases.

To test whether the assignment into populations is a valid one, we want to check whether the population is in equilibrium with the overall potential.

The routine then assigns each particle to one of the two populations, based on its Mg metallicity. $75 \pm 4\%$ of all stars are assigned to the correct underlying distribution. This in turn changes the half-light radius by 110pc and -62pc for initial 390 pc, 730 pc half light radii. These changes are rather high, but the two populations still show distinct half-light radii.

We explicitly assume two populations of stellar tracers in dwarf galaxies, each with Gaussian distributions in metallicity with means μ_1, μ_2. Without requiring a minimum distance $\delta\mu = \mu_2 - \mu_1$ between them, a representation with $\mu_1 \approx \mu_2$ can be found. This model shows a higher χ^2 than other models and is thus disfavoured, but cannot be rejected from a Kolmogorov-Smirnov test on a $p < 0.05$ basis.

Here we show the influence of setting a prior minimal $\delta\mu_{\rm min}$ on the goodness of fit, and the allowed range of $\delta\mu$. We work on the metallicity distribution of one of the mock dwarfs with cusped density profiles described earlier on, setting

$$\mu_1 \in [-1.0, 2.0]; \quad \delta\mu \in [\delta\mu_{\rm min}, 5.0]; \tag{30}$$

with $\delta\mu_{\rm min}$ varying from 0.0 to 0.4. We let the MCMC run for a) 10k iterations with 8k burn-in/discarded models; b) 50k iterations with 40k burn-in. From the accepted models, we compute the mean Gaussian distributions and compare the corresponding overall bimodal distribution to the actual metallicity distribution from the data with a 2 component Kolmogorov-Smirnov test, and take the two-tailed statistics p from 30 drawings. If the $p > 0.05$, then we cannot reject the hypothesis that the distributions of the two samples are the same.

Results for varying the minimal distance between the two Gaussians between 0.0 and 0.4 are shown in fig 8.

All models with $p > 0.05$ give a reasonable fit, with a maximum for 10k iterations at around $\delta\mu = 0.1$. Models with $\delta\mu > 0.2$ give no good fit anymore after 10k iterations.

The goodness of fit is enhanced if we take more iterations, so in the plot for 50k iterations, there is a maximum $p = 0.79$ compared to $p = 0.4$ from 10k iterations only. The whole curve is shifted to higher $\delta\mu$ values. The models with $\delta\mu > 0.3$ are rejected. The restriction of $\delta\mu > 0.4$ (last point to the right) is well-fitting again, but this is due to the fact that the fraction of particles for population 2 was found to be smaller than 10 percent, thus mainly fitting the metallicity distribution with one Gaussian only, plus some skewing from an almost negligible stellar component. Although this

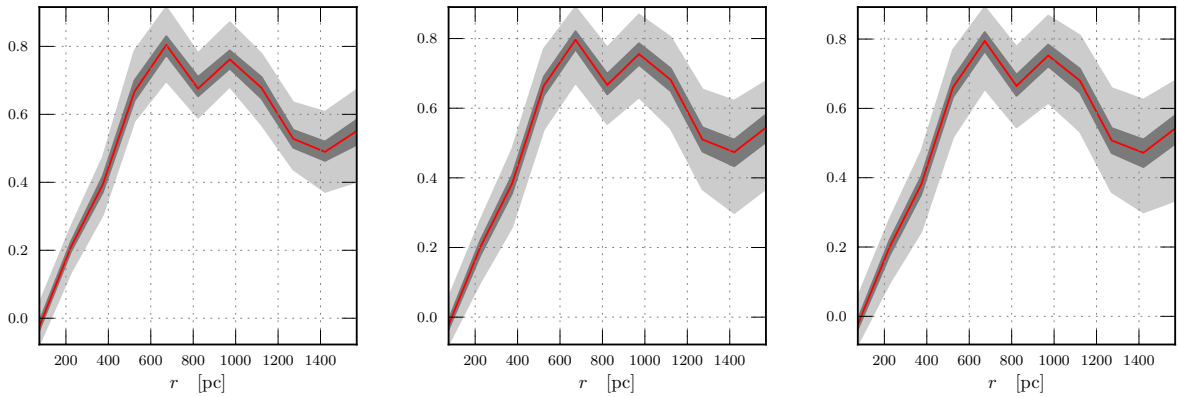

Figure 6. Convergence of δ_1 after (5k,50k,67k) iterations (left to right).

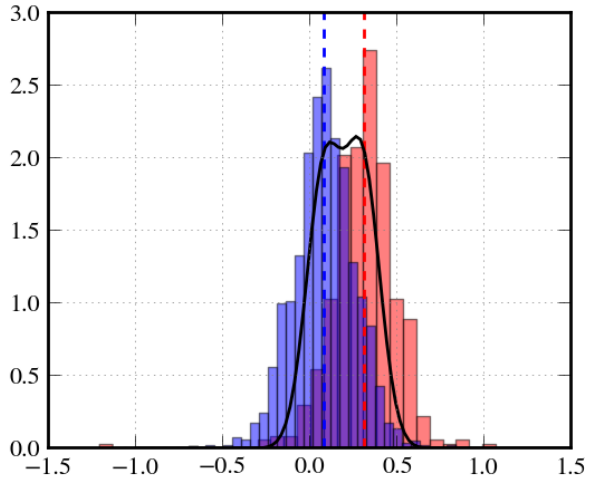

Figure 7. Distribution function of two populations (filled histograms) and the reproduced overall distribution.

model cannot be rejected, it lies above the rejected models at $\delta\mu_{\min} = 0.4$. Furthermore, it does not yield a second component with a scale length distinctly different from the main component, rendering the additional gain from two components obsolete. Thus, we will restrict the MCMC search of the population fraction to the range $f \in [0.3, 0.7]$.

6.7 Effect of Wrong Assignment of Populations

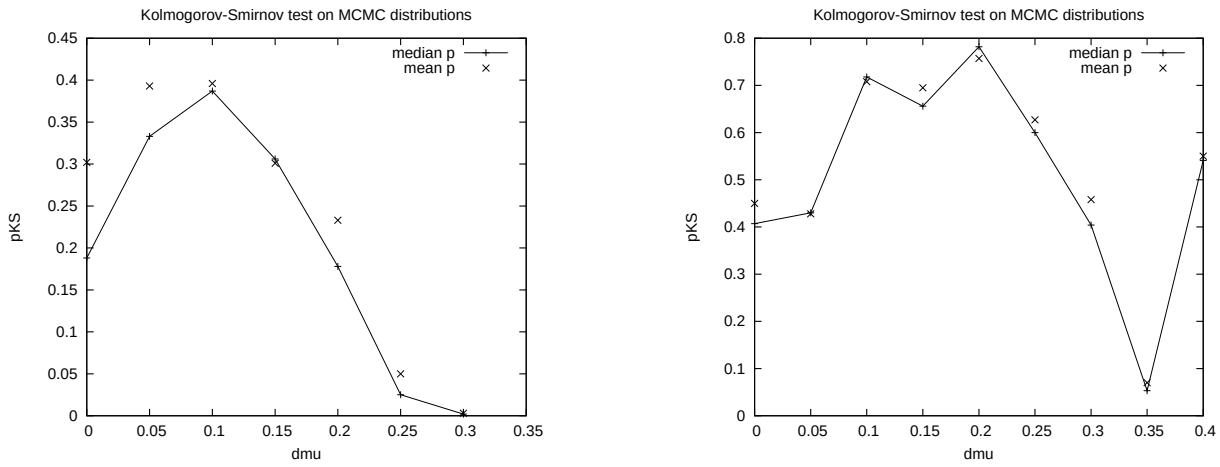

Figure 8. Kolmogorov-Smirnov test for correspondance between models with $\delta\mu > \delta\mu_{\rm min}$ as described in the text.