# Basis Sets in Galactic Dynamics

Edward James Lilley

Churchill College

UNIVERSITY OF
CAMBRIDGE

2020 July

This thesis is submitted for the degree of Doctor of Philosophy

# Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

# Summary


## Basis Sets in Galactic Dynamics

### Edward James Lilley

Since their introduction in 1972 by Clutton-Brock, so-called *biorthogonal* basis sets have become a popular tool in galactic dynamics. They provide a method of solving Poisson's equation that scales linearly with the number of particles in a galaxy or halo simulation, and have also received wide attention in the perturbation theory of disk galaxies.

This thesis begins by discussing the theory behind such basis sets, and the context of their applications. Some introductory results are presented, deriving the possible functional forms that basis sets can take, and identifying the lack of suitably flexible basis sets in the literature as a stumbling block to further usage of the technique.

Subsequently, several new families of basis sets are derived, whose free parameters and resemblance to classic double-power law formulas provide a much needed increase in flexibility for the modelling of realistic galaxies and haloes. Along the way a simple yet under studied spherical double-power model is described that interpolates between the properties of the more famous NFW and Hernquist models.

Finally, we turn to an application of the new basis sets: the problem of efficiently re-simulating cosmological dark matter haloes. The ability to place new objects in a realistic time-evolving gravitational potential is desirable from a modelling point of view, as it permits us to constrain the properties of massive, unseen galactic components (dark haloes) via visible dynamical tracers such as stellar streams.

The thesis concludes by suggesting some further practical applications of the basis set technique in astronomy, as well as some pointers towards future theoretical developments and applications to other areas of physics.

# Contents

# Preface

Chapters 3, 4 and 5 are based, respectively, on the following publications: Lilley et al. (2018a), Lilley et al. (2018b) and Lilley et al. (2018c).

Ch. 7 is based on a draft paper currently submitted to Monthly Notices of the Royal Astronomical Society, co-authored with Wyn Evans, Jason Sanders, Eugene Vasiliev and Denis Erkal. The description and implementation of the spline-based method therein was carried out by Eugene Vasiliev.

One result in Ch. 2 was obtained partially via a question posted on MathOverflow (http://mathoverflow.net): the first two cases of Eq. (2.39) were given by Robert B. Israel[a].

[a] https://mathoverflow.net/questions/263996/an-indefinite-integral-containing-functions-that-are-solutions-to-a-2nd-order-li

# Acknowledgements

First of all, I would like to thank my very patient and helpful supervisors, Prof. Wyn Evans and Dr. Jason Sanders, for guiding me through the whole PhD process from start to finish. Their comments were edifying and their scientific and mathematical contributions were invaluable.

Secondly, thanks to the Streams Group and certain other individuals at the Institute of Astronomy for their useful input along the way. In particular, Vasily Belokurov, Eugene Vasiliev, Denis Erkal, Matt Auger and Donald Lynden-Bell.

Lastly, thanks to my supportive friends: Sam, Finlay, Douglas, Alastair, Moira, Hamish, Simon, Max, Sophia, Luke, Johanna, Dan, Chris, Bess, my sister Helen, and above all to Aïda, without whom this would not have been possible.

# Chapter 1

# Introduction

The main part of this thesis concerns mathematical techniques for modelling self-gravitating astrophysical objects. We are mostly interested in applying these techniques to model the morphology and evolution of dark matter haloes. Therefore to motivate the study of these haloes we briefly recapitulate the standard Λ-CDM theory of cosmology, that describes structure formation in the universe. This prefatory material is based loosely on Dodelson (2003, Ch. 6, 7 & 9) and Mo et al. (2010, Ch. 7).

## 1.1 Λ-CDM and the origin of dark matter haloes

According to the consensus Λ-CDM[a] model of cosmology, the origin of structure in the universe lies during the early stages of the Big Bang, when an inflationary field undergoes quantum fluctuations. This field $\zeta$ is presumed to be an isotropic Gaussian random field, with perturbations $\Delta_\zeta$ defined entirely by a power spectrum $P_\zeta(k)$,

$$P_\zeta(k) \equiv \frac{1}{(2\pi)^3} \left\langle \Delta_\zeta(\mathbf{k})^2 \right\rangle \propto \|\mathbf{k}\|^{n_s-4}, \tag{1.1}$$

where $\mathbf{k}$ is the Fourier-space coordinate and $n_s$ the *spectral index*, which is close to unity and varies weakly as a function of $\|\mathbf{k}\|$.

These primordial fluctuations source near-linear perturbations in the various matter and radiation fields. We are interested in the $\sim 30\%$ of the universe's mass-energy budget that is in matter, most of which is in cold dark matter (DM), so we will describe the matter anisotropies, as they give rise to the massive DM *haloes* in which galaxies reside.

[a] 'Λ' refers to the model of dark energy, 'CDM' (cold dark matter) refers to the theoretical form of matter that simultaneously provides for both the peaks in the cosmic microwave background data, and the observed rotation curves of galaxies.

Like the inflationary field, the matter fields behave like isotropic Gaussian random fields, each however with a power spectrum that has evolved away from the primordial spectrum due to various interactions during the successive epochs of the Big Bang. Each perturbation evolves in a different way based on its scale, matter species and the given epoch. In particular, at late times over-densities of cold dark matter smaller than the cosmological event horizon evolve according to the *Meszaros equation*: each over-density $\delta\rho_m$ eventually grows linearly with the cosmological scale factor $a(t)$.

However, each over-density also experiences its own gravity, so to follow the evolution precisely we must eventually leave the regime of linear perturbation theory. We can track the size of a perturbation of mass $\delta\rho_m$ using the dimensionless quantity

$$\delta \equiv \frac{\delta\rho_m}{\bar{\rho}_m}, \tag{1.2}$$

where $\bar{\rho}_m$ is the background matter density of the universe (evolving according to the Friedmann equations).

When a given perturbation grows in the linear regime to the critical value $\delta \approx 1.69$ it collapses under its own self-gravity, and a halo is formed. As this collapse gets underway, the kinetic energy of in-falling particles is transferred to random disordered motion as the particles describe orbits in their own gravitational potential. The collapse stops when this kinetic energy in random motion $K$ balances the potential self-energy $V$, i.e. when the *virial theorem*

$$V = -2K, \tag{1.3}$$

is satisfied. The halo is then said to have *virialised*. A heuristic argument shows that the density of such a virialised halo is approximately $\rho_{\text{vir}} \approx 200 \times \bar{\rho}_m$. The spatial extent of a given halo is conventionally defined as the region where $\rho \geq \rho_{\text{vir}}$. The size of this region is denoted $r_{\text{vir}}$, and the enclosed mass within $r_{\text{vir}}$ is $M_{\text{vir}}$. The mass of a halo thus depends on its formation time.

The statistics of DM haloes formed in this way are described using the quantity $\sigma^2$, which is the variance of the matter perturbations on a certain length-scale, typically $R = 8\,\text{Mpc}$,

$$\sigma^2 \equiv \frac{1}{2\pi^2} \int P_m(k)\, \hat{W}(k, R)^2\, k^2\, \mathrm{d}k. \tag{1.4}$$

Here $\hat{W}$ is the Fourier transform of a windowing function (e.g. a top hat function) that smooths the density field of perturbations. A larger observed value of $\sigma^2$ corresponds to larger matter fluctuations, indicating that structure formation occurred earlier.

A model known as the *Press-Schechter formalism* (Press & Schechter, 1974) has been developed in order to derive the number distribution of haloes from first principles. The magnitudes of the density peaks are modelled as undergoing random walks, with a probability of eventually passing a given mass threshold. Because the fluctuations of the underlying density field are Gaussian, the statistics can be calculated analytically. The number density of collapsed objects with masses between $m$ and $m + \mathrm{d}m$ is derived to be

$$n(m) \propto \sqrt{\frac{2}{\pi}}\, \frac{\mathrm{e}^{-m^2\bar{\rho}^2/(2\sigma^2)}}{\sigma}, \tag{1.5}$$

so haloes are predicted to be distributed normally with a variance proportional to $\sigma^2$. This variance varies as a power of the mass, and observationally it is

$$\sigma \propto a(t)\, m^{\frac{-1}{6}(n_s-1)}, \tag{1.6}$$

where $a(t)$ is the cosmological scale factor obtained by solving the Friedmann equations. Structure is thus formed hierarchically, with larger DM haloes hosting galaxies, and smaller haloes (or *subhaloes*) furnishing those galaxies with satellites – some of which themselves host smaller 'dwarf' galaxies.

The Press-Schechter formalism gives results impressively close to full-blown numerical simulations of structure formation (White et al., 1993), but says nothing about the actual morphology of the resulting haloes. To understand their formation we must first learn how to simulate the nonlinear collapse of over-densities under their own self-gravity – these are $N$-body methods, which we discuss in Sec. 1.2.

Before we leave behind cosmology, however, we briefly mention a *cosmological* motivation for studying the dynamics of the Milky Way. One of the tensions in $\Lambda$-CDM is the *subhalo problem* (Del Popolo & Le Delliou, 2017). The number of small haloes that orbit a large halo, such as the Milky Way, is over-predicted by theory compared to the number of observed satellites. One way of resolving this tension is to more closely measure the gravitational potential of the Milky Way, looking for the signatures of subhaloes that do not contain any visible matter, and taking into account the time-evolution of the potential. This is almost certainly significant, not least because of the close approach of the Large Magellanic Cloud to the MW over the past few gigayears.

## 1.2 $N$-body methods

Given initial conditions predicted by the linear theory of structure formation outlined in the preceding section, we find the universe populated with massive, self-gravitating objects – dark matter haloes – whose subsequent evolution is computationally expensive to follow. This nonlinear evolution is governed by Poisson's equation, which links the spatial distribution of matter $\varrho$ to its corresponding gravitational potential $\psi$,

$$\nabla^2\psi = 4\pi G\varrho. \tag{1.7}$$

The equation of motion for each individual particle is then the negative of the gradient of the potential,

$$\mathbf{a} = -\boldsymbol{\nabla}\psi. \tag{1.8}$$

While the potential controls the evolution of the system, the full information about the state of the system is contained within the distribution function (DF) $f(\mathbf{x}, \mathbf{v}, t)$, which represents a probability distribution over six-dimensional phase space – three spatial dimensions and three velocity components[b]. The evolution of $f$ is determined by coupling Eq. (1.7) to Eq. (1.8) and

[b] In fact each species of matter has its own distribution function, but here we consider only dark matter; the six dimensional phase space may also be augmented by additional information about the species under consideration, see e.g. Sanders & Binney (2015).

imposing a local conservation law for the flow of particles through phase space, resulting in the *Vlasov equation*,

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \boldsymbol{\nabla} f - \boldsymbol{\nabla}\psi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0. \tag{1.9}$$

For a population of tracer particles in a fixed background potential, we can take moments of Eq. (1.9) with respect to powers of the velocity. If the geometry is simple these reduce to the *Jeans equations*, which can be solved fairly simply.

However, in the case where we expect the mass distribution to self-consistently generate its own gravitational potential, it is much harder to solve Eq. (1.9) directly, so the most generally useful tool to investigate the dynamical behaviour of such systems is the *$N$-body simulation*. The mass distribution is represented on a computer as a cloud of $N$ point sources, the $i$-th particle having mass $m_i$ and position vector $\mathbf{r_i}$. We may then write the density using a Dirac delta function for each particle,

$$\varrho(\mathbf{r}) = \sum_i^N m_i \delta\left(\mathbf{r} - \mathbf{r}_i\right), \tag{1.10}$$

so that the gravitational potential is the sum of the corresponding Keplerian potentials

$$\psi(\mathbf{r}) = -\sum_i^N \frac{G\, m_i}{\|\mathbf{r} - \mathbf{r}_i\|}. \tag{1.11}$$

Thus the acceleration on the $j$-th particle is

$$\mathbf{a}_j = -\sum_{i \neq j}^N \frac{G\, m_i\, m_j \left(\mathbf{r}_i - \mathbf{r}_j\right)}{\|\mathbf{r}_i - \mathbf{r}_j\|^3}. \tag{1.12}$$

The positions of the particles are then updated on the computer using a suitable integration scheme (e.g. the Leapfrog method). Clearly, this *direct* method of simulation requires $\mathcal{O}(N^2)$ operations to calculate all the forces on the particles.

The direct method of Eq. (1.11) is generally limited to around $10^6$ particles even with modern GPU acceleration (Aarseth, 1999, Wang et al., 2015). However, a Milky Way-sized galaxy contains on the order of $10^{11}$ stars, and so using a direct $N$-body simulation for just the galaxy's stars alone is still computationally out of reach; to say nothing of the number of dark matter particles in a galaxy's DM halo (likely $> 10^{70}$).

However, the situation is saved because many astrophysical systems are *collisionless*, meaning that relaxation effects (or collisions) between bodies are negligible. This enables more efficient $N$-body methods to be used. For stellar systems, the collisionless property arises because the characteristic timescale over which stars have close encounters that impulsively change their momentum is around $t_{\text{cross}} \times 0.1N/\log N$. Here $N$ is the number of stars in the system, and $t_{\text{cross}}$ the crossing time – the size of the system divided by the typical velocity of the stars (Binney & Tremaine, 1987). Cold dark matter, on the other hand, is also collisionless, but this is more straightforwardly because the particles have negligible interaction cross-sections with both ordinary matter and other dark matter particles. Thus they experience non-gravitational forces extremely feebly, and collisions can be neglected.

In collisionless systems the particles move solely under the influence of a smooth, mean gravitational field. This fact enables the use of alternative techniques for calculating the force that are more computationally efficient than the direct summation method. Such methods, that we might term *indirect*, come in several classes[c].

Firstly, the *multipole* method (see Sec. 1.3.2.1). Suitable for isolated centrally-concentrated systems, this method expands the gravitational field in a series of harmonics truncated to some order $l_{\max}$. The force computation is $\mathcal{O}(l_{\max}^2 \log N)$, which is then evaluated at $N$ particle locations.

Secondly, *tree codes*, introduced by Barnes & Hut (1986). Particles are divided recursively into cells. For a given particle, the force due to a sufficiently faraway cell is then calculated using a multipole expansion. This method can be improved by considering each cell to interact as a whole via its multipoles (essentially expanding both force sinks and sources using the multipole method). This is known as the *fast multipole* method (FMM) (Greengard & Rokhlin, 1987, Dehnen, 2014). The scaling of the FMM force-evaluation with particle number lies somewhere between $\mathcal{O}(N)$ and $\mathcal{O}(N \log N)$, and there is also an $\mathcal{O}(l_{\max}^2)$ cost in both sinks and sources. The method is also advantageous due to its avoiding certain numerical blow-ups inherent in the basic tree code method.

Thirdly, *particle-mesh* (PM) algorithms. These interpolate the positions of particles onto a grid of cells, and compute the force using a fast Fourier transform (FFT), using $\mathcal{O}(N_{\text{cells}} \log N)$ steps, at each of the $N$ particles. This can be optimised, e.g. by placing additional cells in regions where the density is higher, or by using non-rectangular cells.

Fourthly, the *self-consistent field* (SCF) method (Hernquist & Ostriker, 1992, Hernquist, Sigurdsson & Bryan, 1995). This is conceptually similar to the multipole method, but augments it with a radial expansion (truncated at some order $n_{\max}$), thus avoiding the $\mathcal{O}(\log N)$ particle-sorting step required in the former case. The expansion coefficients are computed once per time-step in $\mathcal{O}(n_{\max} l_{\max}^2 N)$ steps, and then evaluated once for each $N$ particles – thus the method scales ideally with increasing computer core count. It is the SCF method which is the subject of this thesis. Specifically, we investigate the construction of suitable sets of basis functions such that the zeroth-order usefully matches some system of interest. This is important, as a large number of basis functions would be required to accurately represent a mass distribution about which nothing is known a priori, potentially increasing the complexity to $\mathcal{O}(N^2)$ (see e.g. Kalapotharakos et al., 2008). Due to the requirement that the system under study remain at all times fairly close to the basis set used, the SCF method has traditionally been used mostly for long-term perturbation analyses of stable systems. However, this does not preclude it being used in principle in more chaotic dynamical situations. For example, Hozumi et al. (2019) has examined the merger of two galaxies, each of which is represented by a separate basis expansion.

The process by which baryonic matter populates DM haloes to form visible galaxies is beyond the scope of this thesis, but typically either a disk-like or an ellipsoidal structure is formed; disks (like the Milky Way) may also have a central ellipsoidal bulge. The addition of baryonic matter to dynamical simulations of DM haloes typically requires a combination of $N$-body methods combined with techniques such as *smoothed particle hydrodynamics* (SPH), exemplified by the GADGET code

[c] This review is loosely based on Trenti & Hut (2008).

(Springel, 2005) which was used to simulate the haloes considered in Chapters 3–7 of this thesis.

## 1.3 Models of dark haloes

### 1.3.1 Models with spherical symmetry

As described in Sec. 1.1, haloes comprised of dark matter arise from the collapse of over-densities in the initial conditions of the universe, eventually attaining mildly triaxial forms (Jing & Suto, 2002). The exact shape is debated as they can only be observed indirectly via numerical simulations of structure formation.

Dark haloes provide a large majority of a galaxy's mass budget, and galactic rotation curves are found to be flat over a large range of radii. This implies any model of the halo must have an approximately *isothermal* logarithmic slope over this range, i.e.

$$\frac{\mathrm{d}\log\rho}{\mathrm{d}\log r} \approx -2. \tag{1.13}$$

The isothermal terminology comes from the *singular isothermal sphere* model, which has this property everywhere,

$$\rho(r) \propto r^{-2}, \qquad \psi(r) \propto \log r, \tag{1.14}$$

but it is not a realistic model in its own right due to the divergent nature of all quantities as $r \to 0$ and $r \to \infty$.

A simple model of spherical collapse, such as that considered in Sec. 1.1, predicts that over-densities in dark matter particles will collapse to form approximately isothermal spheres (Gunn & Gott, 1972, Shapiro et al., 1999). $N$-body simulations confirm that $\rho \sim r^{-2}$ over a moderate range of $r$, but the density profile steepens at very large radii, to $\rho \sim r^{-\beta}$ with $\beta \approx 3$. While it is required that $\beta > 3$ in order for the total mass to remain finite, the most popular halo model in fact has $\beta = 3$ exactly. This is the Navarro, Frenk & White (1997) (hereafter NFW) model, whose functional form is

$$\begin{aligned} \psi_{\mathrm{NFW}} &= \frac{4\pi G\rho_0 r_\mathrm{s}^3}{r}\log(r+r_\mathrm{s}), \\ \rho_{\mathrm{NFW}} &= \frac{\rho_0 r_\mathrm{s}^3}{r(r+r_\mathrm{s})^2}. \end{aligned} \tag{1.15}$$

Note that for completely spherically symmetric models, the Laplacian $\nabla^2$ reduces to the form (which we will generally denote $\nabla_r^2$)

$$\nabla_r^2 \equiv r^{-2}\frac{\mathrm{d}}{\mathrm{d}r}r^2\frac{\mathrm{d}}{\mathrm{d}r}. \tag{1.16}$$

We also have $M(r) \propto r\,\mathrm{d}\psi/\,\mathrm{d}r$ for the enclosed mass of a spherically symmetric model, so it is easy to see that the total mass as $r \to \infty$ of the NFW model diverges.

Any given halo is parameterised in terms of two dimensionful parameters the *virial mass* $M_{\mathrm{vir}}$ and *virial radius* $r_{\mathrm{vir}}$, which are assigned to a given halo on the basis of cosmology considerations

(see Sec. 1.1). The NFW model then provides a single additional dimensionless parameter $c$, termed the *concentration* (the ratio of $r_{\rm vir}$ to the intrinsic scale-length of the model $r_{\rm s}$), which is varied so as to provide the best fit to the density within $r_{\rm vir}$[d]. This trick of fitting the model to just the part of the halo contained within virial radius is what allows models with infinite total mass, such as the NFW, to be used. When a large number of simulated DM haloes are fit with the NFW model, they are found to lie roughly along a line in $(c, M_{\rm vir})$-space (with some scatter), a finding that is referred to as the *mass-concentration* relation (Dutton & Macciò, 2014, Diemer & Kravtsov, 2015).

While detailed dark matter-only simulations reveal that the NFW model provides a convenient universal form for haloes, there exist variations that claim a better fit. For example the Einasto (1965) model has recently begun to rival the NFW as a possible universal halo form (Navarro et al., 2004).

Another popular model which has a particularly simple analytical form is the Hernquist (1990) model,

$$\begin{aligned} \psi_{\rm Hq} &= \frac{GM}{r+b}, \\ \rho_{\rm Hq} &= \frac{Mb}{2\pi}\frac{1}{r(r+b)^3}. \end{aligned} \tag{1.17}$$

This model was proposed for spherical haloes or galactic bulges, and shares many of the simple analytical properties of the famous Plummer (1911) model. In fact, they form two members of the sequence of models introduced by Veltmann (1979), which were termed the '$\alpha$' models by Zhao (1996) and whose 'hypervirial' dynamical properties were studied by Evans & An (2005). The Hernquist model is now disfavoured as a DM halo model due to the steepness of its outer slope ($\beta = 4$), but still finds use as a simple model of central galactic *bulges* (or *nuclei*). The projected light distributions of such bulges are well-modelled by the empirical law of de Vaucouleurs (1953), with which the Hernquist model agrees fairly well when projected along the line of sight.

There is still debate about the true morphology of haloes. In particular, the *cusp-core controversy* (de Blok, 2010) revolves around the disagreement between properties of DM haloes as inferred from the observed rotation curves of dwarf galaxies, and from simulations of pure DM haloes. Denoting the asymptotic behaviour of the density profile as $r \to 0$ as $r^{-\gamma}$, then the centre of the halo is referred to as a *cusp* if $\gamma > 0$, and a *core* if $\gamma = 0$ (note that we must always have $\gamma < 3$ in order for the enclosed mass to remain finite). The disagreement, therefore, is that cores or mild cusps are implied by the observed rotation curves of dwarf galaxies ($\gamma = 0.2$ or shallower), and steeper cusps are seen in the centres of DM $N$-body simulations ($\gamma \approx 0.8$, though both steeper and shallower slopes have been claimed). This tension is evidence for an interaction between the DM particles and baryons that is not being correctly modelled in the simulations, or else an alternate theory entirely is called for. For example both *warm* and more recently *fuzzy* dark matter (Burkert, 2020) predict that a core will form instead of a cusp, due to non-negligible interactions between the DM particles – however the *warm* DM theory is now disfavoured as the core would be so small that effectively it is always observed as a cusp (Macciò et al., 2012).

[d] See also the discussion on halo fitting in Sec. 3.2.2.

Many of the dynamical properties of a self-gravitating system are accessible if fit using crude models with spherical symmetry. For example, an inversion can be performed to find a self-consistent distribution function (Eddington, 1916, Osipkov, 1979, Merritt, 1985, Cuddeford, 1991). Force calculations are also vastly simplified by any simple analytical model.

Due to the linearity of the Poisson equation, models of more complex astronomical systems can be constructed by simply summing the associated densities and potentials; each component may be formed out of different species of matter, and may have very different approximate geometries. For example, a model for the Milky Way (Bovy et al., 2012) may consist of the sum of a Hernquist (1990) bulge, an NFW dark matter halo, and an exponential disk.

To bring some organising principle to the great profusion of spherical halo and bulge models, we may consider the extremely general *double-power law* family of models, which includes almost all the models mentioned above as sub-cases. These were first considered altogether by Zhao (1996), who writes

$$\rho(r) = \frac{C\, a^{\beta-3}}{r^{\gamma}\ \left(a^{1/\alpha} + r^{1/\alpha}\right)^{\alpha(\beta-\gamma)}}. \tag{1.18}$$

The three exponents are free parameters: the inner slope $\gamma$, the outer slope $\beta$, and a parameter $\alpha$ that affects the width of the transition region between the two power-law regimes (larger $\alpha$ corresponds to wider transition region). There are two remaining parameters: the scale-length $a$; and the normalisation constant $C$, which has units of mass and is proportional to the total mass if it is finite. The potential corresponding to this very general distribution is written by Zhao as the sum of two incomplete beta functions $\mathcal{B}_z(a,b)$[e],

$$\psi(r) = -4\alpha\pi G\, C\, \left[\frac{1}{r}\, \mathcal{B}_{\chi}(\alpha(3-\gamma), \alpha(\beta-3)) + \frac{1}{a}\, \mathcal{B}_{1-\chi}(\alpha(\beta-2), \alpha(2-\gamma))\right], \qquad \chi \equiv \frac{r^{1/\alpha}}{a^{1/\alpha} + r^{1/\alpha}}. \tag{1.19}$$

The simple properties of various subsets of this general parameterisation (e.g. the hypervirial models, the Dehnen (1993) family, etc.) arise because for certain combinations of $(\alpha, \beta, \gamma)$ the beta functions of Eq. (1.19) reduce to elementary functions.

The Einasto model mentioned above as a recent rival to the NFW model is not included in Zhao's double-power law family. It introduces an additional parameter $n$, which aids its flexibility in matching the behaviour of halo density profiles. The functional form of the density is

$$\rho(r) \propto \mathrm{e}^{-k\,(r/a)^{1/n}}, \tag{1.20}$$

with most haloes falling between $n = 5$ and $n = 10$. While the Einasto profile may appear to be far-removed from the traditional double power-law form, An & Zhao (2013) point out that it may be thought of as a particular limiting case of the general parameterisation of the logarithmic slope

$$\frac{\mathrm{d}\log\rho}{\mathrm{d}\log r} = -\frac{\gamma + \beta\,(r/a)^{1/\alpha}}{1 + (r/a)^{1/\alpha}}. \tag{1.21}$$

[e]This special function, along with the many others used in this thesis, is listed in Appendix A.

This integrates to give Eq. (1.18) when $\beta$ is finite, and integrates to the generalised Einasto form

$$\rho \propto r^{-\gamma}\, \mathrm{e}^{-k\,(r/a)^{1/\alpha}}, \qquad k \equiv \alpha\beta \tag{1.22}$$

upon taking a limiting procedure whereby both $\beta, a \to \infty$ while $\beta/a^{1/\alpha}$ is kept finite.

### 1.3.2 Models without spherical symmetry

As discussed above, a galaxy is often modelled as a sum of simple analytical components. However, in reality the picture is not so simple. Adding baryonic components to simulations, as well as distortions due to mergers and interactions between haloes, tends to produce a wider range of halo shapes. Haloes are at least triaxial (stratified on ellipsoids) in a radius-dependent way, and this is affected by the presence of baryonic matter in the simulation (Bryan et al., 2013). Powerful simulations and abundant observational data suggest more flexible models are necessary. Analogues of the haloes found in simulations can then be fit to the observations (e.g. Shao et al., 2020, for a recent example).

It is therefore desirable to find a way to represent non-spherical structure while still retaining some efficiency of computation. Additionally, it should be possible to use the large amount of observational data that is now available to constrain more detailed models of the Milky Way's density and potential (e.g. observations from the Gaia spacecraft Gaia Collaboration et al., 2016, Lindegren et al., 2016).

The simplest way to generalise a spherical model is to directly change the radial dependence of the density to be stratified on ellipsoidal contours; an algorithm for finding the corresponding potential is then given in Binney & Tremaine (1987, Ch. 2). However, for greatest flexibility we may choose to capture the angular detail using spherical harmonics, which are a complete basis for functions on the sphere (DLMF, §14.30). Spherical harmonics are defined by

$$Y_{lm}\,(\theta,\phi) \equiv P_l^{(m)}(\cos\theta)\;\mathrm{e}^{\mathrm{i}m\phi}, \tag{1.23}$$

where the $P_l^{(m)}(x)$ are associated Legendre functions which are the angular eigenfunctions of the Laplacian,

$$\nabla^2 Y_{lm} = \frac{l(l+1)}{r^2} Y_{lm}. \tag{1.24}$$

Additionally, they obey an orthogonality relation

$$\int Y_{lm} Y^*_{l'm'}\,\mathrm{d}\sin\theta\,\mathrm{d}\phi = J_{lm}\delta_{lm}^{l'm'}, \tag{1.25}$$

where $J_{lm}$ is the normalisation constant. In fact following standard practice we will always multiply the definition (1.23) by $1/\sqrt{J_{lm}}$, absorbing the normalisation constant without changing the notation, so that the integral (1.25) has a value of unity. We will also tend to use *real* spherical harmonics, whereby the harmonics with negative values of $m$ are proportional to $\sin(m\phi)$, and those with positive values of $m$ are proportional to $\cos(m\phi)$. This avoids the complex conjugation in Eq. (1.25).

Two useful related functions are the regular ($R_{lm}$) and irregular ($I_{lm}$) *solid harmonics*

$$R_{lm} \equiv r^l \, Y_{lm}, \qquad I_{lm} \equiv r^{-1-l} \, Y_{lm}, \tag{1.26}$$

which are the solutions to Poisson's equation in a vacuum.

In order to create models with more flexible angular dependence, one possibility is to simply write down a sum of terms, each multiplied by a different spherical harmonic. Each term's parameters are permitted to have an $l$ or $m$ dependence, and the whole model is then fit numerically with this increased number of free parameters. Let us write both the density and potential as a sum of terms,

$$\begin{aligned} \rho(\mathbf{r}) &= \sum_{lm} \rho_l(r) \, Y_{lm}(\theta, \phi), \\ \psi(\mathbf{r}) &= \sum_{lm} \Phi_l(r) \, Y_{lm}(\theta, \phi), \end{aligned} \tag{1.27}$$

where the '$lm$' sum is over $l \geq 0$, subject to $|m| \leq l$. Factoring out the spherical harmonics, the equation satisfied by each $\rho_l$ and $\Phi_l$ is

$$\left( \nabla_r^2 - \frac{l(l+1)}{r^2} \right) \Phi_l = 4\pi G \, \rho_l. \tag{1.28}$$

As derived in Sec. 1.4.3, the potential functions are asymptotically required to obey $\Phi_l \sim r^l$ as $r \to 0$, and $\Phi_l \sim r^{-1-l}$ as $r \to \infty$.

Zhao (1996) gives an expression analogous to Eq. (1.19) for the $l$-dependent potentials $\Phi_l$ for the entire double-power law family. While the particular choice of $l$-dependence in $\Phi_l$ is not unique, his form is the only possibility that retains the double-power law form of the associated $\rho_l$ while still obeying the boundary conditions in Eq. (1.50). The modified expression for the potential is

$$\Phi_l(r) \propto \frac{a^l}{r^{l+1}} \mathcal{B}_\chi(\alpha(3+l-\gamma), \alpha(\beta-l-3)) + \frac{r^l}{a^{l+1}} \mathcal{B}_{1-\chi}(\alpha(\beta+l-2), \alpha(2-l-\gamma)) \, . \tag{1.29}$$

The expression for the density $\rho_l$ remains the same as in Eq. (1.18), with the understanding that all the parameters $(C, a, \alpha, \beta, \gamma)$ may optionally now have any $l$ or $m$ dependence as determined by some fitting procedure[f]. Note that terms with $l > 0$ do not contribute to the total mass. Our expressions Eqs (1.19) and (1.29) differ slightly in parameterisation from those written down in Zhao (1996), with the form given here elucidating the connection to the basis sets presented in Chapters 3 and 5.

#### 1.3.2.1 Multipole expansion

We now discuss the construction of potential-density models with arbitrary morphology, whose angular dependence is expressed through spherical harmonics. In order to arrive at the correct

[f] That is, the potential Eq. (1.29) *guarantees* the correct $l$-dependence such that the boundary conditions of Eq. (1.50) will be satisfied automatically; but the parameters may have any additional $l$ and $m$ dependence, determined numerically.

expressions, we first note that solutions to Poisson's equation may be written as a convolution of a density $\rho$ with a Keplerian potential,

$$\psi(\mathbf{r}) = -G \int \mathrm{d}^3\mathbf{r}' \, \frac{\rho(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|}. \tag{1.30}$$

Next we note that there exists an expansion in solid harmonics for the Keplerian potential (DLMF, §18.12). Then, by applying the spherical harmonic addition theorem (DLMF, §14.30.9), we can write the potential as an integral over the radial part combined with a sum over the angular wavenumbers,

$$\psi(\mathbf{r}) = -\sum_{lm} \frac{4\pi G}{2l+1} \, Y_{lm}(\theta, \phi) \, \left[ r^{-l-1} \, f_{lm}(r) + r^l \, g_{lm}(r) \right]. \tag{1.31}$$

We have also defined the *interior* and *exterior multipoles*

$$\begin{aligned} f_{lm}(r) &\equiv \int_0^r \mathrm{d}r' \, r'^{l+2} \int_\Omega \mathrm{d}\sin\theta \, \mathrm{d}\phi \, Y_{lm}(\theta, \phi) \rho(r', \theta, \phi), \\ g_{lm}(r) &\equiv \int_r^\infty \mathrm{d}r' \, r'^{1-l} \int_\Omega \mathrm{d}\sin\theta \, \mathrm{d}\phi \, Y_{lm}(\theta, \phi) \rho(r', \theta, \phi). \end{aligned} \tag{1.32}$$

In numerical applications the angular sum must be truncated at some finite order $l_{\max}$. When the density $\rho$ consists of a cloud of points, as in Eq. (1.10), we can now evaluate the forces on all particles in $\mathcal{O}(l_{\max}^2 N \log N)$ steps by sorting the spherical shells by radius. The truncation of the angular series automatically smooths out the angular part of any relaxation effects arising from particle under-sampling. However, the system is still composed of interacting spherical shells, so relaxation effects can occur radially. This technique has been termed the *multipole expansion*, or MEX (see Meiron et al. (2014) for an overview of its numerical properties).

#### 1.3.2.2 Complete expansions

The MEX method expresses the angular dependence via a truncated series expansion, but must still sort the particles in order to construct the radial shells. A more natural method would be to use a separate series expansion to encompass the entire radial dependence, using a set of *basis functions*.

When used in the context of $N$-body simulations or to replay time-dependent gravitational potentials, such an approach is called a *halo expansion* (HEX) or *self-consistent field* (SCF) method (see e.g. Hernquist et al., 1995, Lowing et al., 2011, Meiron et al., 2014).

It is natural to retain the spherical harmonics as the angular part of the expansion, so there are now three indices of summation, with the additional radial summation index typically denoted $n$ by analogy with the radial quantum number. Alternatively one may consider expansions suitable for disk-like systems, reducing the problem to two dimensions, and use a Fourier basis for the angular component. We also note that any method that expresses the angular part of the solution in terms of spherical harmonics (MEX or SCF) can take advantage of symmetries in the physical system under consideration, in order to reduce computational load.

There is a large literature proposing possible sets of basis functions, as different methods may be appropriate in different circumstances. Such radial functions must be *complete*, preferably *biorthogonal*[g], and perhaps even *analytical*. *Completeness* refers to the existence of a sequence of functions that converges to any desired physically plausible model; *biorthogonality* means that the pairwise self-energy between each potential and density functions must form a diagonal matrix; *analytical* means that all basis functions can be expressed using a finite number of elementary arithmetical operations, possibly augmented by a number of special functions[h].

We will give a short introduction to those basis sets that satisfy all three of the conditions listed above in Sec. 1.4. An in-depth review of the prior literature may be found in Sec. 2.1.

Expansions that drop the biorthogonality requirement include those based on a set of complete functions combined with a weight function that is determined by the choice of zeroth-order model (see e.g. Saha, 1991, 1993, Robijn & Earn, 1996). Alternatively one may use a *local* basis (commonly splines, such as in the approach of Vasiliev (2013)). Any approach that drops the orthogonality requirement must orthogonalise the basis functions by performing an LU-decomposition on the *overlap* matrix[i]

$$D_{ij} \equiv \langle \Phi_i, \Phi_j \rangle \propto \int \mathrm{d}^3\mathbf{r}\ \Phi_i\, \nabla^2 \Phi_j, \tag{1.33}$$

or else multiply the vector of expansion coefficients by the inverse of this matrix. Decomposing the overlap matrix is always possible as it is positive definite (as demonstrated by Eq. (1.39)). The decomposition is equivalent to performing a Gram-Schmidt orthogonalisation on the basis functions. Depending on the form of the basis functions, computing $D_{ij}$ may require a large number of numerical integrals of uncertain numerical stability.

Alternatively, one can use the structure of the Poisson equation more directly to take a purely numerical approach to constructing basis sets. This is the approach of Brown & Papaloizou (1998) (using the integral formulation of the Poisson equation Eq. (1.30)) and of Weinberg (1999) (using the Sturm-Liouville formulation of the Poisson equation, which we will discuss in Ch. 2.2). In both cases, total flexibility of zeroth-order density is achieved at the cost of significant up-front computational costs. Kalapotharakos et al. (2008) criticises approaches that require the basis functions to be numerically tabulated, on the grounds that small errors in the potential are greatly magnified upon finite-differencing the potential to obtain the acceleration field.

## 1.4 Basic properties of biorthogonal basis sets

In the remainder of this thesis we focus on basis sets that are simultaneously complete, biorthogonal and analytical. We now lay out the basic theory behind such expansions.

In this approach, small corrections to a simple underlying zeroth-order model are efficiently captured by higher-order terms in the series (analogous to an expansion in orthogonal polynomials

[g]The terms *biorthogonal*, *biorthonormal*, *orthogonal* and *orthonormal* are used somewhat interchangeably in the literature.

[h]Here we are using the colloquial sense of the term *analytical*, meaning essentially *closed-form*; distinct from the term *analytic* from pure mathematics.

[i]Called the *stiffness matrix* in spline-based methods.

or Fourier modes), truncated at some radial order $n_{\rm max}$ and angular order $l_{\rm max}$,

$$\psi(\mathbf{r}) = \sum_{n=0}^{n_{\rm max}} \sum_{l=0}^{l_{\rm max}} \sum_{m=-l}^{l} C_{nlm}\, \Phi_{nlm}(\mathbf{r}), \tag{1.34}$$
$$\varrho(\mathbf{r}) = \sum_{n=0}^{n_{\rm max}} \sum_{l=0}^{l_{\rm max}} \sum_{m=-l}^{l} C_{nlm}\, \rho_{nlm}(\mathbf{r}),$$

We will often condense the notation by simply writing '$nlm$' under the summation sign. This improvement in flexibility allows axisymmetric, triaxial, lop-sided or distorted density distributions to be built up from the underlying zeroth-order model.

When the forces on the particles are calculated from the truncated series, we automatically avoid some relaxation effects due to under-sampling the number of particles in the system. Because the coefficients are only calculated once, and the series of functions is evaluated once per particle to find the force, such methods are $\mathcal{O}(N)$ in the number of particles, and are $\mathcal{O}(n_{\rm max} l_{\rm max}^2 N)$ overall.

An important limitation, however, is the relative paucity of analytical basis sets described in the literature. Such basis sets are particularly useful to us if the zeroth-order density-potential pair is sufficiently close to a realistic spherically-symmetric density profile[j].

The potential and density basis functions are factorised with respect to the coordinates $(r, \theta, \phi)$,

$$\Phi_{nlm}(\mathbf{r}) \equiv \Phi_{nl}(r)\, Y_{lm}(\theta, \phi), \tag{1.35}$$
$$\rho_{nlm}(\mathbf{r}) \equiv K_{nl}\, \rho_{nl}(r)\, Y_{lm}(\theta, \phi),$$

These functions must obey Poisson's equation,

$$\nabla^2 \Phi_{nlm} = 4\pi G \rho_{nlm} \qquad \Leftrightarrow \qquad \left(\nabla_r^2 - \frac{l(l+1)}{r^2}\right) \Phi_{nl} = 4\pi G\, K_{nl}\, \rho_{nl}, \tag{1.36}$$

and we augment the spherical harmonics' orthogonality relation (Eq. 1.25) by a *bi-orthogonality* relation that is satisfied by the radial part of each potential-density pair, so that the full set of basis functions obeys

$$\int \mathrm{d}^3\mathbf{r}\ \Phi_{nlm}\, \rho_{n'l'm'} = \int r^2\, \mathrm{d}r\, \Phi_{nl}\, \rho_{n'l'} \int Y_{lm} Y_{l'm'}\, \mathrm{d}\Omega = K_{nl} N_{nl} \delta_{nlm}^{n'l'm'}, \tag{1.37}$$

Note the following: 1) the orthogonality in the index $l$ in Eq. (1.37) is set by the integral over the spherical harmonics $Y_{lm}$, so the biorthogonality of the radial functions need only be with respect to the $n$ index; 2) the $K_{nl}$ are some arbitrary constants that are included purely to simplify the expression for the radial density functions $\rho_{nl}$; 3) the boundary conditions derived in Sec. 1.4.3 also apply to each individual radial potential function $\Phi_{nl}$.

The biorthogonality property means that the coefficients $C_{nlm}$ can be calculated by an integral over either $\rho$ or $\Phi$, which we will derive in Sec. 1.4.1. In the case that the density is formed from a cloud of point particles with masses $m_i$ and positions $\mathbf{r_i}$, the integrals reduce to a sum

$$C_{nlm} = \sum_i m_i\, \Phi_{nlm}(\mathbf{r_i}). \tag{1.38}$$

[j]There is also a theory of basis functions appropriate for disk geometries, paralleling that for spherical geometries; this is briefly discussed in Sec. 2.1.1.

For the basis sets presented in this thesis (and indeed all those previously reported in the literature), we do not prove completeness; but we do assume that expansions derived from Sturm-Liouville equations inherit this property, such as those investigated in Sec. 2.2. Note that the basis functions of our generalised family of Ch. 5 are linear combinations of the non-orthogonal basis functions $(\tilde{\Phi}_{nl}, \tilde{\rho}_{nl})$ described in Sec. 5.1 (which are polynomials of degree $n$) and so we assume completeness in that case[k].

### 1.4.1 Expression for the coefficients

We now derive the expression (1.38) for the coefficients of the biorthogonal expansion. For the purposes of this section we write $i$ to stand for the multi-index $nlm$, and we set $G = 1$. Assume that we possess a set of normalised potential basis functions $\Phi_i$ that obey an orthogonality relation with respect to the following inner product,

$$\langle \Phi_i, \Phi_j \rangle = \int \mathrm{d}^3\mathbf{r}\, \boldsymbol{\nabla}\Phi_i \cdot \boldsymbol{\nabla}\Phi_j = \delta_{ij}. \tag{1.39}$$

We also assume that these functions are *complete*, meaning there exists a set of coefficients $C_i$ such that an arbitrary potential $\Phi$ can be represented as

$$\Phi = \sum_i C_i \Phi_i. \tag{1.40}$$

Furthermore, for each potential basis function, the corresponding density basis function is *defined* as

$$\rho_i = \frac{-1}{4\pi} \nabla^2 \Phi_i. \tag{1.41}$$

We regard a sample of $N$ particles as a single realisation of the true potential $\Phi$, and denote it $\hat{\Phi}$. The density corresponding to $\hat{\Phi}$ is a cloud of Dirac delta functions $\hat{\rho}$,

$$\hat{\rho}(\mathbf{r}) = \sum_{j=1}^{N} m_j\, \delta(\mathbf{r} - \mathbf{r}_j). \tag{1.42}$$

We wish to use the single realisation $\hat{\Phi}$ to estimate the $C_i$, such that the following $\chi^2$ error is minimised,

$$\chi^2\left(\{C_i\}\right) = \frac{1}{2}\left\|\Phi - \hat{\Phi}\right\|^2 = \left\langle \Phi - \hat{\Phi}, \Phi - \hat{\Phi} \right\rangle. \tag{1.43}$$

To proceed, we differentiate $\chi^2$ with respect to one coefficient $C_j$, and set the result to zero. We have

$$\frac{\partial \chi^2}{\partial C_j} = \left\langle \Phi_j, \sum_i C_i \Phi_i - \hat{\Phi} \right\rangle = 0, \tag{1.44}$$

and applying (1.39) and rearranging gives

$$C_i = \langle \Phi_i, \hat{\Phi} \rangle. \tag{1.45}$$

[k] That family of basis sets is also continuously related to the Zhao (1996) family by the parameter $\nu$, which is suggestive that the completeness property carries over.

Applying Green's first identity and the definition of $\hat{\rho}$ then gives

$$C_i = (\Phi_i, \hat{\rho}) = \sum_{j=1}^{N} m_j \, \Phi_i(\mathbf{r}_j) \, . \tag{1.46}$$

### 1.4.2 Why biorthogonality?

One may ask why it is necessary to have such an unusual orthogonality relation, that mixes the potential and density functions. Or equivalently, if one regards the density basis functions as being defined by a choice of potential functions, one may ask why the need for the non-standard inner product (1.39), rather than simply applying a standard weighted $L^2$-norm to either the potential or density basis functions. Such an alternative scheme might work as follows. We define

$$\begin{aligned} \rho(\mathbf{r}) &= \sum A_i \varrho_i(\mathbf{r}), \\ \Phi(\mathbf{r}) &= \sum A_i \Psi_i(\mathbf{r}), \end{aligned} \tag{1.47}$$

where $\nabla^2 \Psi_i \propto \varrho_i$, and the $\varrho_i$ are normalised functions obeying an orthogonality relation with respect to the standard $L^2$ inner product,

$$(\varrho_i, \varrho_j) = \int \mathrm{d}^3\mathbf{r} \;\; \varrho_i(\mathbf{r}) \; \varrho_j(\mathbf{r}) = \delta_{ij}. \tag{1.48}$$

Unfortunately, using the $L^2$ inner product $(f, g)$ imposes undesirable restrictions on the density – we lose the ability to represent models with infinite mass or steep central cusps. This may be remedied by inserting a weight function $\Omega(\mathbf{r})$ into Eq. (1.48), turning it into an inner product for the weighted space $L^2_\Omega$. However, this would limit the possible functional forms of the basis functions and would not even guarantee that the density remain representable after dynamical evolution, because the inner or outer slope may change over time. Such a change in outer slope is observed in simulations of dark haloes: see e.g. Gao et al. (2008) and Dutton & Macciò (2014), who show that the best-fit Einasto model steepens with redshift; this also occurs with the halo we study in Ch. 7.

Using the inner product $\langle f, g \rangle$ (Eq. 1.39) instead, corresponding to the Sobolev space $H^1$ excluding the constant functions, guarantees that the acceleration field $\boldsymbol{\nabla}\Phi$ itself is in the $L^2$ space. This is a more desirable property from the point of view of convergence, as $\boldsymbol{\nabla}\Phi$ is the dynamical quantity of interest.

Another reason is more practical: any heuristic method of searching for analytical basis sets according to the alternative scheme proposed above would necessarily involve specifying the density basis functions such that they satisfy the orthogonality relation (1.48); the potential basis functions would then have to be calculated backwards via integration, which is in general a non-trivial operation. By contrast, our standard scheme lends itself to finding basis sets where both the potential and the density are conveniently expressible, as explored systematically in Ch. 2.

However, this does not preclude some usefulness for basis sets corresponding to Eq. (1.48). Natural choices for density functions may involve Zernike, Laguerre or Hermite polynomials

combined with an appropriate weighting factor. However, such an approach seems to be absent from the literature, suggesting that it may not have been found fruitful.

We have given some reasons for the desirability of the *bi* part of *biorthogonality*. But what is the justification for requiring *orthogonality*? After all, as mentioned in Sec. 1.3.2.2, any non-orthogonal but complete set of functions may be used, requiring only that the overlap matrix $D_{ij}$ (Eq. 1.33) be LU-decomposed or inverted. Apart from the additional time required by this computation, which is arguably negligible as all matrix operations can be carried out in advance, there are two downsides. 1) There is no guarantee that the numerical integration of the overlap matrix (followed by its decomposition) is numerically stable. Computing oscillatory integrals using floating-point arithmetic, even when the result is known in closed-form, is fraught with difficulties (see e.g. Appendix D, where we resorted to arbitrary-precision arithmetic in order to accurately compute the highly oscillatory indefinite integral between two basis functions). 2) When computing an expansion with respect to a basis set that is both orthogonal and complete, Parseval's theorem guarantees that the coefficients in the expansion decay to zero; even when the basis set is not complete (for example, perhaps we are trying to represent a cusped halo with an un-cusped basis set), Bessel's inequality guarantees that the expansion coefficients will remain bounded by the total self-energy of the target potential-density. When the basis set is non-orthogonal the expansion coefficients can grow without bound, which may cause numerical issues.

### 1.4.3 Physical boundary conditions

As a brief but crucial aside, we note some requirements on boundary conditions, that will apply to most of the derivations in the remainder of the thesis[1].

Given any multipole method for representing the gravitational potential, with radial components $\Phi_l(r)$, the asymptotic $(l, m)$-dependence of these functions as $r \to 0$ and $r \to \infty$ is necessarily determined. For a given angular order $l$, define the mass enclosed by a shell with inner radius $a$ and outer radius $b$ as

$$M_l(a,b) = 4\pi \int_a^b \mathrm{d}r\, r^2\, \rho_l = \int_a^b \mathrm{d}r \left[\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \frac{\mathrm{d}}{\mathrm{d}r}\Phi_l\right) - l(l+1)\Phi_l\right]. \tag{1.49}$$

In order for the density and potential to describe a physically plausible system, we impose that the mass enclosed by a thin shell goes to zero at the origin and at infinity, and that the potential remain finite everywhere. That is, we require that

$$\lim_{a\to 0} M_l(0,a) = 0, \qquad \lim_{a\to\infty} M_l(a,\infty) = 0, \qquad \Phi_l(r) < \infty. \tag{1.50}$$

These translate into equations that must be satisfied by the derivative of the integrand of (1.49), which is proportional to $\rho_l$. So, examining (1.49) we see that the asymptotic solutions for the potential are $\Phi_l \sim r^l$ as $r \to 0$, and $\Phi_l \sim r^{-1-l}$ as $r \to \infty$. One can also read these conditions off from the expression for the multipole expansion in Eq. (1.31).

[1] Mention of these boundary conditions is made in previous derivations of basis sets (Clutton-Brock, 1973, Hernquist & Ostriker, 1992).

## 1.5 Further applications of basis sets

In addition to the SCF method of $N$-body simulation discussed in Sec. 1.2, we give a brief overview of some other applications of biorthogonal basis sets to problems in galactic dynamics.

### 1.5.1 Representing time-evolving potentials

The potential value of basis function techniques in representing time-evolving potentials was noted in the early work of Hernquist et al. (1995), and they have recently enjoyed a resurgence in this application. Lowing et al. (2011) also pointed out how the time evolution of computationally expensive simulations of galaxy formation could be represented cheaply: snapshots of the original simulation are decomposed with basis functions, saving on storage costs. The entire simulation can then be cheaply replayed many times at will. This gives realistic approximations to the build-up of a halo as a function of redshift. For example, new objects can be inserted into the simulations and their behaviour studied as if they had been present originally, assuming their gravitational influence on the original simulation would have been negligible. This technique has already been exploited in studies of the evolution of tidal streams in the Milky Way (Ngan et al., 2015, 2016). However, the gamut of applications is far broader, including the evolution of accreting subhaloes, and the dynamics of satellite galaxies and globular clusters in dark haloes.

Both Lowing et al. (2011) and Ngan et al. (2015) decomposed numerical halo simulations using the Hernquist & Ostriker (1992) basis function expansion, as the options in the literature were heretofore very limited.

If, as is standard, the same choice of basis set is made for the entire evolution, then the expansion holds an important property: the density expansion is linear in the parameters, so the potential and forces are likewise linear. At any time-step, a standard time-interpolation scheme uses two or more consecutive points: two points in the case of linear interpolation, more for higher-order polynomial interpolants. But because of the linearity property, we can time-interpolate the expansion coefficients in advance, thereby saving on half or more of the potential evaluations.

For example, suppose we have some potential $\Phi$ that is known at times $t_0$ and $t_1$, and we wish to integrate the trajectory of a test particle (with arbitrary initial conditions) between those times, interpolating the gravitational forces linearly in time. The time-dependent potential expansion depends indirectly on time, by summing over the time-dependent coefficients $\{C_i(t)\}$,

$$\Phi(\mathbf{r},t) \equiv \Phi(\mathbf{r};\{C_i(t)\}) = \sum_i C_i(t)\,\Phi_i(\mathbf{r}). \tag{1.51}$$

A naïve interpolation scheme would evaluate the forces twice at each time-step,

$$\mathbf{F}_{\text{interp}} = -\boldsymbol{\nabla}\Phi(\mathbf{r};\{C_i(t_0)\})\left(1-\frac{t-t_0}{t_1-t_0}\right) - \boldsymbol{\nabla}\Phi(\mathbf{r};\{C_i(t_1)\})\left(1-\frac{t_1-t}{t_1-t_0}\right). \tag{1.52}$$

However, because the time interpolation is linear in the coefficients, we can instead get away with a single force evaluation per time step,

$$\mathbf{F}_{\text{interp}} = -\boldsymbol{\nabla}\Phi\left(\mathbf{r};\left\{C_i(t_0)\left(1-\frac{t-t_0}{t_1-t_0}\right) + C_i(t_1)\left(1-\frac{t_1-t}{t_1-t_0}\right)\right\}\right). \tag{1.53}$$

The same argument applies to interpolation involving higher-order polynomials, but with even more computational savings: the interpolant depends on higher powers of $t$ and additional consecutive coefficients $\{C_i(t_j)\}$, but is still linear in the coefficients themselves.

We explore this application of basis expansions in detail in Sec. 7.

### 1.5.2 Secular perturbation theory

An alternative route into the subject arises when studying the stability of collisionless stellar systems. Given a distribution of matter and a self-consistent model for its dynamics (together solving the Vlasov-Poisson equations), a natural question is how the system responds to a small perturbation. If a realistic perturbation grows without bound, then this suggests that the model is not physically justified. Classically a number of models were considered for their stability properties, in particular thin circular or elliptical discs. Certain families of circular disc models were studied by Hunter (1963) and Kalnajs (1972), as well as a family of elliptical discs by Tremaine (1976), giving a profusion of ingenious analytical methods for studying a number of different kinds of perturbation.

One computational approach that permits consideration of arbitrary perturbations is a decomposition into basis functions; and it was for this purpose that Clutton-Brock (1972) derived the first set of disc-like orthogonal basis functions. Having performed the decomposition, the normal modes of the system then correspond to the eigenvectors of a certain matrix equation, leading to this technique being labelled the *matrix method* (e.g. Polyachenko & Shukhman, 1981, Fridman & Polyachenko, 1984, Weinberg, 1989, Saha, 1991, Palmer, 1994, Evans & Read, 1998).

The growing awareness that galaxies are embedded in approximately-spherical DM haloes led to a need for basis functions more suited to the spherical geometry. Stability studies of spherical systems often used spherical Bessel functions as the radial basis functions, alongside spherical harmonics as the angular basis functions. This is a natural choice, as together they form the eigenfunctions of the Laplacian in spherical coordinates. However, spherical Bessel functions form a discrete basis only over a finite radial range, so the underlying galaxy or halo model must have finite extent.

Subsequently, basis function techniques were used to provide algorithms to evolve collisionless stellar systems, partly as a check on the results of linear stability theory. For example, Allen et al. (1990) used a biorthogonal spherical Bessel basis function expansion to study the radial and circular orbit instability in spherical galaxy models. As instabilities in stellar systems often arise from nearly-resonant orbits, accurate modelling of the precession of individual orbits for several orbital periods is very important.

A more recent development is that of Fouvry (2016), who suggested that the collisionless (Vlasov-Poisson) equations ought to be superseded by an alternative set of equations (termed *Fokker-Planck*) that allows for perturbations to be sourced by collective excitations of the entire system. These perturbations can then be studied by a matrix method analogous to that described above.

## 1.6 Outline of thesis

We have motivated the need for biorthogonal basis expansions, and described their application. We now briefly summarise the structure of this thesis.

Ch. 2 reviews the analytical biorthogonal basis sets extant in the literature, and systematically investigates each proposed method for deriving them. In particular we attempt to produce basis sets from all the classical orthogonal polynomials (uncovering one previously-unknown basis set), as well as from a number of different integral transforms.

In Ch. 3 we propose a new simple spherical model for haloes and bulges, intermediate in properties between Hernquist (1990) and NFW, dubbing it the *super-NFW* (sNFW) model. We derive analytical forms for the projected light curve, circular velocity and both isotropic and anisotropic distribution functions.

Ch. 4 presents the first major result on biorthogonal basis functions since that derived in Zhao (1996). It is an extension of the result of Rahmati & Jalali (2009) to a one-parameter family of basis functions, along with a more elegant and computationally-efficient way of expressing both the potential and density functions. One particular member of this new family corresponds to the sNFW model discussed in Ch. 3.

In Ch. 5 we derive a far-reaching generalisation of the results of the previous chapter. Similarities between the basis sets of Ch. 4 and Zhao (1996) are suggestive of a larger family of basis sets with an additional free parameter, and we find this generalised family by extending the original generating function method. This two-parameter family provides basis sets corresponding to almost every well-known double-power law model, including the NFW model and many others.

Ch. 6 comprises some further results on basis sets, inspired by the preceding two chapters. We derive an exceptional one-parameter family of basis sets whose zeroth-orders are cusped-exponential models. We claim that one member of this family in particular, whose zeroth-order density is a Gaussian, deserves further attention, and we discuss applications to the construction of galaxy models from deprojected data.

In Ch. 7 we turn to the practical application of the results of Ch. 5. We investigate the ability of basis function expansions to reproduce the time evolution of a Milky Way-like dark matter halo, extracted from a cosmological zoom-in simulation. For each snapshot, the density of the halo is reduced to a basis function expansion, with interpolation used to recreate the evolution between snapshots. The radial variation may be represented either by biorthonormal basis functions or by splines, and we compare the performance of the basis function method to a quintic spline representation (Vasiliev, 2013, 2019). Naïve calculation of the coefficients of the basis expansion falls prey to an analogue of the Gibbs phenomenon caused by the finite truncation radius of the simulation, but this can be circumvented by adding a linear correction to the expansion coefficients. The motion of the halo centre is affected by the structure on large scales in the original cosmological simulation. Because we centre the coordinate system on the halo, we must take into account an additional spatially-uniform but time-dependent force due to the non-inertial nature of the reference frame, and we show how to calculate this correction. We demonstrate that high fidelity orbit reconstructions are attainable using either method, by comparisons with the trajectories of particles

in the original simulation. We quantify how the error in the reconstructed orbits varies with the order of the expansion, snapshot spacing and number of particles.

Ch. 8 summarises the results of the thesis and discusses possible new directions for research, both within astronomy and in other parts of physics.

# Chapter 2

# A systematic investigation of basis expansions

We begin this chapter with a review of the existing mathematical techniques for obtaining *analytical* orthogonal basis sets – meaning those for which all basis functions can be written using a finite number of elementary or well-known special functions, and for which no additional orthogonalisation is required. There are two major categories, requiring somewhat differing approaches: basis sets suitable for describing razor-thin disks, expressed in cylindrical polar coordinates; and basis sets appropriate for haloes and bulges, expressed in spherical polar coordinates. While the major new results in this thesis (Ch. 4–Ch. 6) all fall into the latter category, there is some overlap between the analytical techniques used, and historically the majority of effort has been spent on the former type. We therefore review both groups of expansions.

In Sec. 2.2 we then focus on one of two main avenues of investigation – the direct method of finding (what we coin) *Clutton-Brock expansions*. This method involves explicitly mapping Poisson's equation to some given Sturm-Liouville system with known eigenfunctions. The existing basis sets of this type have typically been presented in the literature as being the results of informed guesswork, but we attempt the first systematic enumeration of the possible basis sets of this form. We uncover one promising new lead (the 'Laguerre/NFW' basis set of Sec. 2.2.1.1), but ultimately suggest that the possibilities have been largely exhausted by the prior results in the literature.

Then in Sec. 2.3 we turn to the second type of approach – the use of integral transforms. Sec. 2.3.1 describes the *Hankel transform* method, which will be the key tool by which we later derive the main results of the thesis. We also briefly review other integral transform approaches.

Finally, in Sec. 2.4, we mention a tool arising from classical potential theory, the *Kelvin transform*, that can be used to generate new basis sets from old.

As this chapter deals with primarily mathematical results, we set Newton's constant to $G = 1$, and set the scale-length of each potential-density model to $r_s = 1$.

## 2.1 Prior results

### 2.1.1 Disk geometry

In the mid-20th century, galactic dynamics as a discipline focussed on the study of highly-flattened systems, often with cylindrical symmetry; and it is with these systems that the study of biorthogonal basis sets first progressed significantly, with the pioneering paper of Clutton-Brock (1972).

Before directly dealing with basis sets, we detour with a more general discussion of Poisson's equation in the infinitesimally thin disk geometry.

While investigating ways to recover the matter distribution of a disk galaxy from its rotation curve, Toomre (1963) described *Toomre's device*[a]. This is a hybrid solution to Poisson's equation that gives a model of a razor-thin disk as a product of the eigenfunction solution in the plane $(R, \phi)$ with the fundamental solution in the $z$ coordinate,

$$\begin{aligned} \psi_{km}(\mathbf{r}) &= J_m(kR)\ \mathrm{e}^{\pm \mathrm{i}m\phi}\ \mathrm{e}^{-k|z|}, \\ \sigma_{lm}(\mathbf{r}) &= J_m(kR)\ \mathrm{e}^{\pm \mathrm{i}m\phi}\ \delta(z), \\ \nabla^2 \Phi_{km} &= -2k\, \sigma_{lm}. \end{aligned} \tag{2.1}$$

One immediate attraction of this expression is that both sides can be multiplied by an arbitrary function $g(k)$, and the integration performed over $k$. Because integration in $k$ commutes with derivatives in $R$, the result remains a valid potential-density pair. Such integrals involving Bessel functions are called Hankel transforms (see Sec. 2.3.1 for a brief overview).

Taking the simplest case of cylindrical symmetry ($m = 0$) and looking only within the disk ($z = 0$), Toomre (1963) used this property to derive a sequence of simple double-power law models. The lowest-order member of the family (also studied by Kuzmin (1956)) is found by positing for the gravitational potential

$$\psi(R) \propto \frac{1}{R}\left(1 + \frac{R^2}{a^2}\right)^{-1/2}. \tag{2.2}$$

Performing an inverse Hankel transform then gives

$$g(k) = a\, I_{\frac{1}{2}}(ka/2)\, K_{\frac{1}{2}}(ka/2)\,, \tag{2.3}$$

and the surface density $\sigma$ follows from multiplying Eq. (2.1) by $g(k)$ and integrating with respect to $k$, giving

$$\sigma(R) \propto \left(a^2 + R^2\right)^{-3/2}. \tag{2.4}$$

Higher-order models of the family are produced by differentiation with respect to the scale-length $a$ (which also commutes with both $R$-derivation and $k$-integration).

This line of thought encourages us to find models for the potential or rotation curve whose analytical form is amenable to two sequential Hankel transforms. A similar integration for higher harmonics $m$ furnishes models that describe departures from cylindrical symmetry.

[a] This name being coined in Clutton-Brock (1972).

Clutton-Brock (1972) wondered if a sequence of models obtained by Toomre's device might be constructed so as to satisfy an orthogonality property. He noted that the Bessel functions used to construct the eigenfunctions of the Laplacian in cylindrical coordinates are orthogonal on the interval $(0, \infty)$ (by the Fourier-Bessel theorem), with a continuous eigenvalue $k$. This orthogonality property can then be transferred from real space to $k$-space if the auxiliary functions $g(k)$ are chosen appropriately. More precisely, suppose we possess a set of functions $g_{nm}(k)$ with the property that

$$\int_0^\infty \mathrm{d}k\, g_{nm}(k)\, g_{n'm}(k) = \delta_{nn'} \tag{2.5}$$

for each $m$. Then, we define the *potential basis functions* $\psi_{nm}$ and *density basis functions* $\sigma_{nm}$ as follows,

$$\psi_{nm}(R,\phi,z) = \frac{-1}{2\pi}\, \mathrm{e}^{\mathrm{i}m\phi} \int \mathrm{d}k\, J_m(kR)\; g_{nm}(k)\, \mathrm{e}^{-k|z|}, \tag{2.6}$$

$$\sigma_{nm}(R,\phi,z) = \mathrm{e}^{\mathrm{i}m\phi} \int \mathrm{d}k\, k\, J_m(kR)\, g_{nm}(k)\, \delta(z).$$

By Eq. (2.1) these functions satisfy Poisson's equation; and via the Fourier-Bessel theorem the orthogonality of the $g_{nm}(k)$ transfers to an orthogonality property in real space (noting that the integral becomes zero away from the disk),

$$\int_0^R R\,\mathrm{d}R \int_0^{2\pi} \mathrm{d}\phi\, \psi_{nm}\, \sigma_{nm} = \delta_{nn'}\, \delta_{mm'}. \tag{2.7}$$

This permits us to use these functions as a complete basis for any matter density in the plane. Clutton-Brock (1972) dubs such a set of pairs of functions *biorthogonal*[b]. The question now is how to find suitable sets of functions $g_{nm}(k)$ such that their Hankel transforms (2.6) are tractable, and such that their lowest order member describes a suitable realistic galaxy model. Clutton-Brock (1972) made a natural choice,

$$g_{nm}(k) = k^m\, L_n^{(2m)}\,(2k)\; \mathrm{e}^{-k}, \tag{2.8}$$

where $L_n^{(\alpha)}(x)$ is a Laguerre polynomial (DLMF, §18.3). The product of two such functions immediately gives us the standard form for the Laguerre polynomial orthogonality relation. The Hankel transform turns out to be simple, and the resulting basis set exactly matches Toomre's family of potential-density pairs at zeroth order. The functions also all satisfy the boundary conditions described in Sec. 1.4.3, noting that (by convention) the $m$ index in cylindrical coordinates takes the role of the $l$ index in spherical polar coordinates. Clutton-Brock did not in fact explicitly evaluate the integrals (2.6) for every $n$ and $m$, but exploited a generating function for the Laguerre polynomials to express the basis functions indirectly via recurrence relations. The derivation was completed by Aoki & Iye (1978), who observed that the Hankel-transformed Laguerre generating function is exactly the generating function for the Gegenbauer polynomials, enabling the result to be given in closed-form. Their calculation is essentially the same as the one we use in Ch. 4 to derive a similar

[b] See the note in Sec. 1.3.2.2 regarding terminology.

basis set in spherical geometry. Aoki & Iye (1978) also point out that Clutton-Brock's basis set is precisely what would result if one were to *explicitly* carry out a Gram-Schmidt orthogonalisation of the family of potential-density pairs of Toomre (1963). This close link confirms that the existence of convenient basis sets is heavily constrained by the analytical properties of the Hankel transform[c].

Kalnajs (1976) approaches the problem using a different integral transform: first, the radial coordinate is rescaled logarithmically; then a Fourier transform is performed with respect to this quantity, and its conjugate Fourier-space variable is denoted $\alpha$. This method is therefore just a disguised form of the Mellin transform. The fundamental surface density appearing in the integral kernel is

$$K(R,\phi,z) = R^{-3/2}\,\mathrm{e}^{\mathrm{i}m\phi+\mathrm{i}\alpha\log R}\,\delta(z), \tag{2.9}$$

which describes logarithmic spirals in the plane. Kalnajs claims that a sum of spirals of this form is a natural way of expressing the surface density of a spiral galaxy (although in general almost any functional form can result from Mellin-transforming an appropriate auxiliary function, integrating over a continuous distribution of such spirals). Given a function $A_m(\alpha)$, a valid potential-density pair can be expressed using a Fourier series inversion (DLMF, §1.8) combined with a Mellin transform inversion (DLMF, §2.5). The density is that of a razor-thin disk,

$$\sigma(R,\phi,z) = \frac{R^{-3/2}}{4\pi^2}\sum_m \int_{-\infty}^{\infty} \mathrm{d}\alpha\, A_m(\alpha)\,\mathrm{e}^{\mathrm{i}m\phi+\mathrm{i}\alpha\log R}, \tag{2.10}$$

and the potential *on the disk* is

$$\psi(R,\phi,0) = -4\pi\,\frac{R^{-1/2}}{4\pi^2}\sum_m \int_{-\infty}^{\infty} \mathrm{d}\alpha\, K_{|m|}(\alpha)\,A_m(\alpha)\,\mathrm{e}^{\mathrm{i}m\phi+\mathrm{i}\alpha\log R}. \tag{2.11}$$

The function $K_m(\alpha)$ takes a moderately complex analytical form,

$$K_m(\alpha) = \frac{1}{2}\left|\frac{\Gamma\left(\frac{m+1/2+\mathrm{i}\alpha}{2}\right)}{\Gamma\left(\frac{m+3/2+\mathrm{i}\alpha}{2}\right)}\right|^2. \tag{2.12}$$

The full expression for the off-disk potential is so complicated that it is unused and unmentioned in the literature,

$$\begin{aligned}\psi(R,\phi,z) = \frac{-G}{2\pi^{3/2}}\sum_m \int_{-\infty}^{\infty} \mathrm{d}\alpha\, 2^m \left|\Gamma\left(\frac{|m|+\mathrm{i}\alpha+1/2}{2}\right)\right|^2 (R^2+z^2)^{\frac{2\mathrm{i}\alpha-1}{4}} \\ \times\, P^{(-m)}_{-\frac{1}{2}-\mathrm{i}\alpha}\left(\frac{|z|}{\sqrt{R^2+z^2}}\right) A_m(\alpha)\,\mathrm{e}^{\mathrm{i}m\phi+\mathrm{i}\alpha\log R},\end{aligned} \tag{2.13}$$

where $P^{(\mu)}_\nu(z)$ is an associated Legendre function.

Because the self-energy of any potential-density pair is preserved under the transforms (2.10)–(2.11), a family of orthogonal functions $A_{nm}(\alpha)$ may be chosen so as to produce correspondingly biorthogonal potential-density pairs,

$$\int_0^\infty r\,\mathrm{d}r\,\sigma_{nm}(r)\,\psi_{n'm}(r) = \int_{-\infty}^{\infty}\mathrm{d}\alpha\,A_{nm}(\alpha)\,K_m(\alpha)\,A_{n'm}(-\alpha) = \delta_{nn'}. \tag{2.14}$$

[c] Interestingly, the orthogonalisation of an initially *non-orthogonal* basis set also forms a key part of the method of Ch. 5, with the result eventually being expressible in closed-form.

By expressing the potential as a convolution between the inverse Mellin transforms of $K_m(\alpha)$ and $A_m(\alpha)$, Kalnajs writes (2.11) in the form of an Abel transform; then, by considering two similar ways to split up the factor $K_m(\alpha)$ between the potential and density, he re-derives the basis set of Clutton-Brock (1972) along with a new basis set suitable for disks of finite extent. We explore the extension of this Mellin transform method to the spherical geometry in Sec. 2.3.2. Also notable is Kalnajs' mention of the Kelvin transform as a method of producing new basis sets from old, which we discuss in Sec. 2.4.

Qian (1993) rewrites the integral transform (2.11) yet again, and derives further basis sets: one with a Gaussian zeroth order and a family corresponding to the Kuzmin-Toomre disks. Both Kalnajs and Qian's higher-order radial basis functions are expressed using fairly complex recursion relations, albeit involving only elementary arithmetical operations. As with Clutton-Brock, they restrict their attention to the potential on the disk. Finding expressions for the gravitational potential off the disk is in general difficult, although for Clutton-Brock's original Hankel transform approach the approach is clear in principle if not in practice.

Out of the menagerie of integral transform approaches to the disk geometry described above, at least two are applicable with slight modification to spherical systems[d]. In Sec. 2.3.1 we give the most general extension to spherical-geometry of Clutton-Brock's original Hankel transform approach, and show that it works in any number of dimensions; a similar method involving the Mellin transform also exists. In general it seems the integral expressions for the disk geometry – and their accompanying 'simple' basis sets – are somewhat more exceptional than their spherical counterparts. The generalisation of the full results of Ch. 4–Ch. 6 to the disk setting is therefore an ongoing research project.

We note in passing that it is perhaps not altogether shocking that several totally different integral transform methods (Clutton-Brock, 1972, Kalnajs, 1976, Qian, 1993) have proved fruitful in solving Poisson's equation in cylindrical polar coordinates: the Abel, Hankel and Fourier transforms are linked by the *FHA cycle*, which is another way of writing the projection-slice theorem. It states that $FA = H$ for spherically-symmetric functions. This may well be a worthwhile line of inquiry for the derivation of basis sets in spherical polar coordinates. Some relevant discussion may be found in Sections 2.3.1, 2.3.2 and 6.3.2.

### 2.1.2 Spherical geometry

In the spherical geometry the story also begins with a groundbreaking paper from Clutton-Brock: Clutton-Brock (1973), in which he took a substantially different approach from the previous case of cylindrical geometry. This time he solved Poisson's equation directly, by substituting in a product of a spherical harmonic with a Gegenbauer polynomial; thus producing a remarkably simple basis set.

It is perhaps surprising that this is possible, as the cylindrical case requires some degree of mathematical trickery in order to find analytical potential-density pairs (hence the good fortune

[d] Note the crucial distinction, when solving Poisson's equation, between systems embedded in *two-dimensional* space, and infinitely-flattened two-dimensional systems which are nevertheless embedded in *three-dimensional* space.

that the expression (2.1) exists and possesses an orthogonality property). However in the spherical case, matters are simplified somewhat: the potential-density pairs are merely required to satisfy Poisson's equation. Because this is a linear second-order ODE, one can look for a reparameterisation that allows it to be written as an eigenfunction equation in Sturm-Liouville form, to which the solutions are orthogonal. Thus it is this strategy that is employed by Clutton-Brock, who guessed a reparameterisation leading to a basis set whose zeroth order matches the Plummer (1911) model.

This result was then extended by Hernquist & Ostriker (1992) to produce a basis set whose zeroth-order matched the Hernquist (1990) model; before finally being generalised by Zhao (1996) to a family of basis sets corresponding to all the hypervirial models (Evans & An, 2005). In the interest of making explicit comparisons between Zhao's basis set and those developed in the remainder of this thesis, we give it in full:

$$\Phi_{nlm}(\mathbf{r}) = \sqrt{4\pi}\,\frac{r^l}{\left(1+r^{1/\alpha}\right)^{\mu}}\,C_n^{(\mu+1/2)}(\xi)\;Y_{lm}(\theta,\phi), \tag{2.15}$$

$$\rho_{nlm}(\mathbf{r}) = -\sqrt{4\pi}\,K_{nl}\,\frac{r^{l-2+1/\alpha}}{\left(1+r^{1/\alpha}\right)^{\mu+2}}\,C_n^{(\mu+1/2)}(\xi)\;Y_{lm}(\theta,\phi).$$

Here $\mu \equiv \alpha(1+2l)$, the polynomials $C_n^{(\beta)}(\xi)$ are those of Gegenbauer (DLMF, §18.3), and $\xi \equiv \left(r^{1/\alpha}-1\right)/\left(r^{1/\alpha}+1\right)$. The constant $K_{nl}$ appearing in the definition of the density is

$$K_{nl} = \frac{(n+\mu)(n+\mu+1)}{4\pi\alpha^2}, \tag{2.16}$$

and the orthogonality relation is

$$\int \mathrm{d}^3\mathbf{r}\;\Phi_{nlm}\,\rho_{n'l'm'} = K_{nl}\,N_{nl}, \tag{2.17}$$

$$N_{nl} = \frac{-2\pi\alpha\mu(2\mu+1)_n\Gamma(\mu)^2}{n!(2n+2\mu+1)\Gamma(2\mu)}.$$

The Clutton-Brock (1973) basis set is obtained when $\alpha = 1/2$, with a Plummer profile at zeroth-order. The Hernquist & Ostriker basis set has $\alpha = 1$ and a Hernquist zeroth-order.

It is in fact possible to adjust Clutton-Brock (1972)'s cylindrical coordinate method to cover spherical systems, but this has received scant attention in the literature[e]. We shall demonstrate (Sec. 2.3.1 and Ch. 4) that it provides a powerful alternative route into constructing further basis sets. Polyachenko & Shukhman (1981) did consider the Bessel functions over a finite interval $(0,a)$ where $a$ is the radial extent of the system being studied. This gives a biorthogonal basis set with discrete indices (see Sec. 2.2.2.1), although the expansion suffers from the issue that the Bessel functions by themselves do not resemble any particular well-known galactic profile. Unlike Clutton-Brock (1972), they did not consider the case of a continuous eigenvalue on the semi-infinite interval $(0,\infty)$.

However, Rahmati & Jalali (2009) did make the leap to the infinite radial interval, choosing like Clutton-Brock an auxiliary function $g_n(k)$ involving Laguerre polynomials, and deriving a

[e] Robijn & Earn (1996) acknowledge that the method is applicable to spherical systems, but concedes that it "would probably require some ingenious integrations".

basis set whose zeroth-order matches de Zeeuw's perfect sphere. It is this result that we improve upon in Ch. 4. The technique is generalised further in Ch. 5, the results of which finally allow us to make a direct link back to Zhao's original basis set.

## 2.2 The search for new expansions

This section comprises an exhaustive search for simple basis expansions: those expressible via classical orthogonal polynomials or other well-known special functions, combined with a change of variables. We start by heuristically suggesting the functional form of a generic basis function, and from this develop machinery for systematically deriving its exact expression. Our results suggest that other than Zhao's family and a short list of special cases, there are no additional simple basis sets.

### 2.2.1 Clutton-Brock expansions

We seek sequences of potential and density basis functions that obey the relations

$$\nabla^2 \Phi_{nlm} \propto \rho_{nlm}, \tag{2.18}$$

$$\int \mathrm{d}^3\mathbf{r}\ \Phi_{nlm}\, \rho_{n'l'm'} \propto \delta^{n'l'm'}_{nlm}. \tag{2.19}$$

Assuming that we are seeking basis sets expressed in spherical polar coordinates, for each basis function we use a spherical harmonic for the angular part and use a heuristic for the functional form of the radial part (cf. Clutton-Brock, 1973, Hernquist & Ostriker, 1992, Zhao, 1996)

$$\begin{aligned} \Phi_{nlm}(\mathbf{r}) &= \Phi_l(r)\, W_{nl}(r)\, Y_{lm}(\theta,\phi), \\ \rho_{nlm}(\mathbf{r}) &= \rho_l(r)\, K_{nl}\, W_{nl}(r)\, Y_{lm}(\theta,\phi), \end{aligned} \tag{2.20}$$

where $W_{nl}$ are functions that will obey an orthogonality relation in $r$, and the $K_{nl}$ are constants[f]. We further impose that $W_{0l} = 1$, which means that $\rho_l$ follows from the definition of $\Phi_l$,

$$\Phi_l{}'' + \frac{2}{r}\Phi_l' - \frac{l(l+1)}{r^2}\Phi_l = \rho_l K_{0l}. \tag{2.21}$$

Eq. (2.18) otherwise (with $n > 0$) expands to

$$\Phi_l W_{nl}{}'' + 2\left(\Phi_l' + \frac{\Phi_l}{r}\right) W_{nl}{}' + \left(\Phi_l{}'' + \frac{2}{r}\Phi_l' - \frac{l(l+1)}{r^2}\Phi_l - \rho_l K_{nl}\right) W_{nl} = 0. \tag{2.22}$$

We denote sets of basis functions that obey Eq. (2.22) as *Clutton-Brock expansions*, as they closely match the functional form of the basis sets introduced in Clutton-Brock (1973). Now we use (2.21) to immediately simplify (2.22), defining $A_{nl} \equiv K_{0l} - K_{nl}$, giving

$$\Phi_l W_{nl}{}'' + 2\left(\Phi_l' + \frac{\Phi_l}{r}\right) W_{nl}{}' + A_{nl}\rho_l W_{nl} = 0. \tag{2.23}$$

[f] The $K_{nl}$ are freely-adjustable for now, but their values will end up affecting the analytical form of the basis functions once we choose a Sturm-Liouville equation to map to.

This equation defines a Sturm-Liouville operator

$$L \equiv -\frac{d}{dr}\left(\Phi_l^2 r^2 \frac{d}{dr}\right), \tag{2.24}$$

whose eigenfunctions $W_{nl}$ are orthogonal with respect to the weight function

$$\Omega_l(r) \equiv r^2\, \rho_l(r)\, \Phi_l(r). \tag{2.25}$$

The orthogonality relation (2.19) can then be written as

$$\int Y_{lm} Y^*_{l'm'}\, \mathrm{d}\phi \cos\theta\, \mathrm{d}\theta \int W_{nl} W_{n'l'} \Omega_l\, \mathrm{d}r = \delta^{nlm}_{n'l'm'} N_{nl}, \tag{2.26}$$

where $N_{nl}$ normalises the functions $W_{nl}(r)$.

We must now search for valid sets of functions $W_{nl}(r)$, taking advantage of the existing literature on Sturm-Liouville equations to find functions of suitable form (see Everitt, 2005, for a catalogue of such equations). Suppose we have to hand a Sturm-Liouville equation with independent variable $z$, known eigenfunctions $y_n(z)$, and eigenvalues $\lambda_n$,

$$-\frac{d}{dz}\left(p(z)\frac{dy_n}{dz}\right) = \lambda_n \omega(z) y_n. \tag{2.27}$$

Now via a change of variables $z = z(r)$ we set $W_{nl}(r) = y_n(z(r))$. Then, to match solutions of Eq. (2.27) to those of Eq. (2.22) we need $z$ and $\Phi_l$ to satisfy the following two constraints (using dashes for $r$-derivatives and dots for $z$-derivatives, so that $z' = dz/dr = 1/\dot{r}$),

$$r^2 \Phi_l^2 = \frac{p}{z'}, \tag{2.28}$$

$$r^2 \Phi_l \rho_l = \kappa z' \omega, \tag{2.29}$$

where $\kappa \equiv \lambda_n / A_{nl}$ is a freely-adjustable parameter that we will later use to simplify the algebra as much as possible[g]. Because $\rho_l$ is set by Eq. (2.21), we can regard Eq. (2.28) and Eq. (2.29) as a nonlinear third-order system of differential equations for $\Phi_l(r)$ and $z(r)$ (with $r$ as independent variable), or for $\Phi_l(r(z))$ and $r(z)$ (with $z$ as independent variable).

To simplify, we set $\Phi_l(r) = r^l f(r)$ and eliminate $f(r)$ between Eq. (2.28) and Eq. (2.29), thus arriving at

$$\kappa\, \omega\, p = \frac{1}{2}\left(\frac{p}{z'}\right)'' \frac{p}{z'} - \frac{l(l+1)}{r^2}\left(\frac{p}{z'}\right)^2 - \frac{1}{4}\left(\frac{p}{z'}\right)'^2. \tag{2.30}$$

Now we swap $r$-derivatives for $z$-derivatives, using $z' = 1/\dot{r}$ and $\frac{d}{dr} = \frac{1}{\dot{r}}\frac{d}{dz}$, to find

$$\frac{1}{2}\frac{\dddot{r}}{\dot{r}} - \frac{3}{4}\frac{\ddot{r}^2}{\dot{r}^2} - l(l+1)\frac{\dot{r}^2}{r^2} = \kappa\frac{\omega}{p} - \frac{1}{2}\frac{\ddot{p}}{p} + \frac{1}{4}\frac{\dot{p}^2}{p^2} \equiv F(z), \tag{2.31}$$

noting that the RHS is just a known function of $z$ that we have denoted $F(z)$. Now we make the substitutions $v(z) \equiv \sqrt{r/\dot{r}}$ and $\beta \equiv 2l+1$ to find

$$\ddot{v} + F(z) v + \frac{\beta^2}{4} v^{-3} = 0, \tag{2.32}$$

which is a form of the Ermakov-Pinney (hereafter EP) equation (Morris & Leach, 2015). The standard solution method for an EP equation such as Eq. (2.32) is as follows:

[g]The freedom to adjust $\kappa$ is inherited from the definition of $K_{nl}$ in Eq. (2.20).

1) First, solve the following associated homogeneous equation, given by the Schrödinger-like equation
$$\ddot{u} + F(z)u = 0, \tag{2.33}$$
for which we denote the two independent solutions $u_1$ and $u_2$.
2) Find the Wronskian
$$W \equiv Wr[u_1, u_2] = u_1\dot{u}_2 - \dot{u}_1 u_2. \tag{2.34}$$
3) The general solution to Eq. (2.32) is then
$$v(z) = \sqrt{Au_1^2(z) + Bu_2^2(z) + Cu_1(z)u_2(z)}, \tag{2.35}$$
where $A$, $B$ and $C$ are constrained by the relation
$$C^2 = 4AB + \beta^2/W^2.$$

In our case the associated equation (2.33) can be immediately transformed back into the S-L equation (2.27) using the substitution
$$u = \sqrt{p}y, \tag{2.36}$$
which implies that $W$ is always constant. Also, if we know the two independent solutions $(y_1, y_2)$ of Eq. (2.27) that correspond to the eigenvalue $\kappa$, we can immediately write down $(u_1, u_2)$ and hence obtain the solution to the EP equation (2.35). The trick lies in choosing the value of $\kappa$ (allowing negative values) such that we end up with a suitably simple expression for $v(z)$, and hence an invertible function $r(z)$. This is perhaps the most arbitrary part of the method, as a great variety of possible functional forms can result depending on the choice of $\kappa$. However the requirement that we end up with simple analytical basis functions is quite restrictive, so in practice our hand is forced to quite a restricted set of possibilities for $\kappa$.

Now, knowing the EP solution (2.35), we can invert $v(z)$ to write down $r(z)$ as the integral
$$\log r = \int \frac{\mathrm{d}z}{v^2} = \int \frac{\mathrm{d}z}{Au_1^2(z) + Bu_2^2(z) + Cu_1(z)u_2(z)}, \tag{2.37}$$
noting that the constant of integration introduced here turns out to be the scale-length $r_s$, that we set to 1 from now on. Note that $u_1$ and $u_2$ may be labelled in either order. In fact, because only the ratio
$$\frac{u_1}{u_2} = \frac{y_1}{y_2} \tag{2.38}$$
is used in the final result, we can substitute in two independent solutions of Eq. (2.27) directly (these solutions may be easy to find if the S-L system in question is well-studied).

Depending on the parameter choices, the solution to the integral (2.37) will fall under one of the following cases[h],
$$\log r = \begin{cases} \frac{2}{\beta}\tanh^{-1}\left(\frac{2AW}{\beta}\frac{y_2}{y_1} + \frac{CW}{\beta}\right), & \text{if } A \neq 0, \\ \frac{2}{\beta}\tanh^{-1}\left(\frac{2BW}{\beta}\frac{y_2}{y_1} + \frac{CW}{\beta}\right), & \text{if } B \neq 0, \\ \frac{1}{\beta}\log\left(\frac{y_2}{y_1}\right), & \text{if } A = B = 0. \end{cases} \tag{2.39}$$

[h]The first and second cases are symmetric under reversing the labelling of $A$ and $B$.

Therefore if we can choose $\kappa$, $A$ and $B$ such that $y_2/y_1$ is an invertible analytical function, we will be able to write down $z(r)$ in closed form. In addition, we require that $z(0)$ and $z(\infty)$ map to the endpoints of the interval of orthogonality of the S-L equation (2.27). With these conditions, $\Phi_l$ can finally be computed from Eq. (2.28).

#### 2.2.1.1 Laguerre polynomials

For our first application of this method, we will choose the generalised Laguerre polynomials $L_n^{(\alpha)}(z)$ as our Sturm-Liouville system (DLMF, §18.3). These are orthogonal on $(0, \infty)$ and have a single free parameter, here denoted $\alpha$. The relevant functions (to insert into Eq. (2.27)) are

$$\begin{aligned} \omega(z) &= z^\alpha e^{-z}, \\ p(z) &= z^{\alpha+1} e^{-z}, \end{aligned} \tag{2.40}$$

so that the quantity that appears in Eq. (2.33) is

$$F(z) = \frac{-1}{4z^2}\left(z^2 - (4\kappa + 2\alpha + 2)\, z + \alpha^2 - 1\right). \tag{2.41}$$

We now choose to set $\alpha = 1$ and $\kappa = -1$, in order to get the simplest possible form for $F(z)$ – it is now merely a constant, $F = -1/4$. The solutions to Eq. (2.33) are therefore

$$u(z) = e^{\pm z/2}, \tag{2.42}$$

so the general solution to the EP equation (2.32) is

$$v^2 = Ae^z + Be^{-z} + \sqrt{4AB + \beta^2}. \tag{2.43}$$

We pick the solution with $A = 0$ and $B = -\beta$, so that

$$\log r = \frac{1}{\beta}\int \frac{\mathrm{d}z}{1 - e^{-z}} = \log\left(1 - e^z\right), \tag{2.44}$$

and the change of variables can be inverted to find

$$z = \log\left(1 + r^\beta\right). \tag{2.45}$$

Knowing $z(r)$ now lets us use Eq. (2.28) and Eq. (2.29) to find explicit expressions for the radial potential and density basis functions,

$$\begin{aligned} \Phi_{nl}(r) &= \frac{\log\left(1 + r^{2l+1}\right)}{r^{l+1}}\, L_n^{(1)}\left(\log\left(1 + r^{1+2l}\right)\right), \\ \rho_{nl}(r) &= \frac{r^{3l-1}}{(r^{2l+1} + 1)^2}\, L_n^{(1)}\left(\log\left(1 + r^{1+2l}\right)\right), \end{aligned} \tag{2.46}$$

and remarkably we find that at zeroth-order ($n = l = 0$) this reproduces the NFW model. The normalisation and proportionality constants ($N_{nl}$ and $K_{nl}$) can easily be derived from the properties of the Laguerre polynomials.

Investigating the numerical properties of this basis set, we find that it performs acceptably for spherical systems, but when the asphericity is moderate the expansion fails to converge. The high powers of $l$ to which $r$ is raised causes a ringing effect to appear around the scale-length.

Fig. 2.1 shows a comparison between the basis set described above (which we dub the *NFW/Laguerre* basis set) and the basis set of Ch. 5, the latter of which matches the NFW profile at zeroth order when its free parameters are set to $\alpha = 1$ and $\nu = 0$. With both basis sets we expand the same moderately flattened NFW halo. We see a spike in the residuals for the NFW/Laguerre basis set around $r = 1$, which does not decrease in magnitude even at high expansion order.

The severity of the numerical artifacts is lessened when considering the potential and acceleration fields, as oscillations are effectively integrated out, so it remains unclear which basis set has the ultimate advantage, and a comprehensive comparison of the two basis sets is still to be made. However. there is no reason that the $l > 0$ functions of a particular basis set have to come from the *same* basis set as the $l = 0$ functions. The orthogonality in $l$ is set by the angular part of the orthogonality integral (1.36), so a well-behaved basis set could perhaps be constructed taking the $l = 0$ terms from the afore-mentioned NFW/Laguerre set, and the $l > 0$ terms from the Hernquist & Ostriker basis set.

#### 2.2.1.2 Jacobi polynomials

The next most obvious Sturm-Liouville system to try is the Jacobi polynomials $P_n^{(\gamma,\delta)}(z)$ (DLMF, §18.3). These have two free parameters, denoted $\gamma$ and $\delta$, and the functions appearing in Eq. (2.27) are

$$\begin{aligned} \omega(z) &= (1-z)^{\gamma}(1+z)^{\delta}, \\ p(z) &= (1-z)^{\gamma+1}(1+z)^{\delta+1}. \end{aligned} \tag{2.47}$$

We then find that a suitably simple form of $F(z)$ is obtained by setting $\delta = \gamma$ and $\kappa = -\gamma(1+\gamma)$, giving

$$F(z) = \frac{1-\gamma^2}{(1-z^2)^2}. \tag{2.48}$$

Other parameter choices apparently do not give analytically tractable expressions. The corresponding solutions of Eq. (2.33) are then

$$\begin{aligned} u_1 &= \sqrt{1-z^2}\,\exp\left(-\gamma^2 \arctan z\right), \\ u_2 &= \frac{1}{2\gamma}\,(1-z)^{\frac{1-\gamma}{2}}\,(1+z)^{\frac{1+\gamma}{2}}. \end{aligned} \tag{2.49}$$

If we choose $A = B = 0$ in the general solution to Eq. (2.32) then we obtain

$$v^2 = \sqrt{C u_1 u_2} = \frac{\beta}{2\gamma}\,(1-z^2)\,, \tag{2.50}$$

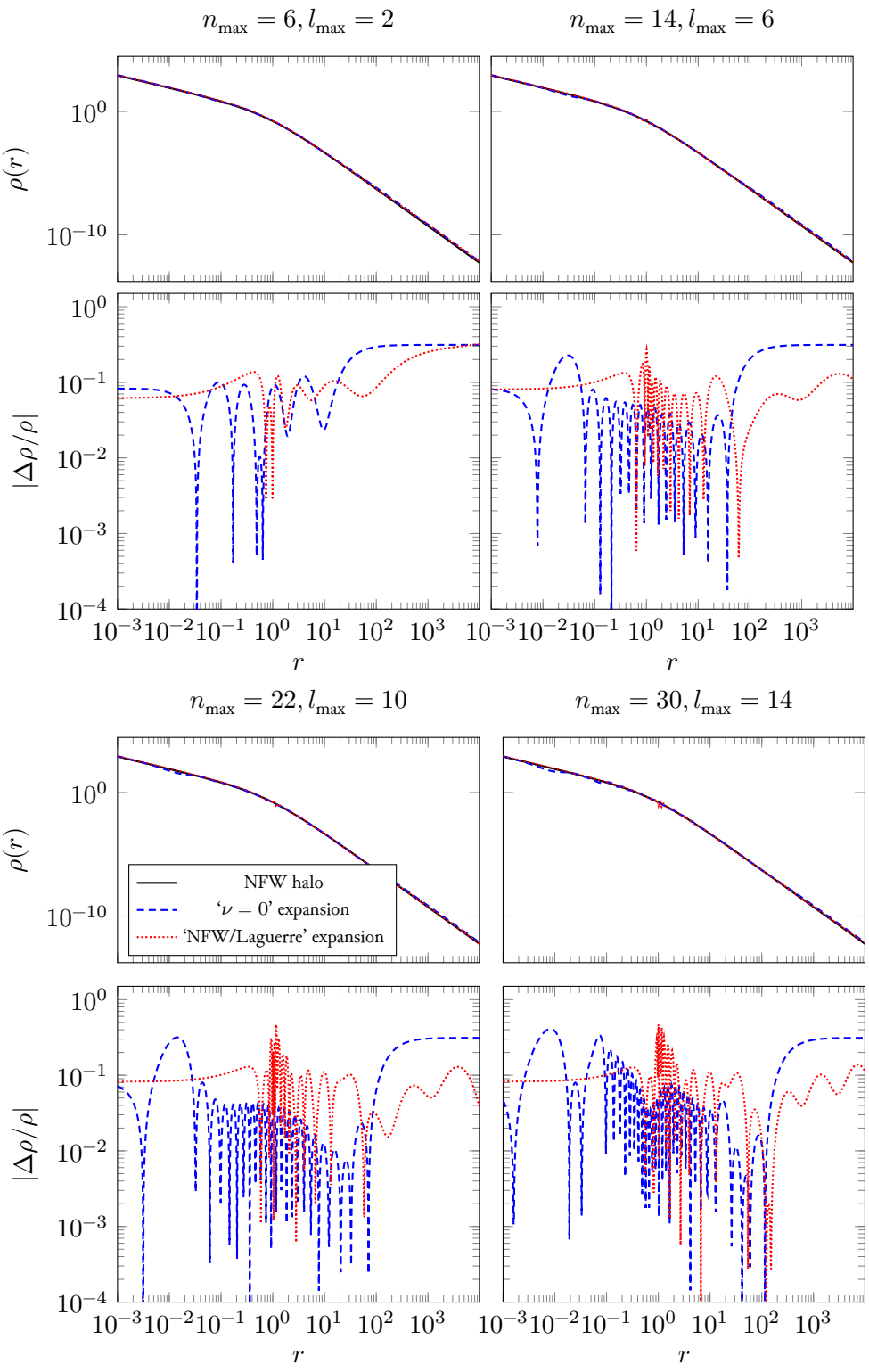


Figure 2.1: Two expansions of a flattened ($q = 0.7$) NFW halo. Each expansion has a spherically-symmetric NFW model as its zeroth order. The $\nu = 0$ expansion (dashed blue) is a particular member of the family described in Ch. 5. The *NFW/Laguerre* expansion (dotted red) is the basis set derived in Sec. 2.2.1.1. The density reconstruction is shown for four choices of expansion order, along with residuals.

again finding that other parameter choices lead to intractable equations. The final result for $r(z)$ is therefore

$$r = \exp\left[\frac{2\gamma}{\beta}\left(\int \frac{\mathrm{d}z}{1-z^2}\right)\right] = \left(\frac{z+1}{z-1}\right)^{\gamma/\beta}. \tag{2.51}$$

Numerical evidence from Sec. 2.2.1.1 suggests that it is undesirable to have an exponential dependence on $l$ in our expression for $z(r)$. Because we have retained the free parameter $\gamma$, we can now choose it to remove the $l$-dependence from the definition of $z(r)$. We set $\gamma/\beta \equiv \alpha$, so that

$$z(r) = \frac{r^{1/\alpha}-1}{r^{1/\alpha}+1}. \tag{2.52}$$

It remains to use Eq. (2.28) and Eq. (2.29) to find the actual functional forms of the potential and density; these are

$$\begin{aligned}
\Phi_{nl}(r) &\propto \frac{r^l}{\left(1+r^{1/\alpha}\right)^{\alpha(1+2l)}} P_n^{(\alpha(1+2l),\alpha(1+2l))}\left(\frac{r^{1/\alpha}-1}{r^{1/\alpha}+1}\right), \\
\rho_{nl}(r) &\propto \frac{r^{1/\alpha-2+l}}{\left(1+r^{1/\alpha}\right)^{\alpha(1+2l)+2}} P_n^{(\alpha(1+2l),\alpha(1+2l))}\left(\frac{r^{1/\alpha}-1}{r^{1/\alpha}+1}\right).
\end{aligned} \tag{2.53}$$

We immediately recognise these as the basis functions of the Zhao expansions, as the Jacobi polynomials with both parameters equal $P^{(\gamma,\gamma)}(z)$ are proportional to the Gegenbauer polynomials. We have thus demonstrated how several previous results (Clutton-Brock, 1973, Hernquist & Ostriker, 1992, Zhao, 1996) could have been mechanically derived from a specified target Sturm-Liouville equation, employing the heuristic method developed above.

### 2.2.2 Pseudo-Clutton-Brock expansions

An alternative set of solutions that obey the conditions (2.18) and (2.19) can be obtained by dropping the assumption that $W_{0l} = 1$. This means that Eq. (2.22) does not simplify, so we must solve

$$-\frac{d}{dr}\left(P\frac{dW_{nl}}{dr}\right) + QW_{nl} = K_{nl}W_{nl}\Omega, \tag{2.54}$$

where we have defined

$$\begin{aligned}
P &\equiv r^2\Phi_l^2, \\
Q &\equiv -r^2\Phi_l\left(\nabla^2 - \frac{l(l+1)}{r^2}\right)\Phi_l, \\
\Omega &\equiv r^2\Phi_l\rho_l.
\end{aligned} \tag{2.55}$$

There is no longer a straightforward connection between $\Phi_l$ and $\rho_l$, because Eq. (2.21) followed from the assumption that $W_{0l} = 1$. They are now just pre-factors to the orthogonal part of the basis functions. We call sets of basis functions that obey Eq. (2.22) *pseudo-Clutton-Brock expansions*.

Compared to Eq. (2.23), we see that Eq. (2.54) has an extra term proportional to $W_{nl}$, so we will try to match it onto a given Sturm-Liouville equation with *three* terms,

$$-\frac{d}{dz}\left(p\frac{dy_n}{dz}\right) + qy_n = \lambda_n \omega\, y_n. \tag{2.56}$$

The procedure is similar to before, but we now have three simultaneous conditions to satisfy:

$$r^2\Phi_l^2 = \frac{p}{z'}, \tag{2.57}$$

$$r^2\Phi_l\left(\nabla^2 - \frac{l(l+1)}{r^2}\right)\Phi_l = -qz', \tag{2.58}$$

$$K_{nl}r^2\Phi_l\rho_l = -\lambda_n z'\omega. \tag{2.59}$$

Using the first two conditions, we can again perform a simplification that results in an EP equation. The key difference to before is changing the definition of $F(z)$ in Eq. (2.32) to

$$F(z) = -\frac{q}{p} - \frac{1}{2}\frac{\ddot{p}}{p} + \frac{1}{4}\frac{\dot{p}^2}{p^2}, \tag{2.60}$$

noting the lack of a free parameter ($\kappa$ in Eq. (2.29)). Therefore, we require two independent solutions (denoted $y_1, y_2$) to

$$-\frac{d}{dz}\left(p\frac{dy}{dz}\right) + qy = 0, \tag{2.61}$$

and the solution for $z(r)$ proceeds via integration as above. Then $\rho_l$ is a straightforward calculation from Eq. (2.59).

#### 2.2.2.1 Bessel functions

The Bessel functions $J_\mu(z)$ are widely-studied special functions (DLMF, §10.2) with commonly-available numerical routines to compute them. They obey a Sturm-Liouville equation of the type (2.56) with one free parameter (denoted $\mu$), with

$$p = \omega = z^{2\mu+1}; \qquad q = 0; \qquad \lambda_n = z_n^2, \tag{2.62}$$

where $z_n$ is the $n$-th zero of the Bessel function. They are orthogonal on the interval $(0, 1)$ with an orthogonality relation[i] that depends on $z_n$ as follows,

$$\int_0^1 z\,\mathrm{d}z\; J_\mu(z\,z_n)\; J_\mu(z\,z_m) = \frac{1}{2}\,\delta_{nm}\,\left(J_{\mu+1}(z_n)\right)^2. \tag{2.63}$$

In this case, the two independent solutions to Eq. (2.61) are

$$y_1 = z^{-2\mu}; \;\; y_2 = \frac{1}{2\mu}, \tag{2.64}$$

[i] There is also a continuous version of this orthogonality relation, valid on $(0, \infty)$, that we shall deal with later.

so to calculate Eq. (2.39) we choose the case $A = B$, to find

$$r^\beta = z^{2\mu}. \tag{2.65}$$

In order to retain a free parameter, we set $\mu = \alpha\beta = \alpha(1 + 2l)$, and then applying Eq. (2.57) and Eq. (2.59) we find

$$\begin{aligned} \Phi_l &\propto r^l, \\ \rho_l &\propto k^2\, r^l, \end{aligned} \tag{2.66}$$

and hence the orthogonal basis functions (with discrete index $n$) are

$$\begin{aligned} \Phi_{nl} &\propto r^{-1/2}\, J_{\alpha(1+2l)}(z(r)\, z_n)\,, \\ \rho_{nl} &\propto r^{1/\alpha-5/2}\, J_{\alpha(1+2l)}(z(r)\, z_n)\,. \end{aligned} \tag{2.67}$$

These are exactly the functions used by Polyachenko & Shukhman (1981) in their pioneering stability analysis of collisionless systems. For our purposes, it is interesting that they also identified the free parameter $\alpha$, which we will find to be more widely applicable (see Ch. 4–Ch. 6). The Bessel basis set is itself less universally useful, as it does not resemble any particular galactic halo profile at zeroth order, and can only be used on an (appropriately rescaled) finite interval.

#### 2.2.2.2 Associated Legendre functions

One final set of promising functions is the associated Legendre functions $P_n^{(m)}(z)$, which are obtained from the Legendre polynomials by differentiating $m$ times and subsequently allowing the index $m$ to be real-valued (DLMF, §14.3). The quantities relating to their Sturm-Liouville equation (2.56) are

$$p = 1 - z^2; \qquad q = m^2/(1 - z^2); \qquad \omega = 1; \qquad \lambda_n = n(n + 1). \tag{2.68}$$

Thus we need two independent solutions to

$$-\frac{d}{dz}\left((1 - z^2)\frac{dy}{dz}\right) + \frac{m^2}{1 - z^2}y = 0, \tag{2.69}$$

which are

$$\begin{aligned} y_1(z) &= \cosh\left(\frac{m}{2}\log\left(\frac{1-z}{1+z}\right)\right), \\ y_2(z) &= \sinh\left(\frac{m}{2}\log\left(\frac{1-z}{1+z}\right)\right). \end{aligned} \tag{2.70}$$

Hence we select the case $B = \beta/2$ in Eq. (2.39) and then choose $A$ such that $C = 0$, to find

$$r^{\beta/2} = \left(\frac{1-z}{1+z}\right)^{m/2}. \tag{2.71}$$

Setting $m = \alpha\beta = \alpha(1+2l)$ we can then easily invert to find a choice of $z(r)$ that does not depend on $l$. Using the relations (2.57)–(2.59) we find $\Phi_l = 1/\sqrt{r}$, so that the potential basis functions are

$$\Phi_{nl}(r) = \frac{1}{\sqrt{r}} P_n^{(\alpha(1+2l))}\left(\frac{1-r^{1/\alpha}}{1+r^{1/\alpha}}\right). \tag{2.72}$$

This expression looks suspiciously like the expression for the Zhao basis functions given in (2.53), and our suspicions are confirmed upon applying the connection relation between associated Legendre functions and Gegenbauer polynomials (DLMF, §14.3(iv)), which demonstrates that the two expansions are in fact the same.

### 2.2.3 Conclusions

Not included in the discussion above are the Hermite polynomials, which do not appear to lead to any useful expressions; and the simple harmonic oscillator equation, which leads to a Fourier transform-based method that is likely numerically inferior (see Appendix B for discussion).

Based on the heuristic method developed in the preceding sections, we therefore conjecture that there are no more straightforward ways of transforming Sturm-Liouville equations into biorthogonal basis sets other than those listed above.

This search was non-exhaustive, as there are likely many other simple analytical Sturm-Liouville eigenfunctions with useful properties. However, the classical orthogonal polynomials are provably the only polynomials which obey a second-order ordinary differential equation (Bochner, 1929). We therefore search for alternative methods of constructing biorthogonal basis sets.

## 2.3 Integral transform methods

From Sec. 2.2 it is clear that we must broaden our search for methods of deriving biorthogonal basis sets. Fortunately, Poisson's equation is widely-studied, so a wide variety of solution techniques exist – in particular we note the existence of *integral transform* methods. We will proceed to give an overview of their use in deriving basis sets, but once again likely do not exhaust the possibilities.

Generically, the integral transform $T$ of a function $f(x)$ gives a new function $g(y)$, by integrating $f(x)$ with some kernel function $K_1(x,y)$,

$$T_x\{f(x)\}(y) = \int \mathrm{d}x\, K_1(x,y)\, f(x) = g(y). \tag{2.73}$$

In many cases an inversion theorem exists, so that (with another appropriate kernel $K_2$) the original function can be recovered,

$$T_y^{-1}\{g(y)\}(x) = \int \mathrm{d}x\, K_2(x,y)\, g(y) = f(x). \tag{2.74}$$

Various integral transforms have special properties that motivate their application to specific problems. In particular, all the transforms below transform the Laplacian operator to a simpler form. Crucial to the success of a given transform method is the ease with which various elementary

or otherwise well-known functions can be transformed; tables are often published for this purpose (e.g. BMP, 1954, G&R, 2007).

### 2.3.1 Hankel transform

We now describe the most promising route to our desired goal of writing down new biorthogonal basis sets: the *Hankel transform*. This method's success depends on two key facts: 1) the kernel function is the comparatively well-studied Bessel function, and it therefore has numerous attractive known analytical properties; 2) the radial eigenfunction of the Laplacian is a spherical Bessel function, which is obtainable from the standard Bessel function by a trivial reparameterisation.

The Hankel transform of order $\alpha$ is given by

$$\mathcal{H}_k\left\{f(k)\right\}(x) \equiv \int_0^\infty k\,\mathrm{d}k\; f(k)\; J_\alpha(kx)\,. \tag{2.75}$$

Notable is the fact that it is its own inverse, i.e.

$$\mathcal{H}_k\left\{\mathcal{H}_q\left\{f(q)\right\}(k)\right\}(x) = f(x). \tag{2.76}$$

We shall avoid this formal notation, and instead generically refer to any integral containing a Bessel $J_\alpha$ function as a 'Hankel transform'.

The Laplacian operator in $d$ spatial dimensions is

$$\nabla^2 \equiv \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_d^2} = \nabla_r^2 + r^{1-d}\frac{\partial}{\partial \hat{\mathbf{r}}}, \tag{2.77}$$

with generalised radial coordinate $r$ and unit vector $\hat{\mathbf{r}}$, and the radial Laplacian $\nabla_r^2$ is defined as

$$\nabla_r^2 \equiv r^{1-d}\frac{d}{dr}\left(r^{d-1}\frac{d}{dr}\right). \tag{2.78}$$

Its eigenfunctions in generalised spherical coordinates satisfy

$$\begin{aligned}\nabla^2 u_{kl\boldsymbol{m}}(\mathbf{r}) &= -k^2\, u_{kl\boldsymbol{m}}(\mathbf{r}),\\ u_{kl\boldsymbol{m}}(\mathbf{r}) &\equiv j_l(kr)\, Y_{l\boldsymbol{m}}(\hat{\mathbf{r}}),\end{aligned} \tag{2.79}$$

where $\boldsymbol{m}$ is a multi-index corresponding to the $d-2$ quantities $m_1 \ldots m_{d-2}$, $Y_{l\boldsymbol{m}}$ is a generalised spherical harmonic[j] and $j_l$ is (our definition of) a spherical Bessel function. Such spherical Bessel functions are related to standard Bessel functions $J_\mu$ by

$$j_l(x) \equiv x^{1-d/2} J_{l+d/2-1}(x)\,. \tag{2.80}$$

Given an arbitrary function $g(k)$, if we multiply both sides of Eq. (2.79) by $g(k)$ and integrate with respect to $k$ we obtain a valid potential-density pair that solves Poisson's equation.

[j] See e.g. Cohl (2013, Appendix B) for a clear description of how to construct spherical harmonics in arbitrary dimension.

The fact that the Hankel transform is its own inverse is a consequence of the fact that Bessel functions also obey a *continuous* version of their orthogonality relation[k],

$$\int_0^\infty z\,\mathrm{d}z\; J_\mu(kz)\;\; J_\mu(qz) = \frac{\delta(k-q)}{k}, \tag{2.81}$$

and so the spherical Bessel functions also obey a similar relation,

$$\int_0^\infty r^{d-1}\,\mathrm{d}r\; j_l(kr)\; j_l(qr) = \frac{\delta(k-q)}{k}. \tag{2.82}$$

This suggests the following method of producing biorthogonal basis sets: find an arbitrary set of functions $g_n(k)$ that obey an orthogonality relation with respect to the weight $k$,

$$\int_0^\infty k\,\mathrm{d}k\; g_n(k)\; g_{n'}(k) \propto \delta_{nn'}, \tag{2.83}$$

and calculate the corresponding potential-density pairs with the Hankel transform.

However, we immediately generalise the procedure above to include an additional degree of freedom, as derived in Sec. 2.2.2.1 and first noted by Polyachenko & Shukhman (1981). We define

$$z \equiv r^{1/(2\alpha)}, \qquad \mu \equiv 2\alpha(l+d/2-1), \qquad \alpha \geq 1/2, \tag{2.84}$$

and calculate the radial potential-density pairs as follows,

$$\begin{aligned} \Phi_{nl}(r) &= r^{1-d/2} \int \mathrm{d}k\; g_n(k)\; J_\mu(kz), \\ \rho_{nl}(r) &= r^{1/\alpha-1-d/2} \int k^2\,\mathrm{d}k\; g_n(k)\; J_\mu(kz). \end{aligned} \tag{2.85}$$

Applying the obvious change of variables to the orthogonality relation (2.81), we see that these slightly more general functions also transfer the $g_n(k)$ orthogonality property (2.83) from $k$-space into $r$-space,

$$\int r^{d-1}\,\mathrm{d}r\; \Phi_{nl}(r)\; \rho_{n'l}(r) \propto \int k\,\mathrm{d}k\; g_n(k)\; g_{n'}(k) \propto \delta_{nn'}. \tag{2.86}$$

The '$\alpha$' degree of freedom will be found valuable as it adjusts several properties of the resulting basis functions: the power-law cusp in the density functions (that is, their asymptotic behaviour as $r \to 0$); the size of the 'turnover' region between the two parts of the double-power law density profile; and the spacing of the roots of the polynomials that appear in the basis functions.

The Hankel transform method is the principle method for basis set construction explored in this thesis. Chapters 4–6 outline a variety of new basis sets discovered (directly or indirectly) via this technique. For the purposes of this thesis we restrict our attention to $d = 3$ spatial dimensions, but note that all our new basis sets are given in terms of the parameter $\mu$, and so they generalise trivially to higher dimensions (but not $d = 2$) simply by changing $d$ in Eq. (2.84).

[k] See Sec. 2.2.2.1 for a description and application of the discrete orthogonality relation.

### 2.3.2 Mellin transform

The Mellin transform is a multiplicative analogue of the Fourier transform, and may be used to solve certain differential equations. In this section we will explore a technique for solving Poisson's equation in spherical polar coordinates that is analogous to the method in cylindrical coordinates mentioned in Sec. 2.1.1 (introduced in Kalnajs, 1971, 1976).

The forward and inverse Mellin transforms are defined by

$$\mathcal{M}_r\left\{f(r)\right\}(s) \equiv \int_0^\infty r^{s-1}\, f(r)\, \mathrm{d}r, \tag{2.87}$$
$$\mathcal{M}_s^{-1}\left\{g(s)\right\}(r) \equiv \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} r^{-s}\, g(s)\, \mathrm{d}s.$$

We shall also make use of the multiplicative convolution property

$$\mathcal{M}_s^{-1}\left\{\mathcal{M}_r\left\{f(r)\right\}(s) \times \mathcal{M}_r\left\{g(r)\right\}(s)\right\}(r) = \int_0^\infty f(r')\; g\left(r/r'\right) \frac{\mathrm{d}r'}{r'}. \tag{2.88}$$

We will first consider the Mellin transform as a general method of recovering the potential corresponding to a given mass distribution; and afterwards consider it more specifically as a technique to manufacture biorthogonal basis sets.

Given a potential density pair $(\Phi, \rho)$, let us expand the angular part of each in spherical harmonics,

$$\Phi(\mathbf{r}) = \sum_{lm} \Phi_l(r)\, Y_{lm}(\theta,\phi); \qquad \rho(\mathbf{r}) = \sum_{lm} \rho_l(r)\, Y_{lm}(\theta,\phi). \tag{2.89}$$

Next, for each $l$ define the *reduced potential* $\varphi_l$ and *reduced density* $\varrho_l$,

$$\varphi_l(r) \equiv r^{1/2}\, \Phi_l(r); \qquad \varrho_l(r) \equiv r^{5/2}\, \rho_l(r), \tag{2.90}$$

and associated Mellin transforms

$$\psi_l(s) \equiv \mathcal{M}_r\left\{\varphi_l(r)\right\}(s); \qquad \omega_l(s) \equiv \mathcal{M}_r\left\{\varrho_l(r)\right\}(s). \tag{2.91}$$

Using the standard properties of the Mellin transform (BMP, 1954, Ch. 6), the radial Poisson equation (1.28) satisfied by $\Phi_l$ and $\rho_l$ translates into the following relation satisfied by $\psi_l$ and $\omega_l$ (defining the multiplicative factor as $K_l(s)$),

$$K_l(s)\psi_l(s) \equiv \left(s - \frac{1}{2} - l\right)\left(s + \frac{1}{2} + l\right)\psi_l(s) = 4\pi G\, \omega_l(s). \tag{2.92}$$

Therefore, using the Mellin inversion formula, we can find a $\psi_l$ for any given $\omega_l$. This is of course most useful when the function we are proposing as a density distribution is in a form that appears in standard reference tables of integral transforms.

For the purposes of finding orthogonal basis sets, we are interested in the inner product between two arbitrary gravitational potentials $\Phi$ and $\Phi'$

$$\langle \Phi, \Phi' \rangle \equiv \int \mathrm{d}^3\mathbf{r}\; \boldsymbol{\nabla}\Phi \cdot \boldsymbol{\nabla}\Phi'. \tag{2.93}$$

Specifically, we would like to write this as a one-dimensional integral in a more symmetric manner, akin to Eq. (2.86). As with the Hankel transform method of Sec. 2.3.1, the angular part of the integral is taken care of by the spherical harmonics. Let us write $\psi_l$ and $\psi'_l$ for the two Mellin-transformed reduced potentials, and additionally define a function $g_l$ corresponding to each reduced potential $\psi_l$,

$$g_l(s) \equiv (s + l + 1/2)\, \psi_l(s) \tag{2.94}$$

Applying the inverse Mellin transform to the radial part of $\Phi$ and $\Phi'$, and substituting in the definitions of $g_l$ and $g'_l$, we see that their inner product reduces to

$$\langle \Phi, \Phi' \rangle \propto \sum_l \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \mathrm{d}s\; g_l(s)\, g_l(-s). \tag{2.95}$$

This is the symmetric form we are seeking. To make contact with Kalnajs (1971), we would choose $g_l(s)$ purely imaginary and make the substitution $s = \mathrm{i} \log r$. Of note is the fact that the auxiliary function $K_l(s)$ that links the potential and the density in $s$-space is comparatively simple (compare this with the analogous function in the disk setting $K_m(\alpha)$ of Eq. (2.12)). To follow further the path of Kalnajs (1971), we now re-express the original potential function in terms of the inverse Mellin transform of $g_l(s)$, which we define to be $G_l(r)$,

$$G_l(r) \equiv \mathcal{M}_s^{-1} \left\{ g_l(s) \right\} (r), \tag{2.96}$$

and let $k_l(r)$ be the inverse Mellin transform of the following function

$$k_l(r) \equiv \mathcal{M}_s^{-1} \left\{ \frac{1}{s + l + 1/2} \right\} (r). \tag{2.97}$$

Using the tables of transforms found in BMP (1954, Ch. 7), and the convolution formula (2.88), we obtain for the potential and density,

$$\begin{aligned} \Phi_l(r) &= r^{-1/2} \int_0^\infty \frac{\mathrm{d}x}{x} G_l(x)\, k_l\left(\frac{r}{x}\right) = r^l \int_0^r G_l(x)\, x^{-l-3/2}\, \mathrm{d}x, \\ \rho_l(r) &= \mathcal{M}_s^{-1} \left\{ (s + l + 1/2)\, g_l(s) \right\} (r) = \frac{-r^{-l-2}}{4\pi G} \frac{d}{dr} \left[ r^{l+1/2}\, G_l(r) \right]. \end{aligned} \tag{2.98}$$

Note that we could have chosen the opposite signs for the prefactors in Eq. (2.94), which would have led us to the expressions

$$\begin{aligned} \Phi_l(r) &= r^{-l-1} \int_0^r G_l(x)\, x^{l-1/2}\, \mathrm{d}x, \\ \rho_l(r) &= \frac{-r^{l-1}}{4\pi G} \frac{d}{dr} \left[ r^{-l-1/2}\, G_l(r) \right]. \end{aligned} \tag{2.99}$$

Like Kalnajs (1971), we obtain two different integral expressions for the potential; but ours are simply equivalent to the first and second terms in the standard multipole expression (1.31). Lastly, we note that the inner product expression (2.95) takes the form of a Mellin convolution, and so we can write (given a suitable normalisation),

$$\langle \Phi_l, \Phi'_l \rangle = \int_0^\infty G_l(r)\, G'_l(r)\, \frac{\mathrm{d}r}{r}. \tag{2.100}$$

This is an attractively symmetric form, and as such may be suited to finding basis sets. However, if we wish to prescribe a particular functional form for the potential (for example when systematically searching for possible basis sets), we would have to allow for a change of variables. This would introduce a further degree of freedom, and the resulting expressions would be equivalent to Eq. (2.19).

### 2.3.3 Confluent hypergeometric functions

We now introduce a generalisation of the Hankel transform method that is apparently new, involving confluent hypergeometric functions ${}_1F_1$. This is not a *true* integral transform method, as the potential and density use different kernel functions, and we do not derive an inversion formula. However, it will nonetheless prove useful in the construction of basis sets, as we shall see later.

Given some auxiliary function $f(t)$ we can construct a valid potential-density pair according to

$$\Phi_l(r) = \frac{r^l}{\Gamma(\mu)} \int_0^\infty \mathrm{d}t\, t^{-1} f(t)\, {}_1F_1\left(\begin{matrix}\mu+\nu \\ \mu+1\end{matrix}\middle| -r^{1/\alpha}\, t\right), \tag{2.101}$$

$$\rho_l(r) = \frac{r^{l-2+1/\alpha}}{\Gamma(\mu+1)} \int_0^\infty \mathrm{d}t\, f(t)\, {}_1F_1\left(\begin{matrix}\mu+\nu+1 \\ \mu+1\end{matrix}\middle| -r^{1/\alpha}\, t\right).$$

As with the Hankel transform method of Sec. 2.3.1, we have used the definition $\mu \equiv \alpha(1+2l)$. This new transform method involves two ${}_1F_1$ functions with slightly different parameters, and is interesting because it naturally introduces the free parameter $\nu$ in addition to the parameter $\alpha$.

We found these expressions during the derivation of a two-parameter family of basis sets (see Sec. 5.2.1). A generalisation of the auxiliary function $f(t)$ suggested a possible path to a three-parameter family of basis sets, which we discuss in Sec. 5.6.

The limitation of this approach is that the inner product between the functions $\Phi_l$ and $\rho_l$ does not reduce to a one-dimensional integral (unlike with the method of Sec. 2.3.1), as there is no orthogonality relation between the two ${}_1F_1$ functions. Instead, we have the double integral

$$\int_0^\infty r^2\, \mathrm{d}r\, \Phi_l(r)\, \rho_l(r) = \frac{\alpha\mu\Gamma(\mu+2\nu)}{\Gamma(\mu+\nu)\,\Gamma(\mu+\nu+1)} \int_0^\infty \mathrm{d}t \int_0^\infty \mathrm{d}s\, \frac{(ts)^{\nu+1}\, f(t)\, f(s)}{(t+s)^{\mu+2\nu}}, \tag{2.102}$$

which is derived by applying Kummer's transformation (DLMF, §13.2.39) to $\Phi_l$ and $\rho_l$, then using the result of Saad & Hall (2003) to evaluate the integral over $r$, followed by the application of a few Appell function reductions (HTF, 1955, Ch. 5.10). By taking the limit $\nu \to \infty$ while keeping $tz^2/\nu$ constant, the ${}_1F_1$ functions in Eq. (2.101) reduce to Bessel functions $J_\mu(x)$, showing that the Hankel transform is a limiting form of this integral transform[1].

## 2.4 Kelvin transform

There is a further degree of freedom that can be obtained immediately, which arises from classical potential theory: the *Kelvin transform*. Let $\Phi_l^A$ and $\rho_l^A$ be a potential-density multipole pair (thus

[1] This property is what links the derivations of Sec. 6.1 and Sec. 6.2.

satisfying Poisson's equation in the form of Eq. (1.28)). Then the Kelvin transform of this pair is also a valid multipole pair, given by

$$\Phi_l^B(r) \equiv r^{2-d}\,\Phi_l^A\left(\frac{1}{r}\right), \qquad (2.103)$$
$$\rho_l^B(r) \equiv r^{-2-d}\,\rho_l^A\left(\frac{1}{r}\right).$$

This is clear upon substituting $r = 1/x$,

$$\begin{aligned}\left[\nabla_r^2 - \frac{l(l+d-2)}{r^2}\right]\Phi_l^B(r) &= r^{-2-d}\left[\left(\Phi_l^A\right)''\left(r^{-1}\right) + (d-1)r\left(\Phi_l^A\right)'\left(r^{-1}\right)\right. \\ &\quad \left. - \left[l(l+d-2)r^2\Phi_l^A\left(r^{-1}\right)\right]\right. \\ &= x^{d+2}\left[\left(\Phi_l^A\right)''(x) + \frac{d-1}{x}\left(\Phi_l^A\right)'(x) - \frac{l(l+d-2)}{x^2}\Phi_l^A(x)\right] \\ &= 4\pi G\, x^{d+2}\,\rho_l^A(x) \\ &= 4\pi G\,\rho_l^B(r).\end{aligned} \qquad (2.104)$$

Furthermore, their total self-energy is the same (in absolute value),

$$\begin{aligned}\int x^{d-1}\,\mathrm{d}x\;\Phi_l^B(x)\rho_l^B(x) &= \int x^{-1-d}\,\mathrm{d}x\;\Phi_l^A\left(x^{-1}\right)\rho_l^A\left(x^{-1}\right) \\ &= -\int r^{d-1}\,\mathrm{d}r\;\Phi_l^A(r)\rho_l^A(r).\end{aligned} \qquad (2.105)$$

A more general expression for the Kelvin transform in fact holds for arbitrary potential-density pairs, not just those expressible as multipole terms, but we omit the proof here.

Specialising to $d = 3$ dimensions, we find that we have a useful technique for generating new basis sets from existing ones. The preservation of self-energy (2.105) ensures that no further orthogonalisation is needed after the applying the transform, up to a change of sign in the normalisation constant. In Ch. 5 we use the Kelvin transform to immediately derive our 'Family B' basis sets from the original 'Family A' derivation.

It is also worth noting that all the classic basis sets in the literature (Clutton-Brock, 1972, 1973, Hernquist & Ostriker, 1992, Zhao, 1996) are *invariant* under the Kelvin transform, and this may be related to their having simple functional forms. In fact, the general double-power law model contains a two-parameter subset that is invariant under the Kelvin transform (this is clear from the density expression Eq. (1.18)).

## 2.5 Concluding remarks

We begun the chapter with a comprehensive review (Sec. 2.1) of the existing results on analytical biorthogonal basis sets. We then attempted (Sec. 2.2) to take the most obvious next step towards the generalisation of Clutton-Brock's original derivation in spherical polar coordinates. We found that new results are likely few and far between, depending entirely on the availability of suitable

Sturm-Liouville equations which can be mapped onto our heuristic form of Poisson's equation (Eq. 2.22).

We therefore turned to other methods, in particular the use of integral transforms, which we reviewed (Sec. 2.3). We identified the Hankel transform (Sec. 2.3.1) as holding the most promise, an idea which we develop in Ch. 4. We also noted (Sec. 2.4) that the Kelvin transform can be used to generate a new basis set from any existing one; we put this result to good use in Ch. 5.

# Chapter 3

# The Super-NFW model

Analytical galaxy profiles provide simplicity for the modelling of galaxy components as well as insight into the dynamics of more realistic galaxies. For instance, the NFW model (Eq. 1.15) is often used to model dark haloes; and the Hernquist model (Eq. 1.17) is often used to represent bulges and elliptical galaxies, as it follows the de Vaucouleurs (1953) profile to a good approximation.

The asymptotic fall-off of the density in the Hernquist model is $\rho_{\rm Hq} \sim r^{-4}$, so it avoids the defect of infinite mass which afflicts the NFW halo with $\rho_{\rm NFW} \sim r^{-3}$. In this chapter, along similar lines, we introduce the *super-NFW* (sNFW) model, which has density and potential pair:

$$\begin{aligned} \psi_{\rm sNFW}(r) &= \frac{GM}{r+a+\sqrt{a}\sqrt{r+a}}, \\ \rho_{\rm sNFW}(r) &= \frac{3M\sqrt{a}}{16\pi}\frac{1}{r(r+a)^{5/2}}, \end{aligned} \tag{3.1}$$

where $a$ is related to the effective radius by $R_{\rm e} = 5.478a$.

Why is it 'super-NFW'? The model provide a good match to cosmological haloes, but it has finite mass, as the density falls off more slowly as $\rho_{\rm sNFW} \sim r^{-3.5}$. This is slower than the Hernquist model, but faster than the NFW. Recent work on the splashback radius (Diemer et al., 2017) suggests that the density of cosmological haloes drops more rapidly than NFW, but slower than Hernquist, at large radii. There are of course other models in the literature that have $\rho \propto r^{-1}$ at the centre and have an asymptotic fall off with logarithmic gradient between $-3$ and $-4$ (see An & Evans, 2006, An & Zhao, 2013). The super-NFW model however has another special property – it is the zeroth-order term of a biorthogonal expansion (see Ch. 4). The potential-density pairs of arbitrarily distorted sNFW models are, therefore, straightforward to construct. This is important as cosmological dark haloes show many deviations from spherical symmetry, and therefore a spherical model is only a crude approximation.

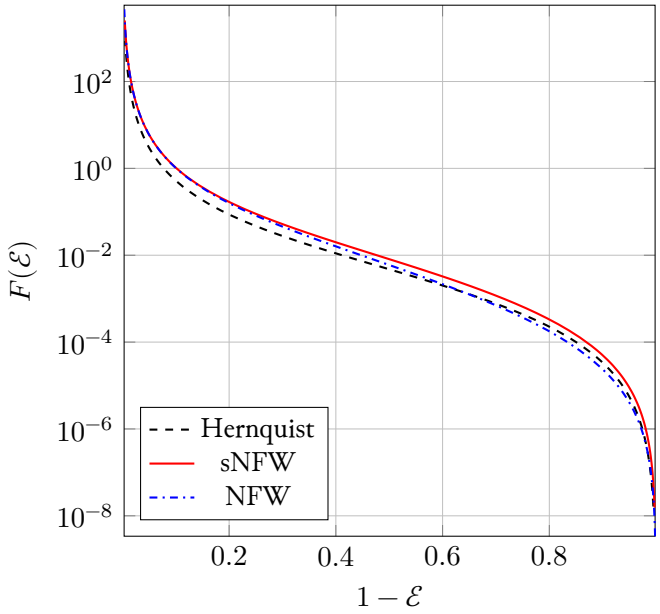


Figure 3.1: The isotropic DFs $F(\mathcal{E})$ of the sNFW model (solid red) compared to the DFs of the Hernquist model (dashed black) and NFW model (dash-dotted blue). The three models have the same central value of the potential and the same halo scalelength $r_s$.

The sNFW model is part of the general double-power law family investigated by Zhao (1996) and subsequently studied in detail by An & Zhao (2013), namely

$$\rho(r) = \frac{C}{r^{\gamma}(1+r^{1/\alpha})^{(\beta-\gamma)\alpha}}, \tag{3.2}$$

where $C$ is a normalisation constant. In Zhao's notation, the sNFW model has $(\alpha, \beta, \gamma) = (1, 7/2, 1)$. Zhao was the first to note that the potential is simple. The sNFW model is also the $b = 7/2$ member of the generalised NFW family

$$\rho(r) = \frac{C}{r(1+r)^{b-1}}. \tag{3.3}$$

These models were studied in Evans & An (2006), who give asymptotic results for the dynamical quantities for the whole family. Before the later developments of Ch. 5, biorthogonal basis sets for this entire family of models were not known.

## 3.1 The model

### 3.1.1 Isotropic distribution functions

Let us use units in which $G = a = 1$ and $M = 2$, so that the central value of the sNFW potential is $\psi_{\rm sNFW}(0) = 1$. The enclosed mass is

$$M(r) = 2 - \frac{2+3r}{(1+r)^{3/2}}, \tag{3.4}$$

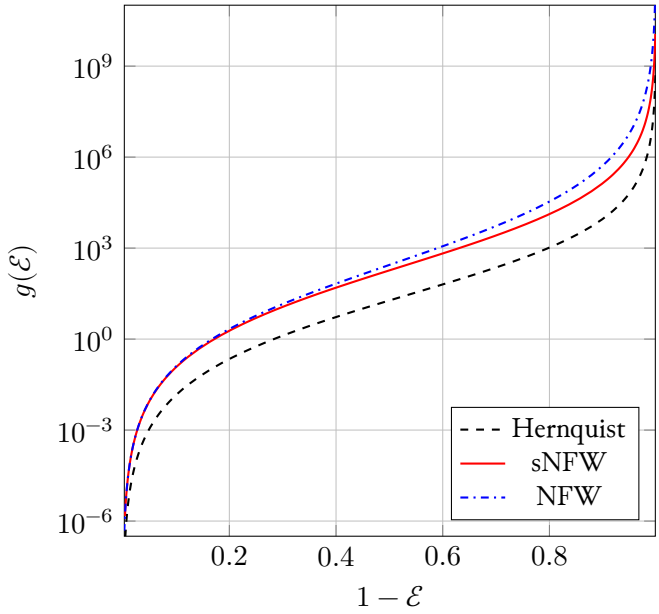


Figure 3.2: The density of states $g(\mathcal{E})$ of the super-NFW model (solid red) compared to the Hernquist (dashed black) and NFW (dash-dotted blue) models.

so that the half-mass radius is $r_{1/2} = x + x^* + 2 \approx 7.29086$, where $x = 1/2\left(37 + \mathrm{i}\sqrt{3}\right)^{1/3}$.

The potential $\psi(r)$ can be inverted simply by setting $x = \sqrt{1+r}$ and solving the resulting quadratic in $x$. The form of $r(\psi)$ is then

$$r(\psi) = \frac{4 - \psi - \sqrt{\psi(8+\psi)}}{2\psi}, \tag{3.5}$$

so that the density $\rho$ can be expressed as

$$\rho(\psi) = \frac{3\psi^{7/2}\left(4 - \psi + \sqrt{\psi(8+\psi)}\right)\left(\sqrt{\psi} + \sqrt{8+\psi}\right)^5}{2^{16}\pi(1-\psi)}. \tag{3.6}$$

The isotropic distribution function (DF) is then given by Eddington (1916) as

$$F(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2}\frac{d}{d\mathcal{E}}\left[\int_0^{\mathcal{E}} \frac{d\rho}{d\psi}\frac{\mathrm{d}\psi}{\sqrt{\mathcal{E}-\psi}}\right]. \tag{3.7}$$

This can be evaluated exactly as

$$\begin{aligned} F(\mathcal{E}) &= \frac{3}{7\cdot 2^{10}\pi^3(8+\mathcal{E})(1-\mathcal{E})^2}\left[252\tfrac{8+\mathcal{E}}{\sqrt{2(1-\mathcal{E})}}\sin^{-1}\left(\sqrt{\mathcal{E}}\right)\right. \\ &+ P_1(\mathcal{E})\sqrt{\tfrac{\mathcal{E}}{2}} + P_2(\mathcal{E})\,E\left(-\tfrac{\mathcal{E}}{8}\right) + P_3(\mathcal{E})\,K\left(-\tfrac{\mathcal{E}}{8}\right) \\ &+ \left. 189(8+\mathcal{E})\,\Pi\left(\mathcal{E}\,\middle|\,-\tfrac{\mathcal{E}}{8}\right)\right], \end{aligned} \tag{3.8}$$

where $E$, $K$ and $\Pi$ are complete elliptic integrals and the $P_i$ are the following polynomials:

$$\begin{aligned} P_1(x) &= -4(32x^6+416x^5+1200x^4-920x^3-2198x^2+399x+504), \\ P_2(x) &= -8(32x^6+352x^5+656x^4-1176x^3-586x^2+173x+360), \\ P_3(x) &= (x+8)(128x^5+512x^4-576x^3-480x^2+56x+171). \end{aligned}$$

The distribution function (3.8) is more complicated than the isotropic DF of the Hernquist model, but simpler than the isotropic DF of the NFW model, which has been numerically constructed (Widrow, 2000, Lokas & Mamon, 2001) and subsequently analytically derived (Evans & An, 2006).

In Fig. 3.1, we show the isotropic DF of this model against that of the Hernquist and NFW models. To compare all three models, we use the *halo* or *isothermal* scalelength $r_{\rm s}$, which is defined as the radius at which the logarithmic slope of the density attains the isothermal value, that is

$$\left.\frac{d\log\rho}{d\log r}\right|_{r=r_{\rm s}} = -2. \tag{3.9}$$

For the Hernquist model we find $r_{\rm s} = b/2$; for sNFW $r_{\rm s} = 2a/3$, and for NFW the initial choice of $r_{\rm s}$ already satisfies this property. As the form of the cusp is the same at small radii ($\rho \sim 1/r$), so the DFs of all three models diverge like $(1-\mathcal{E})^{-5/2}$ as $\mathcal{E} \to 1$. However, for stars close to the binding energy ($\mathcal{E} \to 0$), the Hernquist DF behaves like $\mathcal{E}^{5/2}$, the sNFW like $\mathcal{E}^2$ and the NFW like $\mathcal{E}^{3/2}$. The density of states is

$$g(\mathcal{E}) = (4\pi)^2 \int_0^{r_\mathcal{E}} r^2 \sqrt{2(\mathcal{E}-\psi)}\,{\rm d}r, \tag{3.10}$$

where $r_\mathcal{E}$ is the maximum radius of orbit with energy $\mathcal{E}$ (Binney & Tremaine, 1987, §4.3.1). After some calculation, we obtain

$$g(\mathcal{E}) = \frac{4\sqrt{2}\pi^2}{3\mathcal{E}^2}\left\{12\cos^{-1}\left(\left(\sqrt{\mathcal{E}}\right)\right)\frac{1+\mathcal{E}}{\sqrt{\mathcal{E}}} + 12(1-\mathcal{E})^{3/2}(2\mathcal{E}+3)\right. \tag{3.11}$$

$$\left. + \sqrt{8+\mathcal{E}}\left[(2\mathcal{E}^2+7\mathcal{E}-16)\,E(\phi\,|\,\kappa^2) + \mathcal{E}(1-2\mathcal{E})\,F(\phi\,|\,\kappa^2)\right]\right\}, \tag{3.12}$$

where $F$ and $E$ are incomplete elliptic integrals[a], with argument $\phi \equiv \tan^{-1}\left(\sqrt{(1-\mathcal{E})/\mathcal{E}}\right)$ and modulus $\kappa^2 \equiv 8/(8+\mathcal{E})$. The density of states for the sNFW model is compared to that of the Hernquist and NFW models in Fig. 3.2.

The isotropic velocity dispersion is

$$\begin{aligned} \langle v_r^2\rangle &= \frac{1}{6r(1+r)}\Big[6-9r-176r^2-406r^3-350r^4-105r^5 \\ &- \sqrt{1+r}\,(6-12r-88r^2-120r^3-48r^4) \\ &+ 3r^2(1+r)^{7/2}\left(35\,{\rm csch}^{-1}(\sqrt{r}) - 16\log(1+1/r)\right). \end{aligned} \tag{3.13}$$

[a]Note that we we use the *Mathematica* convention for the arguments of the elliptic functions, so that $E(\phi\,|\,m) \equiv \int_0^\phi {\rm d}\theta\,\sqrt{1-m\sin^2\theta}$.

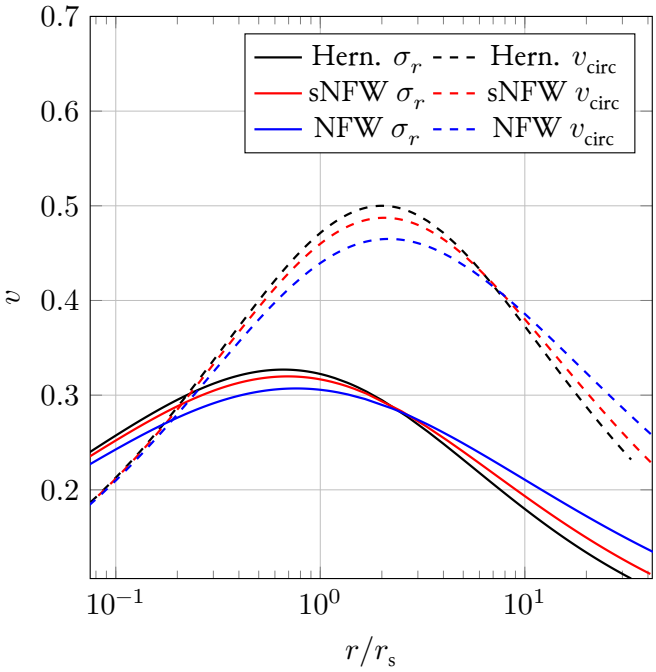


Figure 3.3: The velocity dispersion (full) and rotation curve (dotted) of the sNFW model (red) compared to the Hernquist (black) and NFW (blue) models plotted against radius (in units of $r_{\rm s}$). Note the peaks of the rotation curve and velocity dispersion of the models are comparable, but the decline for the Hernquist model takes place more quickly than for the sNFW and NFW models.

The circular velocity curve (or rotation curve) is

$$v_{\rm circ}^2 = \frac{2(1+r)^{3/2} - 3r - 2}{r(1+r)^{3/2}}. \tag{3.14}$$

In Fig. 3.3, the velocity dispersion profile and the circular velocity curve are shown for this model, as well as for the Hernquist and NFW models. Both the velocity dispersion and the rotation curve of the sNFW model have the desirable feature that they fall off much more slowly than for the Hernquist model. This is useful in modelling elliptical galaxies. For example, Gerhard et al. (2001) found that the circular velocity profiles of giant ellipticals are flat to within 10 to 20 per cent between $0.2R_{\rm e}$ to at least $2R_{\rm e}$, independent of luminosity.

### 3.1.2 Anisotropic distribution functions

Analyses of kinematic data suggest that most elliptical galaxies are close to isotropic (Gerhard et al., 2001). Anisotropy is usually parameterised via

$$\beta = 1 - \frac{\langle v_\theta^2 \rangle + \langle v_\phi^2 \rangle}{2\langle v_r^2 \rangle}, \tag{3.15}$$

where angled brackets denote averages over the DF. Gerhard et al. (2001) find that $-0.5 \lesssim \beta \lesssim 0.5$ in their study of giant ellipticals. Mild radial anisotropy is most common, though some tangential

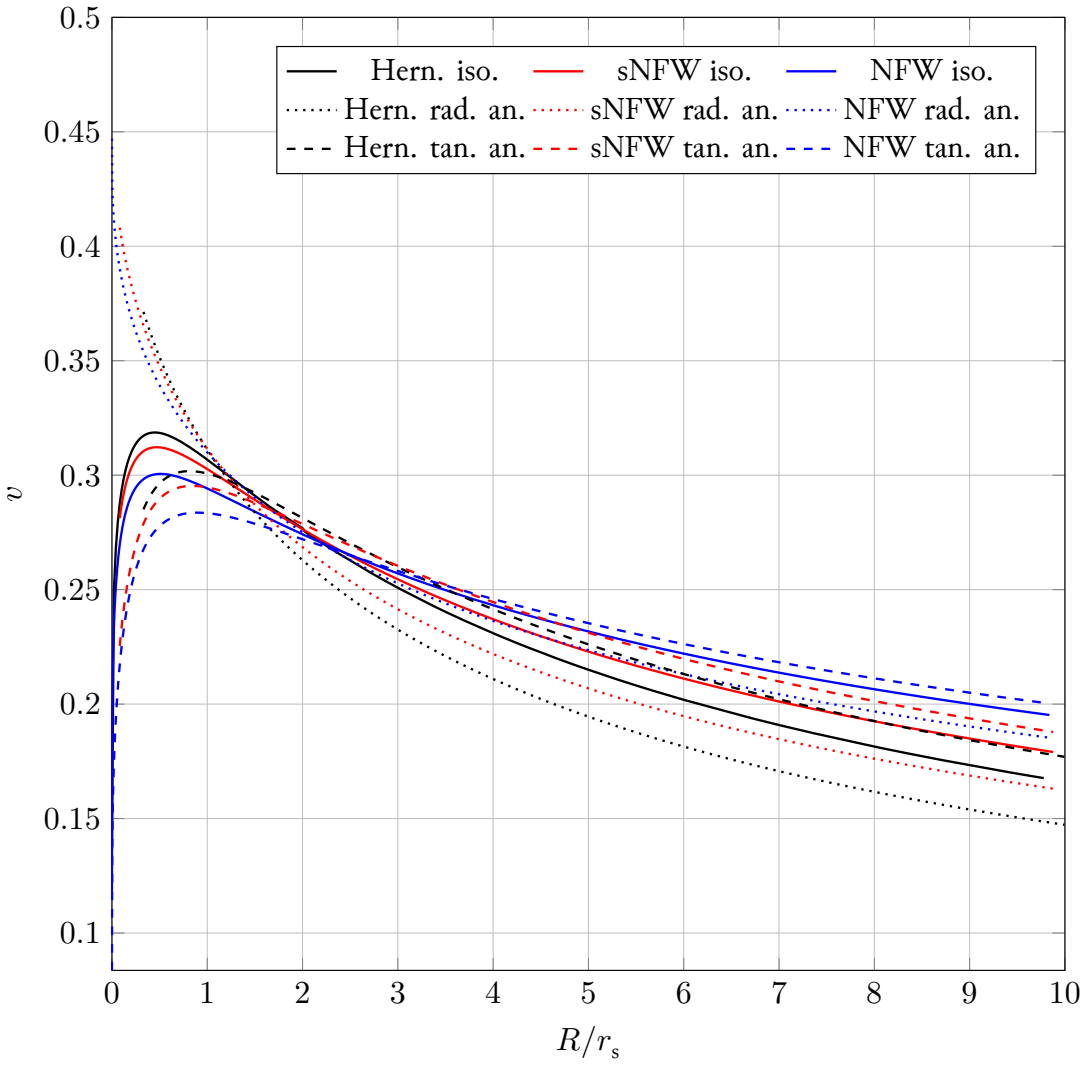


Figure 3.4: The line of sight velocity dispersions for the sNFW, Hernquist and NFW models plotted against projected distance (in units of $r_s$). Full lines are the isotropic model, dotted radially anisotropic ($\beta = \frac{1}{2}$) and dashed tangentially anisotropic ($\beta = -\frac{1}{2}$).

anisotropic elliptical galaxies are known. We now develop two models that bracket the range of relevant anisotropies.

The DF of a spherical system with constant anisotropy is

$$F(\mathcal{E}, L) = L^{-2\beta} f_E(\mathcal{E}). \tag{3.16}$$

As first shown by Cuddeford (1991), the unknown function $f_E(\mathcal{E})$ can be recovered from an integral inversion formula (see also Wilkinson & Evans, 1999, Evans & An, 2006),

$$f_E(\mathcal{E}) = \frac{2^\beta (2\pi)^{-3/2}}{\Gamma(1-\lambda)\Gamma(1-\beta)} \frac{d}{d\mathcal{E}} \int_0^{\mathcal{E}} \frac{\mathrm{d}\psi}{(\mathcal{E}-\psi)^\lambda} \frac{d^n h}{d\psi^n}, \tag{3.17}$$

where $h = r^{2\beta}\rho$ is expressed as a function of $\psi$, and $n = \lfloor(3/2 - \beta)\rfloor$ and $\lambda = 3/2 - \beta - n$ are the integer floor and the fractional part of $3/2 - \beta$. This includes the Eddington (1916) formula for the isotropic DF as a special case ($\beta = 0$).

For the radially anisotropic model when $\beta = 1/2$, the expression for the DF reduces to

$$F(\mathcal{E}, L) = \frac{1}{2\pi^2}\frac{1}{L}\left.\frac{dh}{d\psi}\right|_{\psi=\mathcal{E}} = \frac{f_E(\mathcal{E})}{L}, \tag{3.18}$$

which for our model becomes

$$f_E(\mathcal{E}) = \frac{15\,\mathcal{E}^{3/2}\left(\sqrt{\mathcal{E}} + \sqrt{\mathcal{E}+8}\right)^6}{2^{14}\,\pi\sqrt{\mathcal{E}+8}}. \tag{3.19}$$

The radial velocity dispersion is

$$\begin{aligned}\langle v_r^2\rangle = \frac{1}{6r\sqrt{r+1}} \times \Bigg[ & 30r^3\left(2\sqrt{r+1} - 1\right) + 5r^2\left(28\sqrt{r+1} - 15\right) \\ & + r\left(92\sqrt{r+1} - 55\right) + 12\left(\sqrt{r+1} - 1\right) \\ & - 30r(r+1)^3\left(\log\left(\frac{r}{r+1}\right) + 2\,\mathrm{csch}^{-1}(\sqrt{r})\right)\Bigg].\end{aligned} \tag{3.20}$$

The analogous radially anisotropic ($\beta = 1/2$) DFs for the Hernquist model is very simple and was derived by Baes & Dejonghe (2002) and Evans & An (2005).

For the tangentially anisotropic model when $\beta = -1/2$, the expression for the DF further reduces to

$$F(\mathcal{E}, L) = \frac{L}{2\pi^2}\left.\frac{d^2h}{d\psi^2}\right|_{\psi=\mathcal{E}} = L f_E(\mathcal{E}), \tag{3.21}$$

where

$$\begin{aligned}f_E(\mathcal{E}) = & \frac{3\mathcal{E}^{7/2}\left(\sqrt{\mathcal{E}} + \sqrt{\mathcal{E}+8}\right)^4\left(4 - \mathcal{E} + \sqrt{\mathcal{E}(\mathcal{E}+8)}\right)^2}{2^{20}\pi^3(\mathcal{E}-1)^3(\mathcal{E}+8)} \\ & \times\left[\sqrt{\mathcal{E}+8}\,(5\mathcal{E}^2 + 19\mathcal{E} - 36) + \sqrt{\mathcal{E}}\,(5\mathcal{E}^2 + 31\mathcal{E} - 72)\right]\end{aligned} \tag{3.22}$$

whilst the second moment is

$$\begin{aligned}\langle v_r^2\rangle = \frac{1}{12r\sqrt{r+1}} \times \Bigg[ & 15r^5\left(21\sqrt{r+1} - 8\right) + 15r^4\left(49\sqrt{r+1} - 20\right) \\ & + r^3\left(483\sqrt{r+1} - 220\right) + 5r^2\left(\sqrt{r+1} - 6\right) - 2r\left(5\sqrt{r+1} - 3\right) \\ & + 8\left(\sqrt{r+1} - 1\right) - 15r^3(r+1)^3\left(8\log\left(\frac{r}{r+1}\right) + 21\,\mathrm{csch}^{-1}(\sqrt{r})\right)\Bigg].\end{aligned} \tag{3.23}$$

Fig. 3.4 shows the line of sight velocity dispersions for the sNFW model for the three choices of anisotropy ($\beta = \frac{1}{2}, 0$ and $-\frac{1}{2}$). The equivalent results for the Hernquist and NFW models are

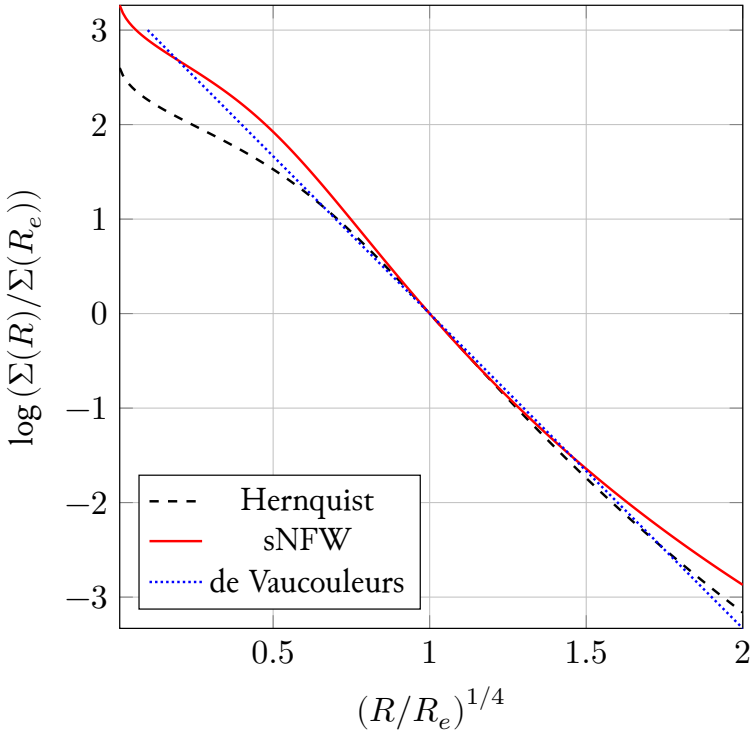


Figure 3.5: The surface density of the sNFW model (red), the Hernquist model (black) and the de Vaucouleurs profile (dashed).

also shown. As expected, radial anisotropy leads to an enhancement of the line of sight dispersion near the centre (where the dotted curves lie above the full curves). Tangential anisotropy causes the line of sight dispersions to be enhanced above the isotropic case in the outer parts (where the dashed curves lie above full curve). Note that the line of sight dispersion profiles of the sNFW model show a more gradual decline with distance than the Hernquist model. This is in good accord with the data on nearly round elliptical galaxies, which show slow declines out to approximately $2R_e$ (Kronawitter et al., 2000).

Overall, the sNFW model gives somewhat more complicated expressions for quantities such as the DFs than the Hernquist model. The pay-back is that the density profile falls off more slowly ($\rho_{\rm sNFW} \sim r^{-3.5}$) and so the rotation curve and velocity dispersion profiles are much flatter. This is much more like the observed behaviour of elliptical galaxies and dark matter haloes.

## 3.2 Comparisons

### 3.2.1 Sersic and de Vaucouleurs Profile

Traditionally, the light profiles of elliptical galaxies and bulges have been fit by the de Vaucouleurs (1953) profile, which is

$$\log_{10}\left(\frac{\Sigma(R)}{\Sigma(R_{\rm e})}\right) = -3.331\left[\left(\frac{R}{R_{\rm e}}\right)^{1/4} - 1\right]. \tag{3.24}$$

Here, $R_{\rm e}$ is the effective radius, or the radius of the isophote that encloses half the luminosity. Caon et al. (1993) examined the photometric profiles of a large sample of elliptical galaxies and argued for use of the slightly more general Sersic (1968) law

$$\log_{10}\left(\frac{\Sigma(R)}{\Sigma(R_{\rm e})}\right) \approx 0.8686(0.1635 - n)\left[\left(\frac{R}{R_{\rm e}}\right)^{1/n} - 1\right], \tag{3.25}$$

where the Sersic index $n$ generally lies between $2 \lesssim n \lesssim 10$. The photometric profiles of bulges were studied by Andredakis et al. (1995), who found that bulges of S0s are well-fit by a de Vaucouleurs profile, whilst bulges of late-type spirals are better fit by an exponential.

To compare the sNFW model against these photometric laws, we must first compute its projected density. This is

$$\Sigma(R) = \frac{(R+1)(R+3)\,K\left(\frac{R-1}{2R}\right) - 8R\,E\left(\frac{R-1}{2R}\right)}{2^{3/2}\pi\sqrt{R}\,(R^2-1)^2}. \tag{3.26}$$

The half-light or effective radius is $R_{\rm e} = 1.81527$. Assuming a constant mass-to-light ratio, Fig. 3.5 shows the surface brightness of the de Vaucouleurs profile, together with the Hernquist and sNFW models between $0.1$ and $2R_{\rm e}$. The sNFW model is a better global fit than the Hernquist model to the de Vaucouleurs profile. Beyond about $5R_{\rm e}$, however, the sNFW surface density falls off rather more slowly than both the Hernquist and de Vaucouleurs profiles.

More formally, we can fit the projected densities of both the Hernquist and sNFW models between $0.1$ and $2R_{\rm e}$ to the Sersic profiles. The Hernquist model gives $n = 3.388$ and the sNFW $n = 4.200$. Given the range of properties of elliptical galaxies and bulges, both profiles are useful. The Hernquist model is a better match to $n \approx 3$ Sersic profiles, whilst the sNFW is a better match to $n \approx 4$ (i.e. the de Vaucouleurs profile).

### 3.2.2 Numerical halo fitting

Dark matter haloes are often parameterised in terms of their virial mass $M_{\rm v}$, virial radius $r_{\rm v}$ and concentration $c$ (Diemer & Kravtsov, 2015). The virial mass is the mass contained within a spherical shell of radius $r_{\rm v}$, that is $M_{\rm v} = M(r_{\rm v})$. Once a particular model is chosen to fit the halo, the length scale is parameterised using the concentration parameter $c = r_{\rm v}/r_{\rm s}$, where $r_{\rm s}$ is the halo scalelength as defined in Eq. (3.9). For the NFW model, in the notation of equation (1.15), we

have

$$\rho_{\rm NFW} = \frac{M_{\rm v}}{4\pi}\left(\log(1+c_{\rm NFW}) - \frac{c_{\rm NFW}}{1+c_{\rm NFW}}\right)^{-1} \frac{1}{r\left(r+r_{\rm v}/c_{\rm NFW}\right)^2}. \tag{3.27}$$

The analogous definition for the sNFW model is

$$\rho_{\rm sNFW} = \frac{3^{3/2}M_{\rm v}}{8\pi}\left(1 - \frac{1+c_{\rm sNFW}}{(1+2c_{\rm sNFW}/3)^{3/2}}\right)^{-1} \frac{(r_{\rm v}/c_{\rm sNFW})^{1/2}}{r\left(2r+3r_{\rm v}/c_{\rm sNFW}\right)^{5/2}}. \tag{3.28}$$

We have fit both the NFW and sNFW models to ten numerically-constructed dark matter haloes extracted from cosmological simulations (for more details on their provenance, see Ch. 4). Four of these fits are shown in Fig. 3.6, and the relation between the derived concentration parameters for the two models, along with a best fit line, is shown in Fig. 3.7. We notice that the sNFW model does at least as good a job as the NFW profile in fitting the shapes of these ten haloes. The concentration $c_{\rm sNFW}$ of the best-fit sNFW model is related to that of the best-fit NFW model via

$$c_{\rm sNFW} = 1.36 + 0.76 c_{\rm NFW}. \tag{3.29}$$

This gives an easy way to transform the mass-concentration relations for NFW models, given for example in Dutton & Macciò (2014, Eq. 8 & 9), to provide a cosmologically-inspired sequence of sNFW dark haloes.

## 3.3 Conclusions

We have introduced the super-NFW (sNFW) model. This is a potential-density pair useful for representing spherical bulges, elliptical galaxies and dark haloes. The density of the sNFW model falls like $\rho_{\rm sNFW} \sim r^{-1}$ near the centre, and like $\rho_{\rm sNFW} \sim r^{-3.5}$ on the outskirts. This gives it two important advantages over competitors such as the Hernquist model: first, it is a better match to the de Vaucouleurs profile in the inner parts, so it is useful for modelling the light profiles of elliptical galaxies and bulges. Its density falls off somewhat more slowly than the Hernquist model ($\rho_{\rm Hq} \sim r^{-4}$), which makes it a better match to the line of sight velocity dispersion profiles of ellipticals at large radii. Secondly, the asymptotic density fall-off is closer to the density profile of numerically-constructed haloes, which approximately follow the Navarro-Frenk-White (NFW) form and have $\rho_{\rm NFW} \sim r^{-3}$ at large radii. The advantage of using a sNFW model rather than an NFW model is that we have provided a suite of DFs (isotropic, radially and tangentially anisotropic) for the sNFW model, whereas the DFs of the NFW model are not elementary (Evans & An, 2006).

There are of course many other models with central density cusps like $\rho \sim r^{-1}$ and with asymptotic decays between $\rho \sim r^{-3}$ and $\rho \sim r^{-4}$ (e.g. An & Evans, 2006) or halo models with finite mass (e.g. Navarro et al., 2004, Zhao & Silk, 2005). In particular, An & Zhao (2013) provide a compendium of properties of spherical double-power-law models, some of which can also provide equally good matches to the density profiles of dark haloes. However, these models do not readily generalise to arbitrary distorted geometries. We show in Ch. 4 that the sNFW model has another remarkable property: it can be used to form a new biorthogonal basis function expansion in a

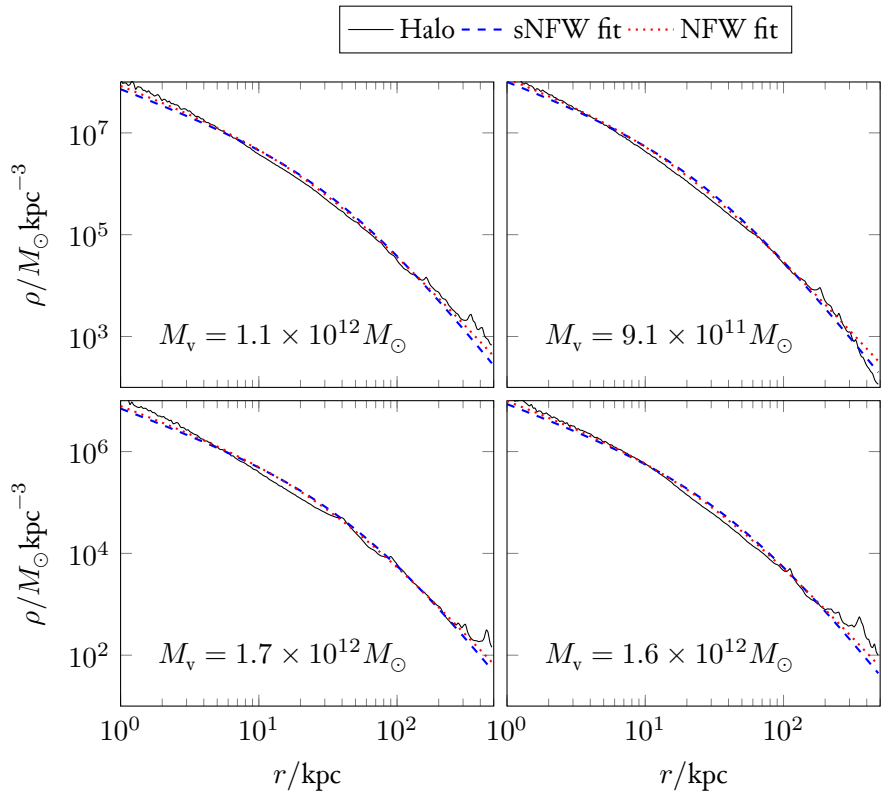


Figure 3.6: Fits of the sNFW and NFW models to the radial density profile of four dark matter haloes extracted from a cosmological $N$-body simulation. The numerical halo data is binned logarithmically, and the virial mass of each halo is inlaid.

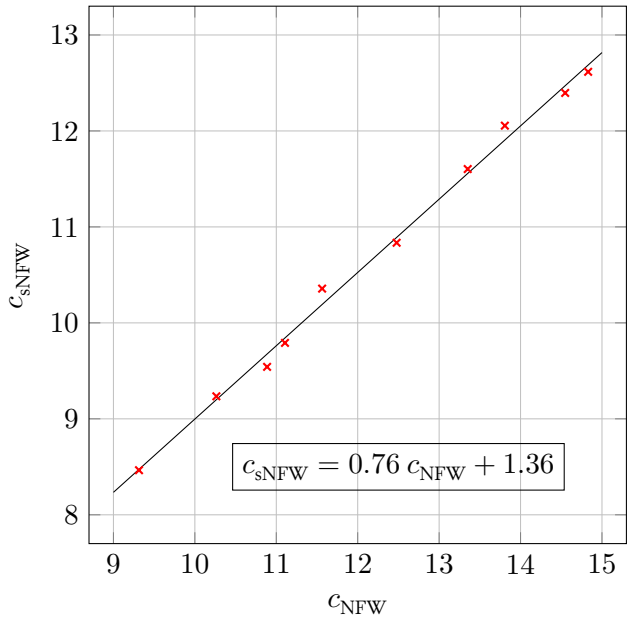


Figure 3.7: Concentrations for 10 numerical haloes as fit by the sNFW and NFW models. The line of best fit is shown.

manner analogous to that discovered by Hernquist & Ostriker (1992) for the Hernquist model itself. Amongst other advantages, this means that the model can be easily extended to flattened, triaxial and lopsided geometries.

The intrinsic properties of our model (such as the DFs and velocity dispersions) are more complicated than the Hernquist model, but less complicated than the NFW model. We conclude that the sNFW model provides an excellent trade-off between simplicity and realism in modelling dark haloes and elliptical galaxies.

Chapter 4

# The Super-NFW family of basis sets

Hernquist & Ostriker (1992) have clearly articulated the advantages of basis function expansions for which the zeroth-order model actually resembles a galaxy. They used the clever method pioneered by Clutton-Brock (1973) to transform the Laplacian into an equation whose eigenfunctions are already known. They tellingly remarked that "for reasons not immediately obvious to us, Clutton-Brock's approach has been virtually ignored in the literature". In the Hernquist & Ostriker (1992) basis function expansion, the zeroth-order model is a Hernquist (1990) density profile, which resembles an elliptical galaxy or dark matter halo with $\rho \sim 1/r$ at small radii and $\rho \sim 1/r^4$ at large radii.

In this chapter we broaden the basis function technique by introducing a new family of biorthogonal basis functions, and discuss its application. Recently, Rahmati & Jalali (2009) explicitly extended the derivation of Clutton-Brock (1972), Aoki & Iye (1978) to the spherical case, making an analogous choice of auxiliary function $g_n(k)$ and using a Hankel transform – hence deriving a basis set whose lowest order density is the perfect sphere (de Zeeuw, 1985). Our work is a generalisation of this result, adding a free parameter to form a whole family of basis sets (similar to how Zhao (1996) generalised the result of Clutton-Brock (1973)). Rahmati & Jalali's basis set corresponds to the lowest member of our new family of biorthogonal pairs. We term the new family of basis sets the 'super-NFW family' as another member of the family matches, at zeroth order, the 'super-NFW' (sNFW) model introduced in Ch. 3.

The mathematical results of this chapter are in fact subsumed by a later development – the more general family of basis sets described in Ch. 5 – but the simpler method here still serves as a useful pedagogical example motivating the generating function approach to the development of biorthogonal basis functions, which was first suggested by Aoki & Iye (1978) in the disk setting.

We begin in Sec. 4.1 by describing the range of zeroth-order models covered by the new family of basis sets. We then show how to derive analytical expressions for the new family in Sec. 4.2.

The numerical implementation is then discussed in Sec. 4.3, and we continue by expanding a series of $N$-body cosmological haloes that are imitations of the Milky Way's dark halo using the new expansions.

## 4.1 Zeroth-order models

We begin by describing some of the zeroth-order models in our sequence of basis function expansions, many of which resemble realistic galactic components.

The models are labelled by a parameter $\alpha$. However, those with $0 < \alpha < 1/2$ have vanishing density at the origin, and so we do not consider them. The behaviour of the density at small and large radii is

$$\lim_{r\to 0} \rho_{000}(r) \sim r^{-2+1/\alpha}, \qquad \lim_{r\to\infty} \rho_{000}(r) \sim r^{-3-1/(2\alpha)}, \tag{4.1}$$

so that the $\alpha = 1/2$ model is cored, whilst the remainder are cusped. The density of all the models falls off with a logarithmic slope between $-3$ and $-4$. We list some of the zeroth-order models that are obtained when $\alpha$ is integer or half-integer, as in these cases the potential reduces to elementary functions; in general real values of $\alpha$ are permitted, but the potential must be evaluated using the incomplete beta function.

1) When $\alpha = 1/2$, the zeroth-order model is the *perfect sphere* of de Zeeuw (1985), which has density and potential

$$\rho_{000} = \frac{1}{\sqrt{2}\,\pi^{3/2}}\,\frac{1}{(1+r^2)^2}, \qquad \Phi_{000} = -\sqrt{\frac{2}{\pi}}\frac{\arctan r}{r}. \tag{4.2}$$

   This is the spherical limit of the perfect ellipsoid, which provides triaxial densities that are close to the luminosity profiles of elliptical galaxies. There is a large literature on these models as the potential is separable or Stäckel, and so the orbits and action-angles can be found as quadratures (e.g., Binney & Tremaine, 1987, Bertin, 2014). Notice that the density has a harmonic core at the centre but falls like $\rho \sim r^{-4}$ at large radii.

2) When $\alpha = 1$, we obtain the sNFW model, whose properties are discussed in Ch. 3 (see also Evans & An, 2006, An & Zhao, 2013). This has a central density cusp with $\rho \propto r^{-1}$ similar to the Hernquist (1990) or NFW models. Asymptotically, the density falls like $\rho \propto r^{-3.5}$ at large radii, which is somewhat faster than the NFW model, but has the distinct advantage of finite total mass. The potential-density pair of the sNFW model is

$$\rho_{000} = \frac{3}{16\pi}\,\frac{1}{r\,(1+r)^{5/2}}, \qquad \Phi_{000} = \frac{-1}{1+r+\sqrt{1+r}}. \tag{4.3}$$

3) More strongly cusped models can also be obtained as zeroth-order models. When $\alpha = 3/2$, we have

$$\rho_{000} = \frac{2\sqrt{2}}{9\pi^{3/2}}\,\frac{1}{r^{4/3}\left(1+r^{2/3}\right)^3}, \quad \Phi_{000} = \sqrt{\frac{2}{\pi}}\left[\frac{1}{r^{2/3}(1+r^{2/3})} - \frac{\arctan(r^{1/3})}{r}\right]. \tag{4.4}$$

The central density cusp is now $\rho \sim r^{-4/3}$, similar to the inner regions of the massive cosmological clusters studied by Diemand et al. (2005).

4) When $\alpha = 2$, we have

$$\rho_{000} = \frac{15}{64\pi} \frac{1}{r^{3/2}\,(1+r^{1/2})^{7/2}}, \qquad \Phi_{000} = -\frac{1+2\sqrt{1+\sqrt{r}}}{\left(1+\sqrt{1+\sqrt{r}}\right)^2 (1+\sqrt{r})^{3/2}}. \tag{4.5}$$

This has an inner density profile $\rho \sim r^{-3/2}$, similar to the very steepest cusps found in cosmological simulations (Moore et al., 1998).

Explicitly, the lowest-order basis functions are

$$\begin{aligned} \rho_{000} &= \frac{2^{\alpha+1}\Gamma(\alpha+3/2)}{\sqrt{\pi}} \frac{1}{r^{2-1/\alpha}(1+r^{1/\alpha})^{\alpha+3/2}}, \\ \Phi_{000} &= -\frac{2^{\alpha-1}\Gamma(\alpha+3/2)}{\sqrt{\pi}} \frac{\mathcal{B}_\chi(\alpha,1/2)}{r}, \qquad \chi \equiv \frac{r^{1/\alpha}}{1+r^{1/\alpha}}, \end{aligned} \tag{4.6}$$

where $\mathcal{B}_x(a,b)$ is the incomplete beta function. The full suite of models has a range of inner density profiles suitable for representing dark haloes with cores and weak or strong cusps, and $\alpha$ can be tuned to match the behaviour of a given halo (see Sec. 4.3.3).

## 4.2 Construction of the new basis set

Using the theoretical framework described in Sec. 2.3.1 for manufacturing biorthogonal basis functions from Hankel transforms, we now show how to construct the associated higher-order basis functions. We introduce a two parameter family of auxiliary functions $g_n(k)$. We derive an expression for the potential basis functions that depends on several free parameters, but find that the requirements of physicality reduce this set to a single parameter family. This reduction of complexity allows us to express the potential and density basis functions in a more succinct fashion in Sections 4.2.1 and 4.2.2.

Inspired by Rahmati & Jalali (2009)[a], and initially following their calculations, we now choose

$$g_n(k) \equiv k^{(\eta-1)/2}\, \mathrm{e}^{-k}\, L_n^{(\eta)}(2k)\,, \tag{4.7}$$

where $L_n^{(\eta)}(x)$ are the generalised Laguerre polynomials (DLMF, §18.3), and $\eta$ is (for now) a free parameter. The Laguerre polynomials are orthogonal on $(0,\infty)$ with respect to a weight function $\omega(x) \equiv x^\eta\ \exp(-x)$,

$$\int_0^\infty \mathrm{d}x\, L_n^{(\eta)}(x)\, L_{n'}^{(\eta)}(x)\, \omega(x) = \delta_{nn'} \frac{\Gamma(n+\eta+1)}{\Gamma(n+1)}\,. \tag{4.8}$$

Our $g_n(k)$ consist of a Laguerre polynomial multiplied by a factor of $\sqrt{\omega(2k)/k}$, hence ensuring the orthogonality on $(0,\infty)$ with respect to $k\,\mathrm{d}k$, as required by the argument in Sec. 2.3.1

[a] Whose choice of $g_n(k)$ was in turn motivated by that of Clutton-Brock (1972).

(in particular, Eq. (2.83)). We can obtain an expression for $\Phi_{nl}$ using the integral BMP (1954, Eq. 8.6(6)),

$$\int_0^\infty \mathrm{d}u\, \mathrm{e}^{-us} u^\nu J_\mu(u) = \frac{\Gamma(\nu+\mu+1)}{(1+s^2)^{(1+\nu)/2}}\, P_\nu^{(-\mu)}\left(\frac{s}{\sqrt{1+s^2}}\right). \tag{4.9}$$

Setting $u = kr^{1/(2\alpha)}$ and $s = r^{-1/(2\alpha)}$, and using the explicit polynomial representation of the Laguerre polynomials

$$L_n^{(\eta)}(x) = \sum_{j=0}^{n} (-1)^j \frac{1}{j!} \binom{n+\eta}{n-j} x^j, \tag{4.10}$$

we find

$$\Phi_{nl} = \sum_{j=0}^{n} A_{njl\eta} \left[r\left(1+r^{1/\alpha}\right)^{(\eta+1)/2+j}\right]^{-1/2} P^{(-\mu)}_{\frac{\eta-1}{2}+j}\left(\frac{1}{\sqrt{1+r^{1/\alpha}}}\right), \tag{4.11}$$

$$A_{njl\eta} \equiv \frac{(-2)^j}{j!}\binom{n+\eta}{n-j}\Gamma\left(j+\frac{\eta+1}{2}+\mu\right).$$

We use the results of Sec. 1.4.3 on asymptotic limits to fix the parameter $\eta$. The asymptotic behaviour of the associated Legendre function (DLMF, §14.8(i)) as $z \to 1$ from above is $P_\nu^{(\mu)}(z) \sim (1-z)^{-\mu/2}$. Therefore, as $r \to 0$, every term in $\Phi_{nl}$ goes as

$$\Phi_{nl} \sim r^{-1/2}\left(1 - \frac{1}{\sqrt{1+r^{1/\alpha}}}\right)^{\mu/2} \propto r^{\mu/(2\alpha)-1/2} = r^l, \tag{4.12}$$

so the first limit implied by Eq. (1.50) is already satisfied. On the other hand, as $r \to \infty$, $1/\sqrt{1+r^{1/\alpha}} \sim r^{-1/(2\alpha)}$, so the associated Legendre function goes to a constant (DLMF, §14.5.1) and we are left with the prefactor

$$\Phi_{nl} \sim \left[r\left(r^{1/\alpha}\right)^{(\eta+1)/2+j}\right]^{-1/2} \to r^{-1/2-(\eta+1)/(4\alpha)}, \tag{4.13}$$

where we have used the fact that the $j = 0$ term dominates the sum as $r \to \infty$. By Eq. (1.50), we require this limit to be $r^{-l-1}$, which we can achieve by setting $\eta = 2\mu - 1$, so the $g_n$-normalisation constant is

$$I_n = \frac{\Gamma(n+2\mu)}{2^{2\mu}\, n!}. \tag{4.14}$$

Now that $\eta$ is no longer a free parameter, we can obtain the basis functions in their most convenient form.

### 4.2.1 Potential basis functions

The sum (4.11) is unsatisfactory for several reasons: 1) it is numerically unstable for high $n$, 2) $n$ and $j$ are coupled in such a way that calculating $n$ basis functions requires $\mathcal{O}(n^2)$ operations , and 3) associated Legendre functions of non-integer or negative degree or order are rarely implemented numerically. Therefore, departing from Rahmati & Jalali (2009), we seek a superior expression, which can be obtained using the recurrence relation for the Laguerre polynomials,

$$n\, L_n^{(\alpha)}(x) = (n+\alpha)\, L_{n-1}^{(\alpha)}(x) - x\, L_{n-1}^{(\alpha+1)}(x). \tag{4.15}$$

Using Eq. (4.7), we can immediately write down

$$n\, g_n(k) = (n + 2\mu - 1)\, g_{n-1}(k) - 2\exp(-k)\, k^\mu\, L_{n-1}^{(2\mu)}(2k)\,. \tag{4.16}$$

So we obtain the following recurrence relation for the basis functions

$$n\Phi_{nl} = (n+2\mu-1)\Phi_{n-1,l} + \frac{2A_{nlm}}{\sqrt{r}} \int_0^\infty k^\mu L_{n-1}^{(2\mu)}(2k) \exp(-k) J_\mu(kz)\, \mathrm{d}k, \qquad \left(z \equiv r^{1/(2\alpha)}\right). \tag{4.17}$$

To evaluate the latter integral is some work. First, we note that a generating function for the Laguerre polynomials is

$$\sum_{n=0}^{\infty} t^n\, L_n^{(\lambda)}(k) = \frac{\exp\left(-tk/(1-t)\right)}{(1-t)^{\lambda+1}}, \tag{4.18}$$

so that, using the following Hankel transform (G&R, Eq. 6.623(1))

$$\int_0^\infty x^\nu\, \mathrm{e}^{-ax}\, J_\nu(xy)\, \mathrm{d}x = \frac{2^\nu\, \Gamma(\nu + 1/2)}{\sqrt{\pi}}\, \frac{y^\nu}{(a^2 + y^2)^{\nu+1/2}}, \tag{4.19}$$

as well as the generating function for the Gegenbauer polynomials $C_n^{(\lambda)}(\xi)$,

$$\sum_{n=0}^{\infty} t^n C_n^{(\lambda)}(\xi) = \frac{1}{(1 - 2\xi t + t^2)^\lambda} = \frac{(1+z^2)^\lambda}{[(1+t)^2 + (1-t)^2 z^2]^\lambda}, \qquad \xi \equiv \frac{z^2-1}{z^2+1} \equiv \frac{r^{1/\alpha} - 1}{r^{1/\alpha} + 1}, \tag{4.20}$$

we can find the Hankel transform of the remaining term in the recurrence relation,

$$\begin{aligned}
\sum_{n=0}^{\infty} t^n \int_0^\infty k^\mu L_n^{(2\mu)}(2k)\, \mathrm{e}^{-k}\, J_\mu(kz)\, \mathrm{d}k &= \int_0^\infty k^\mu \frac{\exp(-k(1+t)/(1-t))}{(1-t)^{2\mu+1}} J_\mu(kz)\, \mathrm{d}k \\
&= \frac{2^\mu\, \Gamma(\mu + 1/2)}{\sqrt{\pi}}\, \frac{z^\mu}{[(1+t)^2 + (1-t)^2 z^2]^{\mu+1/2}} \\
&= \sum_{n=0}^{\infty} t^n\, \frac{2^\mu\, \Gamma(\mu + 1/2)}{\sqrt{\pi}}\, \frac{z^\mu}{(1+z^2)^{\mu+1/2}}\, C_n^{(\mu+1/2)}(\xi).
\end{aligned} \tag{4.21}$$

The recurrence relation becomes

$$n\, \Phi_{nl} = (n + 2\mu - 1)\, \Phi_{n-1,l} + A_{nlm} \frac{2^{\mu+1}\, \Gamma(\mu + 1/2)}{\sqrt{\pi}}\, \frac{r^l\, C_{n-1}^{(\mu+1/2)}(\xi)}{(1 + r^{1/\alpha})^{\mu+1/2}}. \tag{4.22}$$

This means that the $(n+1)$-th basis function can be trivially calculated from the $n$-th by simply adding on a single extra term. Evaluating $n$ basis functions therefore requires $\mathcal{O}(n)$ steps (although see Sec. 4.3.1).

So far we have stated without proof the $n = 0, l = 0$ case. We now obtain an explicit expression for all the potential basis functions, from which the zeroth order ($n = 0$, $l \geq 0$) can be extracted.

We begin by writing our auxiliary function in a form that splits off the constant term in the polynomial from the remaining terms,

$$g_n(k) = k^{\mu-1}\,\mathrm{e}^{-k}\left(L_n^{(2\mu-1)}(0) - 2k\sum_{j=0}^{n-1}\frac{\binom{n+2\mu-1}{n-1-j}}{(n-j)\binom{n}{j}}L_j^{(2\mu)}(2k)\right), \tag{4.23}$$

where again $\mu = \alpha\,(1+2l)$. Aside from the constant term, the Hankel transforms can be done using Eq. (4.21). To evaluate the Hankel transform of the constant term, we use Eq. (4.9) to find

$$\int_0^\infty k^{\mu-1}\,\mathrm{e}^{-k}\,J_\mu(kz)\;\mathrm{d}k = \frac{\Gamma(2\mu)}{(1+z^2)^{\mu/2}}\,P_{\mu-1}^{(-\mu)}\left(\frac{1}{\sqrt{1+z^2}}\right) = \frac{\Gamma(2\mu)}{\Gamma(\mu)\;2^\mu\;z^\mu}\,\mathcal{B}_\chi(\mu,1/2)\,, \tag{4.24}$$

where we have used an alternative hypergeometric representation of the Legendre function $P_{\mu-1}^{(-\mu)}(x)$ (DLMF, §14.3.17) to rewrite it as an incomplete Beta function. So, reassembling the terms, we obtain an expression for the potential as

$$\Phi_{nl}(r) = -A_{nlm}\frac{2^\mu\,\Gamma(\mu+1/2)\;(2\mu)_n}{\sqrt{\pi}\,n!}\left[\frac{\mathcal{B}_\chi(\mu,1/2)}{2\,r^{1+l}} - \frac{2\,r^l}{(1+r^{1/\alpha})^{\mu+1/2}}\sum_{j=0}^{n-1}\frac{j!\,C_j^{(\mu+1/2)}(\xi)}{(2\mu)_{j+1}}\right], \tag{4.25}$$

with $\mu \equiv \alpha\,(1+2l)$, $\chi \equiv \frac{1+\xi}{2} \equiv r^{1/\alpha}/(1+r^{1/\alpha})$ and $(a)_n \equiv \Gamma(a+n)\,/\Gamma(a)$ (the rising factorial, or Pochhammer symbol). The only special functions required to evaluate the potential are the incomplete beta function and the Gegenbauer polynomials, which are standard library functions in many numerical software packages. For example, both are included in the GNU Scientific Library (Galassi, 2003). Using the results above, we see that it is easy to write down and compute a basis set for all real values of $\alpha$. In particular we note that there is no $n$-dependence inside the summation over the Gegenbauer polynomials, which enables the $n$-th potential basis function to be evaluated recursively from the $(n-1)$-th, without having to recompute every term in the sum (the recurrence relation (4.22) is stating essentially the same thing). Along with the introduction of the free parameter $\alpha$, this is the key improvement over the result of Rahmati & Jalali (2009).

### 4.2.2 Density basis functions

A similar calculation can be performed to find expressions for the density basis functions, though it is simpler as the generating function can be applied immediately with no further manipulation of $g_n(k)$. First, we note the Hankel transform BMP (1954, Eq. 8.6(5)),

$$\int_0^\infty x^{\nu+1}\,\exp(-ax)\,J_\nu(xy)\;\mathrm{d}x = \frac{2^{\nu+1}\,\Gamma(\nu+3/2)}{\sqrt{\pi}}\,\frac{a\,y^\nu}{(a^2+y^2)^{\nu+3/2}}. \tag{4.26}$$

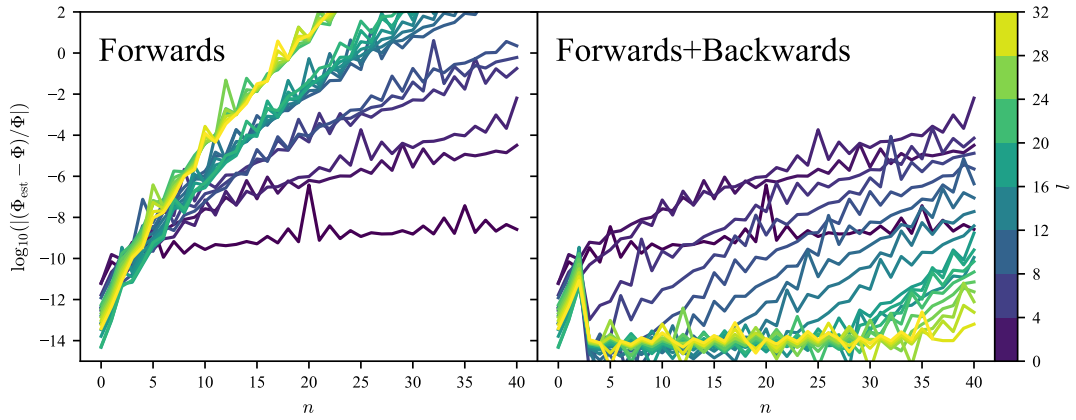


Figure 4.1: Relative error in potential basis function calculation using two methods: on the left the result of using a forward recursion scheme and on the right using a combination of forward and backward recursion. The relative error is the difference between the potential computed by the basis expansion and an arbitrary precision calculation from *Mathematica* computed at $r/r_s = 1.3$ using $\alpha = 1.2$. Note that the floor on the error in the right panel is set by the precision of our *Mathematica* calculation and can be lowered if so desired. We only perform the 'forwards+backwards' procedure for $l > 4$.

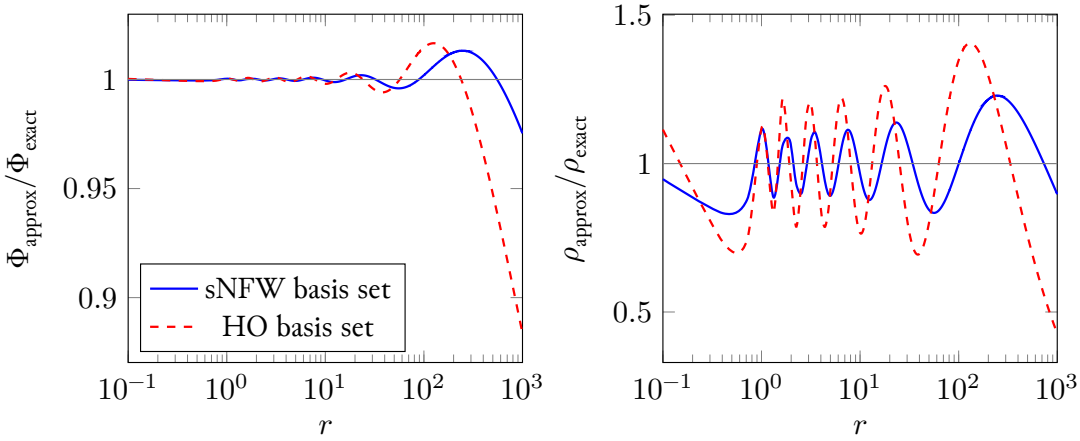


Figure 4.2: Reconstruction of a spherical NFW potential (left) and density (right) with the new super-NFW basis set (blue) and the HO basis set (red). The distance is given in units of the NFW scalelength. Both expansions use radial terms up to $n = 20$ and no angular terms ($l = 0$). Both expansions oscillate around the true value, which is represented by the horizontal grey line.

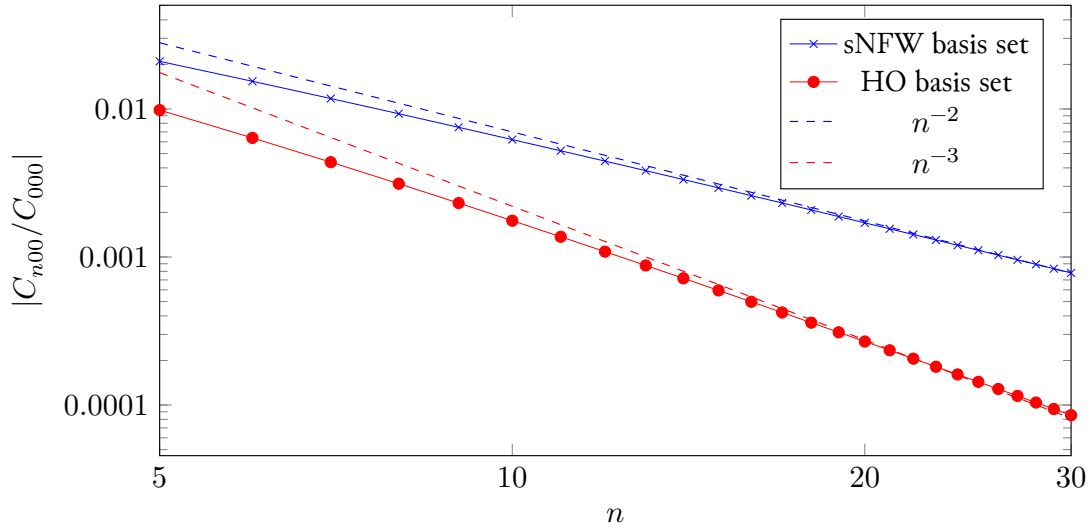


Figure 4.3: The run of the radial expansion coefficients with $n$ for a spherical NFW model using the new basis set (blue) and the HO basis set (red). Also plotted as dashed lines are the $n^{-2}$ and $n^{-3}$ curves, suggesting that the coefficients fall off asymptotically like $n^{-2}$ in our case and like $n^{-3}$ for the HO case.

Then, using (4.18) and (4.20) as above, we have

$$\begin{aligned}\sum_{n=0}^{\infty} t^n\,\rho_{nl} &\propto r^{1/\alpha-5/2}\sum_{n=0}^{\infty} t^n\int_0^{\infty} k^{\mu+1}\,L_n^{(2\mu-1)}(2k)\,\mathrm{e}^{-k}\,J_\mu(kz)\;\mathrm{d}k \qquad (4.27)\\ &= \frac{r^{1/\alpha-5/2}}{(1-t)^{2\mu}}\int_0^{\infty} k^{\mu+1}\,\mathrm{e}^{-k(1+t)/(1-t)}\,J_\mu(kz)\;\mathrm{d}k\\ &= \frac{2^{\mu+1}\,\Gamma(\mu+3/2)}{\sqrt{\pi}}\,r^{1/\alpha-2+l}\,\frac{(1+t)(1-t)^2}{[(1+t)^2+(1-t)^2z^2]^{\mu+3/2}}\\ &= \frac{2^{\mu+1}\,\Gamma(\mu+3/2)}{\sqrt{\pi}}\,\frac{r^{1/\alpha-2+l}}{(1+r^{1/\alpha})^{3/2+\mu}}\\ &\quad\times\sum_{n=0}^{\infty} t^n\,\left[C_n^{(\mu+3/2)}(\xi)-C_{n-1}^{(\mu+3/2)}(\xi)-C_{n-2}^{(\mu+3/2)}(\xi)+C_{n-3}^{(\mu+3/2)}(\xi)\right].\end{aligned}$$

The last line can be simplified by adding together two Gegenbauer recursion relations (G&R, Eq. 8.933(2), 8.933(3)), resulting in

$$\rho_{nl}(r) = A_{nlm}\,\frac{2^{\mu-3}\Gamma(\mu+1/2)}{\pi^{3/2}\alpha^2}\,\frac{r^{1/\alpha-2+l}}{(1+r^{1/\alpha})^{3/2+\mu}}\left[(n+\mu+1/2)\;C_n^{(\mu+1/2)}(\xi)\right.$$
$$\left.-\,(n+\mu-1/2)\;C_{n-1}^{(\mu+1/2)}(\xi)\right]. \qquad (4.28)$$

Just as for the potential basis functions, evaluating $n$ density basis functions requires only $\mathcal{O}(n)$ steps. It is now straightforward to verify that the basis functions satisfy Eq. (1.36). If we set $A_{nlm} = 1$ the overall normalisation constant for the basis functions written in Eqs. (4.25) and (4.28) is

$$N_{nl} = \frac{\Gamma(2\mu+n)}{2^{2\mu}\,n!\,2\alpha}. \qquad (4.29)$$

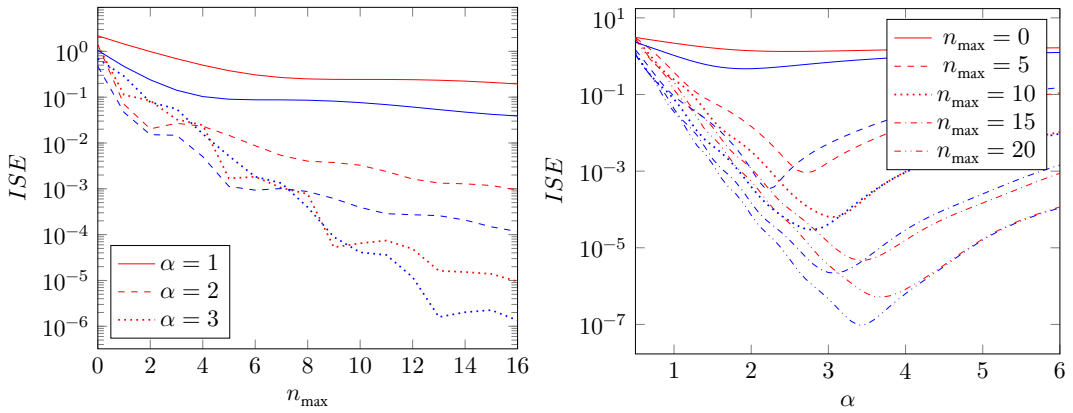


Figure 4.4: Left: The integrated squared error as defined in Eq. (4.43) between the exact NFW profile and the profile reconstructed with radial basis functions up to order $n_{\max}$ for different values of $\alpha$ in Zhao's basis set (red lines) and the new basis set in this chapter (blue lines). Right: The integrated squared error between the exact NFW profile and the profile reconstructed with radial basis functions, against the parameter $\alpha$ that characterises each basis set for different numbers of radial basis functions. Again red lines use Zhao's basis set and blue lines use the new basis set.

For numerical purposes, it may be desirable to redefine $A_{nlm}$ to incorporate more of the $n$ and $l$-dependent prefactors in Eqs. (4.25) and (4.28) (see Sec. 4.3.1).

### 4.2.3 Comparison with existing basis sets

We note immediately the close similarity in form between Zhao's basis set (see Sec. 2.1.2 and in particular Eq. (2.15) for the full expressions) and our new potential and density basis functions in Eq. (4.25) and Eq. (4.28), where the basis functions are also expressed in terms of the Gegenbauer polynomials. Comparing the lowest-order density of the Zhao basis set with that of our new basis set,

$$\rho_{00} \propto \frac{1}{r^{2-1/\alpha}\left(1+r^{1/\alpha}\right)^{\alpha+3/2}}, \qquad \rho_{00}^{\text{Zhao}} \propto \frac{1}{r^{2-1/\alpha}\left(1+r^{1/\alpha}\right)^{\alpha+2}}. \tag{4.30}$$

Evidently, the only difference is the shallower outer slope in the former case. This is significant as popular models for dark matter haloes tend to have outer slopes closer to $r^{-3}$. For example, the generalised NFW profile (Navarro et al., 2004) has $\rho \propto r^{-\gamma}(1+r)^{\gamma-3}$, with values for the inner slope $\gamma$ ranging from 0.7 to 1.5, and outer slope fixed to $r^{-3}$.

In both our and Zhao's expansion families, when the inner slope is fixed ($\gamma \equiv 2 - 1/\alpha$) the asymptotic outer slope is then constrained. To give some examples, if we set $\gamma = 0.7$, then Zhao's outer slope is $r^{-4.3}$ whereas ours is $r^{-3.65}$; and if $\gamma = 1.5$ then Zhao's is $r^{-3.5}$ and ours is $r^{-3.25}$. We therefore expect that our basis functions will be more efficient than Zhao's for describing typical dark matter haloes.

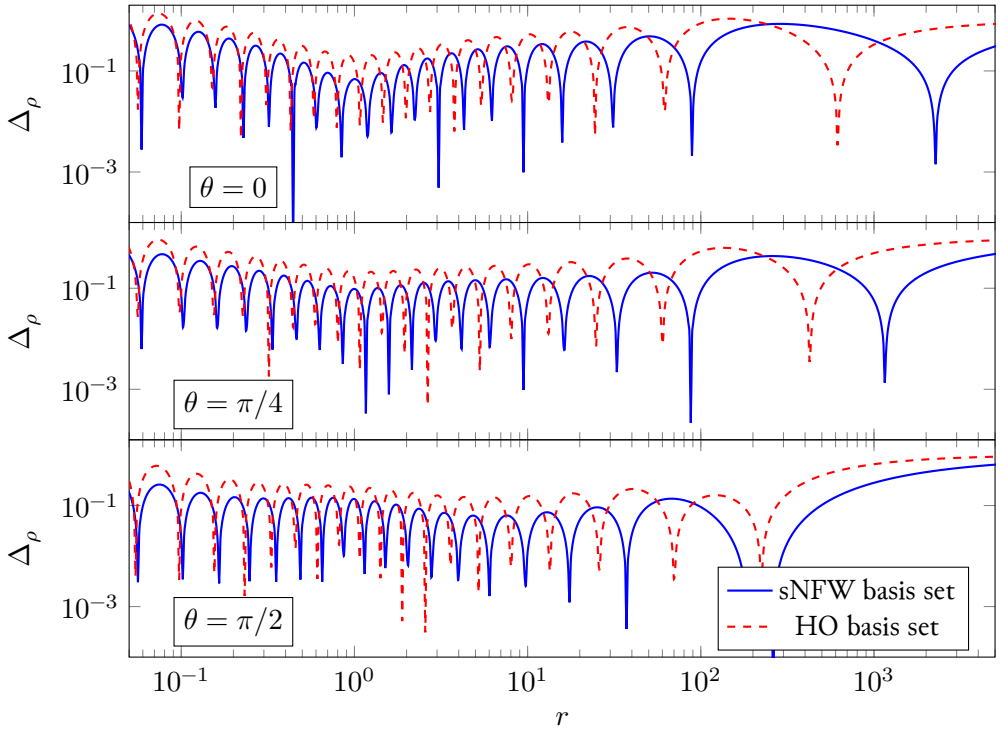


Figure 4.5: Expansion of a flattened ($q = 0.8$) NFW density. Viewing angle $\theta$ is shown on each plot. Both expansions use radial terms up to $n = 20$ and angular terms up to $l = 12$. The error measure is $\Delta_\rho \equiv \log\left|1 - \rho_{\text{approx}}/\rho_{\text{exact}}\right|$ (lower is better; the dips are due to the oscillations around the true value).

## 4.3 Numerical performance of the expansion

We turn to the application of our new basis expansion to the representation of cosmological haloes. Before inspecting both analytic and numerical haloes in Sections 4.3.2 and 4.3.3, we present a formulation of the basis expansion that is computationally friendly.

### 4.3.1 Numerical implementation

It is very important to take advantage of a number of recursion relations for purposes of speed. It is also more efficient to factor out the constant parts of the potential-density pairs such that only the spatially-dependent pieces are calculated for each particle. We therefore write hatted versions of all the quantities involved in the expansion, corresponding to a new normalisation that is numerically

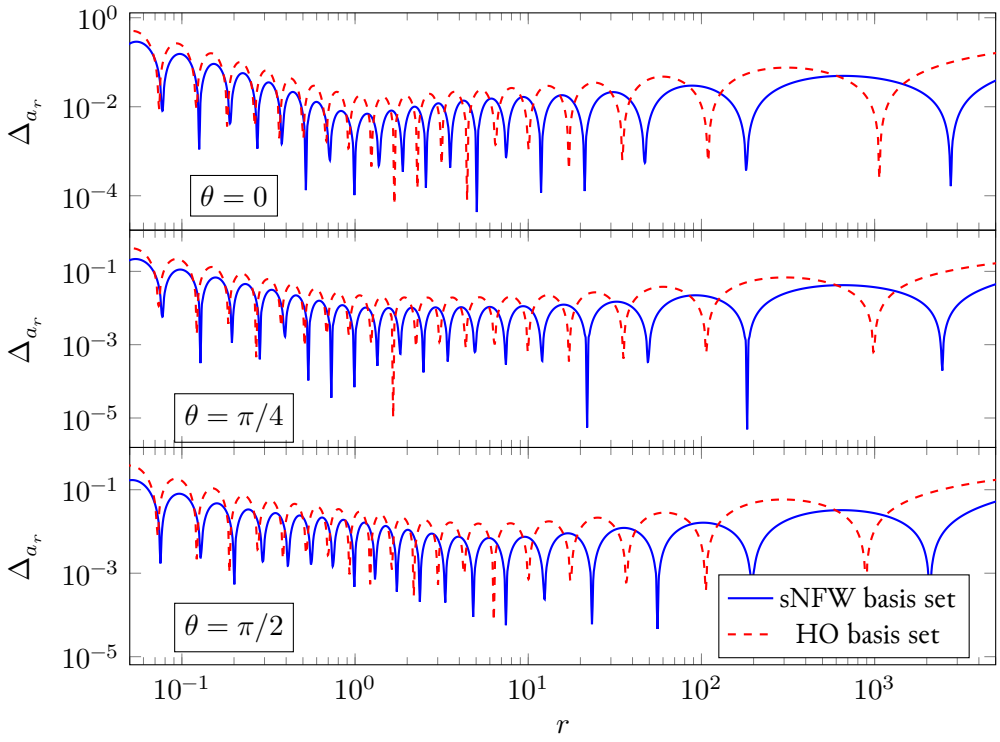


Figure 4.6: Expansion of the radial acceleration in a flattened ($q = 0.8$) NFW potential. Viewing angle $\theta$ is shown on each plot. Both expansions use radial terms up to $n = 20$ and angular terms up to $l = 12$. The error measure is $\Delta_{a_r} \equiv \log\left|1 - a_{r\,\mathrm{approx}}/a_{r\,\mathrm{exact}}\right|$.

more convenient. The basis functions themselves are modified as

$$\hat{\rho}_{nlm}(\mathbf{r}) = \hat{\rho}_{nl}(r)\, Y_{lm}(\theta,\phi), \tag{4.31}$$

$$\hat{\rho}_{nl}(r) = \frac{r^{1/\alpha-2+l}}{(1+r^{1/\alpha})^{\mu+3/2}}\Big((n+\mu+\tfrac{1}{2})C_n^{(\mu+1/2)}(\xi) - (n+\mu-\tfrac{1}{2})C_{n-1}^{(\mu+1/2)}(\xi)\Big),$$

$$\hat{\Phi}_{nlm}(\mathbf{r}) = \hat{\Phi}_{nl}(r)\, Y_{lm}(\theta,\phi),$$

$$\hat{\Phi}_{nl}(r) = \frac{\mathcal{B}_\chi(\mu,1/2)}{2\,r^{1+l}} - \frac{2r^l}{(z^2+1)^{\mu+1/2}}\sum_{j=0}^{n-1}\frac{j!\Gamma(2\mu)}{\Gamma(2\mu+j+1)}\,C_j^{(\mu+1/2)}(\xi),$$

so that Poisson's equation becomes $\nabla^2\hat{\Phi}_{nlm} = 4\pi K_{nl}\hat{\rho}_{nlm}$ with

$$K_{nl} = -\frac{n!\Gamma(2\mu)}{8\pi\alpha^2\Gamma(2\mu+n)}, \tag{4.32}$$

and the normalisation constant

$$\hat{N}_{nlm} = \int \mathrm{d}^3\mathbf{r}\,\hat{\Phi}_{nlm}(\mathbf{r})\,\hat{\rho}_{nlm}(\mathbf{r}) = \frac{\alpha\sqrt{\pi}\,\Gamma(\mu)}{2^{1+2\mu}\,\Gamma\left(\frac{1}{2}+\mu\right)}. \tag{4.33}$$

The Gegenbauer polynomials can be constructed recursively using the following relation (G&R, Eq. 8.933(1))

$$n\, C_n^{(\beta)}(\xi) = 2(\beta + n - 1)\, \xi\, C_{n-1}^{(\beta)}(\xi) - (2\beta + n - 2)\, C_{n-2}^{(\beta)}(\xi), \tag{4.34}$$

where $C_0^{(\beta)}(\xi) = 1$ and $C_1^{(\beta)}(\xi) = 2\beta\xi$. To construct the potential function $\hat{\Phi}_{nl}$, we define

$$A_n = \frac{2\, n!\, \Gamma(2\mu)}{\Gamma(2\mu + n + 1)}, \tag{4.35}$$

which satisfies the recurrence relation

$$\frac{A_{n+1}}{A_n} = \frac{n+1}{n+1+2\mu}, \qquad A_0 = 1/\mu, \tag{4.36}$$

such that the ladder of potential functions at fixed $l$ are given by

$$\hat{\Phi}_{nl} = \hat{\Phi}_{(n-1)l} - \frac{r^l}{(1+z^2)^{\mu+1/2}} A_{n-1} C_{n-1}^{(\mu+1/2)}(\xi). \tag{4.37}$$

A naive implementation of this recursion relation, wherein one builds up the higher-$n$ terms by starting from $\Phi_{0l}$, results in large errors for high $n$. In the left panel of Fig. 4.1, we show the logarithm of the relative difference between the potential computed using this naive approach and an arbitrary precision calculation from *Mathematica*. The error grows with increasing $n$ and $l$, as the computation requires taking a small difference between large numbers. To see this we note that a valid series expansion of the incomplete beta function is[b] (using DLMF (2020, §8.17.8) and Fields & Wimp (1961, Eq. 2.5))

$$\mathcal{B}_\chi(\mu, 1/2) = \frac{\chi^\mu}{\sqrt{1+z^2}} \sum_{j=0}^{\infty} A_j C_j^{(\mu+1/2)}(\xi). \tag{4.38}$$

Comparison with Eq. (4.37) shows that the terms in $\hat{\Phi}_{nl}$ tend to zero as $n \to \infty$, resulting in the aforementioned catastrophic cancellation. This expression, however, provides us with an alternative method of computing $\hat{\Phi}_{nl}$. At some large order $N$, e.g. $N = 2n_{\rm max}$, we assume that $A_N \approx 0$. Then we can write

$$\hat{\Phi}_{nl} \approx \frac{r^l}{(1+z^2)^{\mu+1/2}} \sum_{j=n}^{N} A_j C_j^{(\mu+1/2)}(\xi), \tag{4.39}$$

which requires us to calculate $C_n^{(\mu+1/2)}(\xi)$ and $A_n$ recursively up to order $N$. We then recursively construct the potential basis functions $\hat{\Phi}_{nl}$ downwards using Eq. (4.37) where now all $\hat{\Phi}_{nl}$ are accurate to the magnitude of $A_N$. This procedure[c] results in the reduced errors shown in the right panel of Fig. 4.1. We only perform this procedure for $l > 4$ as for $l \leq 4$ the naive implementation is satisfactory and the decay of $\hat{\Phi}_{nl}$ is weak for small $l$.

[b] A very similar expansion is also valid for the general $\mathcal{B}_\chi(\mu, \nu)$ potential of Ch. 5 in terms of Jacobi polynomials, and the discussion in the remainder of this section also carries over.

[c] It is analogous to Miller's method in numerical analysis, apparently introduced by J.C.P. Miller for the computation of Bessel functions (see e.g. A&S, 1972).

Finally, note that the $\hat{N}_{nl}$ are independent of $n$ and the $K_{nl}$ satisfy the properties

$$\frac{K_{(n+1)l}}{K_{nl}} = \frac{n+1}{n+2\mu}, \qquad K_{0l} = -\frac{1}{8\pi\alpha^2}. \tag{4.40}$$

With these definitions, the set of coefficients corresponding to the new normalisation $\hat{C}_{nlm}$ are computed from a cloud of particles,

$$\hat{C}_{nlm} = \frac{1}{\hat{N}_{nlm}K_{nl}} \sum_i m_i \hat{\Phi}_{nl}(r_i) Y_{lm}(\theta_i, \phi_i), \tag{4.41}$$

allowing the potential and density to be reconstructed as

$$\Phi = \sum_{nlm} \hat{C}_{nlm}\hat{\Phi}_{nlm}, \qquad \rho = \sum_{nlm} K_{nl}\hat{C}_{nlm}\hat{\rho}_{nlm}. \tag{4.42}$$

### 4.3.2 Analytical haloes

#### 4.3.2.1 Spherical NFW models

We first consider the reconstruction of spherical NFW haloes using the radial terms of the expansion. We compare the $\alpha = 1$ member of our family of expansions with the Hernquist & Ostriker (1992) (HO) expansion, which is the $\alpha = 1$ member of the Zhao (1996) family. Both basis sets have lowest-order densities with $1/r$ cusps as $r \to 0$. In Fig. 4.2, we show the expansion of a spherical NFW potential and density, using only radial terms ($n \geq 0$, $l = 0$). With an equal number of terms, our $\alpha = 1$ basis set performs better than the corresponding HO set due to its closer approximation to the NFW profile in the asymptotic fall-off of the lowest-order density, $\rho \sim r^{-7/2}$. The corresponding behaviour of the HO basis set is $\rho \sim r^{-4}$. Note that the $r$-axis of Fig. 4.2 is logarithmically scaled and measured in units of the scalelength, so we in fact get several hundred additional scale-lengths of accurate behaviour. The improved convergence at large radii also reduces the amplitudes of the oscillations at smaller radii. The expansion of the potential is always more accurate than that of the density because the oscillations are integrated over, and therefore effectively smoothed; the accuracy of the radial acceleration expansion lies between that of the potential and the density. According to Fig. 4.3, the coefficients of both expansions follow power laws with respect to the radial order $n$ when used to expand a spherical NFW profile. Specifically, in our new expansion $C_{n00} \sim n^{-2}$ whereas in the HO expansion $C_{n00} \sim n^{-3}$.

To make quantitative statements about the error, we follow Vasiliev (2013) and calculate the *integrated squared error*. This is the mass-weighted fractional density error defined by

$$ISE = \int_{\substack{r>r_{\min} \\ r<r_{\max}}} \frac{\left|\rho_{\text{exact}} - \rho_{\text{approx}}\right|^2}{\rho_{\text{exact}}} \, \mathrm{d}^3\mathbf{r}. \tag{4.43}$$

Here, $\rho_{\text{approx}}$ is the density reconstructed using basis functions up to a radial order of $n_{\max}$. Although other measures of error can be constructed, Vasiliev's suggestion is appealing as it is mass-weighted and does not bias the result towards the outer parts of the model. We use $r_{\min} = 0.01$ and

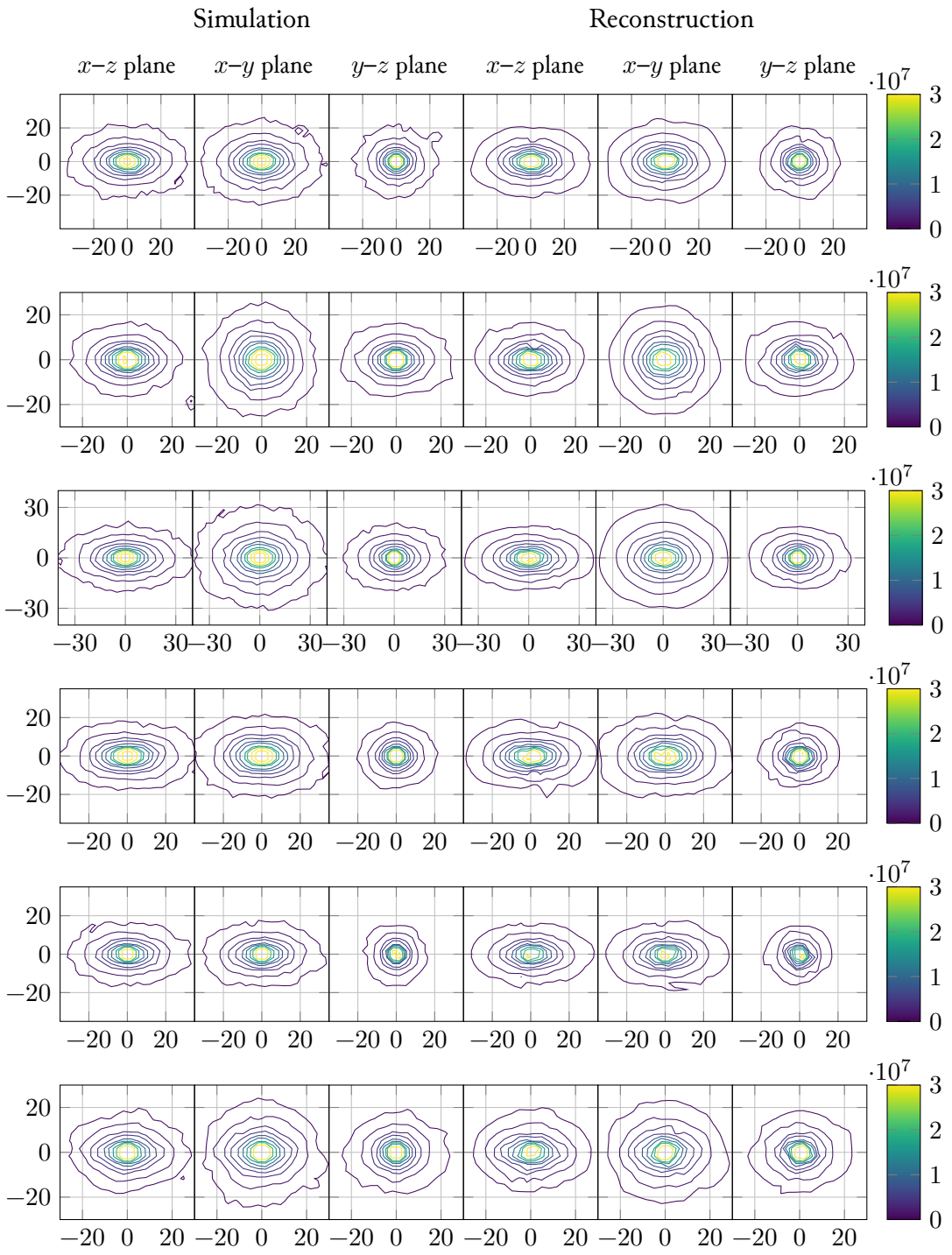


Figure 4.7: From top to bottom: haloes as listed in Table 4.1. From left to right: Density contours in the three principal planes of the original numerical halo, then density contours in the same three principal planes of the halo as reconstructed with basis functions up to $n_{\max} = 20$ and $l_{\max} = 12$. Distances are in kpc, and the densities are in $M_{\odot}\text{kpc}^{-3}$.

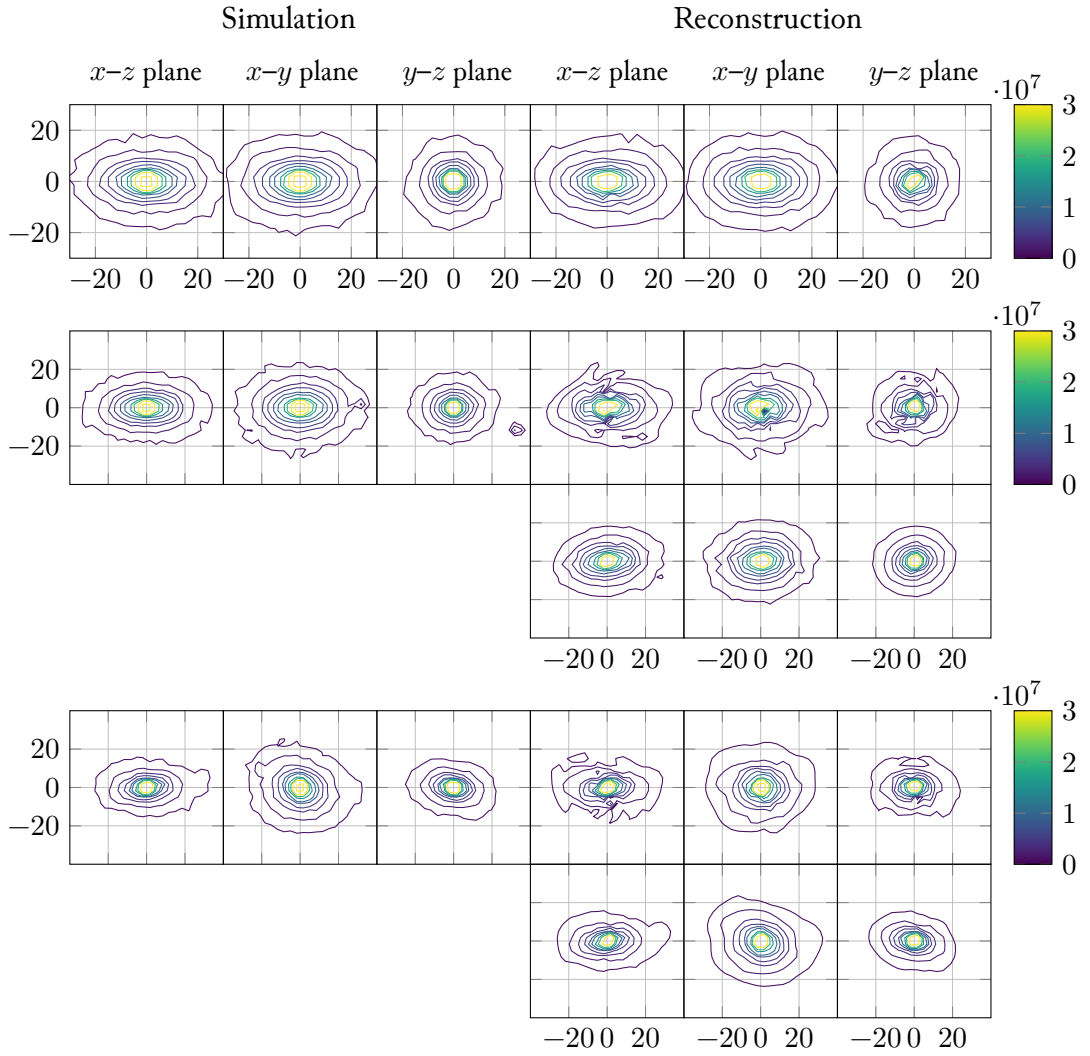


Figure 4.8: Continuation of Fig. 4.7. The latter two haloes possess a large number of small satellites, so for these we display on the right two reconstructions: top, a reconstruction of the full halo; and bottom, one where the particles energetically bound to the 50 most massive satellites have been removed. This illustrates the issues surrounding expansion accuracy in the presence of sub-structure.

$r_{\text{max}} = 100$, as this is the range over which we expect the expansions to be most applicable in astrophysical problems. The run of this error measure with $n_{\text{max}}$ is shown on the left in Fig. 4.4 for the two spherical NFW expansions under consideration. It is clear both that higher values of $\alpha$ give better accuracy, and that for given $\alpha$ and $n_{\text{max}}$ our basis set is more accurate than Zhao's in the task of reproducing NFW haloes. To illustrate this difference, on the right in Fig. 4.4 we show how the optimum $\alpha$ (the minimum of each plot) varies with differing $n_{\text{max}}$, with our basis set performing better than Zhao's in each case.

A detailed error analysis of basis function techniques is carried out in Kalapotharakos et al. (2008), who claim that a significant factor in obtaining high accuracy is choosing a basis set whose

lowest-order density function has the correct inner slope behaviour. For a spherical NFW model, this would imply that the $\alpha = 1$ expansion is best, as the zeroth-order model behave like $\rho \sim r^{-1}$ at small radii. It is claimed in Hernquist & Ostriker that certain combinations of the density basis functions can cause diverging terms at the centre to cancel – Hernquist & Ostriker built a cored density model from the $n = 0$ and $n = 1$ monopole terms of their basis set – however this is misleading, as a perfect balance between angular and radial coefficients is required, and we would expect that any numerical error or even the slightest non-uniformity in the particle distribution would cause this ephemeral core to vanish. To reproduce accurately the precession of orbits that pass close to the centre, the criterion of Kalapotharakos et al. (2008) therefore seems reasonable. However, it is at odds with the plots in Fig. 4.4, which show that models with $\alpha \approx 3$ or $\rho \sim r^{-5/3}$ provide the smallest integrated square error in the density, if $n_{\max} \approx 20$. A similar result was obtained by Vasiliev (2013), who used an identical integrated error measure to the present work and concluded that a somewhat higher value of $\alpha$ is preferable for providing a global fit to cosmological haloes, even at the expense of accuracy of the inner slope exponent near the centre. Clearly, the best choice depends on the application in hand.

#### 4.3.2.2 Flattened NFW models

It is important to consider flattened objects, as the haloes found in $N$-body simulations are generically flattened or triaxial (e.g. Jing & Suto, 2002, Allgood et al., 2006). We test the performance by attempting to reconstruct a flattened density profile and including the angular terms or $(l, m)$ terms in the series. An axisymmetric NFW density profile is $\rho = \bar{m}^{-1}(1 + \bar{m})^{-2}$, where $\bar{m}^2 = x^2 + y^2 + z^2/q^2$. This means the density is stratified on similar concentric spheroids with axis ratio $q$, so the true potential and acceleration can be calculated according to the procedure described in Binney & Tremaine (1987, §2.5). Figs 4.5 and 4.6 shows the expansion of a flattened ($q = 0.8$) NFW density, and the corresponding acceleration due to the potential. In each case, we compare the $\alpha = 1$ member of our family of expansions with the HO expansion, which is the $\alpha = 1$ member of the Zhao (1996) family. Typically, we use $n_{\max} = 20$ radial basis functions and $l_{\max} = 12$ angular basis functions.

Each of the reconstructed quantities is plotted along three polar angles ($\theta$). Note that the convergence is always superior nearer the equator ($\theta = \pi/2$) than at the poles ($\theta = 0$). This is a feature of any expansion involving spherical harmonics and can be remedied by introducing additional angular terms. In the flattened case, both expansions lose accuracy in the very inner and outer parts of the haloes. As the $l$-dependence of both our $\alpha = 1$ set and the HO set is similar, we do not expect either basis set to be favoured in this regard. However, the superior behaviour in the outskirts of our basis set is maintained. Lowing et al. (2011) used the HO basis set to represent haloes, where on the order of tens of terms are used in both the angular and radial directions. They found errors of $< 10\%$ are achieved over a few hundred kiloparsecs. We expect our basis set to provide improved accuracy in this regime.

| $N/10^6$ | $M_v/(10^{12}M_\odot)$ | $r_v$/kpc | $r_s$/kpc | $r_{\rm iso}$/kpc | $\alpha$ | $p$ | $q$ | Figure |
|---|---|---|---|---|---|---|---|---|
| 10.3 | 1.88 | 325 | 65.2 | 33.6 | 1.22 | 0.860 | 0.811 | 4.7 |
| 5.0 | 0.91 | 256 | 37.8 | 20.5 | 1.18 | 0.972 | 0.816 | 4.7 |
| 8.7 | 1.57 | 306 | 47.6 | 25.8 | 1.18 | 0.889 | 0.761 | 4.7 |
| 7.4 | 1.36 | 292 | 52.8 | 25.9 | 1.26 | 0.800 | 0.733 | 4.7 |
| 5.9 | 1.07 | 269 | 58.1 | 30.7 | 1.20 | 0.804 | 0.795 | 4.7 |
| 5.5 | 0.89 | 254 | 46.0 | 23.7 | 1.22 | 0.909 | 0.834 | 4.7 |
| 4.9 | 0.89 | 253 | 35.7 | 19.8 | 1.16 | 0.807 | 0.780 | 4.8 |
| 11.3 | 1.62 | 310 | 123 | 38.1 | 1.59 | 0.858 | 0.769 | 4.8 |
| 7.7 | 1.11 | 273 | 98.5 | 32.9 | 1.54 | 0.926 | 0.779 | 4.8 |
| 7.6 | 1.71 | 315 | 132.6 | 47.6 | 1.49 | 0.861 | 0.730 | 4.10 |

Table 4.1: The properties of each numerical halo in our sample. The haloes are listed in the order that they are displayed in the figures. $N$ is the number of particles, $M_v$ is the virial mass, $r_v$ is the virial radius, $r_s$ is the scalelength, $r_{\rm iso}$ is the distance at which the slope is isothermal (so $r_{\rm iso} = r_s/(1/2+\alpha)^\alpha$), $\alpha$ parameterises the inner and outer slopes of the density basis functions, $p$ is the $y$–$x$ axis ratio, and $q$ is the $z$–$x$ axis ratio.

### 4.3.3 Numerical Haloes

We now analyse a collection of ten Milky Way-like dark matter haloes, extracted from a suite of cosmological $N$-body zoom-in simulations. These simulations are run with the $N$-body part of GADGET-3 which is similar to GADGET-2 (Springel, 2005). The zoom-in strategy follows Oñorbe et al. (2014) and all initial conditions are generated with MUSIC (Hahn & Abel, 2011). Cosmological parameters are taken from the Planck Collaboration et al. (2014) with $h = 0.679$, $\Omega_b = 0.0481$, $\Omega_0 = 0.306$, $\Omega_\Lambda = 0.694$, $\sigma_8 = 0.827$, and $n_s = 0.962$. In order to select our haloes, we first simulate a $50h^{-1}$ Mpc box with $512^3$ particles from $z = 50$ to $z = 0$. We use ROCKSTAR (Behroozi et al., 2013) to identify haloes and select Milky Way-like haloes which have virial masses between $7.5 \times 10^{11}M_\odot - 2 \times 10^{12}M_\odot$, no major mergers since $z = 1$, and no haloes with half the mass of the Milky Way analogue's mass within $2h^{-1}$ Mpc. For each selected halo, we select all particles within 10 virial radii and run an intermediate resolution zoom-in whose maximum resolution is $2048^3$, corresponding to a particle mass of $1.8 \times 10^6 M_\odot$. This intermediate step helps to reduce the contamination from low resolution particles in our final, high resolution zoom-in. For the final zoom-in, we take the intermediate resolution simulation and select all particles within 7.5 virial radii. We then run a zoom-in with a maximum resolution of $4096^3$, corresponding to a particle mass of $2.23 \times 10^5 M_\odot$. All of our high resolution zoom-ins are uncontaminated within $1h^{-1}$ Mpc of the main halo. The detailed properties of each halo are given in Table 4.1. Note that the time-evolving halo considered in Ch. 7 corresponds to Halo #1 at $z = 0$.

For each halo, we take all of the particles within 500 kpc of the main halo in $z = 0$ snapshot. This corresponds to between 5–12 million particles, depending on the halo mass. We wish to investigate the ability of our new basis sets to represent these numerical haloes. To this end, we must first choose the two global parameters that specify the basis set – the scalelength $r_s$ and the parameter $\alpha$. We need not worry at this stage about the overall normalisation (related to the total

mass) because this will be set automatically when performing the sum over particles via Eq. (7.4). To set these, we first fit the zeroth-order density function $\rho_{000}$, which is a spherically-symmetric model. Because we know the virial radius $r_v$ for each halo, we perform the fitting procedure in terms of the dimensionless *concentration* $c \equiv r_v/r_s$, as is standard in the literature. The particles are binned logarithmically in radius, and a non-linear least squares algorithm then adjusts $\alpha$ and $c$ to minimise the difference between the logarithms of the inferred bin density and the model density. For this fitting procedure, we use a form of the zeroth-order density function that is parameterised by the mass enclosed by the virial radius, i.e. that satisfies $\int_0^{r_v} 4\pi\rho(r)r^2\,\mathrm{d}r = M_v$. This is

$$\rho(r) = \left[\mathcal{B}_{\chi_v}(\alpha, 1/2) - \frac{c/\alpha}{\left(1+c^{1/\alpha}\right)^{\alpha+1/2}}\right]^{-1} \frac{M_v\,(\alpha+1/2)}{4\pi\alpha^2\,(r_v/c)^3} \frac{(cr/r_v)^{1/\alpha-2}}{\left(1+(cr/r_v)^{1/\alpha}\right)^{\alpha+3/2}}, \tag{4.44}$$

where $\chi_v \equiv c^{1/\alpha}/\left(1+c^{1/\alpha}\right)$.

We can now perform a full expansion, using for each halo the basis set with the determined best values of $\alpha$ and $r_s$. To compare the accuracy in the reproduction of the density, we draw contour plots in each principal plane, as shown in Figs 4.7, 4.8 and 4.10. The smooth underlying density distribution in each principal plane of the original numerical halo is estimated by taking particles in a slab of width 2 kpc around the plane, then creating a two-dimensional histogram with each bin having an area of approximately 3 kpc$^2$. The density as reconstructed from the basis function expansion is sampled along rays in each principal plane, and then interpolated onto a grid. On each plot we draw the same 12 contours, spaced approximately logarithmically between $10^6$ and $2\times10^8\ M_\odot$ kpc$^{-3}$.

The haloes displayed in Figs 4.7 and 4.8 are rotated such that the principal planes are aligned with the coordinate axes (shortest axis along the $z$-axis). This means that the angular expansion coefficients can be compared meaningfully. Distributions of the expansion coefficients for these nine numerical haloes are shown in Fig. 4.9. One could produce an artificial 'halo' with geometry typical for this family of nine numerical haloes by drawing coefficients from the distributions shown in these plots. From Figs 4.7 and 4.8, we see that key features of the haloes are resolved correctly, including: 1) the orientation of the principal axes, 2) the axis ratios in the three principal planes, and 3) the run of ellipticity with radius. The size of the smallest resolvable feature is limited by the distance between the roots of the polynomial used to define the highest-order function used in the reconstruction.

Two haloes in Fig. 4.8 demonstrate one pitfall of the method. The presence of unresolved but massive sub-haloes can cause a blow-up in the higher-order coefficients of the expansion. This would be cancelled by yet higher-order terms, but as the expansion is truncated these are not present. This problem is generic – it affects other basis sets, such as Zhao's, to a greater or lesser degree – but it can be remedied by removing the unresolved sub-haloes by an automated halo-finding procedure, as demonstrated in the figure. When these sub-haloes are removed the 'optimal' values of $\alpha$ and $r_s$ (as determined by fitting Eq. (4.44)) are reduced to values more characteristic of the other haloes in the sample.

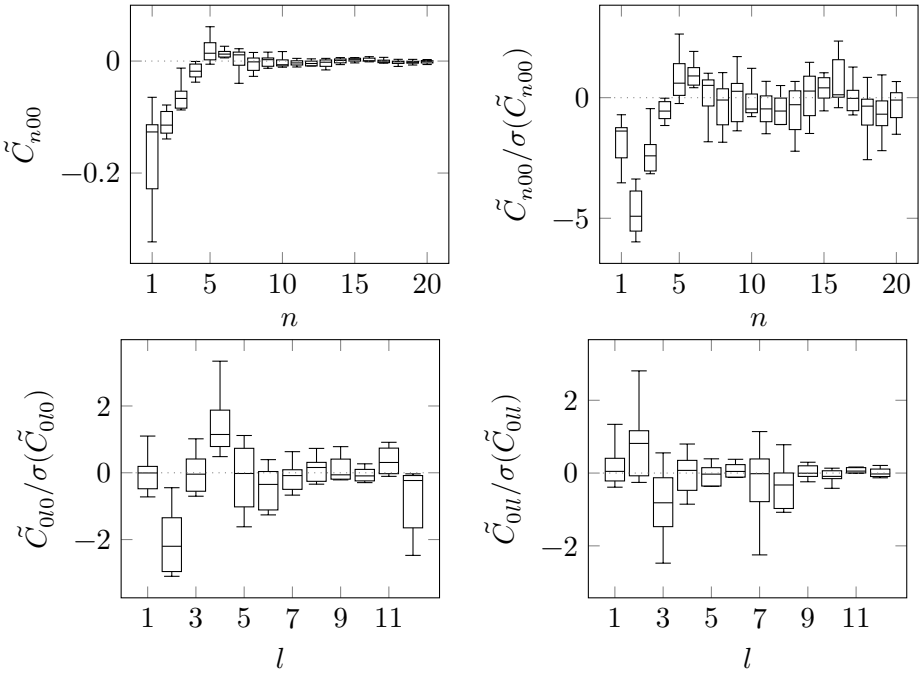


Figure 4.9: Panels 1 and 2: the distribution of radial coefficients $\tilde{C}_{n00}$ for nine numerical haloes. Panel 3: the same for angular coefficients $\tilde{C}_{0l0}$. Panel 4: the same for angular coefficients $\tilde{C}_{0ll}$. The *normalised* coefficients have tildes ($\tilde{C}_{nlm} \equiv C_{nlm}/C_{000}$) and the standard deviation of each $\sigma$. The latter three plots are divided by the standard deviation to make details easier to see. The trend in the $C_{0l0}$ coefficients indicates the flattening of the haloes along the $z$-axis, and the trend in the $\tilde{C}_{0ll}$ coefficients shows the elongation along the $x$-axis.

The final halo, displayed in Fig. 4.10, provides a serious challenge. It is accompanied by a massive close-in satellite with a mass of $1.9 \times 10^{11} M_{\odot}$, and thus has a highly aspherical geometry. We therefore examine this halo in greater detail in Fig. 4.10 using $n_{\rm max} = 20$ and $l_{\rm max} = 12$ (middle panel) and $n_{\rm max} = 40$ and $l_{\rm max} = 40$ (lower panel). Remarkably, the overall structure of both the halo and the large satellite are correctly resolved even with $n_{\rm max} = 20$. This is impressive, as we might have suspected at the outset that two basis function expansions, centred on each object, would be necessary to reproduce the merging structure.

## 4.4 Conclusions

Biorthogonal density and potential basis functions provide useful and flexible ways of describing realistic dark matter haloes and galaxies, which may be aspherical, triaxial or further misshapen. The coefficients of these basis-function expansions can be found easily by summing over the particles in an $N$-body realisation and used to reconstruct both the potential and density. We have discovered a completely new family of biorthogonal potential-density pairs, parameterised in terms of $\alpha$. The zeroth-order model has a density $\rho \sim r^{-2+1/\alpha}$ at small radii and $\rho \sim r^{-3-1/(2\alpha)}$ at large radii. This double-power law profile has a central logarithmic slope between 0 and $-2$

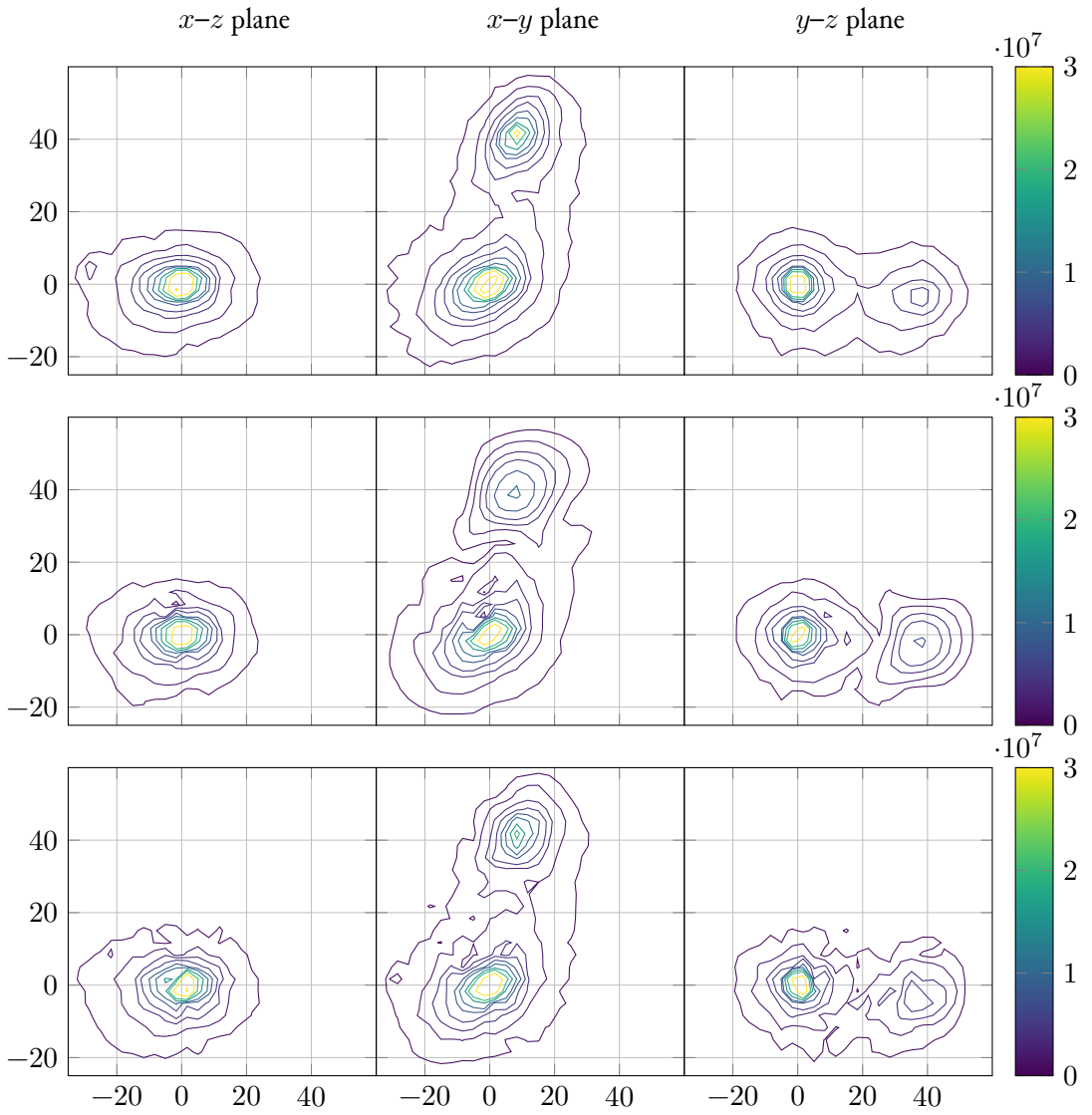


Figure 4.10: A halo with a prominent, massive satellite. From top to bottom: the original halo; the reconstruction with maximum order $n_{\max} = 20$ and $l_{\max} = 12$; the same with $n_{\max} = 40$ and $l_{\max} = 40$.

and an asymptotic logarithmic slope between $-3$ and $-4$, making it perfect for representing dark haloes. The zeroth-order model with $\alpha = 1/2$ has a harmonic core and is the celebrated perfect sphere of de Zeeuw (1985), whilst the model with $\alpha = 1$ has a $1/r$ central density cusp and is the super-NFW model (Ch. 3). For each of these zeroth-order models, we provide a biorthogonal basis function expansion in terms of standard special functions readily available in numerical libraries. This extends the zeroth-order model into the highly realistic regime of flattened, distorted and triaxial density profiles.

Previously, the only known family of biorthogonal potential-density pairs was the one outlined by Zhao (1996), of which the most widely-used member is the HO expansion. The zeroth-order

model has a density $\rho \sim r^{-2+1/\alpha}$ at small radii and $\rho \sim r^{-3-1/\alpha}$ at large radii. Although the central density cusp is the same, the outer density falls off rather more quickly than in our expansion, making Zhao's family less well-matched to modelling dark haloes.

We have demonstrated the capabilities of our basis function expansions by using them to recreate spherical and flattened dark haloes with analytical densities of NFW form. In particular, we showed that our method represents a noticeable improvement over the Hernquist & Ostriker expansion, giving a more accurate reproduction of the density, potential and radial force of NFW-like models. Additionally, we decomposed 10 simulated cosmological haloes using our basis functions, computing the coefficients as simple sums over the particles. This yielded very encouraging results with highly flattened and triaxial dark halo density distributions well-reproduced with typically 20 radial and 12 angular terms (giving a total of $21 \times 13^2 \approx 3500$ terms). Simulated dark haloes can be lopsided, distorted or twisted, especially if there is a nearby companion exerting strong tidal forces (e.g. the Milky Way and the Magellanic Clouds). Particular striking is the ability of our basis function expansion to reproduce the density of a highly asymmetric dark halo which is in the process of accreting a companion large subhalo.

One important area of application is to the fitting of data on the Milky Way galaxy, which has increased substantially in both quality and quantity over the last few years. Models of the Milky Way galaxy are assembled from simple building blocks. Results of calculations using such 'pieces of Lego' are often very troubling. For example, when the position and velocity data of stars in the Sagittarius stream are fit to such models, the conclusion is that the dark halo is triaxial with the short and long axes in the Galactic plane (Law & Majewski, 2010a). This configuration is unstable (see e.g. Debattista et al., 2013) and in conflict with observational data on the disk (Kuijken & Tremaine, 1994). The strong suspicion is that the inflexible model of the Milky Way's potential prevented proper exploration of parameter space and artificially confined the solution for the dark matter distribution into an unrealistic straitjacket. A completely new way of representing the Galactic dark halo is needed with the advent of large-scale photometric, spectroscopic and astrometric surveys of the Galaxy. For model fitting, the dark halo potential should be represented by distributions of the coefficients that can be used as priors in Bayesian inference from the data, rather than say a single number (the flattening) in a predetermined and unadaptable density law.

Of course, the shape of a dark halo depends on the nature of the dark matter particle and on the extent of feedback processes (see e.g. Sellwood, 2004, Macciò et al., 2012). The shape also depends on a host of other factors, including the mass of the halo, its environment (isolated versus group), its recent history (e.g. late in-fall of a large subhalo) and the presence or absence of a disk. It will be interesting to test our new basis function expansion method on the full variety of numerically-constructed haloes, and understand how the distributions of coefficients changes with the underlying physics. The main impediment to efficient exploitation of these ideas is that so few biorthogonal pairs are known. Our new discovery helps in this regard, but it has not exhausted the supply of such expansions. We will show in the following chapter how to extend the methods in this chapter to other cosmologically-inspired dark-halo density laws.

Chapter 5

# A two-parameter family of basis sets

In Ch. 4 we identified a completely new set of analytical biorthogonal expansions based on a lowest order model with density $\rho \sim r^{1/\alpha-2}$ at small radii and $\rho \sim r^{-3-1/(2\alpha)}$ at large radii ($\alpha \geq 1/2$). There are some striking similarities between the two known families of biorthogonal expansions in spherical coordinates (i.e. those of Zhao (1996) and Ch. 4) that strongly suggest that they are part of an underlying and more complete theoretical framework, which we provide in this chapter.

The general analytical double-power law model is discussed in Sec. 1.3.1, and the density and potential are given in Eqs (1.18) and (1.19). Its three shape parameters are $(\alpha, \beta, \gamma)$ which describe the turn-over, outer slope and inner slope respectively. Both of the above-mentioned families of basis functions have double-power law density profiles at lowest order. They lie along completely separate curves in the three-dimensional space spanned by $(\alpha, \beta, \gamma)$. The Zhao (1996) family is defined by $\beta = 3 + 1/\alpha$ and $\gamma = 2 - 1/\alpha$, whilst the family developed in Ch. 4 lies along $\beta = 3 + 1/(2\alpha)$ and $\gamma = 2 - 1/\alpha$.

Here we present a two-parameter family of expansions that encompasses both families. Sec. 5.1 uses the method outlined in Sec. 2.3.1 to construct a non-orthogonal basis set, via Hankel-transforming a sequence of auxiliary functions. In Sec. 5.2, this non-orthogonal set is then diagonalised analytically, producing an orthonormal set.

Special cases, including the cosmologically significant NFW model, are discussed in Sec. 5.4.

## 5.1 A non-orthonormal basis set

### 5.1.1 Family A

Following the methods of Sec. 2.3.1 and Ch. 4, we begin by writing expressing the potential and density basis functions as

$$\begin{aligned}\Phi_{nl}(r) &\propto r^{-1/2}\int_0^\infty \mathrm{d}k\, g_n(k)\, J_\mu(kz),\\ \rho_{nl}(r) &\propto r^{1/\alpha-5/2}\int_0^\infty \mathrm{d}k\, k^2\, g_n(k)\, J_\mu(kz),\end{aligned} \tag{5.1}$$

where $z = r^{1/(2\alpha)}$ and $\mu = \alpha(1+2l)$. Given our particular choice of $g_n(k)$ we will refer to this set of solutions as *Family A* and will present a second family in the next subsection. We arrive at a form for $g_n(k)$ via a heuristic procedure. First, we note that given some density profile $\rho(r)$, the corresponding auxiliary function $g(k)$ is found by inverting the Hankel transform,

$$g(k) = k^{-1}\int_0^\infty \mathrm{d}z\, z\, r^{5/2-1/\alpha}\, \rho(r)\, J_\mu(kz). \tag{5.2}$$

This inversion procedure means we can, in principle, find all the functions $g_n(k)$ corresponding to a particular basis set, even if that basis set was *not* originally derived using the Hankel transform method of Sec. 2.3.1. For instance, taking the zeroth order Zhao basis function,

$$\rho_{0l}(r) \propto r^{-5/2+1/\alpha}\frac{z^\mu}{(1+z^2)^{\mu+2}}, \tag{5.3}$$

the inversion (G&R, Eq. 6.565(4)) gives

$$g_0(k) = k^\mu K_1(k), \tag{5.4}$$

where $K_\nu(k)$ is the modified Bessel function of the second kind[a] (DLMF, §10.25). This leads us to propose a generalised form for $g_0(k)$ as

$$g_0(k) = k^{\mu+\nu-1}K_\nu(k), \tag{5.5}$$

which produces the zeroth order density functions (G&R, Eq. 6.576(7))

$$\rho_{0l}(r) \propto \frac{r^{1/\alpha+l-2}}{(1+r^{1/\alpha})^{\mu+\nu+1}}. \tag{5.6}$$

and corresponding potential functions

$$\Phi_{0l}(r) \propto r^l\, {}_2F_1\left(\begin{matrix}\mu,\mu+\nu\\ 1+\mu\end{matrix}\middle|-z^2\right) \propto \frac{\mathcal{B}_\chi(\mu,\nu)}{r^{l+1}}. \tag{5.7}$$

Here, $\chi \equiv z^2/(1+z^2)$, $\mathcal{B}_x(a,b)$ is the incomplete beta function, and ${}_2F_1$ is the Gaussian hypergeometric function. To obtain the final result we used a Hankel transform (G&R, Eq. 6.576(3))

[a] These functions satisfy the identity $K_{-\nu}(k) = K_\nu(k)$.

and a linear hypergeometric transformation (DLMF, §15.8.1). The potential integral is only valid for $\mu + \nu > 0$, but this constraint is less restrictive than the orthogonality constraint on $\mu$ and $\nu$ (discussed in the following section). The potential basis functions obey the asymptotes derived in Sec. 1.4.3, with $r^l$ behaviour for $r \to 0$ and $r^{-1-l}$ for $r \to \infty$. The inner density slope is $\gamma = 2 - 1/\alpha$ whilst the outer density slope is $\beta = 3 + \nu/\alpha$. For a $\gamma = 1$ cusp, $\alpha = 1$ and $\nu$ controls the outer slope. Slower breaks (e.g. $\alpha = 2$) produce cuspier ($\gamma > 1$) central profiles. To avoid un-physical centrally-vanishing density profiles we require $\alpha \geq 1/2$ and in turn if we require profiles with finite mass then we must have $\beta > 3$ and $\nu > 0$.

In the left panel of Fig. 5.1, we show the range of zeroth-order density profiles encompassed by our Family A models. We see increasing $\alpha$ at fixed $\nu$ 'straightens out' the density profile whilst increasing $\nu$ at fixed $\alpha$ steepens the outer density slope.

We now wish to construct a full basis set with this lowest order potential-density pair. Computing $g_1(k)$ from the first order density basis function of the Zhao (1996) basis set gives

$$g_1(k) = k^\mu (k K_0(k) - \mu K_1(k)), \tag{5.8}$$

suggesting that a full set of solutions can be composed from the set of non-orthonormal basis functions

$$\mathcal{K}_j(k) = k^{\mu+\nu-1+j} K_{\nu-j}(k), \qquad j \geq 0, \;\; j \in \mathbb{Z}. \tag{5.9}$$

We denote the corresponding non-biorthogonal potential-density basis functions $\tilde{\Phi}_{nl}$ and $\tilde{\rho}_{nl}$, and they are computed as (G&R, Eq. 6.576(3))

$$\begin{aligned} \tilde{\Phi}_{nl}(r) &\propto \frac{r^l}{(1+z^2)^{\mu+\nu}} \mathcal{P}_{j-1}^{(\nu)}(\chi), \\ \tilde{\rho}_{nl}(r) &\propto \frac{r^{l+1/\alpha-2}}{(1+z^2)^{\mu+\nu+1}} \mathcal{P}_{j}^{(\nu+1)}(\chi), \end{aligned} \tag{5.10}$$

where we use the shorthand $\mathcal{P}_j^{(\nu)}(\chi)$ for a certain hypergeometric polynomial which can be computed directly as a Jacobi polynomial

$$\mathcal{P}_j^{(\nu)}(\chi) \equiv {}_2F_1\left( \begin{matrix} -j, \mu+\nu \\ 1+\mu \end{matrix} \middle| \chi \right) = \frac{(-1)^j j!}{(\mu+1)_j} P_j^{(\nu-1-j,\mu)}(\xi), \tag{5.11}$$

where $\xi \equiv 2\chi - 1$, and we have made use of the Pochhammer symbol $(z)_n$ (DLMF, §5.2(iii)). The only term in the expressions (5.10) which is not proportional to a polynomial in $\chi$ is the zeroth-order ($j = 0$) of the potential, given by Eq. (5.7) in terms of the incomplete beta function.

### 5.1.2 Family B

A further set of potential-density pairs that solve the Poisson equation can be obtained by applying the Kelvin transform (Sec. 2.4) to the Family A expressions in the preceding section. In terms of

Hankel transforms, this is

$$\begin{aligned}\Phi_{nl}(r) &\propto r^{-1/2}\int_0^\infty \mathrm{d}k\, g_n(k)\, J_\mu(k/z),\\ \rho_{nl}(r) &\propto r^{-1/\alpha-5/2}\int_0^\infty \mathrm{d}k\, k^2\, g_n(k)\, J_\mu(k/z),\end{aligned} \tag{5.12}$$

where the difference to Eq. (5.1) is in the argument of the Bessel functions. With the same choice of auxiliary function $\mathcal{K}_j(k)$ as in Eq. (5.5), we find

$$\rho_{0l}(r) \propto \frac{r^{\nu/\alpha+l-2}}{(1+r^{1/\alpha})^{\mu+\nu+1}}, \tag{5.13}$$

$$\Phi_{0l}(r) \propto r^l\, \mathcal{B}_{1-\chi}(\mu,\nu)\,, \tag{5.14}$$

and the functions for $j>0$ follow similarly. The inner density slope is $\gamma = 2-\nu/\alpha$ whilst the outer density slope is $\beta = 3+1/\alpha$. For cusped models ($0<\nu<2\alpha$), $\alpha$ controls the outer slope but also alters the turn-over of the density profile. We call this family of models *Family B*. The potential integral is only valid for $\mu+\nu>0$. For non-vanishing central density, we require $\nu<2\alpha$. All zeroth-order models have finite mass as $\alpha>0$.

Note that for $\nu=1$, both sets coincide and Zhao's solutions are recovered. However, in general, Family B is distinct from Family A, even if the models have the same inner $\gamma$ and outer $\beta$ density slopes. This is because the gradualness of the transition from inner to outer behaviour is controlled by $\alpha$, which is in general different between the two families.

In the right panel of Fig. 5.1, we show the range of zeroth-order density profiles in Family B. We see that increasing $\alpha$ at fixed $\nu$ 'straightens out' the density profile as with Family B, whilst increasing $\nu$ at fixed $\alpha$ steepens the inner density profile.

## 5.2 An orthonormal basis set

We now show that the basis sets developed above can in fact be orthogonalised in closed-form. That is, by taking linear combinations of the functions $\mathcal{K}_j(k)$, we can find functions $g_n(k)$ that satisfy the orthogonality condition

$$\int_0^\infty \mathrm{d}k\, k\, g_m(k)\, g_n(k) \propto \delta_{mn}. \tag{5.15}$$

### 5.2.1 Method

To find such an orthonormal basis set, we first write down a general linear sum of the non-orthonormal basis,

$$g_n(k) = \sum_{j=0}^{n} c_{nj}\, \mathcal{K}_j(k). \tag{5.16}$$

To evaluate the coefficients $c_{nj}$, we require the integral between each pair of $\mathcal{K}_n$ functions (G&R, Eq. 6.576(4)),

$$\begin{aligned} D_{mn}(\mu,\nu) &\equiv \int_0^\infty \mathrm{d}k\, k\, \mathcal{K}_m(k)\, \mathcal{K}_n(k) \\ &= 2^{m+n+2\mu+2\nu-3}\Gamma(m+\mu+\nu)\Gamma(n+\mu+\nu)\mathrm{B}(m+n+\mu,\mu+2\nu), \end{aligned} \tag{5.17}$$

where $\mathrm{B}(a,b)$ is the beta function. We note that this integral only converges when $\mu > -2\nu$, as each potential-density inner product is required to be finite. To see this directly for the zeroth-order case, the following integral must be finite,

$$\int_0^\infty \mathrm{d}r\, r^2\, \Phi_{00}\, \rho_{00} \propto \int_0^\infty \mathrm{d}r\, \frac{\mathcal{B}_\chi(\alpha,\nu)}{r}\, \frac{r^{1/\alpha}}{(1+r^{1/\alpha})^{\alpha+\nu+1}}. \tag{5.18}$$

As $r \to \infty$, we have $\chi \approx 1 - r^{-1/\alpha}$, so we can approximate the incomplete beta function's defining integral as

$$\mathcal{B}_\chi(\alpha,\nu) \approx \mathrm{B}(\mu,\nu) - r^{-\nu/\alpha}. \tag{5.19}$$

Hence the asymptotic behaviour of the zeroth order potential function is

$$\Phi_{00} \sim \begin{cases} r^{-1}, & \text{if } \nu/\alpha \geq 0 \\ r^{-\nu/\alpha-1}, & \text{otherwise.} \end{cases} \tag{5.20}$$

Inspecting the behaviour of the integrand in Eq. (5.18) as $r \to \infty$ for Family A ($\alpha \geq 1/2$) we find that if $\nu \geq 0$ then the integral clearly converges. However, if $\nu < 0$ then to prevent divergence we must have $\alpha > -2\nu$. An identical constraint on $\alpha$ and $\nu$ is obtained for Family B by considering $r \to 0$.

Although it may appear that a numerical inversion of the matrix (5.17) must be performed, a closed-form expression can in fact be found. Taking advantage of the beta function's integral representation,

$$\mathrm{B}(m+n+\mu,\mu+2\nu) = \int_0^1 \mathrm{d}t\, t^{m+n+\mu-1}(1-t)^{\mu+2\nu-1}, \tag{5.21}$$

and replacing $\mathcal{K}_n$ in Eq. (5.17) by the linear combination $\sum c_{jn}\mathcal{K}_n$, we see that the orthogonality condition (5.15) becomes an orthogonality relation between two polynomials in $t$ with respect to a certain weight function,

$$\int_0^1 \mathrm{d}t\, t^{\mu-1}(1-t)^{\mu+2\nu-1} \left(\sum_{m=0}^{i} c_{im}t^m\right)\left(\sum_{n=0}^{j} c_{jn}t^n\right) \propto \delta_{ij}. \tag{5.22}$$

Fortunately the orthogonal polynomials corresponding to this weight function are well-known: they are simply the Jacobi polynomials under the change of variables $t \mapsto 2t-1$. A simple closed-form expression for these polynomials as a sum over monomials in $t$ can be obtained (G&R, Eq. 8.962(1)),

$$P_n^{(\mu+2\nu-1,\mu-1)}(2t-1) = \frac{(-1)^n\,(\mu)_n}{n!} \sum_{j=0}^{n} \frac{(-n)_j\,(n+2\mu+2\nu-1)_j}{j!\,(\mu)_j}\, t^j. \tag{5.23}$$

Now, writing each $\mathcal{K}_j(k)$ using an integral representation of the modified Bessel function $K_\nu(k)$ (DLMF, §10.32.10), and expressing the weights $c_{nj}$ using the polynomial in Eq. (5.23), we have the following integral expression for the functions $g_n(k)$,

$$g_n(k) = \frac{\mu\, k^{\mu+2\nu-1}}{2^{\mu+2\nu-1}\,\Gamma(\mu+\nu)} \int_0^\infty \mathrm{d}t\, t^{-\nu-1}\, \mathrm{e}^{-t-\frac{k^2}{4t}}\, f_n(t), \tag{5.24}$$

where

$$f_n(t) = {}_2F_2\left( \begin{matrix} -n, n+2\mu+2\nu-1 \\ \mu, \mu+\nu \end{matrix} \middle| t \right) \tag{5.25}$$

which is a polynomial with almost the same coefficients as a Jacobi polynomial (5.23), each term having an additional factor of $(\mu+\nu)_j$ that arises from the integral (5.17). We can now insert these expressions for $g_n(k)$ into Eq. (5.1), evaluating the integral over the Bessel $J$-function (G&R, Eq. 6.631(1)), to obtain integral expressions for $\Phi_{nl}$ and $\rho_{nl}$,

$$\Phi_{nl}(r) = \frac{r^l}{\Gamma(\mu)} \int_0^\infty \mathrm{d}t\, t^{\mu-1}\, \mathrm{e}^{-t}\, f_n(t)\, {}_1F_1\left( \begin{matrix} \mu+\nu \\ \mu+1 \end{matrix} \middle| -z^2 t \right), \tag{5.26}$$

$$\rho_{nl}(r) = \frac{r^{l-2+1/\alpha}}{\Gamma(\mu+1)} \int_0^\infty \mathrm{d}t\, t^{\mu}\, \mathrm{e}^{-t}\, f_n(t)\, {}_1F_1\left( \begin{matrix} \mu+\nu+1 \\ \mu+1 \end{matrix} \middle| -z^2 t \right). \tag{5.27}$$

As a double-check, one can expand $f_n(t)$ as a series in $t$ and evaluate the integral over $t$ for each term (G&R, Eq. 7.621(4)), the overall result being the correct sum over the non-orthogonal basis functions $\tilde{\Phi}_{nl}$ and $\tilde{\rho}_{nl}$.

It is interesting to note that a valid (though not necessarily biorthogonal) potential-density pair would be given by replacing the $t$-dependent part of the integrands in Eqs (5.26) and (5.27) with an arbitrary function. This is the integral transform method which we discussed in Sec. 2.3.3.

Now, as the new expressions for $\Phi_{nl}$ and $\rho_{nl}$ are in integral form, they are not yet useful. However we proceed to evaluate the integrals in Eqs (5.26) and (5.27) *explicitly* via a method based on hypergeometric generating functions. By substituting the appropriate values into Chaundy (1943, Eq. 26), we can find a generating function for $f_n(t)$, noting that the result fortuitously simplifies from a ${}_2F_2$ to a ${}_1F_1$ function,

$$\sum_{n=0}^\infty \frac{(2\mu+2\nu-1)_n}{n!} f_n(t)\, x^n = (1-x)^{1-2\mu-2\nu}\, {}_1F_1\left( \begin{matrix} \mu+\nu-1/2 \\ \mu \end{matrix} \middle| \frac{-4tx}{(1-x)^2} \right). \tag{5.28}$$

This expression can be inserted into (5.26), and the resulting integral over the pair of ${}_1F_1$ functions can be evaluated using Saad & Hall (2003, Eq. 2.2), to give

$$\begin{aligned} \sum_{n=0}^\infty \frac{(2\mu+2\nu-1)_n}{n!} \Phi_{nl} x^n &= \frac{r^l}{(1-x)^{2\mu+2\nu-1}} F_2\left( \begin{matrix} \mu; \mu+\nu-1/2, \mu+\nu \\ \mu, \mu+1 \end{matrix} \middle| \frac{-4x}{(1-x)^2}, -z^2 \right) \\ &= \frac{r^l}{(1+x)^{2\mu+2\nu-1}} F_1\left( \begin{matrix} \mu+\nu; \mu+\nu-1/2, 1/2-\nu \\ \mu+1 \end{matrix} \middle| -\left(\frac{1-x}{1+x}\right)^2 z^2, -z^2 \right), \end{aligned} \tag{5.29}$$

where $F_1$ and $F_2$ are Appell's hypergeometric functions, and the $F_2 \to F_1$ reduction (DLMF, §16.16.3) is justified because the first and fourth arguments of the $F_2$ are equal. Appell's $F_1(a; b_1, b_2; c; x, y)$ function simplifies to Gauss' ${}_2F_1$ function if $b_1 + b_2 = c$ (DLMF, §16.16.1), and we note that the second parameter of the $F_1$ in Eq. (5.29) would need to be increased by 1 in order to satisfy this condition. To accomplish this we make use of the following differential recurrence relation, derivable from the $F_1$ contiguous relations (Mullen, 1966),

$$F_1\left(\begin{matrix} a; b_1+1, b_2 \\ c \end{matrix}\middle| s, t\right) = F_1\left(\begin{matrix} a; b_1, b_2 \\ c \end{matrix}\middle| s, t\right) + \frac{s}{b_1}\frac{\partial}{\partial s}\left[F_1\left(\begin{matrix} a; b_1, b_2 \\ c \end{matrix}\middle| s, t\right)\right]. \tag{5.30}$$

Applying this relation to Eq. (5.29), simplifying both sides of the equation, and applying the now-valid $F_1 \to {}_2F_1$ reduction formula, we obtain

$$\sum_{n=0}^{\infty} \frac{(2\mu+2\nu)_n}{n!}\left(\Phi_{nl} - \Phi_{n+1,l}\right) x^n = \frac{2\, r^l}{(1+x)^{2\mu+2\nu}\,(1+z^2)^{\mu+\nu}} \times {}_2F_1\left(\begin{matrix} \mu+\nu, \mu+\nu+1/2 \\ \mu+1 \end{matrix}\middle| \frac{4x\chi}{(1+x)^2}\right). \tag{5.31}$$

This generating function is also a special case of Chaundy (1943, Eq. 26) and in fact turns out to be a generating function for the Jacobi polynomials (DLMF, §18.12.3), so we finally obtain

$$\Phi_{nl}(r) - \Phi_{n+1,l}(r) = \frac{2\,n!}{(\mu+1)_n}\,\frac{r^l}{(1+z^2)^{\mu+\nu}}\,P_n^{(\mu+2\nu-1,\mu)}(\xi). \tag{5.32}$$

A similar method can be used for $\rho_{nl}$, starting from Eq. (5.27) and applying the generating function above (Eq. 5.28), then integrating using Saad & Hall (2003, Eq. 2.2) and applying the $F_2 \to F_1$ transformation, to give

$$\sum_{n=0}^{\infty} \frac{(2\mu+2\nu-1)_n}{n!}\rho_{nl}x^n = \frac{r^{l-2+1/\alpha}}{(1+z^2)^{\mu+\nu+1}(1-x)^{2\mu+2\nu-1}} \times F_1\left(\begin{matrix} \mu+\nu-1/2; -\nu, \mu+\nu+1 \\ \mu \end{matrix}\middle| \frac{-4x}{(1-x)^2}, \frac{-4x}{(1-x)^2(1+z^2)}\right) \tag{5.33}$$

This time we note that the fourth parameter of the $F_1$ needs to be increased by 1 in order to effect the reduction to a ${}_2F_1$. To accomplish this, we use the following $F_1$ contiguous relation (Mullen, 1966),

$$F_1\left(\begin{matrix} a; b_1, b_2 \\ c \end{matrix}\middle| s, t\right) = \frac{c-a}{c}F_1\left(\begin{matrix} a; b_1, b_2 \\ c+1 \end{matrix}\middle| s, t\right) + \frac{a}{c}F_1\left(\begin{matrix} a+1; b_1, b_2 \\ c+1 \end{matrix}\middle| s, t\right). \tag{5.34}$$

Having applied this, we can use the $F_1 \to {}_2F_1$ transformation on both terms, giving

$$\sum_{n=0}^{\infty} \frac{(2\mu+2\nu-1)_n}{n!}\rho_{nl}x^n = \frac{r^{l-2+1/\alpha}}{(1+z^2)^{\mu+\nu+1}} \times \left[\frac{1/2-\nu}{\mu}(1+x)^{1-2\mu-2\nu}\,{}_2F_1\left(\begin{matrix}\mu+\nu-1/2, \mu+\nu+1\\ \mu+1\end{matrix}\middle|\frac{4x\chi}{(1+x)^2}\right) + \frac{\mu+\nu-1/2}{\mu}\frac{(1-x)^2}{(1+x)^{1+2\mu+2\nu}}\,{}_2F_1\left(\begin{matrix}\mu+\nu+1/2, \mu+\nu+1\\ \mu+1\end{matrix}\middle|\frac{4x\chi}{(1+x)^2}\right)\right]. \quad (5.35)$$

We transform the first ${}_2F_1$ (DLMF, §15.5.15), which turns it into two Chaundy-style generating functions, for the Jacobi polynomials $P_n^{(\mu+2\nu-1,\mu-1)}(\xi)$ and $P_n^{(\mu+2\nu-2,\mu)}(\xi)$ respectively. The second ${}_2F_1$ is a Chaundy-style generating function multiplied by a factor of $(1-x)^2$ and so produces terms proportional to $P_n^{(\mu+2\nu,\mu)}(\xi)$, $P_{n-1}^{(\mu+2\nu,\mu)}(\xi)$ and $P_{n-2}^{(\mu+2\nu,\mu)}(\xi)$. Hence we obtain a sum of five Jacobi polynomials with various parameters. We must then repeatedly apply two more Jacobi recurrence relations (DLMF, §18.9.3, 18.9.5) in order to simplify the expression. The result is

$$\rho_{nl}(r) = \frac{n!(n+\mu)}{\mu(\mu+\nu)(2n+2\mu+2\nu-1)(\mu+1)_n} \times \frac{r^{l-2+1/\alpha}}{(1+z^2)^{\mu+\nu+1}}\left[(n+2\mu+2\nu-1)(n+\mu+\nu)P_n^{(\mu+2\nu-1,\mu)}(\xi) - (n+\mu+2\nu-1)(n+\mu+\nu-1)P_{n-1}^{(\mu+2\nu-1,\mu)}(\xi)\right]. \quad (5.36)$$

Note that the expressions (5.41) and (5.43) in the following section are written using a different normalisation.

Computing the acceleration requires the derivative of the potential. This could be derived by differentiating the expression (5.32), but a more efficient expression can be obtained by differentiating the original $k$-integral (5.1). Using standard properties of Bessel functions, we find that

$$\frac{d\Phi_{nl}}{dr} = -\frac{1+l}{\alpha}\frac{\Phi_{nl}}{r} + \frac{r^{1/(2\alpha)-3/2}}{2\alpha}\int_0^\infty \mathrm{d}k\, k\, g_n(k)\, J_{\mu-1}(kz). \quad (5.37)$$

We can then use a similar approach to that described above, by applying a generating function to $g_n(k)$. The only notable difference is that the second argument in the ${}_1F_1$ function in Eq. (5.26) is $\mu$ rather than $\mu+1$. The Appell functions simplify as before.

### 5.2.2 Result

Here we present the results of the derivation of the previous section in their most convenient forms: we write basis functions $\Phi_{nl}(r)$ and $\rho_{nl}(r)$ that are orthogonal but not normalised to unity, in order to condense the notation and clarify the structure of the expressions. The full basis functions are then

$$\Phi_{nlm}(\mathbf{r}) = \Phi_{nl}(r)\, Y_{lm}(\theta,\phi), \quad (5.38)$$
$$\rho_{nlm}(\mathbf{r}) = K_{nl}\, \rho_{nl}(r)\, Y_{lm}(\theta,\phi).$$

The radial basis functions satisfy

$$\begin{aligned} \int \mathrm{d}r\, r^2 \Phi_{nl} \rho_{n'l} &= \delta_{nn'} N_{nl}, \\ \nabla^2 \left(\Phi_{nl} Y_{lm}\right) &= 4\pi\, K_{nl}\, \rho_{nl}\, Y_{lm}, \end{aligned} \tag{5.39}$$

where the definitions of $N_{nl}$ and $K_{nl}$ are given in Eqs (5.44) and (5.45). This means that $\Phi_{nlm}$ and $\rho_{nlm}$ satisfy

$$\begin{aligned} \int \mathrm{d}^3\mathbf{r}\ \Phi_{nlm}(\mathbf{r})\, \rho_{n'l'm'}(\mathbf{r}) &= \delta^{n'l'm'}_{nlm} N_{nl} K_{nl}, \\ \nabla^2 \Phi_{nlm} &= 4\pi \rho_{nlm}. \end{aligned} \tag{5.40}$$

Note that in order to use physical units, one should use dimensionful values for the length scale and for Newton's gravitational constant. In practice this means making the replacements $r \mapsto r/r_\mathrm{s}$, for some chosen length $r_\mathrm{s}$ that matches the scale on which radial structure exists, and also $K_{nl} \mapsto r_\mathrm{s}^{-2}\, G\, K_{nl}$.

We now give the simplified formulas for the radial basis functions. Recalling the shorthands $\mu = \alpha(1+2l)$, $z = r^{1/(2\alpha)}$, $\chi = z^2/(1+z^2)$ and $\xi = 2\chi - 1$, we have for the potential

$$\begin{aligned} \Phi_{nl} - \Phi_{n+1,l} &= \frac{2\, n!}{(\mu+1)_n} \frac{r^l}{(1+z^2)^{\mu+\nu}} P_n^{(\mu+2\nu-1,\mu)}(\xi), \\ \Phi_{0l} &= \frac{\mu\, \mathcal{B}_\chi(\mu,\nu)}{r^{1+l}}, \end{aligned} \tag{5.41}$$

for the acceleration

$$\begin{aligned} \frac{d\Phi_{nl}}{dr} &= -(1+l)\frac{\Phi_{nl}}{r} + \frac{(1+2l)\, n!}{(\mu)_n} \frac{r^{l-1}}{(1+z^2)^{\mu+\nu}} P_n^{(\mu+2\nu-1,\mu-1)}(\xi) \\ &= -(1+l)\frac{\Phi_{nl}}{r} + \frac{(1+2l)\, n!}{(\mu)_n (2n+2\mu+2\nu-1)} \frac{r^{l-1}}{(1+z^2)^{\mu+\nu}} \\ &\times \left[(n+2\mu+2\nu-1) P_n^{(\mu+2\nu-1,\mu)}(\xi) + (n+\mu+2\nu-1) P_{n-1}^{(\mu+2\nu-1,\mu)}(\xi)\right], \end{aligned} \tag{5.42}$$

(with the second expression using Jacobi polynomials of the same parameters as used in the potential and density[b]). And finally, for the density the normalised closed-form expression is

$$\begin{aligned} \rho_{nl} = \frac{r^{l-2+1/\alpha}}{(1+z^2)^{\mu+\nu+1}} &\left[(n+2\mu+2\nu-1)(n+\mu+\nu) P_n^{(\mu+2\nu-1,\mu)}(\xi)\right. \\ &\left. -(n+\mu+2\nu-1)(n+\mu+\nu-1) P_{n-1}^{(\mu+2\nu-1,\mu)}(\xi)\right]. \end{aligned} \tag{5.43}$$

The normalisation constant, derived from the normalisation of the Jacobi polynomials (5.23), is

$$N_{nl} = \frac{\alpha\, \Gamma(n+\mu+2\nu)\, \Gamma(\mu+1)}{\Gamma(n+2\mu+2\nu-1)}, \tag{5.44}$$

[b] This avoids having to construct two recursive ladders of Jacobi polynomials, which saves on half the floating point operations. This is therefore an improvement over the result published as Lilley et al. (2018c).

and the proportionality constant in Poisson's equation is

$$K_{nl} = -\frac{n!\,\Gamma(\mu+1)}{4\pi\alpha^2\,(2n+2\mu+2\nu-1)\,\Gamma(n+\mu)}. \tag{5.45}$$

In order to maintain finite self-energy and to avoid an un-physical hole in the mass density at the origin, we must have $\alpha > 2\nu$, and also either $\alpha \geq 1/2$ (for Family A) or $\alpha \geq \nu/2$ (for Family B). For the special case $\alpha + \nu = 1/2$, limiting forms of both the basis functions and the associated constants must be used, for which see Sec. 5.3.

The basis functions of Family B can be constructed from those of Family A by the Kelvin transform (Sec. 2.4), which here takes the form: $\chi \mapsto 1-\chi$, $\xi \mapsto -\xi$, $\rho_{nl} \mapsto r^{(\nu-1)/\alpha}\rho_{nl}$, $\Phi_{0l} \mapsto r^{1+2l}\Phi_{0l}$ and $(\Phi_{nl} - \Phi_{0l}) \mapsto r^{\nu/\alpha}(\Phi_{nl} - \Phi_{0l})$. We again emphasise that Families A and B are in general distinct, other than their intersection ($\nu = 1$), which is equal to the family of models given in Zhao (1996).

The family of basis sets described by Eqs (5.41) and (5.43) (and the accompanying 'B' sets) are perhaps the major result of this thesis. By choosing $\alpha$ and $\nu$ appropriately, they can be used to efficiently capture the higher-order corrections to a double-power law model with any combination of inner and outer slopes. The basis sets are analytical – they require no further numerical orthogonalisation, and hence the resulting accuracy is not dependent on the condition number of an overlap or stiffness matrix (compare e.g. Saha (1993), where this orthogonalisation step must be carried out).

In the process of developing the application of this basis set to a realistic numerical time-evolving halo (Ch. 7), we find that we must add a correction to the expansion coefficients to compensate for the truncation in the particle distribution at finite radius. These corrections involve calculating the indefinite integral between the zeroth-order density and the higher-order potential basis functions, and the expressions (involving finite sums of incomplete beta functions) can be found in Appendix D, along with example code.

## 5.3 Limiting forms

In certain cases the density $\rho_{nl}$ and associated constants $N_{nl}$ and $K_{nl}$ must be modified, as they diverge or become zero[c]. Modification is required when two conditions are satisfied: $n = l = 0$, and $\alpha + \nu = 1/2$. Because of the pre-existing constraints on $\nu$ and $\alpha$, this means that the only cases affected are $1/2 \leq \alpha < 1$ and $-1/2 < \nu \leq 0$. We set $n = l = 0$ first, then evaluate the following limits as $\nu \to 1/2 - \alpha$, making use of $\lim_{x\to 0}[x\Gamma(x)] = 1$,

$$\begin{aligned} \lim_{\nu\to 1/2-\alpha}[K_{00}\rho_{00}] &= -\frac{1}{8\pi\alpha}\,\frac{r^{-2+1/\alpha}}{(1+z^2)^{3/2}}, \\ \lim_{\nu\to 1/2-\alpha}[K_{00}N_{00}] &= -\frac{\alpha}{4}\,\csc(\pi\alpha). \end{aligned} \tag{5.46}$$

For these special cases, the orthogonality relation (5.39) must be multiplied through by $K_{00}$ in order to have meaning. Note that the result depends on the order in which the limits $n, l \to 0$

[c]This includes the case when the zeroth order is the modified Hubble profile $\rho \propto (1+r^2)^{-3/2}$.

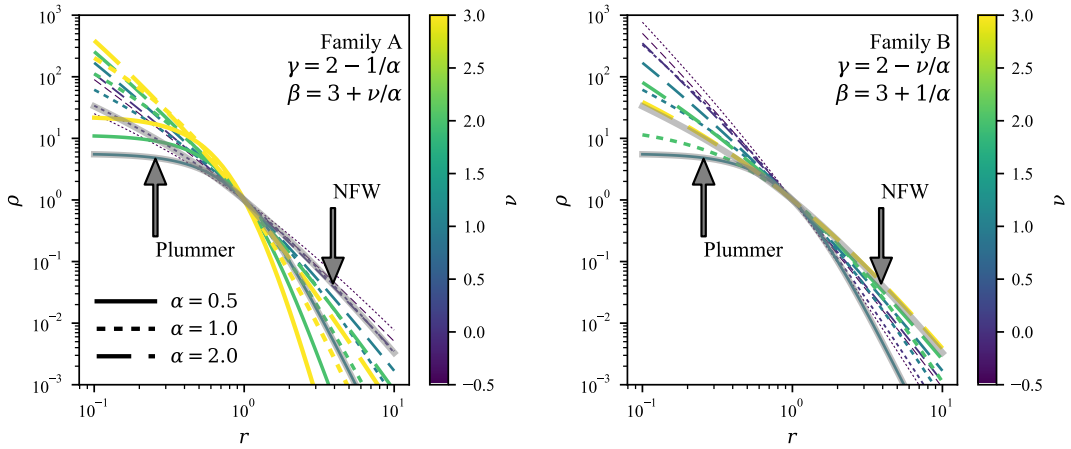


Figure 5.1: The range of zeroth-order density profiles covered by our two families of expansions (A left, B right). Each line is coloured by the value of $\nu$ and the line-styles give the $\alpha$ values. In light grey, we show a Plummer profile and NFW profile.

and $\nu \to 1/2 - \alpha$ were taken, so the same order must be used for both quantities, otherwise the orthogonality relation will not hold.

## 5.4 Special cases

Our two-parameter family of expansions encompasses a number of well-known zeroth-order models as well as all the previously known families of spherical geometry biorthogonal basis expansions as special cases. In Fig. 5.2, we show the range of inner and outer slopes accessible with our two families of models along with the known families and other well-known zeroth-order models. We will discuss each of these known limits before presenting the new special cases encompassed by our family. Each special case is obtained from our general expressions (5.41) and (5.43) by setting the appropriate value of $\nu$.

### 5.4.1 The Zhao sequence

Zhao (1996) gives a family of basis sets whose zeroth orders correspond to his '$\alpha$' family of simple analytical potential-density pairs. This sequence of basis sets (see the expressions in Sec. 2.1.2) fits into our scheme by setting $\nu = 1$ and letting $\alpha$ remain arbitrary in either Family A or Family B.

In Zhao's case, both the density and potential basis functions reduce to Gegenbauer polynomials multiplied by the zeroth order term in the expansion. To see this in the case of the potential, we can apply a recurrence relation (G&R, Eq. 8.961(5)) to the Jacobi polynomial in Eq. (5.32), followed by the reduction to the Gegenbauer polynomials (G&R, Eq. 8.962(4)),

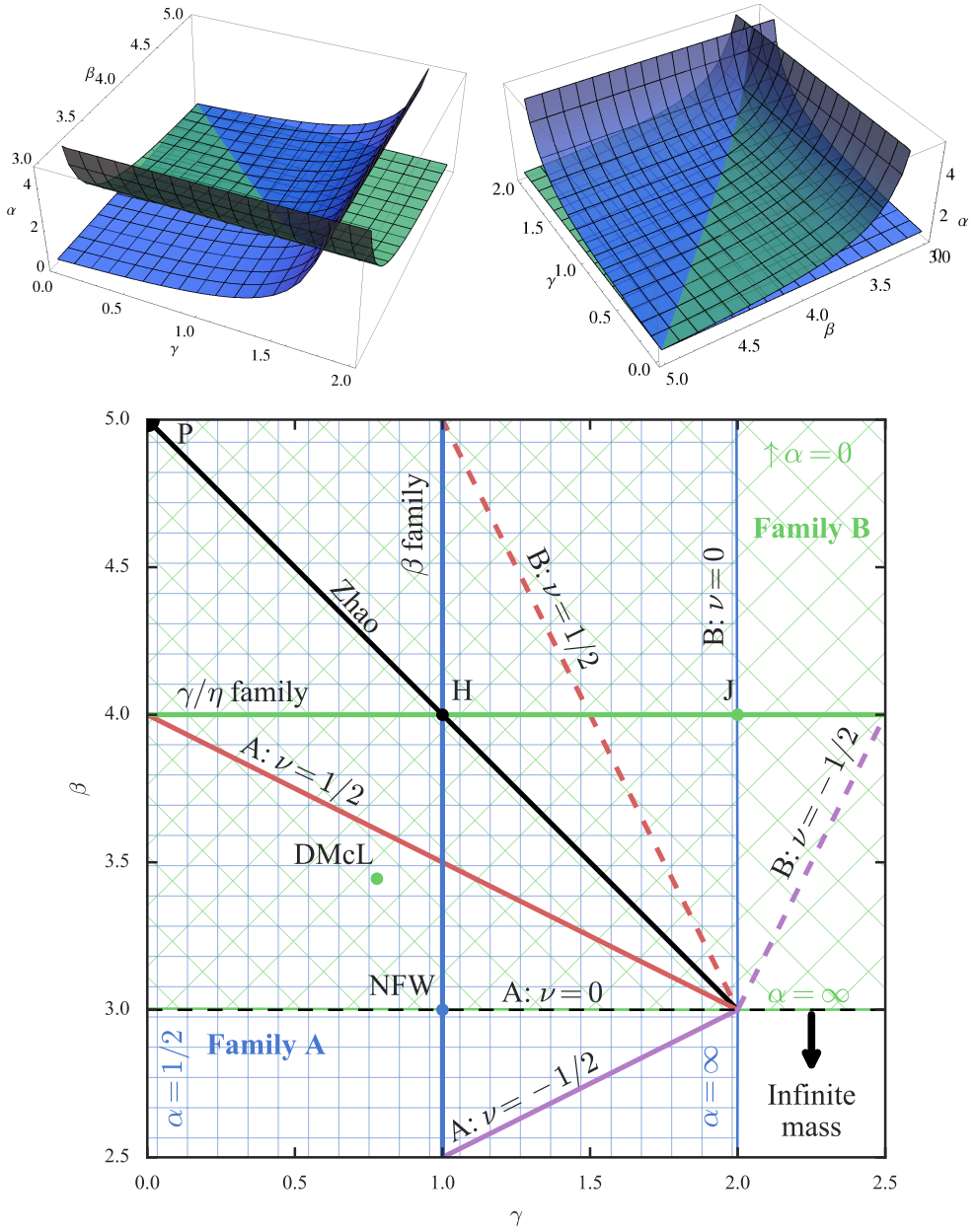


Figure 5.2: Upper Panels: Plots of the surfaces of Family A (blue) and Family B (green) in $(\alpha, \beta, \gamma)$ parameter space. The intersection of the two surfaces is the Zhao (1996) sequence. Lower Panel: Range of inner $\gamma$ and outer $\beta$ slopes encompassed by our basis expansion (blue square shading for Family A and green diagonal shading for Family B). This is the projection of the surfaces in the upper panels into the $(\beta, \gamma)$ plane. Subsets of these families are marked with solid lines: black shows the Zhao (1996) sequence (Family A and B), red and purple shows the Ch. 4 sequence (Family A, $\nu = \pm 1/2$). The red and purple dashed lines show the sequence on Family B with $\nu = \pm 1/2$. The blue vertical line shows Zhao's $\beta$ sequence in Family A, whilst the green horizontal line the $\gamma$ models of Dehnen (1993) and Tremaine et al. (1994) in Family B. Five specific models are shown by points: the NFW, the Plummer (P), the Hernquist (H), the Jaffe (J) and the Dehnen and McLaughlin (DMcL). The colour of the point indicates the Family in which they reside. For all these models, the methods of this chapter allow us to construct biorthogonal basis function expansions.

### 5.4.2 Super-NFW and associated models

When $\nu = \pm 1/2$ in Family A, we recover the expansion derived in Ch. 4. Using the properties of modified Bessel functions of half-integer order (DLMF, §10.47.9, 10.49.16), we see that for $\nu = \pm 1/2$, $\mathcal{K}_0(k)$ is proportional to $k^\mu \mathrm{e}^{-k}$. Up to a factor of $k$ this is the weight function for the generalised Laguerre polynomials, so natural choices for $g_n(k)$ (see (5.1)) are

$$\begin{aligned} g_n(k) &= k^{\mu-1}\, \mathrm{e}^{-k}\, L_n^{(2\mu-1)}(2k), \qquad \nu = 1/2, \\ g_n(k) &= k^{\mu-2}\, \mathrm{e}^{-k}\, L_n^{(2\mu-3)}(2k), \qquad \nu = -1/2. \end{aligned} \tag{5.47}$$

When $\nu = \alpha = 1/2$ the zeroth order is the perfect sphere of de Zeeuw (1985), as first derived by Rahmati & Jalali (2009). When $\nu = 1/2$ and $\alpha = 1$, the zeroth order is the super-NFW model (Ch. 3), which has the cosmologically important $1/r$ density cusp at its centre.

### 5.4.3 NFW and associated models

When we set $\nu = 0$ we obtain Family A expansions whose lowest-order densities all have outer slope $\beta = 3$, and Family B expansions with inner slope $\gamma = 2$. This set encompasses a number of well-studied and astrophysically interesting profiles. For example, when $\alpha = 1$ the beta function in the family A potential can be expressed as a logarithm, revealing the well-known NFW potential and density (Navarro et al., 1997)

$$\rho_{00} \propto \frac{1}{r(1+r)^2}, \qquad \Phi_{00} \propto \frac{-\ln(1+r)}{r}. \tag{5.48}$$

Furthermore, setting $\alpha = 1/2$ for Family A, we produce a basis set whose zeroth order is the modified Hubble profile and setting $\alpha = 1$ for Family B we find the zeroth order model is the Jaffe (1983) profile. See Sec. 5.5.1 for a note on computing the zeroth-order potential for this family of basis sets.

### 5.4.4 Elementary subsets

Zhao (1996) shows that there are four cases when the potentials of the double-power law family (Eq. 1.18) reduce to simpler analytical functions. These occur when combinations of $(\alpha, \beta, \gamma)$ take on integer values (we will use $k$ and $k'$ as integers).

The '$\alpha$' subset is obtained when $(\alpha, \beta, \gamma) = (\alpha, 3 + k'/\alpha, 2 - k/\alpha)$ with the '$\alpha$' family corresponding to $k = k' = 1$. Family A contains the members of the '$\alpha$' subset with $k = 1$ by choosing integer $\nu$; and Family B contains the members with $k' = 1$ also by choosing integer $\nu$. A related subset is obtained when $(\alpha, \beta, \gamma) = (\alpha, 2 + k'/\alpha, 3 - k/\alpha)$. Family A contains the members of this subset with $k' = \alpha + \nu$ and $k = 1 + \alpha$ restricting both $\alpha$ and $\nu$ to integer values. Similarly, Family B contains the members with $k' = 1 + \alpha$ and $k = \alpha + \nu$.

A further elementary subset is the '$\gamma$' subset where $(\alpha, \beta, \gamma) = (k, 3 + k'/k, \gamma)$. This subset contains the special case of the so-called $\gamma$ models (Dehnen, 1993, Tremaine et al., 1994) when $k = k' = 1$. Our Family B encompasses the set of models with $k' = 1$ by setting $\alpha = k$ and

leaving $\nu$ arbitrary. The final elementary subset is denoted the '$\beta$' family by Zhao (1996) where $(\alpha, \beta, \gamma) = (k', \beta, 2 - k/k')$. Family A encompasses the set of models with $k = 1$ by setting $\alpha = k'$ and leaving $\nu$ arbitrary. The special case of the '$\beta$' family when $k' = k = 1$ is discussed in more detail by Zhao.

Although Zhao identifies these further subsets of the double-power law family as possessing elementary potentials, he does not provide the corresponding biorthogonal basis sets. These are now accessible through our work.

Finally, we note that choosing $\alpha = 9/4$ and $\nu = 11/4$ for Family B we reproduce the Dehnen & McLaughlin (2005) models at zeroth order.

## 5.5 Numerical implementation

### 5.5.1 Beta functions

In order to evaluate the zeroth-order potential (5.7) numerically we need a numerical implementation of the incomplete beta function $\mathcal{B}_\chi(\mu, \nu)$ that covers the full parameter space. Common implementations of the incomplete beta function (e.g. GSL) only cover the case of strictly positive parameters $\mu, \nu$; we have $\mu \geq 1/2$ always, but must deal with the cases of zero or negative $\nu$.

When $-1 < \nu < 0$, we can manipulate the incomplete beta function as

$$\mathcal{B}_\chi(p, q) = \mathcal{B}_\chi(p, q+1)\,\frac{\mathrm{B}(p, q)}{\mathrm{B}(p, q+1)} - \frac{\chi^p (1-\chi)^q}{q}, \tag{5.49}$$

and use

$$\mathrm{B}(p, q) = \frac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)}\,\frac{p+q}{q} \text{ for } q < 0. \tag{5.50}$$

For $\nu = 0$, we must use a numerical implementation of the hypergeometric function, using the identity

$$\mathcal{B}_\chi(\mu, 0) = \frac{\chi^\mu}{\mu}\,{}_2F_1\left(\mu, 1; \mu+1; \chi\right), \tag{5.51}$$

or any equivalent transformation (DLMF, §8.17.7), unless $2\alpha$ is an integer (such as in the NFW case), in which case the incomplete beta function reduces to elementary functions at $l = 0$ and the higher-$l$ functions can be found using a recurrence formula (DLMF, §8.17.20).

### 5.5.2 Jacobi polynomials

To evaluate the higher order potential and density basis functions, we require a numerical implementation of the Jacobi polynomials $P_n^{(a,b)}(x)$. Our basis expansions are only valid for $a, b > -1$ which coincides with the domain of applicability in standard numerical implementations. It is efficient to use a recursion relation satisfied by the Jacobi polynomials to construct the ladder of

basis functions (G&R, Eq. 8.961(2))

$$\begin{aligned}&2(n+1)(n+a+b+1)(2n+a+b)P_{n+1}^{(a,b)}(x) = \\ &\quad (2n+a+b+1)\Big[(2n+a+b)(2n+a+b+2)x + a^2 - b^2\Big]P_n^{(a,b)}(x) \\ &\quad - 2(n+a)(n+b)(2n+a+b+2)P_{n-1}^{(a,b)}(x),\end{aligned} \tag{5.52}$$

with the lowest order polynomials given by

$$P_0^{(a,b)}(x) = 1; \quad P_1^{(a,b)}(x) = \frac{1}{2}(a-b+(2+a+b)x). \tag{5.53}$$

For the forces, we require the polynomial $P^{(a,b-1)}(x)$, which can be succinctly expressed as a sum or two polynomials of the form $P^{(a,b)}(x)$ (DLMF, §18.9.5),

$$P_n^{(a,b-1)}(x) = \frac{(n+a+b)P_n^{(a,b)}(x) + (n+a)P_{n-1}^{(a,b)}}{2n+a+b}. \tag{5.54}$$

A full computation of the potential, forces and density thus requires only a single ladder of Jacobi polynomials to be constructed.

### 5.5.3 Numerical aspects of the recurrence relation

The ladder of potential basis functions for increasing $n$ is built up using the inhomogeneous recurrence relation (5.41). As $n \to \infty$ the terms in this relation tend to zero, and the rate at which this happens increases greatly with increasing $l$. This causes the computation of the potential functions to become inaccurate when $n$ is high (due to the accuracy with which the beta function in the zeroth-order basis function can be computed). We can remedy this using the same method as Ch. 4 (see Sec. 4.3 for details). We pick some high order $N_{\max}$ for which the RHS of Eq. (5.41) is presumed to be approximately zero. We then recurse backwards, constructing the ladder of Jacobi polynomials with decreasing $n$. This avoids the issue of cancellation of large terms.

When computing the incomplete Beta function using a recurrence relation (which is possible when $2\alpha \in \mathbb{Z}$), catastrophic cancellation occurs when $r$ and $\nu$ are small and $l$ is high, leading to highly inaccurate results. It is therefore preferable to use a few terms of the Taylor expansion in this parameter regime. See Sec. 7.2.2.2 for a solution to this issue for the $\alpha = 1$ (generalised NFW) case.

## 5.6 A possible three-parameter family

Our new families of expansions lie along surfaces in the three-dimensional $(\alpha, \beta, \gamma)$ space. It is natural to ask whether our approach can be extended to cover the full 3D volume. Here, we suggest how to proceed based on the methodology employed above. We modify the $t$-dependent part of the integrand of Eq. (5.27) to read

$$f(t) = t^{\mu-1}\,\mathrm{e}^{-t}\,{}_1F_1\left(\begin{matrix}\lambda+1\\ \mu+1\end{matrix}\middle|\, t\right). \tag{5.55}$$

Now with $f(t)$ in the integrand of Eq. (5.27) we evaluate the integral (G&R, Eq. 7.622(1)), taking care to apply the Euler transformation (DLMF, §15.8.1(3)) before taking the limit $s \to 1$. The result is

$$\rho_l(r) = A(\mu,\nu,\lambda)\,\frac{r^{l-2-\lambda}}{(1+z^2)^{\mu+\nu-\lambda}}, \tag{5.56}$$

where

$$A(\mu,\nu,\lambda) \equiv \frac{\Gamma(\mu+1)\Gamma(\mu+\nu-\lambda)}{\Gamma(\mu+\nu+1)\Gamma(\mu-\lambda)}. \tag{5.57}$$

This is simply a reparameterisation of Eq. (1.18), so we have a double-power law model whose zeroth-order has inner slope $\gamma = 2+\lambda/\alpha$ and outer slope $\beta = 3+\nu/\alpha$. The potential corresponding to this density can be read off directly from Eq. (1.29),

$$\Phi_l(r) \propto r^{-l-1}\,\mathcal{B}_\chi(\mu-\lambda,\nu) + r^l\,\mathcal{B}_{1-\chi}(\mu+\nu,-\lambda)\,. \tag{5.58}$$

Note that the subset that remains invariant under the $A \leftrightarrow B$ transformation (Sec. 2.4) is defined by $\lambda = -\nu$. This is a two-dimensional surface, in which lies the line $\nu = 1$ that corresponds to Zhao's basis set.

Now, we can make a slightly more general choice of for our $t$-space auxiliary function, adding the parameter $j \in \mathbb{Z}$ to index a set of *non*-orthogonal basis functions. After a lengthy but detailed computation (see Appendix C.1), we obtain a surprisingly simple result:

$$\tilde{\rho}_{jl}(r) = A(\mu,\nu,\lambda)\frac{r^{l-2-\lambda/\alpha}}{\left(1+r^{1/\alpha}\right)^{\mu+\nu-\lambda}}\;{}_2F_1\left(\left.\begin{matrix}-j,\mu+\nu-\lambda\\ \mu-\lambda\end{matrix}\right|\frac{r^{1/\alpha}}{1+r^{1/\alpha}}\right), \tag{5.59}$$

which is proportional to the double-power law density (Eq. 5.56) multiplied by a polynomial in $\chi$. This is clearly analogous to our original two-parameter non-orthonormal density function (5.10), and reduces to it when $\lambda = -1$.

We have therefore generalised our non-biorthogonal density functions to a natural three-parameter family whose zeroth-orders cover the full range of double-power law forms. However, the continuation of our previous method requires that the overlap integral $\int r^2 \mathrm{d}r\,\rho_{nl}(r)\Phi_{n'l}(r)$ be expressible in a form that can be easily diagonalised, and this may be challenging – the evaluation of Eq. (2.102) for the new choice of $\tilde{f}_j(t)$ is a double hypergeometric series (see Appendix C.2). In fact, finding the non-orthogonal potential functions corresponding to $\tilde{f}_j(t)$ is already much more involved than for the density functions, requiring reductions of Appell's $F_2$ function – a natural consequence of the fact that the general double-power law potential (Eq. 5.58) required a sum of two terms.

## 5.7 Conclusions

The biorthogonal expansion series discovered by Hernquist & Ostriker (1992) has sometimes seemed miraculous. It has found widespread applications in astronomy (e.g. Barnes & Hernquist, 1992, Lowing et al., 2011, Ngan et al., 2015). This is because the zeroth order potential-density pair is the spherical Hernquist (1990) model, which is a reasonable representation of galaxies and

dark haloes. The expansion enables us to describe deviations from sphericity (like triaxiality or lopsidedness) very easily. The biorthogonality ensures that the expansion coefficients for both the potential and the density can be calculated easily from an $N$-body realisation.

We have studied the existence of biorthogonal basis function expansion methods for the general double-power law family of density profiles. They are parameterised by $(\alpha, \beta, \gamma)$, where $\beta$ and $\gamma$ are the (negative) logarithmic gradients of the central and asymptotic profile, whilst $\alpha$ controls the rate of transition between the inner and outer behaviour. We have presented an algorithm for constructing biorthogonal basis function expansions for two distinct families in $(\alpha, \beta, \gamma)$ space and provided closed analytical forms for the basis functions which may be efficiently computed via recursion relations. These results systematise all previously known biorthogonal basis function expansions for the spherical geometry, as discovered by Clutton-Brock (1973), Hernquist & Ostriker (1992), Zhao (1996), Rahmati & Jalali (2009), and the additional family described in Ch. 4. It also provides new expansions for a host of familiar models, including the $\gamma$ models of Dehnen (1993) and Tremaine et al. (1994), the Dehnen & McLaughlin (2005) models and the Jaffe (1983) model. Particularly significant in view of its cosmological importance is the Navarro, Frenk & White (1997) (NFW) model.

We have considered the possibility that our methodology can be followed to construct biorthogonal expansions for still more general zeroth-order potential-density pairs. In addition to the Bessel function solutions to the Poisson equation (Eq. 5.1), we have demonstrated that it can also be solved by a integral involving confluent hypergeometric functions (this led to the general discussion in Sec. 2.3.3). The form of this novel integral transform has guided us to some preliminary results (Sec. 5.6) that suggest that there is still more freedom in zeroth-order basis functions to be exploited. Therefore, it seems likely that – in addition to our Families A and B – further sequences exist for which the orthogonalisation procedure can be analytically carried out.

# Chapter 6

# A cuspy-exponential basis set

In Ch. 4 and Ch. 5 we exploited the integral transform methods developed in Ch. 2 to derive many new basis sets. We now show that those results lead naturally to a further family of basis sets, with properties unlike any described before in the literature

Specifically, we derive a family of basis sets that correspond to density models with a central cusp and an exponentially-decaying fall-off at large radii. Spherical models with an exponential factor are often used in dynamical models of dark haloes, arising either naturally or by giving a pre-existing model a smooth exponential truncation (Ciotti, 1991, Navarro et al., 2004, Einasto, 1965, Merritt et al., 2006). This basis set may be appropriate for density profiles that resemble such models at lowest order – one member of the new family strongly resembles an NFW model multiplied by an exponential factor.

Our new result is both 1) the result of taking the general basis set described in Ch. 5 in the limits $\nu, r_{\rm s} \to \infty$; and 2) the result of the method in Ch. 4 when making an alternative, related, choice of auxiliary function $g_n(k)$. We write the basis set using the same notational conventions as Sec. 5.2.2. The free parameters of the basis set family are the dimensionless quantity $\alpha$ and the scalelength $r_{\rm s}$. Explicitly, the radial potential basis functions are

$$\Phi_{0l} = \frac{\mu\,\gamma\big(\mu, (r/r_{\rm s})^{1/\alpha}\big)}{(r/r_{\rm s})^{l+1}}, \tag{6.1}$$

$$\Phi_{nl} - \Phi_{n+1,l} = \frac{2n!(-1)^n}{(\mu+1)_n}\left(\frac{r}{r_{\rm s}}\right)^l \mathrm{e}^{-(r/r_{\rm s})^{1/\alpha}} L_n^{(\mu)}\left(2(r/r_{\rm s})^{1/\alpha}\right), \tag{6.2}$$

and the density basis functions

$$\rho_{nl} = 2\,(-1)^n \left(\frac{r}{r_{\rm s}}\right)^{l-2+1/\alpha} \mathrm{e}^{-(r/r_{\rm s})^{1/\alpha}} \left[L_n^{(\mu)}\left(2(r/r_{\rm s})^{1/\alpha}\right) + L_{n-1}^{(\mu)}\left(2(r/r_{\rm s})^{1/\alpha}\right)\right], \tag{6.3}$$

where $\gamma(\alpha, z)$ is the (lower) incomplete Gamma function and $L_n^{(\alpha)}(x)$ are the Laguerre polynomials. The relevant constants are

$$N_{nl} = \frac{\alpha\Gamma(\mu+1)}{2^{\mu-1}}, \qquad K_{nl} = \frac{-n!\Gamma(\mu+1)}{8\pi\alpha^2\Gamma(n+\mu)}. \tag{6.4}$$

We now present two ways of deriving the functional form of this basis set; the first deriving from the method of Ch. 5, the second from that of Ch. 4.

## 6.1 Derivation via hypergeometric confluence

Starting from the Family 'A' basis expansion of Ch. 5, we take both $\nu \to \infty$ and $r_s \to \infty$, while keeping $r_s/\nu^\alpha$ constant. This result makes use of *hypergeometric confluence*, that takes a Gaussian ${}_2F_1$ hypergeometric function to a confluent ${}_1F_1$ hypergeometric function in a certain limit of its parameters,

$$\lim_{b\to\infty}\left\{{}_2F_1\left(\begin{matrix}a,b\\c\end{matrix}\middle|\frac{x}{b}\right)\right\} = {}_1F_1\left(\begin{matrix}a\\c\end{matrix}\middle|x\right). \tag{6.5}$$

To apply this to the Family 'A' basis functions, we introduce a scale-length that depends on $\nu$, making the replacement $r \to \nu^{-\alpha} r/r_t$; and then take the limit $\nu \to \infty$. Note that we must sometimes pre-multiply by factors of $\nu$ before taking the limit. We build up to the main result by steps. First, for the zeroth-order potential, use

$$\begin{aligned}\lim_{\nu\to\infty}\left\{z^{-2\mu}\mathcal{B}_\chi(\mu,\nu)\right\} &= \lim_{\nu\to\infty}\left\{\frac{1}{\mu}{}_2F_1\left(\begin{matrix}\mu,\mu+\nu\\\mu+1\end{matrix}\middle|\frac{-(r/r_t)^{1/\alpha}}{\nu}\right)\right\}\\ &= {}_1F_1\left(\begin{matrix}\mu\\\mu+1\end{matrix}\middle|-\left(\frac{r}{r_t}\right)^{1/\alpha}\right)\\ &= \left(\frac{r}{r_t}\right)^{1+2l}\gamma\left(\mu,(r/r_t)^{1/\alpha}\right),\end{aligned} \tag{6.6}$$

where $\gamma(\alpha, z)$ is the (lower) incomplete Gamma function. For the Jacobi polynomial terms, bearing in mind that $\chi \to z^2$ as $r_s \to \infty$, we use:

$$\begin{aligned}\lim_{\nu\to\infty}\left\{P_n^{(\mu+2\nu-1,\mu)}(\xi)\right\} &= \frac{(-1)^n(\mu+1)_n}{n!}\lim_{\nu\to\infty}\left\{{}_2F_1\left(\begin{matrix}-n,n+2\mu+2\nu\\\mu+1\end{matrix}\middle|\nu^{-1}\frac{(r/r_t)^{1/\alpha}}{1+\nu^{-1}(r/r_t)^{1/\alpha}}\right)\right\}\\ &= \frac{(-1)^n(\mu+1)_n}{n!}{}_1F_1\left(\begin{matrix}-n\\\mu+1\end{matrix}\middle|2\left(\frac{r}{r_t}\right)^{1/\alpha}\right)\\ &= (-1)^n L_n^{(\mu)}\left(2(r/r_t)^{1/\alpha}\right),\end{aligned} \tag{6.7}$$

where $L_n^{(\alpha)}(x)$ are the Laguerre polynomials. For the power-law terms, we use:

$$\lim_{\nu\to\infty}\left\{\left(1+\nu^{-1}(r/r_t)^{1/\alpha}\right)^{-\mu-\nu-1}\right\} = \mathrm{e}^{-(r/r_t)^{1/\alpha}}. \tag{6.8}$$

The overall result is therefore

$$\lim_{\nu\to\infty}\left\{\nu^{\alpha(l+1)}\Phi_{0l}\right\} = \frac{\mu\,\gamma\big(\mu,(r/r_t)^{1/\alpha}\big)}{(r/r_t)^{l+1}}, \tag{6.9}$$

$$\lim_{\nu\to\infty}\left\{\nu^{-\alpha l}\left(\Phi_{nl}-\Phi_{n+1,l}\right)\right\} = \frac{2n!(-1)^n}{(\mu+1)_n}\left(\frac{r}{r_t}\right)^l \mathrm{e}^{-(r/r_t)^{1/\alpha}} L_n^{(\mu)}\left(2(r/r_t)^{1/\alpha}\right), \tag{6.10}$$

$$\lim_{\nu\to\infty}\left\{\nu^{-\alpha(l-2)-3}\rho_{nl}\right\} = 2(-1)^n\left(\frac{r}{r_t}\right)^{l-2+1/\alpha} \mathrm{e}^{-(r/r_t)^{1/\alpha}} \tag{6.11}$$

$$\times\left[L_n^{(\mu)}\left(2(r/r_t)^{1/\alpha}\right)+L_{n-1}^{(\mu)}\left(2(r/r_t)^{1/\alpha}\right)\right], \tag{6.12}$$

$$\lim_{\nu\to\infty}\left\{\nu^{\alpha(2l+1)-1}N_{nl}\right\} = \frac{\alpha\Gamma(\mu+1)}{2^{\mu-1}}, \tag{6.13}$$

$$\lim_{\nu\to\infty}\left\{\nu K_{nl}\right\} = \frac{-n!\Gamma(\mu+1)}{8\pi\alpha^2\Gamma(n+\mu)}, \tag{6.14}$$

## 6.2 Derivation via Hankel transform

We could instead use the method described in Sec. 2.3.1 directly. We in fact closely follow the derivations of Ch. 4, by making a distinct but related choice[a] for the functions $g_n(k)$. We use $\alpha$ as a free parameter, write $z = r^{1/(2\alpha)}$ and $\mu = \alpha\,(1+2l)$, and define the basis functions $\Phi_{nl}$ and $\rho_{nl}$ as in Sec. 2.3.1. But here we choose a different $g_n(k)$,

$$g_n(k) \equiv L_n^{(\mu-1)}\left(k^2/2\right)\; k^{\mu-1}\exp(-k^2/4). \tag{6.15}$$

Using the change of variables $x = k^2/2$, we find that the potential and density basis functions are biorthogonal,

$$\int_0^\infty \mathrm{d}r\, r^2\,\Phi_{nl}\,\rho_{n'l} \propto \int_0^\infty \mathrm{d}k\, k\, g_n(k)\, g_{n'}(k) \propto \int_0^\infty \mathrm{d}x\, x^{\mu-1}\,\mathrm{e}^{-x}\, L_n^{(\mu-1)}(x)\, L_{n'}^{(\mu-1)}(x) \propto \delta_{nn'}. \tag{6.16}$$

The development of representations for $\Phi_{nl}$ and $\rho_{nl}$ mimics that of Sec. 4.2.

First we derive the potential basis functions. We note the following Hankel transforms (G&R, 2007, Eq. 6.631(4) & Eq. 6.631(5))

$$\int_0^\infty x^{\nu+1}\;\exp(-ax^2)\;J_\nu(bx) = \frac{b^\nu}{2^{\nu+1}\;a^{\nu+1}}\;\exp\left(-b^2/(4a)\right), \tag{6.17}$$

$$\int_0^\infty x^{\nu-1}\;\exp(-ax^2)\;J_\nu(bx) = 2^{\nu-1}\,b^{-\nu}\,\gamma\left(\nu,b^2/(4a)\right).$$

where $\gamma(s,x)$ is the lower incomplete gamma function[b]. We split the Laguerre polynomial into two parts, one containing the constant term and one containing the remainder

$$g_n(k) = k^{\mu-1}\;\exp(-k^2/4)\;\left[\binom{n+\mu-1}{n} - \frac{k^2}{2}\sum_{j=0}^{n-1}\frac{\binom{n+\mu-1}{n-1-j}}{(n-j)\,\binom{n}{j}}\,L_j^{(\mu)}(k^2/2)\right]. \tag{6.18}$$

[a] In fact this choice of $g_n(k)$ follows also from the method in Ch. 5, by using hypergeometric confluence on Eqs (5.26) and (5.27).

[b] Note that the second equation is given in error in the 2007 edition of G&R; other editions give the correct expression.

The constant term becomes

$$r^{-1/2}\,\binom{n+\mu-1}{n}\int_0^\infty k^{\mu-1}\,\exp(-k^2/4)\,J_\mu(kz)\,\mathrm{d}k = 2^{\mu-1}\,\binom{n+\mu-1}{n}\,\frac{\gamma(\mu,z^2)}{r^{1+l}}, \tag{6.19}$$

which has the correct limits given in Eq. (1.50) because $\gamma(s,x)\sim x^s/s$ as $x\to\infty$. To evaluate the other terms, we use the generating function Eq. (4.18) to transform Laguerre polynomials in $k^2/2$ into Laguerre polynomials in $2z^2$,

$$\sum_{n=0}^\infty t^n\int_0^\infty k^{\mu+1}\mathrm{e}^{-k^2/4}L_n^{(\mu)}\left(\frac{k^2}{2}\right)J_\mu(kz)\,\mathrm{d}k = \int_0^\infty k^{\mu+1}\frac{\exp\left(-k^2(1+t)/(4(1-t))\right)}{(1-t)^{\mu+1}}J_\mu(kz)\,\mathrm{d}k \tag{6.20}$$

$$= 2^{\mu+1}\,z^\mu\,\mathrm{e}^{-z^2}\frac{\exp\left(2tz^2/(1+t)\right)}{(1+t)^{\mu+1}}$$

$$= \sum_{n=0}^\infty(-t)^n\,2^{\mu+1}\,z^\mu\,\exp(-z^2)\,L_n^{(\mu)}(2z^2). \tag{6.21}$$

So putting the two parts together, we have

$$\Phi_{nl}(r) = -\frac{2^\mu\,(\mu)_n}{n!}\left[\frac{\gamma(\mu,r^{1/\alpha})}{r^{1+l}} - \frac{2}{\mu}\sum_{j=0}^{n-1}\frac{(-1)^j\,j!}{\Gamma(\mu+1)_j}\,r^l\,L_j^{(\mu)}(2\,r^{1/\alpha})\,\mathrm{e}^{-r^{1/\alpha}}\right]. \tag{6.22}$$

Under the summation sign, there is no coupling between $n$ and $j$, so each additional term in the potential does not need to be recomputed when we go from order $n$ to order $n+1$. Alternatively, we can use the Laguerre polynomial recurrence relation (Eq. 4.15) to write down the recurrence relation

$$n\,\Phi_{nl} = (n+\mu-1)\,\Phi_{n-1,l} + 2^\mu\,(-1)^n\,r^l\,\mathrm{e}^{-r^{1/\alpha}}\,L_n^{(\mu-1)}(2\,r^{1/\alpha}). \tag{6.23}$$

For the density basis functions, we can use the generating function (4.18) and Hankel transform (6.17) directly as

$$\sum_{n=0}^\infty t^n\,\rho_{nl}(r) = r^{1/\alpha-5/2}\sum t^n\int_0^\infty k^{\mu+1}\,\exp(-k^2/4)\,L_n^{(\mu-1)}(k^2/2)\,J_\mu(kz)\,\mathrm{d}k \tag{6.24}$$

$$= \frac{r^{1/\alpha-5/2}}{(1-t)^\mu}\int_0^\infty k^{\mu+1}\,\exp\left(-k^2(1+t)/4((1-t)\right)\,J_\mu(kz)\,\mathrm{d}k$$

$$= r^{1/\alpha-2+l}\,2^{\mu+1}\,\exp(-z^2)\,(1-t)\,\frac{\exp\left(2tz^2/(1+t)\right)}{(1+t)^{\mu+1}}$$

$$= r^{1/\alpha-2+l}\,2^{\mu+1}\,\exp(-z^2)\,(1-t)\sum_{n=0}^\infty t^n\,(-1)^n\,L_n^{(\mu)}(2z^2).$$

So the simplest form of the density basis functions is

$$\rho_{nl}(r) = \frac{2^{\mu+1}\,(-1)^n}{16\,\alpha^2\,\pi}\,r^{1/\alpha-2+l}\,\mathrm{e}^{-r^{1/\alpha}}\,\left[L_n^{(\mu)}(2\,r^{1/\alpha}) + L_{n-1}^{(\mu)}(2\,r^{1/\alpha})\right]. \tag{6.25}$$

The normalisation constant for the biorthonormal set, analogous to Eq. (4.29), is

$$N_{nl} = \frac{\alpha\,2^\mu\,\Gamma(n+\mu)}{n!}, \tag{6.26}$$

## 6.3 Properties of the family

The lowest-order density function is

$$\rho_{000} \propto r^{1/\alpha-2} \exp(-r^{1/\alpha}). \tag{6.27}$$

This is a member of the *generalised Einasto* family of profiles (see An & Zhao (2013) and our discussion in Sec. 1.3.1). There are several parameter values that give interesting functional forms of the basis functions; in particular, the models with $\alpha \approx 1/2$ most resemble the Einasto profiles (Einasto, 1965, Merritt et al., 2006).

Noting that the Sersic (1968) surface brightness law has functional form

$$\Sigma_{\rm Se}(R) \propto \mathrm{e}^{-kR^{1/n}}, \tag{6.28}$$

we also consider the *deprojected* Sersic profile $\rho_{\rm Se}(r)$ (Ciotti, 1991), which cannot be expressed in simple functions. It has the following asymptotic limits,

$$\rho_{\rm Se}(r) \longrightarrow \begin{cases} r^{\frac{1-m}{m}} \exp\left(-r^{1/m}\right) & \text{as } r \longrightarrow 0, \\ r^{\frac{1-2m}{2m}} \exp\left(-r^{1/m}\right) & \text{as } r \longrightarrow \infty. \end{cases} \tag{6.29}$$

Another approximation to this deprojection, which has significant errors only at the very centre, is provided by Prugniel & Simien (1997),

$$\rho_{\rm Se} \approx r^{-\gamma} \exp(-r^{1/m}), \qquad \gamma \approx 1 - \frac{0.59}{m} + \frac{0.055}{m^2}. \tag{6.30}$$

One might choose $\alpha$ in Eq. (6.27) to tune the inner cusp to match one or other of these limits, at the expense of the asymptotic behaviour of the other limit.

Resemblances to Einasto profiles notwithstanding, the $\alpha = 1/2$ basis set warrants further attention, so we devote the following section to it.

### 6.3.1 Gaussian basis set

Setting $\alpha = 1/2$ gives a zeroth-order density that is exactly a Gaussian, $\rho_{00} \propto \exp(-r^2)$. This is notable as sums of *different* Gaussians (also known as multi-Gaussian expansions) are often used to model galaxies (Emsellem et al., 1994).

In order to compare the predictions from dynamical modelling with observation, we must project the model's density distribution (as well other relevant quantities, e.g. the velocity dispersion) onto the sky. However, despite their universal popularity as 3D density profiles, projecting double-power law density distributions is fraught with analytical difficulties. The projection integral,

$$\Sigma(R) = \int_R^\infty \mathrm{d}x \frac{2x\,\rho(x)}{\sqrt{x^2 - R^2}}, \tag{6.31}$$

tends to give simple answers only in rather special cases. For example, the Plummer model projects (Dejonghe, 1987) to

$$\Sigma_{\rm Plu}(R) = \frac{1}{\pi} \frac{1}{(1+R^2)^2}. \tag{6.32}$$

Projecting similar models where the denominator contains a factor of $1 + r^2$ generally gives results which are elementary functions of $R^2$. Alternatively, models whose denominators involve a factor of $1 + r$, such as the Hernquist, NFW or super-NFW models, have projections given in terms of inverse trigonometric or even elliptical functions (see e.g. Eq. (3.26) for the projection of the sNFW model).

In summary we find that projecting the fairly general density $\rho \propto r^{-\gamma}\ (1 + r^2)^{\gamma-\beta}$ requires Gauss' ${}_2F_1$ hypergeometric function, and projecting the density $\rho \propto r^{-\gamma}\ (1 + r)^{\gamma-\beta}$ requires Clausen's ${}_3F_2$ hypergeometric function. The projection of the general three-parameter density (Eq. 1.18) has no solution in the realm of ordinary special functions, and as with the projection of the Einasto profile we are forced to consider ever more general families of special functions (such as the Meijer $G$-function, whose implementation is rarely found in standard numerical libraries). This is before even considering projecting any *basis sets* associated to these simple spherical models.

However, with the new Gaussian basis set described above, we have a solution: not only is the projection of the underlying zeroth-order density trivial (the Gaussian form is left unchanged), but the projection of every higher-order density basis function is analytically tractable, even incorporating the projection of the spherical harmonic factor. Projecting each density function along the $z$-axis involves computing an integral proportional to

$$\int_{-\infty}^{\infty} \rho_{nlm}(r,\theta,\phi)\,\mathrm{d}z \propto \int_{-\infty}^{\infty} \mathrm{d}z\ \left(R^2+z^2\right)^{l/2}\,\mathrm{e}^{-R^2-z^2}\left[L_n^{(l+1/2)}\left(2R^2+2z^2\right)\right.$$
$$\left.+L_{n-1}^{(l+1/2)}\left(2R^2+2z^2\right)\right]\,P_l^{(m)}\left(\frac{z}{\sqrt{R^2+z^2}}\right)\mathrm{e}^{im\phi}, \quad (6.33)$$

and we see that this integral consists of a sum of terms proportional to

$$\int_{-\infty}^{\infty} \mathrm{d}z\ \left(R^2+z^2\right)^{l/2}\ \mathrm{e}^{-z^2-R^2}\ z^{2j}\,P_l^{(m)}\left(\frac{z}{\sqrt{R^2+z^2}}\right), \qquad 0 \le j \le n. \quad (6.34)$$

Combining a hypergeometric definition of the associated Legendre function (DLMF, §14.3.18) with a connection formula (DLMF, §14.9.13), this integral is found to be proportional to

$$R^m\,\mathrm{e}^{-R^2}\int_{-\infty}^{\infty}\mathrm{d}z\,\mathrm{e}^{-z^2}\ z^{2j+l-m}\ {}_2F_1\left(\left.\begin{matrix}\frac{m-l}{2},\frac{m-l+1}{2}\\ m+1\end{matrix}\right|\frac{-R^2}{z^2}\right), \quad (6.35)$$

where, as we have $l \in \mathbb{Z}$ and $-l < m < l$, this new hypergeometric function is a terminating polynomial in $-R^2/z^2$, so integration with respect to $z$ is no issue. Therefore, projecting the basis functions gives us a result of the form

$$\int_{-\infty}^{\infty} \rho_{nlm}(r,\theta,\phi)\,\mathrm{d}z = p_{nm}(R^2)\ R^m\,\mathrm{e}^{-R^2}\,\mathrm{e}^{im\phi}, \quad (6.36)$$

where $p_{nm}(x)$ is some polynomial of degree $n$ with known coefficients determined by the simple integration procedure outlined above. Note that we are not limited to projection along the $z$-axis, as rotating spherical harmonics is a linear operation involving multiplication by the appropriate Wigner $d$-matrix. It would then be a simple operation to project the 3D basis functions $\rho_{nlm}(r,\theta,\phi)$

onto a suitable 2D basis $\Sigma_{nm}(R,\phi)$ – one natural choice (from the perspective of the computation above) for such a 2D basis set might be a combination of a Laguerre polynomial and a Gaussian weight,

$$\Sigma_{nm}(R,\phi) \propto R^m \, \mathrm{e}^{-R^2} \, L_n^{(m)}(2R^2) \, \mathrm{e}^{im\phi}. \tag{6.37}$$

The integral between these 2D functions (Eq. 6.37) and the projected 3D basis (Eq. 6.36) is clearly analytically tractable, and the matrix given by the overlap integral between each pair of 2D/3D basis functions would encode complete information about the projection degeneracies. Deprojection of an arbitrary model would involve computing the pseudo-inverse of this overlap matrix, and the entire range of deprojected models consistent with any given 2D data would then be obtained via standard methods of linear algebra.

Such an approach would usefully complement existing work on deprojection (van den Bosch et al., 2008, Van De Ven et al., 2008), which generally involves approximating density profiles as multi-Gaussians; and modelling line-of-sight velocity distributions using Hermite polynomials or other non-parametric methods (van der Marel & Franx, 1993, Merritt & Saha, 1993). In fact, performing a combined series expansion (in Gaussians multiplied by Hermite/Laguerre polynomials) for both the velocity and spatial parts of the distribution function would be very powerful, and is not without precedent, e.g. in the plasma physics literature (Manzini et al., 2016).

### 6.3.2 Razor-thin disks

Qian (1993) derived a basis set suitable for Gaussian disks using an unusual integral-transform method derived from that of Kalnajs (1976) (see Sec. 2.1.1 for our discussion). One may ask if the choice of auxiliary function $g_n(k)$ made in Sec. 6.2 has a natural analogue in the disk setting, giving us a Gaussian basis set via Clutton-Brock's original Hankel-transform method. The answer is a qualified yes, as we shall now demonstrate. Compared to the method of Sec. 6.2, one additional integral must be used (G&R, Eq. 6.631(1)),

$$\int_0^\infty x^\mu \, \exp(-ax^2) \, J_\nu(bx) = \frac{b^\nu \, \Gamma\left(\frac{\nu+\mu+1}{2}\right)}{2^{\nu+1} \, a^{\frac{\mu+\nu+1}{2}} \, \Gamma(\nu+1)} \, {}_1F_1\left(\begin{matrix}\frac{\nu+\mu+1}{2} \\ \nu+1\end{matrix}\middle|\, \frac{-b^2}{4a}\right). \tag{6.38}$$

Recalling the expressions Eq. (2.6) for the potential-density functions on a razor-thin disk, and making the choice of auxiliary function

$$g_n(k) \equiv k^m \, \mathrm{e}^{-k^2/2} \, L_n^{(m)}(k^2)\,, \tag{6.39}$$

we are led straightforwardly to the zeroth-order potential and density,

$$\begin{aligned}\psi_{0m}(R) &= \frac{\Gamma(m+1/2)}{\sqrt{2}\,\Gamma(m+1)} \, R^m \, {}_1F_1\left(\begin{matrix}m+1/2 \\ m+1\end{matrix}\middle|\, \frac{-R^2}{2}\right), \\ \sigma_{0m}(R) &= R^m \, \mathrm{e}^{-R^2/2}.\end{aligned} \tag{6.40}$$

The potential $\psi_{0m}$ appears complicated due to the presence of the confluent hypergeometric function ${}_1F_1$. However, for any angular order $m$ this expression can be expressed as a sum of $m$

Bessel $I_\alpha$ functions, using the expansion given by Luke (1959, Eq. 1.8). Using results for the asymptotic behaviour of the modified Bessel functions (DLMF, §10.30), we can verify that this potential obeys the correct boundary conditions (1.50) – note these are the three-dimensional boundary conditions, *not* the two-dimensional ones, because although the expansion is confined to a disk we are still solving for the full three-dimensional gravitational potential, even though we are ultimately only evaluating it at $z = 0$.

The density basis functions for $n > 0$ have a very simple analytical form. Following a method analogous to that of Sec. 6.2, we apply the standard Laguerre polynomial generating function (4.18), and find that

$$\sum_{n=0}^{\infty} t^n \, \sigma_{nm}(R) = R^m \, (1+t)^{-m-1} \, \exp\left(-R^2 \, \frac{1-t}{1+t}\right) = R^m \, \mathrm{e}^{-R^2} \sum_{n=0}^{\infty} (-t)^n \, L_n^{(m)}\left(2R^2\right), \tag{6.41}$$

meaning the density functions can be written in closed-form,

$$\sigma_{nm}(R) = (-1)^n \, R^m \, \mathrm{e}^{-R^2} \, L_n^{(m)}\left(2R^2\right). \tag{6.42}$$

Unfortunately, attempting the same method for the $n > 0$ potential basis functions results in Laguerre polynomials of half-integer order. As with the zeroth-order potential (Eq. 6.40) these can in principle be expressed in series of modified Bessel functions, but for now the question of finding a more efficient way of expressing them than Qian (1993) remains open.

One might also hope that the method of Ch. 5 might be applied in the disk setting, essentially writing $g_n(k)$ as a linear combination of Bessel $K_\nu$ functions. And indeed, we find that that the analogous zeroth orders are precisely the Kuzmin-Toomre disks – such a basis set being Qian's other major result. However, the derivation of a closed-form expression for the higher-order basis functions does not fall to quite the same generating function approach as Ch. 5, so this must also be left to a future effort.

Chapter 7

# Efficient representation of evolving dark matter haloes

Cosmological dark matter haloes follow a density law of approximate double-power form. This was first suggested by Dubinski & Carlberg (1991), but made famous by Navarro et al. (1997), who introduced the eponymous NFW density profile(see e.g. Mo et al., 2010, for a useful summary). Subsequent work showed that the slopes of the inner and outer power laws have some scatter about the canonical NFW values (Moore et al., 1998, Klypin et al., 2001, Diemer & Kravtsov, 2014, Dekel et al., 2017), and even that the logarithmic gradient of the density slope may change with radius leading to an Einasto profile rather than double-power laws (Einasto & Haud, 1989, Merritt et al., 2006).

Even so, these laws are really no more than convenient fitting formulas that provide a zeroth order approximation to the dark halo density. Cosmological simulations have long shown that dark haloes are more complicated than simple spherical models. Triaxiality, shape or ellipticity variations with radius, substructure and lopsidedness are all manifestations of the hierarchical assembly of galaxies via merging and accretion (e.g. Moore et al., 1999, Jing & Suto, 2002, Prada et al., 2019). This has detectable consequences – for example, streams caused by dwarf galaxies and globular clusters disrupting in lumpy haloes have markedly different morphologies to those disrupting in smooth haloes with idealised profiles (Ngan et al., 2015). Observationally, too, there are now clear indications that dark haloes have rich and complex shapes, which encode the physical processes that made them. The modelling of long thin streams in the Milky Way halo such as the Orphan Stream has shown the importance of the gravitational effects of the Large Magellanic Cloud (Erkal et al., 2019). This large satellite galaxy is in the process of merging with the Milky Way, and its gravitational pull causes both tidal distortions in the halo and reflex motion of the halo centre. Equally, the stream from the disrupting Sagittarius galaxy in the Milky Way cannot be fit by a potential with fixed triaxial shape (see Law & Majewski, 2010b, Belokurov et al., 2014),

but requires the dark halo shape to change from oblate to triaxial in the outer parts (Vera-Ciro & Helmi, 2013). Similarly, the absence of fanning in the Palomar 5 stream suggests a nearly spherical potential in the inner Milky Way, whereas some triaxiality is required to reproduce the morphology of the Sagittarius stream (Pearson et al., 2015). The description of the kinematics of stars and dark matter in the Milky Way galaxy then requires a much more elaborate dark matter potential than just a static, symmetric halo model with fixed shape.

This chapter develops the idea of describing time-evolving dark matter haloes using basis function expansions with time-varying parameters. The potential and density are written generically as

$$\begin{aligned} \Phi(\mathbf{r}) &= \sum_{lm} B_{lm}(r)\, Y_{lm}(\theta,\phi), \\ \rho(\mathbf{r}) &= \sum_{lm} A_{lm}(r)\, Y_{lm}(\theta,\phi). \end{aligned} \tag{7.1}$$

Two approaches for representing the radial dependence of the spherical-harmonic coefficients, $A_{lm}(r)$ and $B_{lm}(r)$, have been explored in the literature: either as a weighted sum of orthonormal functions involving polynomials of degree $n$ in a scaled radial variable (as discussed extensively in the preceding chapters of this thesis), or as interpolated functions defined by values at $n$ radial grid points. Both of these methods of expansion can encode complex shape variation, together with arbitrary inner and outer density fall-offs for the halo.

The variety of applications for any basis function method is very substantial, as already articulated clearly by Lowing et al. (2011) and Ngan et al. (2015). Once different snapshots of a numerical simulation are expressed in basis function expansions, the time evolution of the simulation can be recreated using interpolation. The simulations can then be replayed efficiently many times with completely new objects inserted. This makes it ideal for studying myriads of problems in galaxy evolution and near-field cosmology, including the disruption of satellites and subhaloes, the precession of tidal streams and the build-up of the stellar halo. Provided the mass ratio of accreted object to host halo is less than 0.1, the effects of dynamical friction are unimportant (Boylan-Kolchin et al., 2008) and the inserted object has no back-reaction on the rest of the simulation. However, to convert this powerful idea into an efficient working tool requires addressing a number of questions:

1) Cosmological haloes participate in the large-scale Hubble flow. They are not isolated but feel the external tidal forces from larger scale structure, as well as the buffeting of frequent accretion events. The integration of orbits in the basis function expansion must also take account of these effects, if the orbits in the simulation are to be recovered accurately. How should they be modelled?
2) Suppose snapshots of an evolving numerical halo are available at fixed times as basis function expansions. An approximation to the state of the halo at intermediate times is recovered by interpolating the coefficient of each basis function between the preceding and following snapshots (Lowing et al., 2011). How do the choices of time interval between snapshots and interpolation scheme affect the orbit recovery? This can be answered by comparing

orbits integrated in this time-varying basis function approximation with the original $N$-body trajectories.

3) Which expansion is optimal for a given simulation? Previous applications of this idea have routinely used the familiar Hernquist-Ostriker biorthogonal expansion (e.g. Lowing et al., 2011, Ngan et al., 2015), but there are now many more options available (Ch. 4–Ch. 6; also e.g. Vasiliev (2013)). This necessitates the development of an error measure for the evolving haloes based on the fidelity of orbit reconstruction, to assess the competing methods.

This chapter is arranged as follows. Sec. 7.1 recapitulates the biorthogonal and spline expansion methods. Sec. 7.2 explains in detail the construction of both basis function expansions for one numerical halo, describing their usage in a time-evolving setting. Finally, Sec. 7.3 discusses the accuracy of the resulting halo representations, and which parameters can be adjusted in order to achieve the optimal speed/accuracy trade-off.

## 7.1 Basis function methods

There are two choices for the radial dependence of the spherical-harmonic coefficients $A_{lm}(r), B_{lm}(r)$ introduced in Eq. (7.1). Both methods express the coefficients as a convergent series indexed by $n$, truncated at some order $n = n_{\max}$. The first method represents each term as a weighted sum of biorthogonal functions of degree $n$, expressed in terms of a scaled radial variable. The second method uses interpolating functions defined by values at an arbitrary set of $n$ radial grid points. We now detail the specifics of each approach.

### 7.1.1 Biorthogonal expansions

In the first approach, using an biorthogonal basis function expansion or basis set expansion (BSE), we write Eq. (7.1), including time-dependence, as

$$\Phi(\mathbf{r};t) = \sum_{nlm} C_{nlm}(t)\,\Phi_{nlm}(r)\,Y_{lm}(\theta,\phi), \tag{7.2}$$

$$\rho(\mathbf{r};t) = \sum_{nlm} C_{nlm}(t)\,\rho_{nlm}(r)\,Y_{lm}(\theta,\phi). \tag{7.3}$$

The basis functions remain time-independent, so they satisfy the Poisson and orthogonality properties discussed previously (Eqs (1.35) and (1.36)).

This is efficient if the expansion captures at zeroth-order the spherically-averaged density profile of a cosmological dark-matter halo. Any deviations are then succinctly described by a small number of the higher-order terms in the basis set.

Biorthogonal functions have some considerable advantages: 1) The recurrence relations for such polynomials enable the higher order basis functions to be calculated rapidly from the low order ones. 2) The reconstructed potential and density are represented by the same coefficients, which are easy to compute as weighted sums over the number of particles. When the density is formed from a cloud of point particles of mass $m_i$ as in a numerical simulation, the integral for the

coefficients reduces to a sum (see Sec. 1.4)

$$C_{nlm}(t) = \int \mathrm{d}^3\mathbf{r} \;\; \Phi_{nlm}(\mathbf{r})\, \rho(\mathbf{r};t) = \sum_i m_i \, \Phi_{nlm}(\mathbf{r_i}(t)). \tag{7.4}$$

The biorthogonality ensures that all the calculations are linear with respect to the particles. 3) Because the potential, forces and density are all linear with respect to the same set of coefficients, the time-interpolated forces can be calculated directly from the time-interpolated coefficients, thus halving the number of force evaluations required.[a] 4) The basis functions are infinitely differentiable, which means arbitrarily-high order integration schemes may be used (we do not take advantage of this in the current work however).

### 7.1.2 Spline expansions

The second approach is to represent the radial dependence of each term in the expansion as numerically-interpolated functions on a radial grid. This idea has its roots in $N$-body simulations (e.g. Aarseth 1967, McGlynn 1984, Sellwood 2003, Meiron et al. 2014), and as a computationally inexpensive way of solving the Poisson equation in Schwarzschild or made-to-measure modelling (Valluri et al. 2004, de Lorenzi et al. 2007, Siopis et al. 2009), often with a restriction to axisymmetry. The coefficients of the angular spherical harmonic expansion are evaluated at a small number of radial grid points, and the radial dependence of forces is then interpolated (typically linearly) between grid nodes.

Vasiliev (2013) suggested using splines to represent the radial basis functions $A_{lm}(r)$ and $B_{lm}(r)$ in Eq. (7.1). In the most recent version of the algorithm (Vasiliev, 2019), the potential coefficients are represented by quintic splines in a suitably scaled radial coordinate, so that the derivatives of the potential up to second order are twice continuously differentiable. The number and positions of nodal points can be chosen arbitrarily (typically a logarithmic radial grid is used), so the method is in principle very flexible. To construct a potential from a given smooth density profile, the latter is expanded in spherical harmonics, and then the Poisson equation is solved by 1D radial integration of each term. When an $N$-body snapshot is used as an input, the spherical-harmonic expansion of its density profile is constructed by penalised least-square fitting, as detailed in the appendix of Vasiliev (2018). Differently from a biorthogonal basis function expansion, the evaluation of the potential and forces at a given point depends only on the coefficients at a few nearby nodes rather than on the whole basis set. However, the computation of the expansion is no longer linear, due to the need to carry out a penalised least squares computation to calculate the parameters of each spline; and due to various scaling transformations designed to improve the accuracy of interpolation.

[a] Half the force evaluations are saved for linear interpolation; the savings are more dramatic for higher-order interpolation schemes.

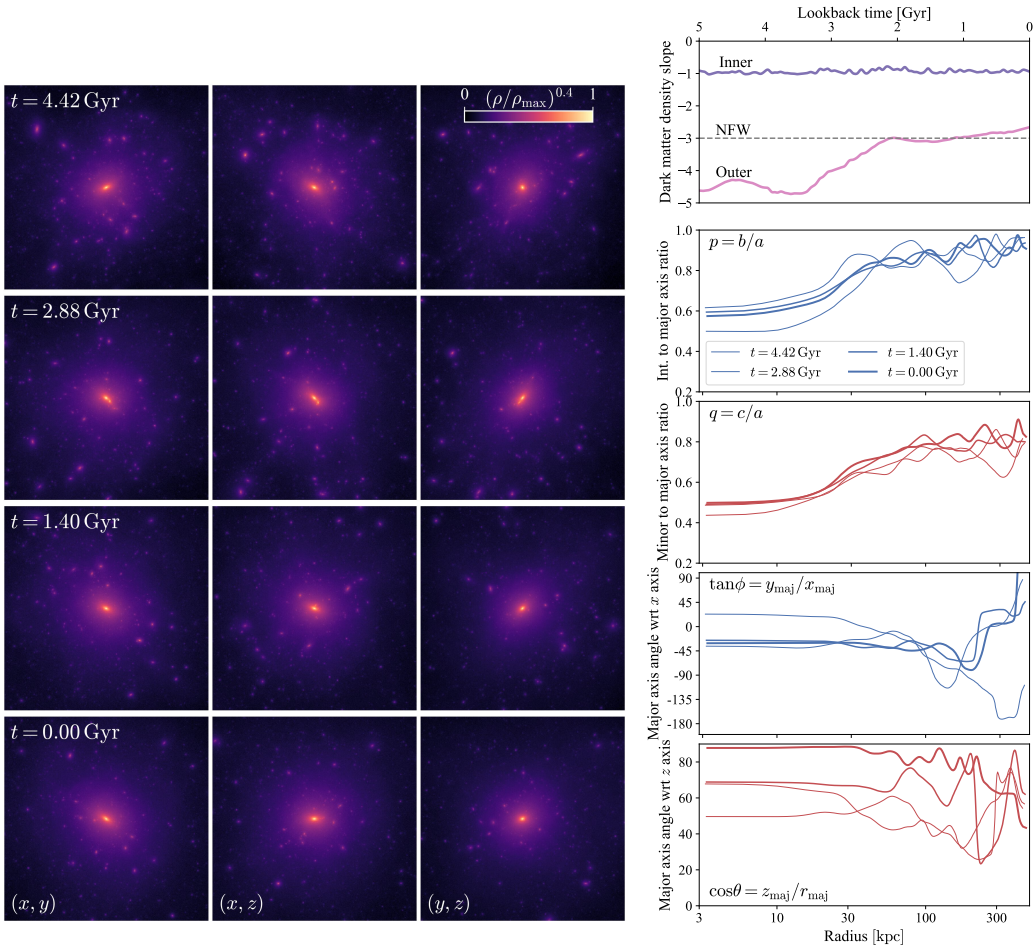


Figure 7.1: Density of the studied dark matter halo at four snapshots. The left set of images shows the projected density. Each column displays a 500 kpc × 500 kpc projection of the halo (left: $(x, y)$, middle: $(x, z)$, right $(y, z)$). Each row is labelled by the look-back time. Note the time dependence of the large scale morphology. On the right we display the dark matter density slope as a function of time in the top panel and in the bottom four panels we show the axis ratios and the direction of the major axis at each radius for the four snapshots (thicker lines are later times).

## 7.2 Application to a time-evolving halo

We now turn to the application of the two methods to a simulated dark matter halo. We first describe the details of the simulation, before describing specific implementation choices for the two methods.

### 7.2.1 A Milky Way-like dark matter halo

#### 7.2.1.1 The density of the halo

From the several simulated Milky Way-like haloes described in Sec. 4.3.3, we select a single halo as our benchmark model and we analyse it in detail. The halo we select corresponds to Halo #1

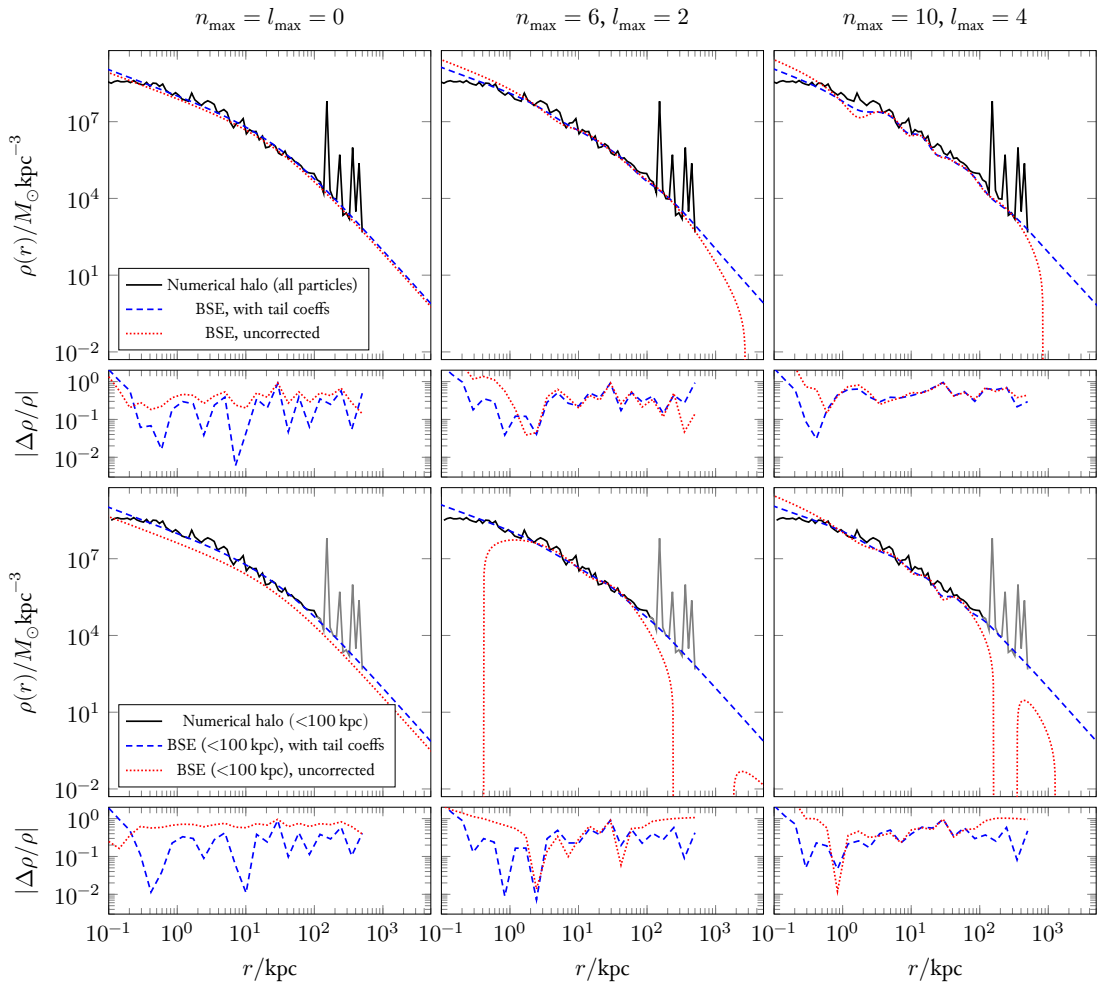


Figure 7.2: Radially-averaged density profiles of the halo at 11.8 Gyr (at which point it resembles an NFW profile), along with reconstructions using the BSE with parameters $\nu = 0$, $r_s = 30$ (the NFW case). The large spikes in the original halo density between 100–500 kpc correspond to substructure. The upper three panels show successively more detailed reconstructions, using $n_{max} = l_{max} = 0$ (a single term), $l_{max} = 2$ (63 terms) and $l_{max} = 10$ (275 terms). The residuals are shown in the sub-panels underneath. Notice that, without the tail correction, there are oscillatory artefacts in the density caused by the finite extent of the simulation data. The lower three panels magnify this effect by using approximately $1/4$ of the particles – those from within $< 100$ kpc – to compute the BSE coefficients. Neglecting the unresolved substructure, the corrected coefficients still provide a reasonable approximation outside of this range.

from that chapter, and its properties at $z = 0$ are given in the first row of Table 4.1. We focus on the final 5 Gyr of evolution, as prior to this the evolution was more tumultuous and dominated by significant merger events. The halo's density profiles at four different snapshots over the last 5 Gyr are shown in Fig. 7.1. The halo contains $1.3 \times 10^7$ particles and has a virial scale-length $r_{\rm vir} = 325$ kpc and a concentration $c = 9.6$. We measure the inner and outer density slopes from a histogram of particles, using 0.08–4 kpc, and 200–500 kpc respectively. As shown in Fig. 7.1, the inner slope is cusped with inner density slope $\gamma \approx 1$, consistent with the NFW model of Eq. (1.15), whilst the outer slope evolves from a steeper fall-off of $\beta \approx 4.5$ to the NFW value of $\beta = 3$ for the last 2 Gyr of evolution. At all times, the halo has an approximately triaxial density distribution characterised by axis ratios $p$ in the $(x, y)$ plane and $q$ in the $(x, z)$ plane. The semiaxes $p$, $q$ and the direction of the major axis as a function of radius at four snapshots are shown in Fig. 7.1. These quantities are computed from the moment of inertia for particles binned by their local density (employing a local density threshold to remove subhaloes). At all times, the central parts of the halo are more flattened ($p \approx 0.6$ and $q \approx 0.5$) than the outer parts ($p \approx 0.9$ and $q \approx 0.8$). The shape evolution over time is quite mild but the alignment of the major axis shows the halo tumbles significantly over the last $\sim 4$ Gyr.

#### 7.2.1.2 The forces on the particles

Any method that reconstructs the force on each particle in the halo must contend with the fact that the halo is a non-inertial reference frame, as the centre of the coordinate system is at each step centred according to the cusp of the density distribution as found using ROCKSTAR. We detail here the computation of the fictitious force arising from the non-inertial frame.

The comoving coordinate of the halo centre is $\mathbf{x}(t) \equiv \mathbf{r}(t)/a(t)$, where $a(t)$ is the cosmological scale factor. The peculiar velocity (the physical velocity $\dot{\mathbf{r}}$ minus the Hubble flow) is $\mathbf{u} = \dot{\mathbf{r}} - H(t)\,\mathbf{r} = \dot{\mathbf{x}}\,a(t)$, where $H(t) \equiv \dot{a}(t)/a(t)$ is the Hubble parameter. Finally, the acceleration of the reference frame associated with the halo centre is simply $\dot{\mathbf{u}}$, which is calculated numerically. The force on the test particle is therefore

$$\mathbf{F}(\mathbf{x}, t) = -\boldsymbol{\nabla}\Phi(\mathbf{x}, t) - \dot{\mathbf{u}}(t). \tag{7.5}$$

where $\Phi(\mathbf{x}, t)$ is the halo potential (as reconstructed by a basis function expansion). Eq. (7.5) now accounts for the forces on the halo overall, including those due to large-scale structure in the cosmological simulation. It neglects tidal effects at the scale of the halo itself, since the corresponding term $\mathbf{x}(t)\,\ddot{a}(t)/a(t)$ is several orders of magnitude smaller than the total force. We stress the importance of taking the acceleration due to the non-inertial reference frame into account: without the second term in the above equation, the agreement between the trajectories computed in the smooth halo potential and the original $N$-body simulation is much worse. When comparing trajectories of test particles to those taken from the original simulation, we correct each original particle's velocity by subtracting off the velocity $\mathbf{u}$ of the halo reference frame.

### 7.2.2 Implementation of the basis set method

#### 7.2.2.1 Choice of expansion parameters

Ch. 5 showed that there exist two families ('A' and 'B' ) of biorthogonal basis functions, lying on distinct, intersecting surfaces in the $(\alpha, \beta, \gamma)$ parameter space of the general double-power law model (Eq. 1.18). The biorthogonal families each have two parameters, $\alpha$ and $\nu$. Here, $\alpha$ corresponds exactly to the double power law $\alpha$ parameter, whilst $\beta$ and $\gamma$ are related to $\nu$ via $\gamma = 2 - 1/\alpha$ and $\beta = 3 + \nu/\alpha$ for Family 'A' and $\gamma = 2 - \nu/\alpha$ and $\beta = 3 + 1/\alpha$ for 'B'. Certain combinations of parameter values give basis sets which can be expressed entirely with elementary functions (algebraic operations as well as the logarithm and gamma functions):

1) The one-parameter family of Zhao (1996) arises as the intersection of the 'A' and 'B' families. It is obtained from either family by setting $\nu = 1$ and leaving $\alpha$ arbitrary.
2) The subset of the 'A' family obtained by setting $\alpha = 1$ gives basis sets corresponding to the 'generalised NFW' (Evans & An, 2006) models, lying along $(\alpha, \beta, \gamma) = (1, 3 + \nu, 1)$. This gives basis sets with flexible outer slopes, including the important NFW case ($\nu = 0$).
3) Similarly, the subset of the 'B' family obtained by setting $\alpha = 1$ gives basis sets corresponding to the '$\gamma$' models (Dehnen, 1993, Tremaine et al., 1994), lying along $(\alpha, \beta, \gamma) = (1, 4, 2-\nu)$. These basis sets have a flexible inner slope.

These three one-parameter sets of basis sets stand out from the wider 'A' and 'B' families. They have the enormous advantage of speediness, as there is no need to call a computationally-expensive special function for each potential function evaluation.

The value of $\alpha$ controls not just the width of the turn-over region, but also the spacing of the zeroes of the polynomials used in the higher-order terms of the expansion. The argument of the polynomials is $r^{1/\alpha}/(1 + r^{1/\alpha})$ giving rise to a spacing of the zeros of $\Delta \ln r \sim \alpha$. This heuristic argument shows that achieving optimum accuracy requires $\alpha \approx 1$, with an acceptable range of around $\alpha = 0.7$–$2$, outside of which the expansions become inefficient. In practice, this limits the flexibility of the Zhao expansions (apart from the widely-used Hernquist & Ostriker expansion, which corresponds to the choice $\alpha = 1$ and so does obey this constraint). The halo we analyse here has a roughly constant inner slope of $r^{-1}$, so the generalised NFW models are expected to provide superior performance compared to the $\gamma$ and Zhao families. Sec. 7.2.2.2 summarises the essential formulas for the generalised NFW expansions as specialised from Ch. 5, which we make use in the rest of the chapter.

There is a final independent parameter, the scale-length $r_s$. The accuracy of a truncated expansion is not strongly dependent on the choice of $r_s$. In our experiments, we find that the scale-length in the expansion $r_s$ must be set to a reasonable value, $r_s \approx (\nu + 1) r_{\rm iso}$, where $r_{\rm iso} = r_{\rm vir}/c$ is the radius at which the logarithmic slope of the (spherically averaged) halo density attains the isothermal value of $-2$. The expansion becomes severely inaccurate if $r_s$ is less than a few percent of $r_{\rm vir}$, but otherwise the exact choice of $r_s$ is not important.

#### 7.2.2.2 Generalised NFW basis functions

Here we give the potential and density basis functions of the generalised NFW basis set. They are derived from the result of Ch. 5 by setting the parameter $\alpha = 1$.The zeroth-order potential is given by

$$\Phi_{00} = \begin{cases} \dfrac{\log(1+r)}{r}, & \text{if } \nu = 0 \\ \dfrac{1-(1+r)^{-\nu}}{\nu\, r}, & \text{otherwise.} \end{cases} \tag{7.6}$$

so that $\nu = 0$ corresponds to the NFW model. At fixed $r$, higher-order terms in $l$ of the potential are given by the recurrence relation

$$\begin{aligned} \Phi_{0,l+1} &= \frac{f_l}{r}\left\{\Phi_{0l} - \frac{r^l}{(1+r)^{1+2l+\nu}}\left[\frac{1+2l+\nu}{2+2l}\,\frac{r}{1+r}+1\right]\right\}, \\ f_l &= \frac{(3+2l)(2+2l)}{(1+2l+\nu)(2+2l+\nu)}. \end{aligned} \tag{7.7}$$

Note that, while formally correct, this recurrence relation suffers from catastrophic cancellation when $l$ is high and $r$ is low. In these situations it is therefore more accurate to use a few terms of the following Taylor expansion

$$\begin{aligned} \Phi_{0l} \approx \frac{r^l}{(1+r)^{2l+\nu}}\Bigg[1 - \frac{(1-\nu)}{2+2l}r + \frac{(1-\nu)(2-\nu)}{(2+2l)(3+2l)}r^2 \\ \ldots + \frac{(1-\nu)\ldots(j-\nu)}{(2+2l)\ldots(j+1+2l)}(-r)^j\Bigg]. \end{aligned} \tag{7.8}$$

A suitable algorithm to compute $\Phi_{0l}$ to at least 6 digits of accuracy over the entire parameter space covered in this chapter would be

$$\Phi_{0l}(r) = \begin{cases} \text{if } r \lesssim 10^{-4/(l+1)}, \text{use (7.8), keeping terms up to } j = 4, \\ \text{otherwise use (7.7).} \end{cases} \tag{7.9}$$

Higher-order terms in $n$ of the potential are given by the recurrence relation

$$\Phi_{n+1,l} = \Phi_{nl} - \frac{2n!}{(2+2l)_n}\,\frac{r^l}{(1+r)^{1+2l+\nu}}\,P_n^{(2l+2\nu,2l+1)}(\xi), \tag{7.10}$$

where $P_n^{(\alpha,\beta)}(x)$ are the Jacobi polynomials and $\xi = (r-1)/(r+1)$. Similarly, the radial component of acceleration is given by

$$\begin{aligned} \Phi'_{nl} = -\frac{(1+l)\Phi_{nl}}{r} + \frac{A_{nl}r^{l-1}}{(1+r)^{1+2l+\nu}}\Big[(n+4l+2\nu+1)P_n^{(2l+2\nu,2l+1)}(\xi) \\ -(n+2l+2\nu)P_{n-1}^{(2l+2\nu,2l+1)}(\xi)\Big], \end{aligned} \tag{7.11}$$

where $A_{nl} \equiv (1+2l)/(2n+4l+2\nu+1)$; and the density functions are given by

$$\rho_{nl} = \frac{r^{l-1}}{(1+r)^{2+2l+\nu}}\Big[(n+4l+2\nu+1)(n+2l+\nu+1)P_n^{(2l+2\nu,2l+1)}(\xi) - (n+2l+2\nu)(n+2l+\nu)P_{n-1}^{(2l+2\nu,2l+1)}(\xi)\Big]. \tag{7.12}$$

In this way, the potential, acceleration and density functions may be constructed (for a given $l$) from a single ladder of recursively-computed Jacobi polynomials $P_n^{(2l+2\nu,2l+1)}(\xi)$. In principle one could find a recursion relation that connects basis functions of consecutive $l$ (at constant $n$). However this would not lead to any additional savings, as recursion in $l$ at $n = 0$ followed by recursion in $n$ is already optimal, requiring $nl$ steps in total.

The associated constants $N_{nl}$ and $K_{nl}$ are

$$N_{nl} = \frac{(2l+1)!}{(n+2l+2\nu+1)_{2l}}, \qquad K_{nl} = -\frac{n!(2l+1)}{4\pi(2n+4l+2\nu+1)(2l+1)_n}. \tag{7.13}$$

For these models the inner slope is fixed at $\gamma = 1$, and the parameter $\nu$ adjusts the outer slope $\beta$, so we could alternatively use this as the free parameter, writing $\beta = 3+\nu$. Note that the 'B' basis sets, with fixed outer slope $\beta = 4$ and variable inner slope $\gamma = 2-\nu$, can be obtained by the transformations $\Phi_{nl}(r) \mapsto r^{-1}\Phi_{nl}(r^{-1})$ and $\rho_{nl}(r) \mapsto r^{-5}\rho_{nl}(r^{-1})$.

#### 7.2.2.3 The outer tails of the expansion

In practice, a snapshot of a simulated halo has a truncation radius $r_{\rm t}$, beyond which there are no particles. This is artificially introduced due to our cutout scheme (our halo data is truncated at $r_{\rm t} = 500\,$kpc). The naive use of biorthogonal expansions on this data results in artefacts: spikes of negative density at very large and very small radii are produced at higher expansion orders ($n_{\rm max} > 10$), in a manner analogous to the Gibbs phenomenon that occurs when a finite number of terms in a Fourier series is used to resolve a jump discontinuity. There is also a severe under-estimate of the radial acceleration when using just the first few series coefficients ($n_{\rm max} \lesssim 5$). This arises as much of the mass of the expansion lies outside of $r_{\rm t}$, but the total mass is by construction the same as that of the data. Examples of these artefacts are visible in the 'uncorrected' curves in the upper panels of Fig. 7.2. The lower three panels amplify this effect by using only the $3.7\times10^6$ particles found within $<100\,$kpc to compute the coefficients. The lower panels also illustrate a second important pitfall: the basis expansion tries to reproduce the hard cut-off at $r_{\rm t}$, rather than the desired asymptotic power-law behaviour.

Our strategy for solving this problem is the extrapolation of the $N$-body data beyond the truncation radius $r_{\rm t} = 500\,$kpc, assuming it follows a power law. This is accomplished by adding to each coefficient a fixed quantity $T_{nlm}$. Multiple evaluations of the series do not require any additional calculations, so this computational effort scales only with the number of terms in the truncated series. Denoting the 'uncorrected' coefficients by $C_{nlm}^{\rm orig}$, the corrected coefficients are

$$C_{nlm} = C_{nlm}^{\rm orig} + \mathcal{A}T_{nlm}, \tag{7.14}$$

where $\mathcal{A}$ is a normalisation constant that ensures that the mass interior to a chosen radius matches that of the $N$-body data $M_{\rm enc}^{\rm N-body}$. This quantity is given by

$$\mathcal{A} = \frac{M_{\rm enc}^{\rm N-body}(r_{\rm t}) - M_{\rm enc}^{\rm orig}(r_{\rm t})}{M_{\rm enc}^{\rm tail}(r_{\rm t})}, \tag{7.15}$$

where $M_{\rm enc}^{\rm orig}(r_{\rm t})$ is the mass interior to $r_{\rm t}$ in the naïve, uncorrected expansion; and $M_{\rm enc}^{\rm tail}(r_{\rm t})$ is the mass contribution due to correction coefficients $T_{nlm}$. An argument motivating the method can be found in Sec. 7.2.2.4, and expressions for the quantities $T_{nlm}$ may be found in Appendix D.1.

We show the results of applying this procedure to the halo in Fig. 7.2, noting how the outer tail of the expansion is more reasonably handled. At all radii the density error is reduced (particularly noticeable when only considering particles with $r < 100\,\mathrm{kpc}$).

This ruse of extrapolating the asymptotic power-law behaviour of the density beyond the truncation radius allows for the use of infinite-extent basis functions on a finite region. Previously the only basis functions for use on a finite region were the spherical Bessel functions (Polyachenko & Shukhman, 1981), these having the disadvantage that they do not resemble any simple halo or bulge profile.

#### 7.2.2.4 Derivation of tail coefficients

Here we derive expressions for the above-mentioned adjustments to the expansion coefficients, correcting for the hard truncation in the particle distribution.

Our $N$-body halo consists of $N$ particles with masses $m_j$ at positions $\mathbf{r}_j$ (with $r_j \leq r_{\rm t}$). For the region $r_{\rm t} < r < \infty$, we affix to the $N$-body realisation $\hat{\rho}(\mathbf{r})$ an analytical 'tail' density corresponding to the underlying zeroth-order density model of our chosen basis set:

$$\hat{\rho}(\mathbf{r}) = \begin{cases} \sum_j m_j \delta^3(\mathbf{r} - \mathbf{r}_j), & \text{if } r \leq r_{\rm t} \\ \mathcal{A}\rho_{000}(r), & \text{if } r > r_{\rm t}, \end{cases} \tag{7.16}$$

where $\mathcal{A}$ is a constant that sets the normalisation of this tail profile. Because Eq. (7.16) defines a linear adjustment to the data, the coefficients of the basis expansion can now simply be linearly corrected to take into account the tail density. We denote the original 'uncorrected' coefficients by $C_{nlm}^{\rm orig}$, and the coefficients corresponding to the tail density by $T_{nlm}$:

$$\begin{aligned} C_{nlm}^{\rm orig} &\equiv \sum_i m_i \Phi_{nlm}(\mathbf{r}_i), \\ T_{nlm} &\equiv \int_{r>r_{\rm t}} \Phi_{nlm}(\mathbf{r})\, \rho_{000}(\mathbf{r})\, \mathrm{d}^3\mathbf{r}\ , \\ C_{nlm} &\equiv C_{nlm}^{\rm orig} + \mathcal{A} T_{nlm}. \end{aligned} \tag{7.17}$$

In order to fix the parameter $\mathcal{A}$, we pick a radius $R$ and constrain the expansion to have the same mass interior to $R$ as the $N$-body halo – this could be any radius, but in practice we use $R = r_{\rm t}$. Denoting by $M_{nlm}(R)$ the mass enclosed at radius $R$ by the $n$-th basis function (an analytical

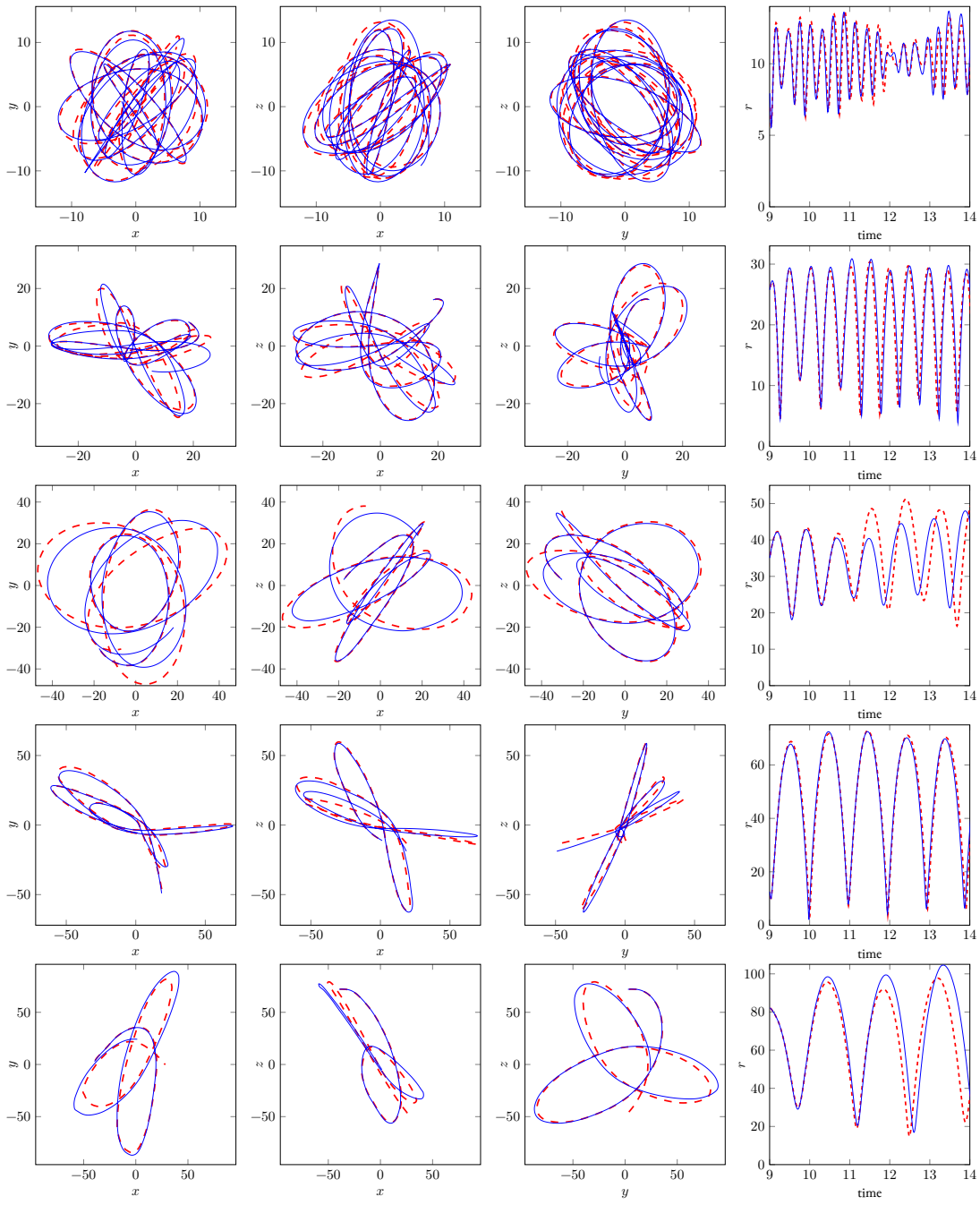


Figure 7.3: Examples of reconstructed orbits (solid blue lines) compared to the original trajectories of particles in the simulation (red dashed lines), for the spline method with $l_{\max} = 10$. Each row plots a single orbit, with the first three columns showing its projections on three principal planes, and the last column – time evolution of the galactocentric radius. Orbital period increases from top to bottom, and we illustrate both good cases (rows 1, 2 and 4), which are more common, and occasional bad reconstructions, usually caused by a single scattering event.

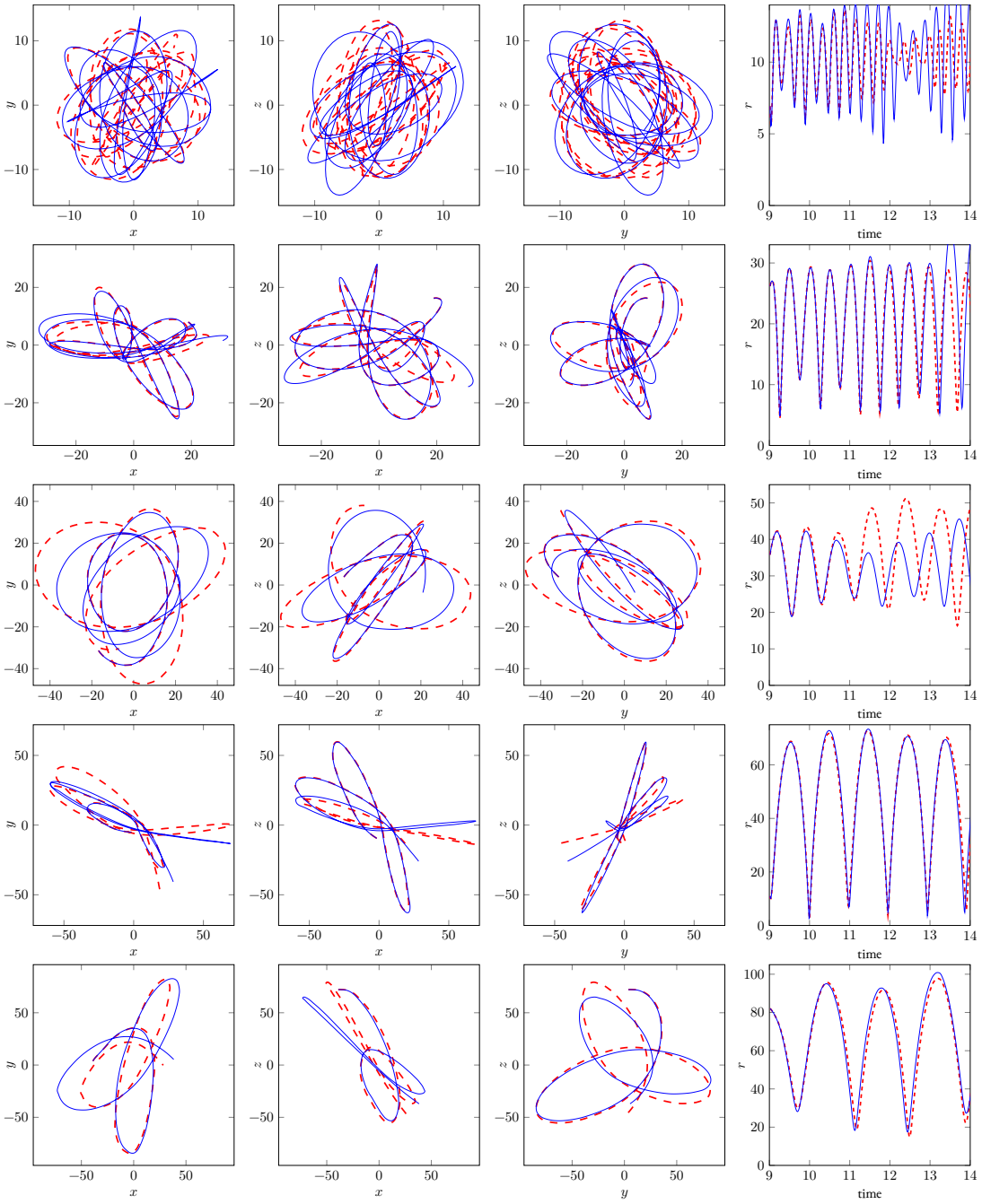


Figure 7.4: Similar to Fig. 7.3, but showing the results of the BSE reconstruction with $l_{\max} = 10$ and $n_{\max} = 22$ (solid blue lines) compared to the original trajectories (red dashed lines). Top to bottom show 5 different particles with increasing orbital period. Left to right show the three principal planes followed by galactocentric radius. As with the spline method there is a mix of good and bad reconstructions. The overall performance is very similar to the spline method.

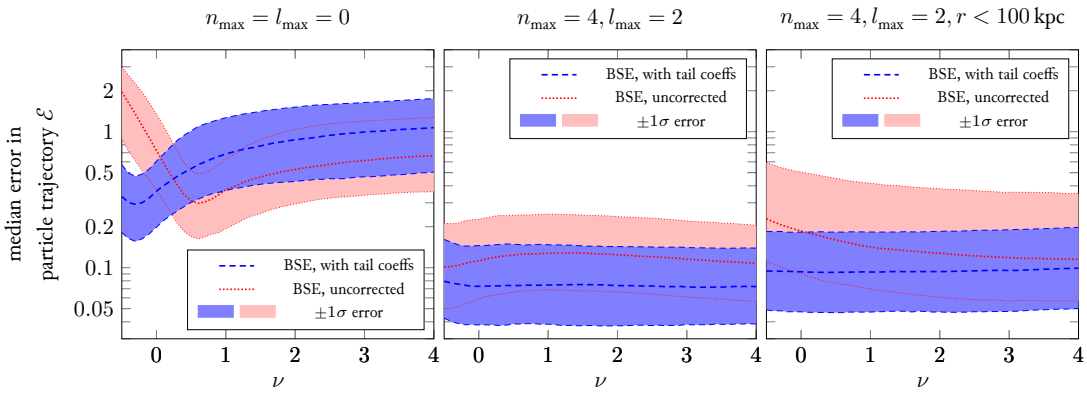


Figure 7.5: The run of the error measure $\mathcal{E}$ with the free parameter of the basis set $\nu$. The outer logarithmic slope of the density profile is $3 + \nu$ such that an NFW halo has $\nu = 0$. For each one we fix the scale-length such that the isothermal length $r_{\mathrm{iso}} = 30\,\mathrm{kpc}$. The shaded region covers the 16th–84th percentile in $\mathcal{E}$. The dotted red lines refer to expansions which omit the 'tail' correction discussed in Sec. 7.2.2.4. First two panels: the basis expansion as calculated from the full halo. Third panel: the expansion as calculated from the halo truncated at $r_{\mathrm{t}} = 100\,\mathrm{kpc}$. For this figure we start the particle trajectories at $t \approx 11.8\,\mathrm{Gyr}$ when the halo strongly resembles an NFW halo.

quantity that is non-zero only when $m = l = 0$), we have:

$$M_{nlm}(R) \equiv \int_{r<R} \rho_{nlm}\,\mathrm{d}^3\mathbf{r} \;= \delta_{l0}\delta_{m0}R^2\,\frac{\mathrm{d}\Phi_{n0}(r)}{\mathrm{d}r}\bigg|_{r=R}. \tag{7.18}$$

Then let $M_{\mathrm{enc}}^{\mathrm{true}}(R)$ and $M_{\mathrm{enc}}^{\mathrm{orig}}(R)$ be the mass interior to $R$ of the corrected and uncorrected expansions, respectively:

$$\begin{aligned} M_{\mathrm{enc}}^{\mathrm{true}}(R) &\equiv \int_{r<R} \rho(\mathbf{r})\,\mathrm{d}^3\mathbf{r} \;= \sum_n C_{n00}M_{n00}(R), \\ M_{\mathrm{enc}}^{\mathrm{orig}}(R) &\equiv \int_{r<R} \rho^{\mathrm{orig}}(\mathbf{r})\,\mathrm{d}^3\mathbf{r} \;= \sum_n C_{n00}^{\mathrm{orig}}M_{n00}(R). \end{aligned} \tag{7.19}$$

Let $M_{\mathrm{enc}}^{\mathrm{tail}}(R)$ be the (unnormalised) mass of the tail portion of the density profile

$$M_{\mathrm{enc}}^{\mathrm{tail}}(R) \equiv \sum_n T_{n00}(r_{\mathrm{t}})M_{n00}(R), \tag{7.20}$$

so we clearly have

$$\begin{aligned} M_{\mathrm{enc}}^{\mathrm{true}}(R) &= \sum_n \left[C_{n00}^{\mathrm{orig}} + \mathcal{A}T_{n00}\right] M_{n00}(R), \\ &= M_{\mathrm{enc}}^{\mathrm{orig}}(R) + \mathcal{A}M_{\mathrm{enc}}^{\mathrm{tail}}(R). \end{aligned} \tag{7.21}$$

And finally let $M_{\mathrm{enc}}^{\mathrm{N-body}}(R)$ be the mass of the halo (interior to $R$) as found by counting the $N$ particles in the simulation,

$$M_{\mathrm{enc}}^{\mathrm{N-body}}(R) \equiv \sum_{r_j<R} m_j. \tag{7.22}$$

Then to fix the value of $\mathcal{A}$, we simply require that $M_{\rm enc}^{\rm true}(R) = M_{\rm enc}^{\rm N-body}(R)$, giving

$$\mathcal{A} = \frac{M_{\rm enc}^{\rm true}(R) - M_{\rm enc}^{\rm orig}(R)}{M_{\rm enc}^{\rm tail}(R)}. \tag{7.23}$$

The quantities $T_{nlm}$ can be calculated in advance, and formulas may be found in Appendix D.1 which hold good for the generalised NFW basis set used in the main body of this chapter. Analogous formulas that cover the full parameter space of possible basis sets may be found in Appendix D.2, and an example implementation written in C may be found in Appendix D.3.

### 7.2.3 Implementation of the spline method

Unlike the biorthogonal expansion, in the spline-interpolated multipole approach implemented in Agama (Vasiliev, 2019), the radial dependence of each spherical-harmonic term of the density expansion is represented by its values at a predefined grid of points in radius. There is still a considerable freedom in assigning the location of grid nodes, but the most natural choice is to use a uniformly-spaced grid in $\log r$ with a fixed ratio between successive grid points $\mathcal{R} \equiv r_{i+1}/r_i$. In this case, the 'relative resolution' (the radial extent of the smallest representable feature divided by its distance from origin) is constant across the entire system. In particular, the radial and angular resolutions roughly match when $\log \mathcal{R} \approx 2.5/l_{\rm max}$. The minimum/maximum grid radii are usually chosen to enclose almost all particles in the system, leaving out only a few dozen particles – just enough to reliably estimate the asymptotic slope of the density profile at small or large radii. The density is extrapolated as a power-law in radius outside the grid. Typical grid sizes are 20–30 radial points covering several decades in radius, and the accuracy starts to deteriorate remarkably when using fewer than 15 points.

To construct the smooth density profile from an $N$-body snapshot, Agama uses penalised spline fits with an automatic choice of smoothing parameters. Namely, for the $l = 0$ (spherically symmetric) term, the logarithm of the density $\log \rho_{000}(r)$ is represented as a cubic spline in $\log r$, with the coefficients and the smoothing parameter (which penalises large fluctuations) determined by minimising the leave-one-out cross-validation score. The higher-order multipole terms are normalised by the value of the $l = 0$ term, and a penalised smoothing spline over the same radial grid is constructed from the multipole coefficients of each particle. These procedures are detailed in the appendix of Vasiliev (2018), and their cost is linear in both the number of particles and the size of the grid.

After a smooth multipole representation $A_{lm}(r)$ of the density is constructed, the corresponding potential terms $B_{lm}(r)$ and their radial derivatives at each grid node $r_i$ are computed by a one-dimensional integration:

$$B_{lm}(r_i) = \frac{4\pi G}{2l+1} \left[ r_i^{-l-1} \int_0^{r_i} A_{lm}(r)\, r^{l+2}\, dr + r_i^{l} \int_{r_i}^{\infty} A_{lm}(r)\, r^{1-l}\, dr \right]. \tag{7.24}$$

The multipole terms of the potential are interpolated as 1D quintic splines in $\log r$ defined by their values and derivatives at grid points. In doing so, the $l \neq 0$ terms are additionally scaled by the

value of the $l = 0$ term, and the latter is logarithmically scaled. Another, more efficient 2D quintic interpolation scheme is used when $l_{\text{max}} > 2$, representing each azimuthal Fourier harmonic term $B_m(r, \theta)$ on a 2D grid in $(r, \theta)$. All these scalings, together with the use of penalised spline fits for the density, break the linearity of the potential representation, but in practice the effect of this is negligible for a large enough $N$-body system.

### 7.2.4 Time evolution

Following the simulation over a range of times requires an approach to interpolating the potential expansions between the fitted snapshots.

In the biorthogonal expansion approach, we consider all the time-dependence in the gravitational force to be due to the series coefficients

$$\mathbf{F}_{\text{BSE}}(\mathbf{x}, t) = -\sum_{nlm} C_{nlm}(t) \boldsymbol{\nabla} \Phi_{nlm}(\mathbf{x}), \tag{7.25}$$

and so in order to get the force at intermediate times (say between halo snapshots at $t_1$ at $t_2$), we interpolate the coefficients,

$$C_{nlm}(t) = \tau(t) C_{nlm}(t_1) + (1 - \tau(t))\, C_{nlm}(t_2), \tag{7.26}$$

where $\tau(t)$ is a function that satisfies $\tau(t_1) = 1$ and $\tau(t_2) = 0$. So for linear interpolation, we use

$$\tau(t) = (t - t_2)/(t_1 - t_2). \tag{7.27}$$

This can be straightforwardly extended to higher-order interpolation, which remains linear in the coefficients $C_{nlm}$. For example, for cubic interpolation, the evaluations are taken at four consecutive times $t_0, t_1, t_2, t_3$, and the interpolated coefficient is $C_{nlm}(t_1)$ when $\tau = 1$ and $C_{nlm}(t_2)$ when $\tau = 0$, but otherwise depends on all four values of $t$ via the Lagrange interpolating polynomial.

As the acceleration is linear in the coefficients, the force due to interpolating the coefficients is equal to that which would result if we calculated the forces first and then interpolated. The fictitious force due to the halo reference frame $\dot{\mathbf{u}}$ is known in advance, and so is simply interpolated in the same way as the coefficients and added on at every time-step.

In the spline approach, the force is calculated first and then linearly interpolated (although higher-order schemes are equally possible):

$$\mathbf{F}_{\text{spline}}(\mathbf{x}, t) = \tau(t) \mathbf{F}_{\text{spline}}(\mathbf{x}, t_1) + (1 - \tau(t)) \mathbf{F}_{\text{spline}}(\mathbf{x}, t_2). \tag{7.28}$$

and the fictitious force is treated in the same way as above.

## 7.3 Performance of the expansions

With the implementation details established, we now turn to the question of how successfully the potential expansions can emulate properties of the simulation. In general, we want any expansion

to successfully reproduce the paths of particles in the simulation, at least in a statistical sense. We therefore opt to inspect a fixed, but representative, sampling of particles within the simulation and test whether their orbits are reproduced (see Lowing et al., 2011, for a similar discussion of the Aquarius simulations). We begin by defining our orbit sample.

### 7.3.1 Orbits

We consider a subset of particles from the original simulation satisfying the following criteria:

- the galactocentric radius never exceeds 200 kpc and is below 100 kpc in the last snapshot;
- the orbital period is less than 3 Gyr;
- the particle does not belong to any subhalo at the initial moment ($t \simeq 9$ Gyr), meaning that it is neither gravitationally bound to it, nor resides within 10 scale radii of the subhalo.

Approximately 20% of all particles in the simulation satisfy these conditions, from which we randomly pick approximately 2000 particles.

Fig. 7.3 shows five example orbits from the simulation, compared to the reconstructed orbits using the Spline expansion at the highest order, with $n_{\max} = 40$, $l_{\max} = 10$. The BSE method with $n_{\max} = 22$, $l_{\max} = 10$ gives extremely similar results, and is shown in Fig. 7.4. From visual inspection, the majority of orbits in our sample are reproduced fairly well over many orbital periods, at least when considering overall orbit parameters such as the peri- and apocentre radii, although the actual trajectories start to diverge due to slight phase differences at later times. The spline method tends to perform slightly better on short-period orbits, and the BSE method on those with long periods.

Occasionally, a particle from the original simulation may experience a close encounter with a subhalo or some other feature not reproduced by the reconstructed potential, after which the two trajectories diverge more strongly. Even though we illustrate these cases in two of the five rows of Figs 7.3 and 7.4, these strong perturbations are actually much rarer.

We now quantitatively inspect the reproduction of our chosen orbit sample, concentrating on 1) the difference between the two expansions, 2) the variation in accuracy with specific parameter choices in the potential expansions, 3) the accuracy with which different types of orbits are reproduced. For this discussion, we require the introduction of a measure of the quality of orbit recovery.

### 7.3.2 Error measure

The error in potential/density approximations has previously been studied with the mean integrated square error or MISE (e.g. Hall, 1983, Silverman, 1986, Vasiliev, 2013). This involves integrating the squared magnitude of the absolute difference in the density or the acceleration field between the reconstructed and original halo over its entire spatial extent. This is best suited to static haloes rather than evolving ones.

Therefore, in order to test the fidelity of a given potential expansion of a time-evolving halo, we instead use the *relative position error* of reconstructed orbits. We define the relative position

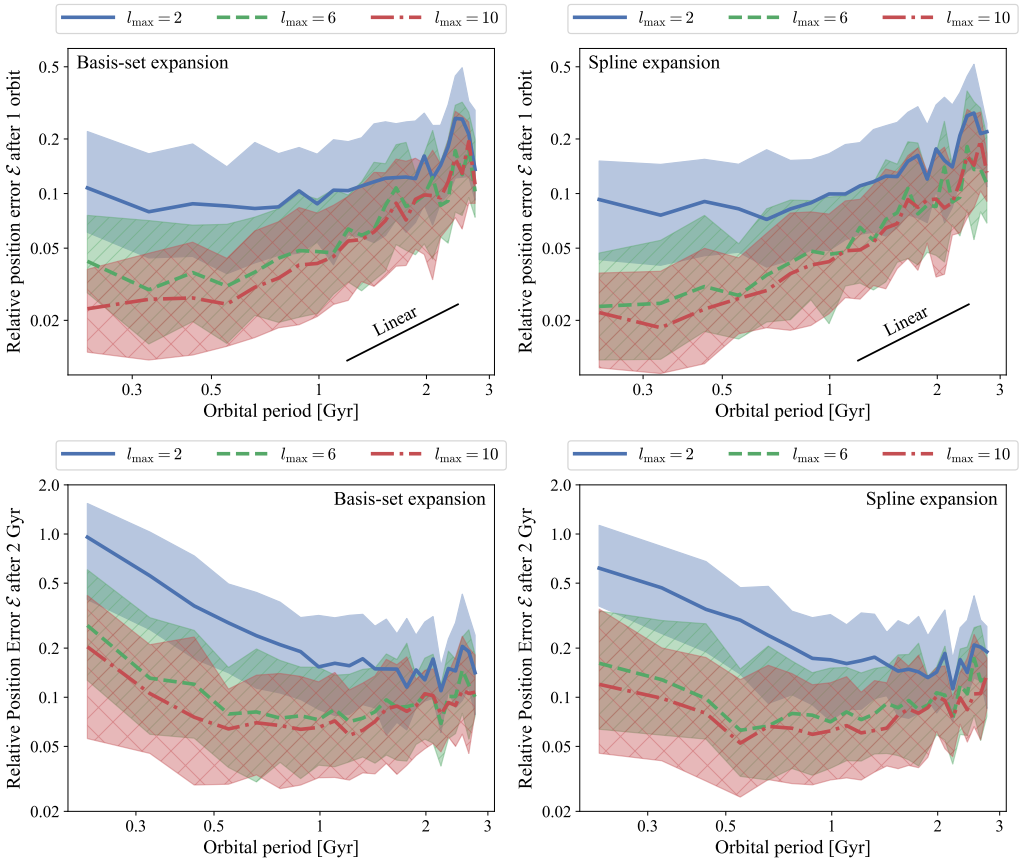


Figure 7.6: Median and $\pm 1\sigma$ of relative position error after a single orbit (top) and after 2 Gyr (bottom) as a function of orbital period. The left panels show results for the basis function expansion and the right for the spline expansion. In each panel we show three sets of results: solid blue for $l_{max} = 2$, dashed green for $l_{max} = 6$ and dash-dotted red for $l_{max} = 10$. The corresponding number of radial terms ($n_{max}$) is described in the text. The small black line shows a linear relation.

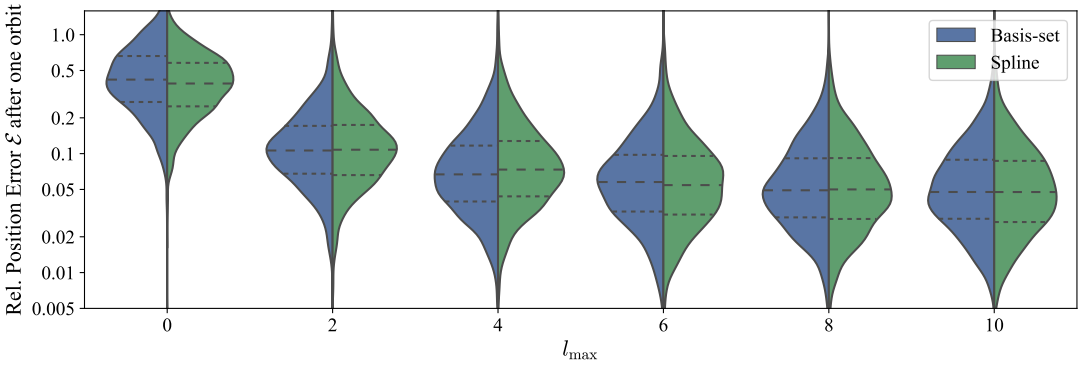


Figure 7.7: 'Violin' plots of the relative position error after one orbit: the probability distributions of are shown explicitly, printed vertically, so that each pair of distributions can be easily visually compared. The left-pointing blue distributions are for the basis expansion, right-pointing green for the spline expansion. The lines give the quartiles of the distributions (short-dashed are 25th and 75th and long-dashed 50th).

error of the reconstructed spatial path of the $i$-th orbit $\mathbf{r}_{\text{recon},i}(t)$ from the truth $\mathbf{r}_{\text{orig},i}(t)$ after a time interval $t_i$ as

$$\mathcal{E}_i = \frac{\left\|\mathbf{r}_{\text{orig},i}(t_i) - \mathbf{r}_{\text{recon},i}(t_i)\right\|}{r_{\text{orig},i}}. \tag{7.29}$$

$r_{\text{orig},i}$ is the time-averaged radius of the $i$th orbit. We choose to perform the comparison after a single period $t_i = T_i$ for each orbit, although we will see this choice is somewhat arbitrary and using a fixed comparison time for all orbits produces qualitatively similar conclusions. $T_i$ is computed by taking a (zero-padded) fast Fourier Transform of the particle's original trajectory and computing one cycle with respect to the dominant frequency. If this time lies outside the simulated interval then the time of the final snapshot is substituted.

#### 7.3.2.1 Choice of biorthogonal expansion

With a well-defined error measure selected, we are in a position to quantitatively select the optimal parameters for the biorthogonal basis expansion. We have made preliminary choices already in Sec. 7.2.2.1. Specifically, we argued that the expansion based on the generalised-NFW models at zeroth order offered a good trade-off between speed and realism. This is a one parameter family with $(\alpha, \beta, \gamma) = (1, 3 + \nu, 1)$, so there remains a single parameter $\nu$ to be freely chosen.

The upper two panels of Fig. 7.5 show the median and $\pm 1\sigma$ spread of $\mathcal{E}$ for our sample of orbits as a function of $\nu$. We show the results for the 'tail corrected' (blue) and uncorrected (red) expansions, and consider evolution over the final 2 Gyr of the simulation. With just the zeroth order term ($n_{\text{max}} = 0$, $l_{\text{max}} = 0$), we expect the NFW model or $\nu = 0$ to be preferred (see Fig. 7.1) – and such is the case for the corrected expansion. As the number of terms in the expansion increases to $n_{\text{max}} = 4$ and $l_{\text{max}} = 2$, the blue band becomes very flat, so the choice of $\nu$ is not at all important. There is no significant gain in using the expansion with the NFW model at zeroth order as compared to the Hernquist & Ostriker expansion ($\nu = 1$), for this particular numerical halo. The main effect of the tail coefficients is to improve the median error, though there is also a slight reduction in the width of the $1\sigma$ shaded region.

The lower two panels show the effect of using just the particles in our sample that are within 100 kpc to construct the expansion. This exaggerates the effect of the artefacts in the uncorrected expansion, so we see larger discrepancies between blue and red bands. However, it is interesting that for the corrected coefficients with $n_{\text{max}} = 4$ and $l_{\text{max}} = 2$, there is little difference between the upper and lower right hand panels – showing that we can use fewer particles ($3.7 \times 10^6$ of the particles are retained when truncating at 100 kpc, about a quarter of the total).

Based on these results, we opt to use the Hernquist & Ostriker expansion ($\nu = 1$) for the main orbit integration, and proceed to examine its performance versus the spline expansion.

#### 7.3.2.2 Comparison of the methods

In Fig. 7.6, we display the median and $\pm 1\sigma$ spread of $\mathcal{E}$ for the considered sample of orbits as a function of their orbital period. For the two methods, we inspect the results for three choices of the number of angular terms: $l_{\text{max}} = (2, 6, 10)$. This corresponds to $n_{\text{max}} = (6, 14, 20)$ radial

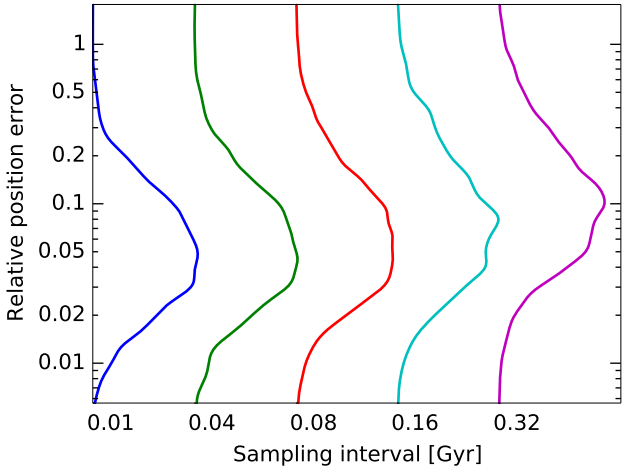


Figure 7.8: Errors in reconstructed particle trajectories as functions of the sampling interval for the potential. The curves show the distribution of relative position errors after one orbital period, as in Fig. 7.7 (here we consider only the Spline method with $l_{\rm max} = 10$). The five curves show the effect of changing the sampling interval for the potential. It is apparent that the accuracy starts to deteriorate once the interval exceeds $\sim 0.05$ Gyr.

terms for the basis-function expansion and $(15, 25, 40)$ radial grid points for the spline expansion. We note the increase in accuracy (reduction in $\mathcal{E}$) for increasing $l_{\rm max}$ particularly for the most bound orbits. We also observe that both methods perform similarly – although this is partly by design, as in both cases the $n_{\rm max}$ is chosen so as to saturate the possible accuracy available at a given $l_{\rm max}$. This behaviour is further illustrated by Fig. 7.7 which shows the full distributions for $\mathcal{E}$ vs $l_{\rm max}$. We see a rapid improvement in accuracy from $l_{\rm max} = 0$ to $l_{\rm max} = 4$ and a much slower improvement for higher $l_{\rm max}$. In general, the distributions of $\mathcal{E}$ are similar for the two methods and generically appear approximately Gaussian but with fatter tails particularly to high $\mathcal{E}$, probably due to particles scattered by subhaloes.

The generic shape of the curves in Fig. 7.6 (rising with increasing orbital period) is a result of our choice of time interval used in the evaluation of $\mathcal{E}$. Longer period orbits have their deviations measured over longer timescales so naturally accumulate more error. This is demonstrated by the approximate linear scaling of $\mathcal{E}$ with period. In Fig. 7.6 we also display the distributions of $\mathcal{E}$ using a fixed time interval of $t_i = 2$ Gyr. For this choice, we find the run of $\mathcal{E}$ is essentially flat for high periods and rises weakly at lower periods. However, the conclusions on the relative performance of difference expansions are unchanged.

#### 7.3.2.3 Dependence on the sampling interval

The $N$-body snapshots in our baseline scenario were stored rather frequently – with a sampling interval of only 10 Myr. We now explore how the error $\mathcal{E}$ depends on this interval. In fact, there are two different time-dependent properties of the system: the potential $\Phi$ and the spatially uniform acceleration associated with the non-inertial reference frame (Eq. 7.5). In principle, each of them

may be sampled at different intervals, so we explore the effect of changing these intervals separately.

Looking at Fig. 7.8, it turns out that frequent sampling is much more important for the non-inertial acceleration than for the potential. Even a tenfold increase of the snapshot spacing (hence the potential sampling) makes the accuracy only somewhat worse, unlike the equivalent increase in the acceleration spacing.

#### 7.3.2.4 Quality of orbit reproductions

In addition to the error measure $\mathcal{E}$, which is useful for comparing the quality of different potential approximations, we can also inspect the overall success of our potential expansion methods through inspection of approximate integrals of motion. Despite the asphericity and time-dependence of the potential, the energy and angular momentum are still useful quantities for summarising a given orbit. In particular, we can check the quality of the orbit recovery by inspecting how well changes in these quantities are reproduced for our sample of orbits (Lowing et al., 2011).

In Fig. 7.9 we show some summary statistics for the changes in the integrals for our orbit sample using the $l_{\mathrm{max}} = 10$ basis-function expansion. As the zero-point of the potential is not well defined, we choose to match the median potential of the expansion to the median potential of the simulation evaluated at the location of all the inspected particles at each time-step. We observe that the distributions of the energies of the orbits at the end of the simulation are very satisfactorily recovered. The median of both the differences in the energy changes and the difference in the angular momentum changes lie around zero at all times with a spread that grows steadily over time such that the dispersion is a few per cent in energy and a few tens of per cent in the components of angular momentum. An alternative way of displaying this information is to look at the energy changes over all 2 Gyr for all orbits (each orbit contributes multiple values). The majority of orbits lie along the one-to-one line, with a small fraction forming clumps far off the line. The most common cause of this is subhalo scattering in the simulation.

Finally, we split the difference in energy changes by orbital eccentricity (defined simply as $[\max(r) - \min(r)]/[\max(r) + \min(r)]$) and find that there is a weak trend for higher eccentricity orbits to be more poorly reproduced. These high eccentricity orbits are naturally more sensitive to successful reproduction of the potential over a wide range of radii, in particular the inner regions.

### 7.3.3 Computational Cost

Our previous discussion has focused on the accuracy of orbit reproduction for the basis expansion and spline expansion without any reference to the computational efficiency of the approaches. As we have demonstrated that both methods produce very similar results at similar order of expansion, it is then natural to ask which method is computationally cheaper. We concentrate on the evaluation costs as opposed to the setup costs: both methods require significant and comparable one-time upfront costs to find either sets of coefficients or spline fits. However, with these in place, a single evaluation of the potential is swift.

In Fig. 7.10 we display the cost of a single force evaluation using each method with varying order of expansion. Force evaluation using the basis expansion scales approximately cubically with

$l_{\rm max}$ as we require $(n_{\rm max}+1)(l_{\rm max}+1)(2l_{\rm max}+1)$ basis function evaluations and we have imposed $n_{\rm max} = 2l_{\rm max}+2$. On the other hand, the spline expansion method (for $l_{\rm max} > 2$) requires summing the 2D interpolated $(r,\theta)$ potential contribution $B_m(r,\theta)$ from each azimuthal order $m$ so scales approximately as $\mathcal{O}(2l_{\rm max}+1)+\mathcal{O}(\log n_{\rm max}+\log l_{\rm max})$ (the log terms corresponding to the bisection algorithm used to locate the grid segment, and in practice are completely negligible for realistic orders of expansion). This means that for large numbers of terms the spline expansion method will always be more efficient per force evaluation.

However, there is a crucial difference in the two expansions when computing the time-dependent forces: the basis function method is both more efficient at lower orders, and also allows for the coefficients to be interpolated such that an interpolated force computation is no more expensive than a single force computation. A similar procedure is not possible for the spline method, and instead one must interpolate the force computation, which requires at least two force computations. For this reason, we have found that for our orbit reconstructions the basis function expansion is always more computationally efficient.

In more complex applications, we may have to evaluate the self-gravity of the re-simulated system (typically via a tree code) which if using GYRFALCON scales as $\mathcal{O}(N)$ in the number of particles (Dehnen, 2000) and takes 1–2$\mu$s per particle: a similar computational cost to the expansion methods.

## 7.4 Conclusions

In recent years, there has been growing awareness that galaxies do not have simple shapes and are not in equilibrium. This has been driven by high resolution simulations – for example, the shapes of the dark haloes in the Auriga project show twisting and this often correlates with recent accretion or merger events (Prada et al., 2019). Observational evidence for disequilibrium is abundant for the Milky Way galaxy. A prominent example is the impending encounter of the Large Magellanic Cloud with the Milky Way, which affects the dynamics of stellar streams (Gómez et al., 2015, Erkal et al., 2019) and which distorts the structure of the dark halo by an induced response (Garavito-Camargo et al., 2019, Belokurov et al., 2019). This has stimulated renewed attention on basis function methods, which have the flexibility to reproduce very general, time-varying gravitational fields, whether for dark haloes (Besla & Garavito-Camargo, 2020, Cunningham et al., 2020) or other Galactic components like bars (Petersen et al., 2016a,b).

A calibration of the performance of different basis function expansions against static galaxy models has already been performed by Vasiliev (2013). Here, we have provided a similar comparison, but for the harder problem of time-evolving models. Suppose we are given snapshots of an $N$-body simulation. For each snapshot, we represented the gravity field by basis function expansions. Interpolating between the expansions at each snapshot gives us a description of the evolving gravity field. How do the $N$-body orbits compare to the reconstructed orbits using the basis function expansions? To answer this question, we introduced a new error measure, based on the fidelity of the reconstructions. For each orbit, we computed the relative position error after a single period.

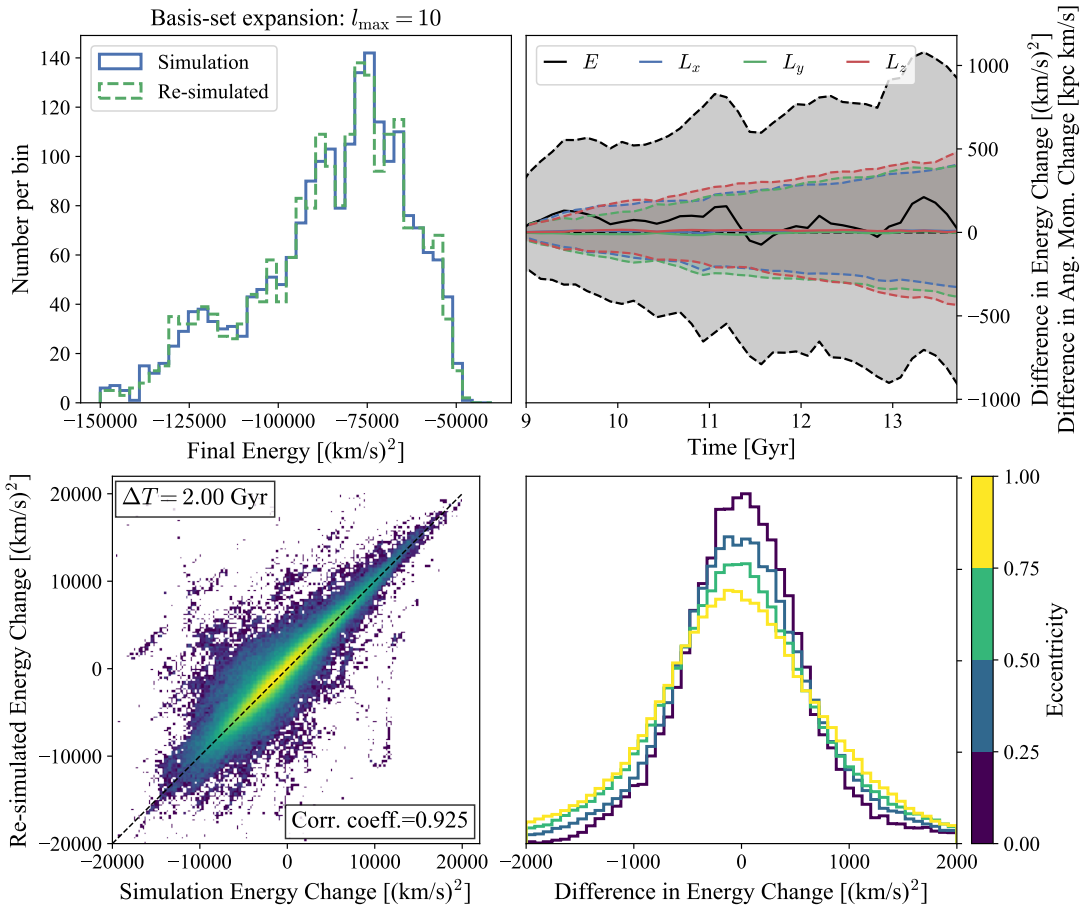


Figure 7.9: Evolution of energy and angular momentum for a sample of $\sim$ 2000 orbits: the top right panel shows the distributions of the final energies from the simulation (solid blue) and the basis-set expansion (dashed green). The top right panel shows the evolution of the difference in change in energy (black) and the components of the angular momentum (colours: blue $x$, green $y$, red $z$) between simulation and basis-set expansion. The median and $\pm 1\sigma$ over orbits are shown. The bottom left panel shows the log-density of the energy changes in the simulation vs basis-set expansion over 2 Gyr time intervals for all particles. The bottom right panel shows the difference in the energy change over 2 Gyr time intervals for orbits separated into quartiles of eccentricity. The more eccentric orbits are more poorly reproduced.

Given a sample of orbits, the median and the spread of relative position errors allow us to quantify the performance of different expansions.

We examined two basis function methods in detail. The first uses biorthogonal expansions to represent the radial variation of the density and potential. The most familiar example is the Hernquist & Ostriker (1992) expansion. However, there are other possibilities in the literature (Zhao, 1996), whilst recent work (see Ch. 4, Ch. 5) has provided an abundance of further such expansions. At zeroth order, the expansions have different density slopes at the centre and the outer parts, raising the possibility that the expansion can be tailored for any numerical halo. The second uses splines, an idea developed by Vasiliev (2013). In its most recent manifestation in the software

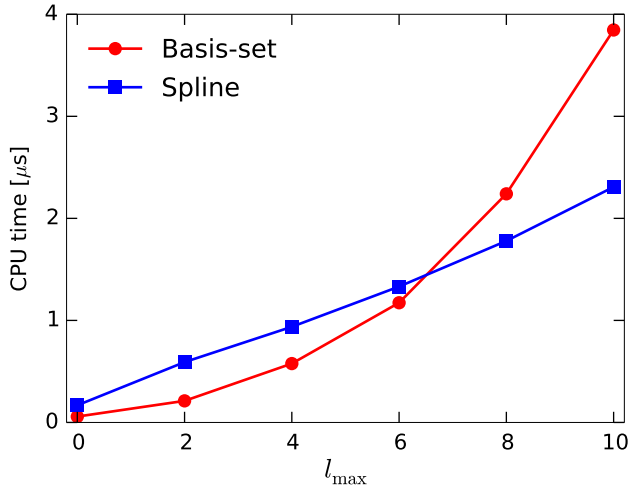


Figure 7.10: Cost of a single force evaluation as a function of the order of expansion: red circles – basis-set expansion with $n_{\max} = 2\,l_{\max} + 2$, blue boxes – spline-interpolated multipole expansion. The latter scales nearly linearly with the number of Fourier terms, while the former scales cubically, but is cheaper for low orders. Note that in a time-dependent potential, the basis-set method interpolates coefficients and evaluates the force only once, while the spline method computes the force for two adjacent moments of time and interpolates between them, thus the actual cost is twice higher than shown in this plot, and remains above that of the basis-set method for all practically relevant orders of expansion. In practice, when simulating a self-gravitating system embedded in an external potential, these costs are further amortised by the need to compute the inter-particle forces, which takes another $\sim 1-2\,\mu$s per particle in GYRFALCON.

package AGAMA (Vasiliev, 2019), quintic splines are employed with nodal points on a logarithmic radial grid. In both cases, spherical harmonics describe the angular dependence.

Our main conclusions are as follows:

1) Interpolation between $N$-body snapshots requires careful attention to the changing acceleration of the reference frame. The simulation particles are not just subject to the forces due to the halo itself, but also those due to the large-scale structure exterior to the halo. We found that numerical computation of the acceleration of the halo centre for each snapshot, followed by interpolation, performs satisfactorily in our orbit reconstructions; this takes account of the motion of the halo due to nearby cosmic structure, and the coordinate-system recentring that is an integral part of basis set approaches. Strictly speaking, this approach neglects any tidal effects on the scale of an individual halo, which may be important for orbits with large apocentric distances.

2) As regards the variety of biorthogonal expansions, the orbit reconstructions for our chosen halo seem largely immune to any particular choice. All the biorthogonal expansions are complete, so can in principle reproduce any smooth density, but we might have expected fewer terms are needed if the zeroth order model is appropriately chosen to mimic the properties of the numerical halo. In fact, despite the greater flexibility afforded at zeroth order by the expansions of Ch. 5, when using more than a few radial and angular terms

we have found no reasons to use anything other than the Hernquist & Ostriker (1992) expansion. The behaviour of the error measure versus the free parameter of the biorthogonal expansions is essentially flat once even a modest number of terms is used, as shown in Fig. 7.5. This is an unexpected conclusion as previous literature (e.g. Kalapotharakos et al., 2008) has stressed the importance of matching the zeroth-order basis function to the system being studied. We have not yet determined the ultimate reason for this parameter insensitivity, but it may be related to the non-inertial reference frame force dominating the gravitational potential in terms of error contribution. The Hernquist & Ostriker expansion is also slightly computationally simpler and faster than the other (generalised NFW) basis sets, although the performance is similar once all special function evaluations are avoided (cf Sec. 7.2.2.2).

3) The coefficients in any biorthogonal expansions can require tail corrections to avoid numerical artefacts caused by the edge, or the finite truncation radius, of numerical haloes. The corrections are important for low order expansions or for orbits that pass close to, or outside, the edge.

4) The spline and biorthogonal basis function methods are very comparable in terms of accuracy, and there is no compelling reason to prefer one over the other – provided a reasonable number of terms are used. Our $N$-body halo has a minor to major axis ratio that varies from $0.5$ to $0.8$ over a Galactocentric range of $200$ kpc in radius. We find a reasonable number of angular terms is $l_{\rm max} \approx 10$ for either method. In terms of total number of parameters, the spline method is greedier, requiring $n_{\rm max} \approx 40$ as compared to $n_{\rm max} \approx 20$ for the Hernquist-Ostriker expansion. For both methods, the performance of individual orbit reconstructions in terms of pericentres, apocentres and eccentricities are normally fine over many orbital periods, but errors in the phase do gradually accumulate. This conclusion is evident from Figs 7.6 and 7.7. Longer period orbits in general tend to be less well reconstructed. In part, this is just a consequence of the fact that the relative position error is measured over a longer time for such orbits. However, both the effects of tidal forces and subhaloes are more important for larger apocentric orbits. A small number of orbits are poorly reconstructed, and this is usually due to scattering by subhaloes in the original simulation. These reconstructions of course only aim to reproduce the smooth underlying halo and do not account for small-scale substructure – although this may in fact be possible if using very large numbers of basis functions, as is suggested in Ch. 4.

5) The computational costs of the spline and basis function methods are similar although force evaluation scales differently with the number of terms in the expansion (cubic for the basis function expansion and linear for the spline expansion). For the re-simulation of time-dependent systems, the basis function expansion is particularly efficient as the coefficients can be interpolated instead of the forces – leading to fewer force evaluations.

# Chapter 8

# Conclusions and further work

We sum up the accomplishments of the thesis, and describe further directions of research, both within astronomy and in other branches of physics.

## 8.1 Results of the thesis

In Ch. 2, we systematically reviewed all known methods for deriving analytical biorthogonal basis sets. These methods fall into two categories: 1) direct substitution into a known Sturm-Liouville equation (Sec. 2.2), in which we uncovered one additional candidate basis set, but otherwise show that there are almost certainly no new basis sets derivable just from standard orthogonal polynomials or special functions; and 2) integral transform methods (Sec. 2.3), in which we provided the most general possible form of the Hankel transform method, and also described further transform methods (including one totally new method based on confluent hypergeometric functions that generalises the Hankel transform) and discussed their advantages and shortcomings. The upshot of this exploration was our realisation that integral transform methods, specifically the Hankel transform method of Ch. 2, were likely to be the most fruitful path to further development in the field.

With that in mind, in Ch. 4 we applied our new method for deriving basis sets based on generating functions, extending the work of Rahmati & Jalali (2009). We derived improved expressions for their singular result, and incorporated it into a totally new one-parameter family of basis sets. The key innovations were: 1) adding the $\alpha$-parameter that affects both the slope and turn-over of the basis sets – this exists due to an additional degree of freedom that exists in the Hankel transform formalism, first noticed by Polyachenko & Shukhman (1981) and rediscovered by us in the course of writing Sec. 2.2; 2) finding a recurrence relation for the basis functions, leading to improved numerical properties – Rahmati & Jalali (2009) had instead expanded the

Laguerre polynomials using the power series definition, an approach that is liable to numerical difficulties due to the alternating signs of the terms of orthogonal polynomials[a].

Along the way we found that one member of our new family has a zeroth-order model whose dynamical properties can be expressed analytically, for which we gave expressions in Ch. 3. Thus the two chapters, Ch. 3 and Ch. 4 form a pair analogous to the classic papers of Hernquist (1990) and Hernquist & Ostriker (1992) (or indeed to Eddington (1916) and Clutton-Brock (1972)).

Having shown that the generating function approach to deriving basis sets is fruitful, we enormously extended the scope of this approach in Ch. 5, deriving a two-parameter family of basis sets which encompasses all known previous results for the spherical geometry. Notably this includes a basis set corresponding to the NFW model, the Jaffe model and the $\gamma$ models – among many others. Our method consists of: 1) heuristically writing down a set of *non*-orthogonal basis functions; 2) calculating the linear combinations of these functions that give an orthogonal basis set; 3) re-summing the non-orthogonal functions analytically, using a generating function to find simple closed-form expressions or recurrence relations for the orthogonal basis functions.

The new two-parameter family of basis sets also gives rise, via a limiting procedure, to a further family that we term the *cuspy-exponential* basis sets, as described in Ch. 6. These are distinguished from all previous results (both in this thesis and in the literature) because they correspond to models with an exponential fall-off in density at large radius – no basis sets with this property have previously been reported, apart from one example of a Gaussian basis set for thin disks (Qian, 1993).

The novel aspects of the method of Ch. 5 also provide some hints to the existence of an as yet larger three-parameter family (Sec. 5.6) that would encompass the full range of double-power laws at zeroth order – thus far we have only written down a suggestive set of non-orthogonal functions.

In Ch. 7 we applied our new family of basis sets to the problem of representing a time-evolving halo potential, extracted from a cosmological simulation that has been run in advance. Comparing the accuracy of orbit reconstruction between the basis set method (BSE) and the competing spline-based method of Vasiliev (2013), we find the following: 1) the maximum attainable accuracy of each method – given the resolution of our numerical halo – is essentially identical; 2) at low expansion orders the BSE method is computationally cheaper – the spline method requires many more radial nodes to represent a smooth underlying model that the BSE method can represent with just one basis function; 3) at high expansion orders (higher than was recoverable from Ch. 7's simulation) the spline method eventually scales better than the BSE method; 4) the BSE method has better time-interpolation properties – the BSE coefficients can be interpolated in advance, so at each time-step only one evaluation of force is required per particle – whereas the spline method must interpolate the force directly, requiring two (for linear interpolation) or more (for higher-order interpolation) force evaluations per time-step. In practice the multiplication of the spline method's computational cost implied by this last property negates its theoretical high-order advantage – this advantage only manifests at high orders that are inaccessible to our particular halo reconstruction.

[a] See e.g. the discussion in Weniger (2011).

We also find, counter-intuitively, that the precise choice of parameters for the specific basis set used in the BSE method made a negligible difference to the overall accuracy of the method at all but the very lowest expansion orders. This finding is at odds with the literature (Dehnen, 2001, Kalapotharakos et al., 2008, Dehnen & Read, 2011) but it may be related to the observation that an accurate time-sampling of the fictitious force due to the halo's non-inertial reference frame – which is tracked based on the location of the DM density cusp – contributes much more error to the reconstructed orbits than do the details of the particular gravitational potential expansion.

When implementing the BSE method, we in fact used only a one-parameter subset of the basis sets available from Ch. 5 – those corresponding to the *generalised-NFW* models. While the general family of Ch. 5 requires the evaluation of a special function to compute the zeroth-order potential, which in principle would greatly worsen the computational performance of the BSE method, our chosen subset avoids this by having all relevant quantities reducible to elementary arithmetical operations. We therefore find a performance close (within about $1.5\times$) to the pre-existing basis sets based on Gegenbauer polynomials (Hernquist & Ostriker, 1992, Zhao, 1996).

Along the way we also developed a numerical method to correct for the ringing effect in the reconstructed potential and density induced by a hard truncation in the available $N$-body data; and we developed a heuristic to determine when to compensate for poor numerical performance of the potential function's recurrence relation as $r \to 0$, by replacing the potential with the first few terms of its Taylor expansion. We have thus provided the $N$-body community with a totally analytical basis set appropriate for all the generalised-NFW models, with performance comparable to the classic basis sets in the literature. Such basis sets, particularly those based on the NFW model, are in great demand from the point of view of applications (e.g. Holley-Bockelmann et al., 2005, Dai et al., 2018); up till now numerically-constructed basis sets have typically been resorted to, via the method of Weinberg (1999)[b].

## 8.2 Directions for future work

The work presented in this thesis has significantly broadened the field of biorthogonal basis sets within galactic dynamics, and has revealed a number of interesting applications and extensions, which we now discuss.

In principle one should be able to use phase-space data of stellar streams to fit non-parametric models of the Milky Way's DM halo. Multipole expansions were considered for this purpose in Bonaca & Hogg (2018), but a full radial expansion has yet to be carried out. Basis expansions of the sort considered in this thesis would appear to be natural, as appending further terms to the truncated expansion is a flexible way to add details to an underlying simple spherically-symmetric model. In addition, given that the parameters are simply the expansion coefficients, such a model is totally linear in its parameters, which may simplify some aspects of the inference problem. The

[b] Indeed, Kalapotharakos, Efthymiopoulos & Voglis (2008) convincingly argues in favour of finding more flexible basis sets whose asymptotic power law behaviour more closely matches that of the system being studied; but also criticises the idea of constructing them numerically. However, their own somewhat exotic method of deriving basis sets 'quantum-mechanically' has seen little traction.

problem set-up would be as follows: suppose we observe $n$ stream orbits (labelled $i = 0 \dots n$), with $\{\mathbf{x}_i\}$ being the trajectory of the $i$-th orbit; we then attempt to fit a model for the gravitational potential that is equal to a basis expansion with coefficients $\{C_j\}$ (for concision we collapse the full $nlm$ index set into a single index $j$). Then the aim is to find a potential $\Phi\big(\mathbf{r}; \{C_j\}\big)$ in which the observed orbits $\{\mathbf{x}_i\}$ best match a set of numerically integrated orbits $\{\mathbf{y}_i\}$ that have the same initial conditions. That is,

$$\ddot{\mathbf{y}}_i(t) \equiv -\boldsymbol{\nabla}\Phi\big(\mathbf{y}_i(t); \{C_j\}\big)\,, \quad \text{with } \mathbf{y}_i(t_0) = \mathbf{x}_i(t_0), \quad \text{and } \dot{\mathbf{y}}_i(t_0) = \dot{\mathbf{x}}_i(t_0), \tag{8.1}$$

and we want to minimise the following error measure with respect to each of the coefficients $C_j$,

$$\mathcal{E}\big[\{\mathbf{x}_i\}; \{C_j\}\big] \equiv \sum_{i=1}^{n} \int \big\|\mathbf{x}_i(\tau) - \mathbf{y}_i(\tau; \{C_j\})\big\|^2 \,\mathrm{d}\tau. \tag{8.2}$$

The time coordinate is unobservable, so we write $\tau$ to indicate some parameterisation of the trajectory, e.g. fraction of the total arc length. Making some assumptions about convergence, it seems possible that this minimisation problem could be solved efficiently by first minimising with respect to $C_0$ (corresponding to the spherical zeroth-order model, e.g. a plain NFW profile), then holding $C_0$ constant and minimising with respect to $C_1$, and so on. We would also have to marginalise over any additional uncertainties in the model, e.g. the observational error in the spatial coordinates of the streams. In the limit of a short time interval and a large number of observed trajectories, the minimisation of Eq. (8.2) clearly reduces to a version of the procedure described in Sec. 1.4.1 where the acceleration field is sampled rather than the mass density. However, in practice we are far from that limit, and must take into account many confounding effects, such as time-evolution of the potential.

Another astronomical setting which involves Poisson's equation is gravitational lensing under the thin lens approximation. In principle a biorthogonal basis set ought to be a natural non-parametric method for modelling lenses. Arbitrary morphology is desirable, and both multipole expansions and wavelets have already seen use in this respect (Evans & Witt, 2003, Birrer et al., 2015). Such approaches bridge the gap between parametric and pixel-based modelling. In Sec. 2.3.1 we noted that the Hankel transform-based approach to deriving basis sets extends trivially to dimensions $d > 3$. However, preliminary results show that when $d = 2$, the methods of Ch. 4–Ch. 6 break down, because the self-energy inner product diverges at either the origin or at infinity. It is likely that a form of regularisation is required to fix this defect, such as *Hadamard* regularisation (Estrada & Kanwal, 1989).

The Hankel-transform formalism developed in Sec. 2.3.1 also generalises to the fractional-order Laplacian operator. This has physical applications; for example, a scheme has recently been proposed (Giusti, 2020) to approximate the effect of MOND by using a fractional-order Laplacian $\nabla^{2s}$; a MOND-like result is obtained when $s \to 3/2$. Considering Eq. (2.84), we find that setting $\alpha = 1/2$ and $\mu = \sqrt{l^s(l+1)^s + 1/4}$ allows us to define basis functions for this fractional-order Laplacian. The weight in the orthogonality relation (2.86) also changes from $k$ to $k^{2s-1}$. We suspect there is enough freedom in the formalism developed in Chapters 4 and 5 to enable analytically convenient fractional basis sets to be written down.

A direct transference of the solution methods of Ch. 4–Ch. 6 to the disk setting[c] remains a possibility. In fact, it seems likely that the disk basis sets of Qian (1993) may be expressible directly via the generating function approach of Ch. 5, and an initial attempt in this direction for the Gaussian disk is made in Sec. 6.3.2. If successful, this approach would provide basis sets in closed-form for both the Gaussian disk and the Kuzmin-Toomre models.

There are some intriguing possible applications for disk-like basis sets – associated to either a genuine thin disk, or arising from projecting a spherical basis set – whose possible applications we now outline. Using integrated field spectroscopy we can make detailed observations of the projected kinematics and visible matter density of nearby galaxies. In the case of low surface brightness galaxies, rotation curves have been fit with simple spherically-symmetric models in order to probe their associated DM haloes (Kuzio de Naray et al., 2006). However, to constrain the structure and kinematics in detail, more flexible methods involving self-consistent distribution functions must be used (van den Bosch et al., 2008, Van De Ven et al., 2008). Van der Marel & Franx (1993) proposed a technique to model line-of-sight velocity profiles non-parametrically, using functions related to Hermite polynomials. In view of this, we paid special attention in Sec. 6.3.1 to the properties of of a new basis set we derived which has a Gaussian density at zeroth-order. The convenient behaviour of Gaussians with respect to projection along an axis makes them well-suited to problems involving data in projection. Combining this with a Gauss-Hermite expansion for the velocity-space part of the distribution function may provide a powerful non-parametric method. In fact, this approach – series expanding both the spatial and kinematic parts of the distribution function – is used in plasma physics, where it goes by names such as the *Legendre-Fourier* or *Hermite-Fourier*[d] expansions (Manzini et al., 2016).

Returning to the general discussion of $N$-body methods of Sec. 1.2, we conjecture that it might be possible to improve the performance of the self-consistent field (SCF) method by importing techniques from other algorithms. For example, the particle-mesh (PM) algorithm interpolates particle positions onto a grid in order to take advantage of the performance of the fast Fourier transform (FFT). A algorithm analogous to the FFT for rapid computation of expansions in classical orthogonal polynomials has recently been developed (Iserles, 2010, Cantero & Iserles, 2012), which would be applicable to all the basis sets considered in this thesis and the literature – they are constructed from the Laguerre, Gegenbauer and Jacobi polynomials. Such an algorithm would be suited to the case where the basis expansion is being computed from and evaluated at a single set of points at each time-step of the $N$-body simulation (as is the case in the SCF method).

The fast multipole method (FMM) also provides some inspiration: it takes advantage of the addition theorem for the spherical harmonics (DLMF, §14.30.9) in order to relate multipole expansions around widely-separated points to each other. We suggest that the SCF and FMM could be hybridised to give a method that treats the radial structure in an SCF analogously to how the spherical shells are handled in the FMM; such a method would therefore be fully decoupled from sums over particles or shells, with each angular and radial component interacting directly

[c] That is to say, a thin disk embedded in three-dimensional space, as opposed to the genuine two-dimensional geometry of a gravitational lens.

[d] In plasma physics the spatial part typically has periodic boundary conditions and hence a Fourier expansion is used.

with the other components. The necessary breakthrough would be the development of addition theorems for the basis functions considered in this thesis. Most of our new basis sets are derived via Bessel functions, which obey various useful addition theorems (DLMF, §10.23(ii)); there are also other routes, see for example Sack (1964), and the discussion at the end of this section.

Other than the representation of time-evolving haloes discussed in Ch. 7, the other major application of basis expansions that has recently seen attention is the *matrix method* approach to the perturbation theory of self-gravitating systems (Hamilton et al., 2018, Fouvry & Bar-Or, 2018). The extension from the disk to the spherical geometry (e.g. Fouvry, 2016, Ch. 4.D) requires the use of an appropriate set of basis functions, normally taken to be the Bessel functions on a finite interval, as described in Sec. 2.2.2.1. However, as the method relies on computing the actions and angles of the underlying smooth model (necessary to express perturbations in the Hamiltonian perturbation formalism), one would ideally have a basis set whose zeroth-order has totally analytical actions and angles. Unfortunately, the only realistic spherical model with totally analytical actions and angles is the isochrone (Binney & Tremaine, 1987, Ch. 3), and as yet no corresponding basis set exists. There are two potential roads forward: firstly, one of the extended integral transform or generating function-based approaches outlined in Ch. 2–Ch. 5 may, by some stroke of luck, prove applicable to the isochrone. Secondly, and more probably, no such simple route exists, and one must commit to computing either the actions and angles or the basis functions numerically. While the latter approach has been developed extensively (Weinberg, 1999), the former seems relatively unstudied, despite semi-analytical methods for computing the actions and angles of arbitrary spherical potentials now being available (Lynden-Bell, 2010, 2015).

Moving away from astronomy, it is perhaps worth noting that the surprising success of the generating function-based approach (Ch. 4–Ch. 6) may have implications for similar problems involving the Laplacian operator. In particular, there is a vast literature concerning numerical solution methods for the Schrödinger equation, much of which revolves around series expansions using orthogonal functions. The position-space wavefunction for the Hydrogen atom is (Bethe & Salpeter, 1957, Eq. 3.16)

$$\Psi_{nlm}(\mathbf{r}) \propto \mathrm{e}^{-Zr/n}\, r^l\, L_{n-l-1}^{(2l+1)}\left(\frac{2Zr}{n}\right)\, Y_{lm}(\theta,\phi), \tag{8.3}$$

where $L_n^{(\alpha)}(x)$ is a Laguerre polynomial, and the momentum-space wavefunction is (Bethe & Salpeter, 1957, Eq. 8.8),

$$\Phi_{nlm}(\mathbf{p}) \propto \frac{p^l}{(n^2p^2+1)^{l+2}}\, C_{n-l-1}^{(l+1)}\left(\frac{n^2p^2-1}{n^2p^2+1}\right)\, Y_{lm}(p_\theta,p_\phi), \tag{8.4}$$

where $C_n^{(\alpha)}(x)$ is a Gegenbauer polynomial. The similarity to some expressions in Ch. 4 is striking, with $\Psi_{nlm}$ strongly resembling the auxiliary function $g_n(k)$ (Eq. 4.7), and $\Phi_{nlm}$ strongly resembling the density basis function $\rho_{nl}(r)$ (Eq. 4.28). Both sets of functions are related by Fourier transforms, although with the roles of position and momentum reversed between the two cases. One wonders whether the somewhat more elaborate method of Ch. 5 also gives rise to a

solution of the Schrödinger equation – although this would necessitate finding an eigenfunction equation satisfied by a function similar to the $g_n(k)$ of Eq. (5.16).

However it would probably be more fruitful to begin a more general investigation. Numerical methods for the many-body Schrödinger equation routinely employ sets of basis functions that are not individually the wavefunction of any particular system, but which nevertheless have convenient analytical and numerical properties – such that suitable combinations of these functions can approximately represent a wide range of physical systems. A useful survey may be found in Weniger (1985), where *B-functions* $B_{nlm}$, *Λ-functions* $\Lambda_{nlm}$ and *Sturmian functions* $\Psi_{nlm}$ are defined:

$$B_{nlm}(\mathbf{r}) \propto r^{l+n+1/2}\, K_{n+1/2}(r)\, Y_{lm}(\theta,\phi), \tag{8.5}$$

$$\Lambda_{nlm}(\mathbf{r}) \propto \mathrm{e}^{-r}\, r^{l}\, L^{(2l+2)}_{n-l-1}(2r)\; Y_{lm}(\theta,\phi), \tag{8.6}$$

$$\Psi_{nlm}(\mathbf{r}) \propto \mathrm{e}^{-r}\, r^{l}\, L^{(2l+1)}_{n-l-1}(2r)\; Y_{lm}(\theta,\phi). \tag{8.7}$$

The latter two functions satisfy convenient orthogonality properties; the function $B_{nlm}$ is defined so that it has a convenient Fourier transform,

$$\bar{B}_{nlm}(\mathbf{p}) = (2\pi)^{-3/2} \int \mathrm{d}^3\mathbf{r}\; \mathrm{e}^{-\mathrm{i}\mathbf{p}\cdot\mathbf{r}} B_{nlm}(\mathbf{r}) \propto \frac{p^l}{(1+p^2)^{n+l+1}}\, Y_{lm}(\theta_p,\phi_p), \tag{8.8}$$

and we realise that in Ch. 5 (in particular Eq. (5.6)) we rediscovered the usefulness of this particular Fourier transform. The Fourier transforms of $\Lambda_{nlm}$ and $\Psi_{nlm}$ functions involve Gegenbauer and Jacobi polynomials respectively, as is expected given our previous results in Ch. 4 and Ch. 5 for functions of very similar form.

Among the problems in computational chemistry that these basis sets are applied to is the derivation of addition theorems. These are expansions of a function of interest $f(\mathbf{r})$ around two centres $(\mathbf{r},\mathbf{r}')$ such that the coordinates appear factored in the terms of the sum; this kind of factorisation is very useful when computing interaction potentials between different bodies in a molecular simulation. The preeminent representative of this technique is the multipole expansion for the Coulomb potential (Sec. 1.3.2.1). It is a *two-range* addition theorem, as two different forms of the expansion are used depending on whether $r < r'$ or $r > r'$. There has also been intense work on the derivation of *one-range* addition theorems for various physically-motivated functions (Weniger, 1985, Weniger & Steinborn, 1989, Filter & Steinborn, 1980, Homeier et al., 1992). For a given function $f(\mathbf{r})$, a one-range addition theorem would typically take the form

$$f(\mathbf{r}\pm\mathbf{r}') = \sum_{nlm}\sum_{n'l'm'} D^{n'l'm'}_{nlm}[f]\, \psi_{nl}(r)\, \psi_{n'l'}(r')\, Y_{lm}(\theta,\phi)\, Y_{l'm'}(\theta',\phi'), \tag{8.9}$$

where the $\psi_{nl}$ are some appropriate basis functions and the coefficients $D^{n'l'm'}_{nlm}[f]$ depend on $f(\mathbf{r})$ by some integration procedure. One-range addition theorems are often computationally and analytically more convenient than two-range addition theorems; when inserted into integrals (such as those computing interaction potentials), the domain of integration no longer needs to be split up into two parts. However, such expansions are fraught with mathematical difficulties, and the derivation of one-range addition theorems, even for some of the most commonly used functions such as the Coulomb potential, remains an active area of research (Weniger, 2009, 2011).

The class of orthogonal basis functions studied in this thesis may find an application to the development of one-range addition theorems for the Coulomb potential. Indeed, given any biorthogonal basis set satisfying Eqs (1.36) and (1.37), we immediately have

$$\frac{-1}{4\pi \left\|\mathbf{r}-\mathbf{r}'\right\|^2} = \sum_{nlm} \frac{\Phi_{nlm}(\mathbf{r})\Phi_{nlm}(\mathbf{r}')}{K_{nl}N_{nl}}, \tag{8.10}$$

which may be easily (albeit formally) verified by convolving this expression with the expansion of an arbitrary density (Eq. 1.34) – the result being the corresponding representation of the potential – because Eq. (8.10) is essentially the Green's function for the Laplacian[e]. The convergence properties of our basis functions are yet to be established, so it is unknown if expansions of the form (8.10) outperform those already proposed (and rejected) in the computational chemistry literature (see Weniger, 2007, 2011, and extensive references therein). However, because every biorthogonal basis set we have considered gives rise to a distinct Green's function expansion of the form Eq. (8.10), it seems plausible that there is some utility to be mined here.

[e]This was apparently first pointed out in Dehnen & Read (2011), with some minor errors.

Appendix A

# Table of mathematical symbols

We briefly summarise the mathematical symbols mentioned in this thesis, and provide links to standard reference works that list their properties. Some of these functions vary in their definitions in the literature; in general we favour the definitions found in the DLMF (2020), and give references to other sources only when they list additional relevant properties.

| Name | Symbol | Reference |
|---|---|---|
| Gamma and Beta functions | $\Gamma(z)$, $\mathrm{B}(a,b)$ | DLMF (2020, Ch. 5) |
| Pochhammer symbol | $(z)_n \equiv \Gamma(z+n)/\Gamma(z)$ | see note below |
| Bessel functions | $J_\alpha(x)$, $Y_\alpha(x)$ | DLMF (2020, §10.2), HTF (1955, Ch. 7) |
| Modified Bessel functions | $I_\alpha(x)$, $K_\alpha(x)$ | DLMF (2020, §10.25) |
| Jacobi polynomial | $P_n^{(\alpha,\beta)}(x)$ | DLMF (2020, §18.3) |
| Gegenbauer polynomial | $C_n^{(\alpha)}(x)$ | 〃 |
| Laguerre polynomial | $L_n^{(\alpha)}(x)$ | 〃 |
| Hermite polynomial | $H_n(x)$ | 〃 |
| Legendre functions (polynomial) | $P_\nu^{(\mu)}(x)$ | DLMF (2020, §14.3) |
| Spherical harmonics[a] | $Y_{lm}(\theta,\phi)$, $Y_{lm}(\hat{\mathbf{r}})$ | DLMF (2020, §14.30) |
| Lower incomplete Gamma function | $\gamma(\alpha,z)$ | DLMF (2020, §8.2) |
| Incomplete Beta function | $\mathcal{B}_z(a,b)$ | DLMF (2020, §8.17) |
| Gaussian hypergeometric function | ${}_2F_1$ | DLMF (2020, §15.2) |
| Confluent hypergeometric function | ${}_1F_1$ | DLMF (2020, §13.2) |
| Generalised hypergeometric function | ${}_pF_q$ | DLMF (2020, §16.2) |

[a] We use unit-normalised spherical harmonics throughout this thesis.

| | | |
|---|---|---|
| Appell's series (functions) | $F_1$, $F_2$ | DLMF (2020, §16.13), HTF (1955, §5.7–5.11) |
| Elliptic integrals | $E(k)$, $K(k)$, $\Pi(\alpha^2 \,|\, k)$ | DLMF (2020, §19.2) |
| Incomplete Elliptic integrals[b] | $E(\phi \,|\, k)$, $F(\phi \,|\, m)$ | DLMF (2020, §19.2) |
| Mellin transform | $\mathcal{M}_x \{f(x)\}(s)$, $\mathcal{M}_s^{-1} \{g(s)\}(x)$ | BMP (1954, Ch. 6) |

The most comprehensive reference on the properties of the factorial, Gamma function, Pochhammer symbol and associated functions is perhaps the following webpage: `https://functions.wolfram.com/GammaBetaErf/Factorial/introductions/FactorialBinomials/ShowAll.html`.

Useful references on generating functions and hypergeometric functions of multiple variables are Slater (1966), Srivastava & Manocha (1984) and Srivastava & Karlsson (1985). In particular, the generating function that forms the key ingredient of the method of Ch. 5 may be found in Srivastava & Manocha (1984, Ch. 2), along with many other hypergeometric-type generating functions that are likely to be applicable to similar problems.

[b] See footnote in Sec. 3.1.1.

# Appendix B

# Fourier basis set

In Sec. 2.2 we developed the mathematical machinery to manufacture a biorthogonal expansion (which we termed *Clutton-Brock* expansions) out of a given Sturm-Liouville equation with known eigenfunctions. We now apply that method to the harmonic oscillator equation,

$$-\frac{d^2 y_\lambda}{dz^2} = \lambda^2\, y_\lambda. \tag{B.1}$$

The resulting radial potential-density pair (labelled with the continuous eigenvalue $k$) is

$$\begin{aligned} \Phi_{kl}(r) &= r^{-1-l}\, e^{\mathrm{i}k\, r^{1+2l}}, \\ \rho_{kl}(r) &= k^2 r^{3l-1}\, e^{\mathrm{i}k\, r^{1+2l}}. \end{aligned} \tag{B.2}$$

We might therefore pick auxiliary functions $g_n(k)$ and produce a basis set according to

$$\begin{aligned} \Phi_{nl}(r) &= \int_{-\infty}^{\infty} \mathrm{d}k\, g_n(k)\, \Phi_{kl}(r) = r^{-1-l} \int_{-\infty}^{\infty} \mathrm{d}k\, g_n(k)\, e^{\mathrm{i}kr^{1+2l}}, \\ \rho_{nl}(r) &= \int_{-\infty}^{\infty} \mathrm{d}k\, g_n(k)\, \rho_{kl}(r) = r^{3l-1} \int_{-\infty}^{\infty} \mathrm{d}k\, g_n(k)\, k^2\, e^{\mathrm{i}kr^{1+2l}}. \end{aligned} \tag{B.3}$$

The orthogonality relation that these functions must obey is

$$\int_0^{\infty} \mathrm{d}r\, r^2 \Phi_{nl}(r)\rho_{nl}(r) = \int_{-\infty}^{\infty} \mathrm{d}k\, k^2\, g_n(k)\, g_{n'}(-k) \propto \delta_{nn'}. \tag{B.4}$$

An obvious choice for $g_n(k)$ would be to take the Hermite polynomials $H_n(k)$ multiplied by the weight $k^{-1}\mathrm{e}^{-k^2/2}$. The lowest order density would be the Gaussian profile $\mathrm{e}^{-r^2/2}$. However, higher orders in $l$ would be proportional to $\exp\left(r^{2+4l}/2\right)$, i.e. containing an extremely steep power of $r$. The observations of Sec. 2.2.1.1 concerning the numerical behaviour when $r$ has an exponential dependence on $l$ suggest that such basis sets would be unsatisfactory in practice.

Appendix C

# Preliminary results on the three-parameter family

## C.1 Method for density basis functions

The computation described here is referenced in the discussion in Sec. 5.6. We choose as our auxiliary function

$$\tilde{f}_j(t) \equiv t^{\mu-1+j}\, \mathrm{e}^{-t}\, {}_1F_1\left(\begin{matrix}\lambda+1\\ \mu+1+j\end{matrix}\middle| t\right). \tag{C.1}$$

To compute the non-orthogonal density functions $\tilde{\rho}_{jl}$ we carry out some rather involved computations, giving however a surprisingly simple result.

$$\begin{aligned}
\tilde{\rho}_{jl}(r) &= \frac{r^{l-2+1/\alpha}}{\Gamma(\mu+1)} \int_0^\infty \mathrm{d}t\, t^{j+\mu}\, \mathrm{e}^{-t}\, {}_1F_1\left(\begin{matrix}\lambda+1\\ \mu+1+j\end{matrix}\middle| t\right) {}_1F_1\left(\begin{matrix}\mu+\nu+1\\ \mu+1\end{matrix}\middle| -z^2 t\right) \\
&= (\mu+1)_j\, r^{l-2+1/\alpha}\, F_2\left(\begin{matrix}\mu+j+1;\lambda+1,\mu+\nu+1\\ \mu+j+1,\mu+1\end{matrix}\middle| 1,-z^2\right) \\
&= \frac{(\mu+1)_j\, r^{l-2+1/\alpha}}{(1+z^2)^{\mu+j+1}}\, F_2\left(\begin{matrix}\mu+j+1;-\nu,\lambda+1\\ \mu+1,\mu+j+1\end{matrix}\middle| \chi,1-\chi\right) \\
&= \frac{(\mu+1)_j\, r^{l-2+1/\alpha}}{(1+z^2)^{\mu+j+1}\,\chi^{\lambda+1}} F_1\left(\begin{matrix}-\nu;\mu+j-\lambda,\lambda+1\\ \mu+1\end{matrix}\middle| \chi,1\right) \\
&= (\mu+1)_j\, A(\mu,\nu,\lambda)\, \frac{r^{l-2+1/\alpha}\,(1-\chi)^{\nu-j}}{(1+z^2)^{\mu+j+1}\,\chi^{\lambda+1}}\, {}_2F_1\left(\begin{matrix}-\nu,\mu+j-\lambda\\ \mu-\lambda\end{matrix}\middle| \chi\right) \\
&= (\mu+1)_j\, A(\mu,\nu,\lambda)\, \frac{r^{l-2-\lambda/\alpha}}{(1+z^2)^{\mu+\nu-\lambda}}\, {}_2F_1\left(\begin{matrix}-j,\mu+\nu-\lambda\\ \mu-\lambda\end{matrix}\middle| \chi\right).
\end{aligned} \tag{C.2}$$

These five steps used, in succession,

1) Saad & Hall (2003, Eq. 2.2);
2) DLMF, §16.16.9;
3) DLMF, §16.16.3;
4) HTF, Eq. 5.10.10;
5) DLMF, §15.8.1(1).

The corresponding expression for the non-orthogonal potential functions $\tilde{\Phi}_{jl}$ cannot follow quite the same path. We find that some contiguous relations for the Appell $F_2$ function must be applied, causing multiple terms to appear in the result.

## C.2 Overlap integral

The first step towards a complete solution to the orthogonalisation problem (as carried out in Sec. 5.2.1 for the two-parameter family) is to compute the overlap matrix $D_{ij}$ and perform an LU-decomposition. The double-integral Eq. (2.102) provides us with an expression for the inner product between two basis functions in terms of a $t$-space auxiliary function. Taking Eq. (C.1) as this auxiliary function, if we write the ${}_1F_1$ functions as power series in $s$ and $t$, and change variables according to $s \mapsto xy$, $t \mapsto x(1-y)$, we obtain the following result.

$$
\begin{aligned}
D_{ij} &= \int r^2 \,\mathrm{d}r\, \tilde{\Phi}_{il}\, \tilde{\rho}_{jl} \qquad\qquad (C.3)\\
&= \Gamma(\mu+i+j)\,\mathrm{B}(\mu+\nu+i, \mu+\nu+j) \\
&\qquad \times \sum_{n,m=0}^{\infty} \frac{(\mu+i+j)_{m+n}(\lambda+1)_m(\lambda+1)_n(\mu+\nu+i)_m(\mu+\nu+j)_n}{m!n!(2\mu+2\nu+i+j)_{m+n}(\mu+1+i)_m(\mu+1+j)_n} \\
&= \Gamma(\mu+i+j)\,\mathrm{B}(\mu+\nu+i, \mu+\nu+j) \\
&\qquad \times F^{1:2,2}_{1:1,1}\left( \begin{matrix} \mu+i+j : \lambda+1, \mu+\nu+i; \lambda+1, \mu+\nu+j \\ 2\mu+2\nu+i+j : \mu+1+i; \mu+1+j \end{matrix} \middle| 1, 1 \right) \\
&= \Gamma(\mu+i+j) \int_0^1 \mathrm{d}t\, t^{\mu+\nu+i-1}(1-t)^{\mu+\nu+j-1}\, F_2\left( \begin{matrix} \mu+i+j; \lambda+1, \lambda+1 \\ \mu+1+i, \mu+1+j \end{matrix} \middle| t, 1-t \right).
\end{aligned}
$$

The symbol $F^{1:2,2}_{1:1,1}$ in the penultimate expression is a Kampé de Fériet function, in the modernised notation of Srivastava & Karlsson (1985). The final expression involving an Appell $F_2$ function was obtained by writing three of the parameters of the $F^{1:2,2}_{1:1,1}$ as a beta function, then expressing the beta function as an integral and reversing the order of integration and summation. Transformations of $F^{1:2,2}_{1:1,1}$ series of unit argument are not unknown (see e.g. Srivastava & Karlsson, 1985, Ch. 9, Eq. 246), and the general technique of beta-integral reduction – especially as applied to the reduction of Kampé de Fériet and Appell series – remains an active area of research (Wei et al., 2013, Rathie & Pogány, 2020), so it remains possible that Eq. (C.3) admits a simplification.

# Appendix D

# Tail coefficients

Here we present some results required in Sec. 7.2.2.4. These 'tail coefficients' are effectively the indefinite integral between the zeroth-order density function of some basis set, and the higher-order potential functions. To simplify the calculation we take advantage of Green's second identity to change the integral to that between the zeroth-order potential and the higher-order density functions.

We give the somewhat simpler results for the $\alpha = 1$ generalised NFW family in Appendix D.1, and then the result for the full parameter set in Appendix D.2. In both cases we give the result for the Family 'A' basis sets, and analogous result for Family 'B' can be found as follows:

$$T^{\mathrm{B}}_{nlm}(R) = \delta_{l0}\delta_{m0}\delta_{n0}N_{00} - T^{\mathrm{A}}_{nlm}\left(R^{-1}\right). \tag{D.1}$$

## D.1 Result for generalised NFW basis set

We require the integral over the interval $(R, \infty)$ between the zeroth-order density and each potential basis function, which is non-zero only when $m = l = 0$. Assuming unit-normalised spherical harmonics and working in units where $G = 1$ and the scale-length $r_{\mathrm{s}} = 1$, we have the result

$$\begin{aligned} T_{n00}(R) &= \int_R^\infty r^2\,\mathrm{d}r\,\Phi_{n0}\rho_{00} \\ &= \delta_{n0}N_{00} - \frac{1}{4\pi K_{00}}R^2\left(\frac{\mathrm{d}\Phi_{00}}{\mathrm{d}r}\Phi_{n0} - \frac{\mathrm{d}\Phi_{n0}}{\mathrm{d}r}\Phi_{00}\right)_{r=R} \\ &\quad - \frac{K_{n0}}{K_{00}}\int_0^R r^2\,\mathrm{d}r\,\Phi_{00}\rho_{n0}. \end{aligned} \tag{D.2}$$

It remains to evaluate the last integral in Eq. (D.2), which is

$$\int_0^R r^2 \,\mathrm{d}r\, \Phi_{00}\rho_{n0} = (n+1)I_n(\chi) - nI_{n-1}(\chi), \tag{D.3}$$

where we have defined $\chi \equiv R/(1+R)$, and

$$\begin{aligned} I_n(\chi) &\equiv (n+\nu+1)(n+2\nu+1)\sum_{j=0}^{n} \frac{(-1)^{n-j}(n+2\nu+2)_j}{(j+1)!}\binom{n}{j} Q_j(\chi), \\ Q_j(\chi) &\equiv \begin{cases} \dfrac{1}{1+j}\left[\chi^{1+j}\left(\dfrac{1}{1+j} - \log(1-\chi)\right) - \mathcal{B}_\chi(1+j,0)\right], & \text{if } \nu = 0 \\ \dfrac{1}{\nu}\left[\mathcal{B}_\chi(1+j,1+\nu) - \mathcal{B}_\chi(1+j,1+2\nu)\right], & \text{otherwise.} \end{cases} \end{aligned} \tag{D.4}$$

Here, $\mathcal{B}_z(a,b)$ is the incomplete beta function and we have made use of the Pochhammer symbol $(z)_j$ to indicate the falling factorial.

## D.2 Result for arbitrary parameter values

Here we evaluate the tail coefficients $T_{nlm}$ for the family of basis sets given in Ch. 5, for general values of the parameters $\alpha$ and $\nu$. For the 'Family A' basis sets the expression is identical to Eq. (D.2), with the exception of the final integral, which is now given by

$$\begin{aligned} \int_0^R r^2 \,\mathrm{d}r\, \Phi_{00}\rho_{n0} = \alpha^2 \big[&(n+2\alpha+2\nu-1)(n+\alpha+\nu)I_n(\chi) \\ &-(n+\alpha+2\nu-1)(n+\alpha+\nu-1)I_{n-1}(\chi)\big], \end{aligned} \tag{D.5}$$

where we have defined $\chi \equiv R^{1/\alpha}/(1+R^{1/\alpha})$ and

$$\begin{aligned} I_n(\chi) &\equiv \frac{(-1)^n(\alpha+1)_n}{n!}\sum_{j=0}^{n}\frac{(-n)_j(n+2\alpha+2\nu)_j}{j!(\alpha+1)_j}Q_j(\chi), \\ Q_j(\chi) &\equiv -\mathcal{B}_\chi(\alpha,\nu)\,\mathcal{B}_{1-\chi}(\alpha+\nu,j+1) + \sum_{k=0}^{j}\frac{(-j)_k\mathcal{B}_\chi(\alpha,\alpha+2\nu+k)}{k!(\alpha+\nu+k)}. \end{aligned} \tag{D.6}$$

As $T_{nlm}(R)$ is a highly oscillating function, at high orders it becomes necessary to use arbitrary-precision arithmetic to ensure that catastrophic cancellation of terms is avoided. In the following section we provide an example program using the `arb` library.

## D.3 Example implementation

The following C code (`tailcoeffs.c`) computes the quantities $T^{\mathrm{A}}_{n00}(R)$, for $n = 0 \ldots n_{\mathrm{max}}$. It requires the `arb` (Johansson, 2017), `flint` (Hart et al., 2013) and `gsl` (Galassi, 2003) libraries. Compile with

```
gcc tailcoeffs.c -o tailcoeffs -larb -lflint -lgsl -lgslcblas -lm
```

and run with

```
./tailcoeffs digits alpha nu r_t r_s nmax
```

where `digits` is the required number of accurate digits, `alpha` and `nu` are the parameters of the basis set, and `r_s` is the scalelength $r_{\mathrm{s}}$ (not assumed to be unity here). Note that this code uses a convention where the potential-density normalisation constant is $\left(K_{nl} A_{nl} r_{\mathrm{s}}^2\right)^{-1}$, and the density basis functions differ by a factor of $-4\pi/r_{\mathrm{s}}^2$ from those defined above.

The output consists of two columns, the first an arbitrary-precision calculation using `arb` (corresponding to the subroutines beginning with `arb_`), the second a standard double-precision calculation using `gsl`. We find that calculations with e.g. $\sim$100 digits of precision have negligible overhead compared to the sum over particles in the main part of the algorithm, for which only double precision is required.

```c
#include "stdio.h"
#include "math.h"
#include "arb.h"
#include "arf.h"
#include "arb_hypgeom.h"
#include <gsl/gsl_sf_gegenbauer.h>
#include <gsl/gsl_sf_gamma.h>
#include <gsl/gsl_sf_hyperg.h>

double arb_get_d(arb_t x) {
  arf_srcptr xx;
  xx = arb_midref(x);
  return arf_get_d(xx, ARF_RND_NEAR);
}

int minusonepow(int n){
  return (1 - 2*(n % 2));
}

double beta(double a, double b, double x){
  if(b==0.) return 1/a*pow(x,a)*gsl_sf_hyperg_2F1(a,1,a+1,x);
  return gsl_sf_beta(a,b)*gsl_sf_beta_inc(a,b,x);
}

void arb_beta(arb_t res, arb_t aa, arb_t bb, arb_t xx, slong prec) {
  arb_hypgeom_beta_lower(res, aa, bb, xx, 0, prec);
}

double curlyQ(double al, double nu, int j, double rt){
  double summand = 0;
  double prefactor = 1;
  double chit = pow(rt,1/al)/(1 + pow(rt,1/al));
  for (int k = 0; k <= j; k++) {
    prefactor = gsl_sf_poch(((double) (-j)),(double) k)/gsl_sf_fact(k);
    summand += prefactor*beta(al, al+2*nu+((double) k), chit)/(al+nu+((double) k));
  }
  return summand - beta(al,nu,chit)*beta(al+nu,((double) j)+1,1-chit);
}
```

```
void arb_curlyQ(arb_t res, arb_t al, arb_t nu, int jj, arb_t rt, slong prec) {
  arb_t j; arb_t summand; arb_t prefactor; arb_t chit;
  arb_t ali; arb_t x1; arb_t k; arb_t x2; arb_t x3; arb_t x4;
  arb_init(j); arb_init(summand); arb_init(prefactor); arb_init(chit);
  arb_init(ali); arb_init(x1); arb_init(k); arb_init(x2); arb_init(x3); arb_init(x4);

  arb_set_si(j, jj);
  arb_zero(summand); arb_zero(prefactor);
  arb_inv(ali, al, prec);
  arb_pow(chit, rt, ali, prec);
  arb_one(x1);
  arb_add(chit, chit, x1, prec);
  arb_inv(chit, chit, prec);
  arb_pow(x1, rt, ali, prec);
  arb_mul(chit, chit, x1, prec);

  for (unsigned int kk = 0; kk <= jj; kk++) {
    arb_set_ui(k, kk);
    arb_zero(x2);
    arb_neg(x2, j); // x2 = -j
    arb_zero(prefactor);
    arb_rising_ui(prefactor, x2, kk, prec); // prefactor = (-j)_k
    arb_zero(x1);
    arb_fac_ui(x1, kk, prec); // x1 = k!
    arb_inv(x1, x1, prec); // x1 = 1/k!
    arb_mul(prefactor, prefactor, x1, prec); // prefactor = (-j)_k / k!
    arb_zero(x2);
    arb_add(x2, x2, al, prec); arb_add(x2, x2, nu, prec);
    arb_add(x2, x2, k, prec); arb_inv(x2, x2, prec);
    arb_zero(x3);
    arb_add(x3, x3, nu, prec); arb_add(x3, x3, nu, prec);
    arb_add(x3, x3, al, prec); arb_add(x3, x3, k, prec);
    arb_beta(x1, al, x3, chit, prec);
    arb_mul(prefactor, prefactor, x2, prec);
    arb_mul(prefactor, prefactor, x1, prec);
    arb_add(summand, summand, prefactor, prec);
  }

  arb_zero(x1); arb_beta(x1, al, nu, chit, prec); arb_zero(x2);
  arb_add(x2, x2, al, prec); arb_add(x2, x2, nu, prec);
  arb_one(x3); arb_add(x3, x3, j, prec); arb_one(x4);
  arb_sub(x4, x4, chit, prec); arb_beta(x2, x2, x3, x4, prec);
  arb_mul(x1, x1, x2, prec); arb_sub(res, summand, x1, prec);
}

double curlyI(double al, double nu, int n, double rt){
  double summand = 0;
  double prefactor = minusonepow(n)*gsl_sf_poch(al+1,(double) n)/gsl_sf_fact(n);

  for (int j = 0; j <= n; j++) {
    double sumfactor = gsl_sf_poch((double) (-n),(double) j)*
      gsl_sf_poch(n+2*al+2*nu,(double) j)/(gsl_sf_fact(j)*gsl_sf_poch(al+1,(double) j));
    summand += sumfactor*curlyQ(al,nu,j,rt);
  }

  return prefactor*summand;
}

void arb_curlyI(arb_t res, arb_t al, arb_t nu, int nn, arb_t rt, slong prec) {
  arb_t n; arb_t ali; arb_t x1; arb_t x2; arb_t x3; arb_t x4;
  arb_t summand; arb_t prefactor; arb_t sumfactor;
  arb_init(n); arb_init(ali); arb_init(x1); arb_init(x2); arb_init(x3);
  arb_init(x4); arb_init(summand); arb_init(prefactor); arb_init(sumfactor);
  unsigned int un = (unsigned int) nn;

  arb_set_si(n, nn);
  arb_zero(summand);
  arb_set_si(prefactor, minusonepow(nn)); // prefactor = (-1)^n
```

```
  arb_one(x1);
  arb_add(x1, x1, al, prec); // x1 = al + 1
  arb_rising_ui(x1, x1, un, prec); // x1 = (al + 1)_nn
  arb_mul(prefactor, prefactor, x1, prec); // prefactor = (-1)^n * (al + 1)_nn
  arb_zero(x2);
  arb_fac_ui(x2, un, prec); // x2 = n!
  arb_inv(x2, x2, prec); // x2 = 1/n!
  arb_mul(prefactor, prefactor, x2, prec); // prefactor = (-1)^n * (al + 1)_nn / n!

  for (unsigned int j = 0; j <= nn; j++) {
    arb_one(sumfactor);
    arb_set_si(x1, -nn);
    arb_rising_ui(x1, x1, j, prec);
    arb_mul(sumfactor, sumfactor, x1, prec);
    arb_zero(x1);
    arb_add(x1, x1, n, prec);
    arb_add(x1, x1, al, prec); arb_add(x1, x1, al, prec);
    arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec);
    arb_rising_ui(x1, x1, j, prec);
    arb_mul(sumfactor, sumfactor, x1, prec);
    arb_one(x1); arb_add(x1, x1, al, prec);
    arb_rising_ui(x1, x1, j, prec);
    arb_zero(x2); arb_fac_ui(x2, j, prec);
    arb_mul(x1, x1, x2, prec); arb_inv(x1, x1, prec);
    arb_mul(sumfactor, sumfactor, x1, prec);
    arb_zero(x1); arb_curlyQ(x1, al, nu, j, rt, prec);
    arb_mul(x1, x1, sumfactor, prec);
    arb_add(summand, summand, x1, prec);
  }
  arb_mul(res, prefactor, summand, prec);
}

double Nnl(double al, double nu, int n, int l){
  double mu = al*(1+2*l);
  return al*exp(lgamma(n+mu+2*nu)+lgamma(mu+1)-lgamma(n+2*mu+2*nu-1));
}

void arb_Nnl(arb_t res, arb_t al, arb_t nu, int nn, int ll, slong prec) {
  arb_t n; arb_t l; arb_t mu; arb_t x1;
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1);

  arb_set_si(n, nn); arb_set_si(l, ll); arb_one(mu);
  arb_add(mu, mu, l, prec); arb_add(mu, mu, l, prec);
  arb_mul(mu, mu, al, prec); // mu = al*(1 + 2*l)
  arb_zero(res); arb_one(x1); arb_neg(x1, x1); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec); // x1 = n+2mu+2nu-1
  arb_lgamma(x1, x1, prec);
  arb_sub(res, res, x1, prec); // res -= log(gamma(n+2mu+2nu-1))
  arb_one(x1); arb_add(x1, x1, mu, prec);
  arb_lgamma(x1, x1, prec);
  arb_add(res, res, x1, prec); // res += log(gamma(mu+1))
  arb_zero(x1); arb_add(x1, x1, n, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec);
  arb_lgamma(x1, x1, prec);
  arb_add(res, res, x1, prec); // res += log(gamma(n+mu+2nu))
  arb_exp(res, res, prec);
  arb_mul(res, res, al, prec);
}

double Anl(double rs, double al, double nu, int n, int l){
  double mu = al*(1+2*l);
  return al/pow(rs,2)*(2*n+2*mu+2*nu-1)*
    exp(lgamma(n+mu)+lgamma(n+2*mu+2*nu-1)-lgamma(n+mu+2*nu)-2*lgamma(mu)-lgamma(n+1));
}

void arb_Anl(arb_t res, arb_t rs, arb_t al, arb_t nu, int nn, int ll, slong prec) {
  arb_t n; arb_t l; arb_t mu; arb_t x1;
```

```
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1);

  arb_set_si(n, nn); arb_set_si(l, ll);
  arb_one(mu); arb_add(mu, mu, l, prec); arb_add(mu, mu, l, prec);
  arb_mul(mu, mu, al, prec); // mu = al*(1 + 2*l)
  arb_zero(res); arb_one(x1); arb_add(x1, x1, n, prec);
  arb_lgamma(x1, x1, prec);
  arb_sub(res, res, x1, prec); // res -= log(gamma(n+1))
  arb_zero(x1); arb_lgamma(x1, mu, prec);
  arb_sub(res, res, x1, prec); arb_sub(res, res, x1, prec); // res -= 2*log(gamma(mu))
  arb_zero(x1); arb_add(x1, x1, n, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec); arb_lgamma(x1, x1, prec);
  arb_sub(res, res, x1, prec); // res -= log(gamma(n+mu+2*nu))
  arb_one(x1); arb_neg(x1, x1); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec);
  arb_lgamma(x1, x1, prec); arb_add(res, res, x1, prec); // res += log(gamma(n+2*mu+2*nu-1))
  arb_zero(x1); arb_add(x1, x1, n ,prec); arb_add(x1, x1, mu, prec);
  arb_lgamma(x1, x1, prec); arb_add(res, res, x1, prec); // res += log(gamma(n+mu))
  arb_exp(res, res, prec);
  arb_one(x1); arb_neg(x1, x1);
  arb_add(x1, x1, n, prec); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec); // res *= (2n+2mu+2nu-1)
  arb_mul(res, res, x1, prec); arb_mul(res, res, al, prec); // res *= al
  arb_div(res, res, rs, prec); arb_div(res, res, rs, prec); // res /= rs^2
}

double Nlm(int l, int m){
  return (2*l+1) * exp(lgamma(l-abs(m)+1)-lgamma(l+abs(m)+1));
}

void arb_Nlm(arb_t res, int ll, int mm, slong prec) {
  arb_t l; arb_t m; arb_t x1;
  arb_init(l); arb_init(m); arb_init(x1);

  arb_set_si(l, ll); arb_set_si(m, abs(mm));
  arb_zero(res); arb_one(x1);
  arb_add(x1, x1, m, prec); arb_add(x1, x1, l, prec);
  arb_lgamma(x1, x1, prec);
  arb_sub(res, res, x1, prec); // res -= log(l+abs(m)+1);
  arb_one(x1); arb_sub(x1, x1, m, prec);
  arb_add(x1, x1, l, prec); arb_lgamma(x1, x1, prec);
  arb_add(res, res, x1, prec); // res += log(l-abs(m)+1);
  arb_exp(res, res, prec); arb_one(x1);
  arb_add(x1, x1, l, prec); arb_add(x1, x1, l, prec);
  arb_mul(res, res, x1, prec);
}

double Knl(double al, double nu, int n, int l){
  double mu = al*(1+2*l);
  return pow(al,-2)*
    exp(lgamma(n+1)+lgamma(mu+1)-lgamma(mu+n+1)) * (n+mu)/(2*n+2*mu+2*nu-1)/mu;
}

void arb_Knl(arb_t res, arb_t al, arb_t nu, int nn, int ll, slong prec) {
  arb_t n; arb_t l; arb_t mu; arb_t x1;
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1);

  arb_set_si(n, nn); arb_set_si(l, ll); arb_one(mu);
  arb_add(mu, mu, l, prec); arb_add(mu, mu, l, prec);
  arb_mul(mu, mu, al, prec); // mu = al*(1 + 2*l)
  arb_one(x1); arb_add(x1, x1, n, prec); arb_lgamma(x1, x1, prec);
  arb_zero(res); arb_add(res, res, x1, prec);
  arb_one(x1); arb_add(x1, x1, mu, prec); arb_lgamma(x1, x1, prec);
  arb_add(res, res, x1, prec);
  arb_one(x1); arb_add(x1, x1, mu, prec); arb_add(x1, x1, n, prec);
  arb_lgamma(x1, x1, prec);
```

```
  arb_sub(res, res, x1, prec); arb_exp(res, res, prec);
  arb_zero(x1); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec);
  arb_mul(res, res, x1, prec);
  arb_one(x1); arb_neg(x1, x1);
  arb_add(x1, x1, n, prec); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec); arb_add(x1, x1, mu, prec);
  arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec);
  arb_div(res, res, x1, prec); arb_div(res, res, mu, prec);
  arb_div(res, res, al, prec); arb_div(res, res, al, prec);
}

double jacobi (double x, int n, double a, double b) {
  if (n==0) {
    return 1.0;
  } else if (n==1) {
    return 0.5 * (a - b + (a + b + 2.0)*x);
  } else {
    double p0, p1, a1, a2, a3, a4, p2=0.0;
    int i;
    p0 = 1.0;
    p1 = 0.5 * (a - b + (a + b + 2)*x);
    for(i=1; i<n; ++i){
      a1 = 2.0*(i+1.0)*(i+a+b+1.0)*(2.0*i+a+b);
      a2 = (2.0*i+a+b+1.0)*(a*a-b*b);
      a3 = (2.0*i+a+b)*(2.0*i+a+b+1.0)*(2.0*i+a+b+2.0);
      a4 = 2.0*(i+a)*(i+b)*(2.0*i+a+b+2.0);
      p2 = 1.0/a1*( (a2 + a3*x)*p1 - a4*p0);
      p0 = p1;
      p1 = p2;
    }
    return p2;
  }
}

void arb_jacobi(arb_t res, const int n, const arb_t a,
                const arb_t b, const arb_t z, slong prec) {
  arb_t nn;
  arb_init(nn);
  arb_set_si(nn, n);
  arb_hypgeom_jacobi_p(res, nn, a, b, z, prec);
}

double Phinl(double al, double nu, int n, int l, double r){
  double mu = al*(1+2*l);
  double z2 = pow(r,1/al);
  double chi = z2/(1 + z2);
  double xi = (z2 - 1)/(z2 + 1);

  if (n == 0)
    return pow(r,-l-1)*beta(mu,nu,chi);

  return Phinl(al,nu,n-1,l,r) -
    2 * 1/mu*exp(lgamma(n)-lgamma(mu+n)+lgamma(mu+1))*
    pow(r,l)/pow(1+z2,mu+nu) * jacobi(xi,n-1,mu+2*nu-1,mu);
}

void arb_Phinl(arb_t res, arb_t al, arb_t nu, int nn, int ll, arb_t r, slong prec){
  arb_t n; arb_t l; arb_t mu; arb_t x1; arb_t z2;
  arb_t xi; arb_t chi; arb_t x2; arb_t x3;
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1); arb_init(z2);
  arb_init(xi); arb_init(chi); arb_init(x2); arb_init(x3);

  arb_set_si(n, nn); arb_set_si(l, ll); arb_one(mu);
  arb_add(mu, mu, l, prec); arb_add(mu, mu, l, prec);
  arb_mul(mu, mu, al, prec); // mu = al*(1 + 2*l)
  arb_zero(z2); arb_add(z2, z2, al, prec);
  arb_inv(z2, z2, prec); arb_pow(z2, r, z2, prec); // z2 = r^(1/al)
```

```
  arb_one(chi); arb_add(chi, chi, z2, prec);
  arb_div(chi, z2, chi, prec); // chi = z2/(1 + z2)
  arb_set_si(xi, -1);
  arb_add(xi, xi, chi, prec);
  arb_add(xi, xi, chi, prec); // xi = 2*chi - 1

  if (nn == 0) {
    unsigned int x = ll+1;
    arb_zero(x1);
    arb_pow_ui(x1, r, x, prec);
    arb_inv(x1, x1, prec);
    arb_beta(res, mu, nu, chi, prec);
    arb_mul(res, res, x1, prec);
    return;
  }

  arb_zero(x1); arb_zero(x2);
  arb_lgamma(x1, n, prec);
  arb_add(x2, x2, x1, prec);
  arb_zero(x1); arb_add(x1, x1, n, prec);
  arb_add(x1, x1, mu, prec); arb_lgamma(x1, x1, prec);
  arb_sub(x2, x2, x1, prec);
  arb_one(x1); arb_add(x1, x1, mu, prec);
  arb_lgamma(x1, x1, prec); arb_add(x2, x2, x1, prec);
  arb_exp(x2, x2, prec); arb_div(x2, x2, mu, prec);
  // x2 = 1/mu * gamma(n)*gamma(mu+1)/gamma(mu+n)
  arb_zero(x1);
  arb_add(x1, x1, mu, prec); arb_add(x1, x1, nu, prec);
  arb_one(x3);
  arb_add(x3, x3, z2, prec); arb_pow(x3, x3, x1, prec);
  arb_inv(x3, x3, prec); arb_mul(x2, x2, x3, prec);
  // x2 = 1/mu * gamma(n)*gamma(mu+1)/gamma(mu+n)/(1+z2)^(mu+nu)
  arb_zero(x1); arb_add(x1, x1, r, prec);
  arb_pow(x1, x1, l, prec); arb_mul(x2, x2, x1, prec);
  // x2 = 1/mu * gamma(n)*gamma(mu+1)/gamma(mu+n)/(1+z2)^(mu+nu) * r^l
  arb_zero(x3); arb_zero(x1);
  arb_set_si(x3, -1); arb_add(x3, x3, mu, prec);
  arb_add(x3, x3, nu, prec); arb_add(x3, x3, nu, prec);
  arb_jacobi(x1, nn-1, x3, mu, xi, prec);
  arb_mul(x2, x2, x1, prec); arb_add(x2, x2, x2, prec);
  arb_zero(res); arb_Phinl(res, al, nu, nn-1, ll, r, prec);
  arb_sub(res, res, x2, prec);
  return;
}

double DPhinl(double al, double nu, int n, int l, double r){
  double mu = al*(1+2*l);
  double z2 = pow(r,1/al);
  double chi = z2/(1 + z2);
  double xi = (z2 - 1)/(z2 + 1);

  if (n == 0)
    return 1/al*pow(r,l-1)/pow(1+z2,mu+nu) - (1+l)*pow(r,-l-2)*beta(mu,nu,chi);

  return DPhinl(al,nu,n-1,l,r) -
    2/(al*mu)*exp(lgamma(n)-lgamma(mu+n)+lgamma(mu+1))*pow(r,l-1)/pow(1+z2,mu+nu+2)*
    ((n+2*mu+2*nu-1)*z2*jacobi(xi,n-2,mu+2*nu,mu+1) +
    ↪  (1+z2)*(al*l*(1+z2)-z2*(mu+nu))*jacobi(xi,n-1,mu+2*nu-1,mu));
}

void arb_DPhinl(arb_t res, arb_t al, arb_t nu, int nn, int ll, arb_t r, slong prec){
  arb_t n; arb_t l; arb_t mu; arb_t x1; arb_t z2;
  arb_t xi; arb_t chi; arb_t x2; arb_t x3;
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1); arb_init(z2);
  arb_init(xi); arb_init(chi); arb_init(x2); arb_init(x3);

  arb_set_si(n, nn); arb_set_si(l, ll); arb_one(mu);
  arb_add(mu, mu, l, prec); arb_add(mu, mu, l, prec);
```

```
  arb_mul(mu, mu, al, prec); // mu = al*(1 + 2*l)
  arb_zero(z2); arb_add(z2, z2, al, prec);
  arb_inv(z2, z2, prec); arb_pow(z2, r, z2, prec); // z2 = r^(1/al)
  arb_one(chi); arb_add(chi, chi, z2, prec);
  arb_div(chi, z2, chi, prec); // chi = z2/(1 + z2)
  arb_set_si(xi, -1); arb_add(xi, xi, chi, prec);
  arb_add(xi, xi, chi, prec); // xi = 2*chi - 1

  if (nn == 0) {
    arb_zero(res); arb_set_si(res, ll-1);
    arb_pow(res, r, res, prec); arb_div(res, res, al, prec); // res = r^(l-1)/al
    arb_one(x1); arb_zero(x2); arb_add(x1, x1, z2, prec);
    arb_add(x2, x2, mu, prec); arb_add(x2, x2, nu, prec);
    arb_pow(x1, x1, x2, prec); // x1 = (1 + z2)^(mu+nu)
    arb_div(res, res, x1, prec); // res = 1/al*r^(l-1)/(1 + z2)^(mu+nu)
    arb_set_si(x1, 1+ll); arb_set_si(x2, -ll-2);
    arb_pow(x2, r, x2, prec); arb_beta(x3, mu, nu, chi, prec);
    arb_mul(x3, x3, x2, prec); arb_mul(x3, x3, x1, prec);
    arb_sub(res, res, x3, prec);
    return;
  }

  arb_one(x1); arb_add(x1, x1, z2, prec); arb_mul(x1, x1, l, prec);
  arb_mul(x1, x1, al, prec); arb_zero(x2); arb_add(x2, x2, mu, prec);
  arb_add(x2, x2, nu, prec); arb_mul(x2, x2, z2, prec);
  arb_sub(x1, x1, x2, prec); // x1 = (al*l*(1+z2)-z2*(mu+nu))
  arb_one(x2); arb_add(x2, x2, z2, prec); arb_mul(x1, x1, x2, prec);
  arb_one(x3); arb_neg(x3, x3); arb_add(x3, x3, mu, prec);
  arb_add(x3, x3, nu, prec); arb_add(x3, x3, nu, prec);
  arb_jacobi(x2, nn-1, x3, mu, xi, prec); arb_mul(x1, x1, x2, prec);
  // x1 = (1+z2)*(al*l*(1+z2)-z2*(mu+nu))*jacobi(xi,n-1,mu+2*nu-1,mu)
  arb_zero(x3); arb_add(x3, x3, mu, prec);
  arb_add(x3, x3, nu, prec); arb_add(x3, x3, nu, prec); arb_one(x2);
  arb_add(x2, x2, mu, prec); arb_jacobi(x3, nn-2, x3, x2, xi, prec);
  arb_set_si(x2, nn-1); arb_add(x2, x2, mu, prec); arb_add(x2, x2, mu, prec);
  arb_add(x2, x2, nu, prec); arb_add(x2, x2, nu, prec);
  arb_mul(x3, x3, x2, prec); arb_mul(x3, x3, z2, prec);
  // x3 = (n+2*mu+2*nu-1)*z2*jacobi(xi,n-2,mu+2*nu,mu+1)
  arb_add(x1, x3, x1, prec);
  // x1 = ((n+2*mu+2*nu-1)*z2*jacobi(xi,n-2,mu+2*nu,mu+1) +
  // (1+z2)*(al*l*(1+z2)-z2*(mu+nu))*jacobi(xi,n-1,mu+2*nu-1,mu))
  arb_zero(x2); arb_one(x3); arb_add(x3, x3, mu, prec);
  arb_lgamma(x3, x3, prec); arb_add(x2, x2, x3, prec);
  arb_zero(x3); arb_add(x3, x3, mu, prec);
  arb_add(x3, x3, n, prec); arb_lgamma(x3, x3, prec);
  arb_sub(x2, x2, x3, prec); arb_lgamma(x3, n, prec);
  arb_add(x2, x2, x3, prec);
  arb_exp(x2, x2, prec); // x2 = gamma(n)*gamma(mu+1)/gamma(mu+n)
  arb_mul(x1, x1, x2, prec); arb_set_si(x2, ll-1);
  arb_pow(x2, r, x2, prec); // x2 = r^(l-1)
  arb_mul(x1, x1, x2, prec);
  arb_one(x2); arb_add(x2, x2, z2, prec);
  arb_set_si(x3, 2); arb_add(x3, x3, mu, prec);
  arb_add(x3, x3, nu, prec);
  arb_pow(x2, x2, x3, prec); // x2 = (1+z2)^(mu+nu+2)
  arb_div(x1, x1, x2, prec); arb_div(x1, x1, al, prec);
  arb_div(x1, x1, mu, prec); arb_add(x1, x1, x1, prec);
  arb_DPhinl(res, al, nu, nn-1, ll, r, prec);
  arb_sub(res, res, x1, prec); // res = DPhinl(n-1) - ...
  return;
}

double PhinlDeriv(double al, double nu, int n, int l, double r){
  if (al < 0) {
    return -pow(r,-2)*Phinl(fabs(al),nu,n,l,1/r) -
      pow(r,-3)*DPhinl(fabs(al),nu,n,l,1/r);
  } else {
    return DPhinl(al,nu,n,l,r);
```

```
  }
}

void arb_PhinlDeriv(arb_t res, arb_t al, arb_t nu, int nn, int ll, arb_t r, slong prec){
  arb_t n; arb_t l; arb_t mu; arb_t x1; arb_t z2;
  arb_t xi; arb_t chi; arb_t x2; arb_t x3;
  arb_init(n); arb_init(l); arb_init(mu); arb_init(x1); arb_init(z2);
  arb_init(xi); arb_init(chi); arb_init(x2); arb_init(x3);
  arb_set_si(n, nn); arb_set_si(l, ll);

  if (arb_is_negative(al)) {
    arb_abs(x1, al); arb_inv(x2, r, prec);
    arb_Phinl(res, x1, nu, nn, ll, x2, prec);
    arb_mul(res, res, x2, prec); arb_mul(res, res, x2, prec);
    arb_DPhinl(x3, x1, nu, nn, ll, x2, prec);
    arb_mul(x3, x3, x2, prec); arb_mul(x3, x3, x2, prec);
    arb_mul(x3, x3, x2, prec); arb_add(res, res, x3, prec);
    arb_neg(res, res);
  } else {
    arb_DPhinl(res, al, nu, nn, ll, r, prec);
  }
  return;
}

double phirhoIntegralLower(double rs, double al, double nu, int n, double rt){
  // this is the integral from r=0 to r=rt/rs of r^2*phi_00*rho_n0
  if (n == 0) {
    return al*(2*al+2*nu-1)*(al+nu)*curlyI(al,nu,0,rt/rs);
  }
  return al*((n+2*al+2*nu-1)*(n+al+nu)*curlyI(al,nu,n,rt/rs) -
  ↪  (n+al+2*nu-1)*(n+al+nu-1)*curlyI(al,nu,n-1,rt/rs));
}

void arb_phirhoIntegralLower(arb_t res, arb_t rs, arb_t al,
                             arb_t nu, int nn, arb_t rt, slong prec){
  arb_t n; arb_t x1; arb_t x2; arb_t x3;
  arb_init(n); arb_init(x1); arb_init(x2); arb_init(x3);
  arb_set_si(n, nn);

  if (nn == 0) {
    arb_div(x1, rt, rs, prec);
    arb_curlyI(res, al, nu, 0, x1, prec);
    arb_zero(x1); arb_add(x1, x1, al, prec); arb_add(x1, x1, nu, prec);
    arb_mul(res, res, x1, prec); arb_set_si(x1, -1);
    arb_add(x1, x1, al, prec); arb_add(x1, x1, al, prec);
    arb_add(x1, x1, nu, prec); arb_add(x1, x1, nu, prec);
    arb_mul(res, res, x1, prec); arb_mul(res, res, al, prec);
    return;
  }
  arb_div(x1, rt, rs, prec); arb_curlyI(res, al, nu, nn, x1, prec);
  arb_zero(x2); arb_add(x2, x2, n, prec);
  arb_add(x2, x2, al, prec); arb_add(x2, x2, nu, prec);
  arb_mul(res, res, x2, prec); arb_one(x2); arb_neg(x2, x2);
  arb_add(x2, x2, n, prec); arb_add(x2, x2, al, prec);
  arb_add(x2, x2, al, prec); arb_add(x2, x2, nu, prec);
  arb_add(x2, x2, nu, prec); arb_mul(res, res, x2, prec);
  // res = (n+2*al+2*nu-1)*(n+al+nu)*curlyI(al,nu,n,rt/rs)
  arb_curlyI(x3, al, nu, nn-1, x1, prec); arb_set_si(x2, -1);
  arb_add(x2, x2, n, prec); arb_add(x2, x2, al, prec);
  arb_add(x2, x2, nu, prec); arb_mul(x3, x3, x2, prec);
  arb_set_si(x2, -1); arb_add(x2, x2, n, prec);
  arb_add(x2, x2, al, prec); arb_add(x2, x2, nu, prec);
  arb_add(x2, x2, nu, prec); arb_mul(x3, x3, x2, prec);
  // x3 = (n+al+2*nu-1)*(n+al+nu-1)*curlyI(al,nu,n-1,rt/rs)
  arb_sub(res, res, x3, prec); arb_mul(res, res, al, prec);
  return;
}
```

```
double phirhoIntegralUpper(double rs, double al, double nu, int n, double rt){
  // this is the integral from r=rt/rs to r=inf of phi_n00*rho_000
  if (al < 0) {
    if (n == 0) {
      return pow(rs,2)*Knl(fabs(al),nu,0,0)*
        phirhoIntegralLower(1/rs,fabs(al),nu,0,1/rt);
    } else {
      return pow(rs,2)*Knl(fabs(al),nu,n,0)*phirhoIntegralLower(1/rs,fabs(al),nu,n,1/rt) -
        pow(rs,4)*pow(rt,-2)*(DPhinl(fabs(al),nu,0,0,rs/rt)*Phinl(fabs(al),nu,n,0,rs/rt) -
                              Phinl(fabs(al),nu,0,0,rs/rt)*DPhinl(fabs(al),nu,n,0,rs/rt));
    }
  }
  if (n == 0) {
    return 1/Anl(rs,fabs(al),nu,0,0) -
      pow(rs,2)*Knl(fabs(al),nu,0,0)*phirhoIntegralLower(rs,fabs(al),nu,0,rt);
  } else {
    return pow(rt,2)*(DPhinl(fabs(al),nu,0,0,rt/rs)*Phinl(fabs(al),nu,n,0,rt/rs) -
                      Phinl(fabs(al),nu,0,0,rt/rs)*DPhinl(fabs(al),nu,n,0,rt/rs)) -
      pow(rs,2)*Knl(fabs(al),nu,n,0)*phirhoIntegralLower(rs,fabs(al),nu,n,rt);
  }
}

void arb_phirhoIntegralUpper(arb_t res, arb_t rs, arb_t al,
                             arb_t nu, int nn, arb_t rt, slong prec){
  arb_t n; arb_t x1; arb_t x2; arb_t x3; arb_t a; arb_t x4;
  arb_init(n); arb_init(x1); arb_init(x2);
  arb_init(x3); arb_init(x4); arb_init(a);
  arb_set_si(n, nn);
  arb_abs(a, al);

  if (arb_is_negative(al)) {
    if (nn == 0) {
      arb_inv(x1, rs, prec); arb_inv(x2, rt, prec);
      arb_phirhoIntegralLower(res, x1, a, nu, nn, x2, prec);
      arb_Knl(x1, a, nu, 0, 0, prec); arb_mul(res, res, x1, prec);
      arb_mul(res, res, rs, prec); arb_mul(res, res, rs, prec);
      return;
    }
    arb_div(x1, rs, rt, prec); arb_DPhinl(x2, a, nu, 0, 0, x1, prec);
    arb_Phinl(x3, a, nu, nn, 0, x1, prec); arb_mul(x2, x2, x3, prec);
    arb_Phinl(x3, a, nu, 0, 0, x1, prec); arb_DPhinl(x4, a, nu, nn, 0, x1, prec);
    arb_mul(x3, x3, x4, prec); arb_sub(x2, x2, x3, prec);
    arb_div(x2, x2, rt, prec); arb_div(x2, x2, rt, prec);
    arb_mul(x2, x2, rs, prec); arb_mul(x2, x2, rs, prec);
    arb_mul(x2, x2, rs, prec); arb_mul(x2, x2, rs, prec);
    arb_inv(x1, rs, prec); arb_inv(x3, rt, prec);
    arb_phirhoIntegralLower(res, x1, a, nu, nn, x3, prec);
    arb_Knl(x1, a, nu, nn, 0, prec);
    arb_mul(res, res, x1, prec); arb_mul(res, res, rs, prec);
    arb_mul(res, res, rs, prec); arb_sub(res, res, x2, prec);
    return;
  }
  if (nn == 0) {
    arb_Anl(x1, rs, a, nu, 0, 0, prec);
    arb_inv(x1, x1, prec);
    arb_phirhoIntegralLower(res, rs, a, nu, 0, rt, prec);
    arb_Knl(x2, a, nu, 0, 0, prec); arb_mul(res, res, x2, prec);
    arb_mul(res, res, rs, prec); arb_mul(res, res, rs, prec);
    arb_sub(res, x1, res, prec);
    return;
  }
  arb_div(x1, rt, rs, prec);
  arb_DPhinl(x2, a, nu, 0, 0, x1, prec);
  arb_Phinl(x3, a, nu, nn, 0, x1, prec);
  arb_mul(x2, x2, x3, prec);
  arb_Phinl(x3, a, nu, 0, 0, x1, prec);
  arb_DPhinl(x4, a, nu, nn, 0, x1, prec);
  arb_mul(x3, x3, x4, prec); arb_sub(x2, x2, x3, prec);
```

```
  arb_mul(x2, x2, rt, prec); arb_mul(x2, x2, rt, prec);
  arb_Knl(x3, a, nu, nn, 0, prec);
  arb_phirhoIntegralLower(x4, rs, a, nu, nn, rt, prec);
  arb_mul(x3, x3, x4, prec); arb_mul(x3, x3, rs, prec);
  arb_mul(x3, x3, rs, prec); arb_sub(res, x2, x3, prec);
  return;
}

int main(int argc, char *argv[]) {

  if (argc == 1) {
    printf("Usage: ./arb_tail_coeffs digits alpha nu r_t r_s nmax\n");
    return 0;
  }

  arb_t alal; arb_t nunu; arb_t rtrt; arb_t rsrs; arb_t res;
  arb_init(alal); arb_init(nunu); arb_init(rtrt); arb_init(res); arb_init(rsrs);

  slong prec, digits, condense;
  digits = atol(argv[1]);
  prec = digits * 3.3219280948873623 + 5;

  double al = atof(argv[2]);
  double nu = atof(argv[3]);
  double rt = atof(argv[4]);
  double rs = atof(argv[5]);
  int n = atoi(argv[6]);

  arb_set_d(alal, al);
  arb_set_d(nunu, nu);
  arb_set_d(rtrt, rt);
  arb_set_d(rsrs, rs);

  for (int i = 0; i <= n; i++) {
    arb_phirhoIntegralUpper(res, rsrs, alal, nunu, i, rtrt, prec);
    printf("%.20f\t%.20f\n", arb_get_d(res), phirhoIntegralUpper(rs,al,nu,i,rt));
  }

  flint_cleanup();
  return 0;
}
```

# References


[A&S] Abramowitz M., Stegun I. A. 1972, Handbook of Mathematical Functions. Dover

Aarseth S., 1967, in Les Nouvelles Méthodes de la Dynamique Stellaire. p. 47

Aarseth S. J., 1999, PASP, 111, 1333

Allen A. J., Palmer P. L., Papaloizou J., 1990, MNRAS, 242, 576

Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781

An J. H., Evans N. W., 2006, AJ, 131, 782

An J., Zhao H., 2013, MNRAS, 428, 2805

Andredakis Y. C., Peletier R. F., Balcells M., 1995, MNRAS, 275, 874

Aoki S., Iye M., 1978, PASJ, 30, 519

[BMP] Erdélyi, A. and Bateman, H. et al. (eds) 1954, Tables of integral transforms. Calif. Inst. Technol. Bateman Manuscr. Project, McGraw-Hill, New York, NY

Baes M., Dejonghe H., 2002, A&A, 393, 485

Barnes J. E., Hernquist L., 1992, ARA&A, 30, 705

Barnes J., Hut P., 1986, Nature, 324, 446

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109

Belokurov V., et al., 2014, MNRAS, 437, 116

Belokurov V., Deason A. J., Erkal D., Koposov S. E., Carballo-Bello J. A., Smith M. C., Jethwa P., Navarrete C., 2019, MNRAS, 488, L47

Bertin G., 2014, Dynamics of Galaxies. Cambridge University Press

Besla G., Garavito-Camargo N., 2020, in Valluri M., Sellwood J. A., eds, IAU Symposium Vol. 353, IAU Symposium. pp 123–127, doi:10.1017/S1743921319008494

Bethe H. A., Salpeter E. E., 1957, Quantum Mechanics of One- and Two-Electron Atoms. Dover

Binney J., Tremaine S., 1987, Galactic dynamics. Princeton, NJ, Princeton University Press, 1987, 747 p.

Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102

Bochner S., 1929, Mathematische Zeitschrift, 29, 730

Bonaca A., Hogg D. W., 2018, ApJ, 867, 101

Bovy J., et al., 2012, ApJ, 759, 131

Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93

Brown M. J. W., Papaloizou J. C. B., 1998, MNRAS, 300, 135

Bryan S. E., Kay S. T., Duffy A. R., Schaye J., Dalla Vecchia C., Booth C. M., 2013, MNRAS,

429, 3316
Burkert A., 2020, arXiv e-prints, p. arXiv:2006.11111
Cantero M. J., Iserles A., 2012, SIAM Journal on Numerical Analysis, 50, 307
Caon N., Capaccioli M., D'Onofrio M., 1993, MNRAS, 265, 1013
Chaundy T. W., 1943, The Quarterly Journal of Mathematics, os-14, 55
Ciotti L., 1991, A&A, 249, 99
Clutton-Brock M., 1972, Ap&SS, 16, 101
Clutton-Brock M., 1973, Ap&SS, 23, 55
Cohl H. S., 2013, SIGMA, 9, 042
Cuddeford P., 1991, MNRAS, 253, 414
Cunningham E. C., et al., 2020, arXiv e-prints, p. arXiv:2006.08621
[DLMF] Olver F. W. J., Olde Daalhuis A. B. et al. (eds) 2020, NIST Digital Library of Mathematical Functions, Release 1.0.26, `http://dlmf.nist.gov/`
Dai B., Robertson B. E., Madau P., 2018, ApJ, 858, 73
Debattista V. P., Roškar R., Valluri M., Quinn T., Moore B., Wadsley J., 2013, MNRAS, 434, 2971
Dehnen W., 1993, MNRAS, 265, 250
Dehnen W., 2000, ApJ, 536, L39
Dehnen W., 2001, MNRAS, 324, 273
Dehnen W., 2014, Computational Astrophysics and Cosmology, 1, 1
Dehnen W., McLaughlin D. E., 2005, MNRAS, 363, 1057
Dehnen W., Read J. I., 2011, European Physical Journal Plus, 126, 55
Dejonghe H., 1987, MNRAS, 224, 13
Dekel A., Ishai G., Dutton A. A., Maccio A. V., 2017, MNRAS, 468, 1005
Del Popolo A., Le Delliou M., 2017, Galaxies, 5, 17
Diemand J., Zemp M., Moore B., Stadel J., Carollo C. M., 2005, MNRAS, 364, 665
Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1
Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108
Diemer B., Mansfield P., Kravtsov A. V., More S., 2017, ApJ, 843, 140
Dodelson S., 2003, Modern cosmology, 1 edn. Academic Press
Dubinski J., Carlberg R. G., 1991, ApJ, 378, 496
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Eddington A. S., 1916, MNRAS, 76, 572
Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87
Einasto J., Haud U., 1989, A&A, 223, 89
Emsellem E., Monnet G., Bacon R., 1994, A&A, 285, 723
Erkal D., et al., 2019, MNRAS, 487, 2685
Estrada R., Kanwal R., 1989, Journal of Mathematical Analysis and Applications, 141, 195
Evans N. W., An J. H., 2005, MNRAS, 360, 492
Evans N. W., An J. H., 2006, Phys. Rev. D, 73, 023524
Evans N. W., Read J. C. A., 1998, MNRAS, 300, 106
Evans N., Witt H., 2003, Mon. Not. Roy. Astron. Soc., 345, 1351

Everitt W. N., 2005, A Catalogue of Sturm-Liouville Differential Equations. Birkhäuser Basel, Basel, doi:10.1007/3-7643-7359-8_12
Fields J. L., Wimp J., 1961, Math. Comp., 15, 390
Filter E., Steinborn E. O., 1980, Journal of Mathematical Physics, 21, 2725
Fouvry J.-B., 2016, PhD thesis. Université Pierre et Marie Curie
Fouvry J.-B., Bar-Or B., 2018, MNRAS, 481, 4566
Fridman A. M., Polyachenko V. L., 1984, Physics of gravitating systems. I. Equilibrium and stability.. Springer Verlag
[G&R] Gradshteyn I., Ryzhik I. 2007, Table of Integrals, Series and Products (Seventh Edition). Academic Press
Gaia Collaboration et al., 2016, A&A, 595, A2
Galassi M., 2003, GNU scientific library : reference manual. Network Theory, Bristol
Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, MNRAS, 387, 536
Garavito-Camargo N., Besla G., Laporte C. F. P., Johnston K. V., Gómez F. A., Watkins L. L., 2019, ApJ, 884, 51
Gerhard O., Kronawitter A., Saglia R. P., Bender R., 2001, AJ, 121, 1936
Giusti A., 2020, Phys. Rev. D, 101, 124029
Gómez F. A., Besla G., Carpintero D. D., Villalobos Á., O'Shea B. W., Bell E. F., 2015, ApJ, 802, 128
Greengard L., Rokhlin V., 1987, Journal of Computational Physics, 73, 325
Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1
[HTF] Erdélyi, A. and Bateman, H. et al. (eds) 1955, Higher transcendental functions. Calif. Inst. Technol. Bateman Manuscr. Project, McGraw-Hill, New York, NY
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hall P., 1983, Ann. Statist., 11, 1156
Hamilton C., Fouvry J.-B., Binney J., Pichon C., 2018, MNRAS, 481, 2041
Hart W., Johansson F., Pancratz S., 2013, FLINT: Fast Library for Number Theory
Hernquist L., 1990, ApJ, 356, 359
Hernquist L., Ostriker J. P., 1992, ApJ, 386, 375
Hernquist L., Sigurdsson S., Bryan G. L., 1995, ApJ, 446, 717
Holley-Bockelmann K., Weinberg M., Katz N., 2005, MNRAS, 363, 991
Homeier H. H. H., Weniger E. J., Steinborn E. O., 1992, International Journal of Quantum Chemistry, 44, 405
Hozumi S., Iwasawa M., Nitadori K., 2019, ApJ, 875, 20
Hunter C., 1963, MNRAS, 126, 299
Iserles A., 2010, Numerische Mathematik, 117, 529
Jaffe W., 1983, MNRAS, 202, 995
Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Johansson F., 2017, IEEE Transactions on Computers, 66, 1281
Kalapotharakos C., Efthymiopoulos C., Voglis N., 2008, MNRAS, 383, 971

Kalnajs A. J., 1971, ApJ, 166, 275
Kalnajs A. J., 1972, ApJ, 175, 63
Kalnajs A. J., 1976, ApJ, 205, 745
Klypin A., Kravtsov A. V., Bullock J. S., Primack J. R., 2001, ApJ, 554, 903
Kronawitter A., Saglia R. P., Gerhard O., Bender R., 2000, A&AS, 144, 53
Kuijken K., Tremaine S., 1994, ApJ, 421, 178
Kuzio de Naray R., McGaugh S. S., de Blok W. J. G., Bosma A., 2006, ApJS, 165, 461
Kuzmin G. G., 1956, Publications of the Tartu Astrofizica Observatory, 33, 75
Law D. R., Majewski S. R., 2010a, ApJ, 714, 229
Law D. R., Majewski S. R., 2010b, ApJ, 714, 229
Lilley E. J., Evans N. W., Sanders J. L., 2018a, MNRAS, 476, 2086
Lilley E. J., Sanders J. L., Evans N. W., Erkal D., 2018b, MNRAS, 476, 2092
Lilley E. J., Sanders J. L., Evans N. W., 2018c, MNRAS, 478, 1281
Lindegren L., et al., 2016, A&A, 595, A4
Lokas E. L., Mamon G. A., 2001, MNRAS, 321, 155
Lowing B., Jenkins A., Eke V., Frenk C., 2011, MNRAS, 416, 2697
Luke Y. L., 1959, Mathematical Tables and Other Aids to Computation, 13, 261
Lynden-Bell D., 2010, MNRAS, 402, 1937
Lynden-Bell D., 2015, MNRAS, 447, 1962
Macciò A. V., Paduroiu S., Anderhalden D., Schneider A., Moore B., 2012, MNRAS, 424, 1105
Manzini G., Delzanno G., Vencels J., Markidis S., 2016, Journal of Computational Physics, 317, 82
McGlynn T. A., 1984, ApJ, 281, 13
Meiron Y., Li B., Holley-Bockelmann K., Spurzem R., 2014, ApJ, 792, 98
Merritt D., 1985, AJ, 90, 1027
Merritt D., Saha P., 1993, ApJ, 409, 75
Merritt D., Graham A. W., Moore B., Diemand J., Terzić B., 2006, AJ, 132, 2685
Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge University Press, Cambridge
Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Morris R. M., Leach P. G. L., 2015, arXiv e-prints, p. arXiv:1510.08992
Mullen J. A., 1966, SIAM Journal on Applied Mathematics, 14, 1152
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Navarro J. F., et al., 2004, MNRAS, 349, 1039
Ngan W., Bozek B., Carlberg R. G., Wyse R. F. G., Szalay A. S., Madau P., 2015, ApJ, 803, 75
Ngan W., Carlberg R. G., Bozek B., Wyse R. F. G., Szalay A. S., Madau P., 2016, ApJ, 818, 194
Oñorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, MNRAS, 437, 1894
Osipkov L. P., 1979, Pisma v Astronomicheskii Zhurnal, 5, 77
Palmer P. L., ed. 1994, Stability of collisionless stellar systems: mechanisms for the dynamical structure of galaxies Astrophysics and Space Science Library Vol. 185, doi:10.1007/978-94-017-

3059–4.
Pearson S., Küpper A. H. W., Johnston K. V., Price-Whelan A. M., 2015, The Astrophysical Journal, 799, 28
Petersen M. S., Katz N., Weinberg M. D., 2016a, Phys. Rev. D, 94, 123013
Petersen M. S., Weinberg M. D., Katz N., 2016b, MNRAS, 463, 1952
Planck Collaboration et al., 2014, A&A, 571, A16
Plummer H. C., 1911, MNRAS, 71, 460
Polyachenko V. L., Shukhman I. G., 1981, Soviet Ast., 25, 533
Prada J., Forero-Romero J. E., Grand R. J. J., Pakmor R., Springel V., 2019, MNRAS, 490, 4877
Press W. H., Schechter P., 1974, ApJ, 187, 425
Prugniel P., Simien F., 1997, A&A, 321, 111
Qian E. E., 1993, MNRAS, 263, 394
Rahmati A., Jalali M. A., 2009, MNRAS, 393, 1459
Rathie A., Pogány T., 2020, Scientia, Series A, Math. Sci., 30
Robijn F. H. A., Earn D. J. D., 1996, MNRAS, 282, 1129
Saad N., Hall R. L., 2003, Journal of Physics A: Mathematical and General, 36, 7771
Sack R. A., 1964, Journal of Mathematical Physics, 5, 252
Saha P., 1991, MNRAS, 248, 494
Saha P., 1993, MNRAS, 262, 1062
Sanders J. L., Binney J., 2015, MNRAS, 449, 3479
Sellwood J. A., 2003, ApJ, 587, 638
Sellwood J. A., 2004, in Ryder S., Pisano D., Walker M., Freeman K., eds, IAU Symposium Vol. 220, Dark Matter in Galaxies. p. 27 (arXiv:astro-ph/0401398)
Sersic J. L., 1968, Atlas de galaxias australes. Cordoba, Argentina: Observatorio Astronomico, 1968
Shao S., Cautun M., Deason A. J., Frenk C. S., 2020, arXiv e-prints, p. arXiv:2005.03025
Shapiro P. R., Iliev I. T., Raga A. C., 1999, MNRAS, 307, 203
Silverman R., 1986, Density Estimation. Chapman and Hall, London
Siopis C., et al., 2009, ApJ, 693, 946
Slater L., 1966, Generalized Hypergeometric Functions. Cambridge University Press
Springel V., 2005, MNRAS, 364, 1105
Srivastava H. M., Karlsson P. W., 1985, Multiple Gaussian hypergeometric series. Ellis Horwood, Chichester
Srivastava H. M., Manocha H. L., 1984, A treatise on generating functions. E. Horwood Halsted Press, Chichester, West Sussex, England New York
Toomre A., 1963, ApJ, 138, 385
Tremaine S. D., 1976, MNRAS, 175, 557
Tremaine S., Richstone D. O., Byun Y.-I., Dressler A., Faber S. M., Grillmair C., Kormendy J., Lauer T. R., 1994, AJ, 107, 634
Trenti M., Hut P., 2008, Scholarpedia, 3, 3930
Valluri M., Merritt D., Emsellem E., 2004, ApJ, 602, 66

Van De Ven G., De Zeeuw P. T., Van Den Bosch R. C. E., 2008, MNRAS, 385, 614
Vasiliev E., 2013, MNRAS, 434, 3174
Vasiliev E., 2018, AGAMA: Action-based galaxy modeling framework (arxiv:1802.08255)
Vasiliev E., 2019, MNRAS, 482, 1525
Veltmann U. I. K., 1979, Soviet Ast., 23, 551
Vera-Ciro C., Helmi A., 2013, ApJ, 773, L4
Wang L., Spurzem R., Aarseth S., Nitadori K., Berczik P., Kouwenhoven M. B. N., Naab T., 2015, MNRAS, 450, 4070
Wei C., Wang X., Li Y., 2013, Advances in Difference Equations, 2013, 360
Weinberg M. D., 1989, MNRAS, 239, 549
Weinberg M. D., 1999, AJ, 117, 629
Weniger E. J., 1985, Journal of Mathematical Physics, 26, 276
Weniger E. J., 2007, arXiv e-prints, p. arXiv:0704.1088
Weniger E. J., 2009, International Journal of Quantum Chemistry, 109, 1706
Weniger E. J., 2011, arXiv e-prints, p. arXiv:1107.4670
Weniger E. J., Steinborn E. O., 1989, Journal of Mathematical Physics, 30, 774
White S. D. M., Efstathiou G., Frenk C. S., 1993, MNRAS, 262, 1023
Widrow L. M., 2000, ApJS, 131, 39
Wilkinson M. I., Evans N. W., 1999, MNRAS, 310, 645
Zhao H., 1996, MNRAS, 278, 488
Zhao H., Silk J., 2005, Phys. Rev. Lett., 95, 011301
de Blok W. J. G., 2010, Advances in Astronomy, 2010, 789293
de Lorenzi F., Debattista V. P., Gerhard O., Sambhus N., 2007, MNRAS, 376, 71
de Vaucouleurs G., 1953, MNRAS, 113, 134
de Zeeuw T., 1985, MNRAS, 216, 273
van den Bosch R. C. E., van de Ven G., Verolme E. K., Cappellari M., de Zeeuw P. T., 2008, MNRAS, 385, 647
van der Marel R. P., Franx M., 1993, ApJ, 407, 525