

Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

René Andrae

born in: Berlin, Germany

Oral examination: October 12, 2011

Alignment of Disc Galaxies

Testing the theory of angular-momentum acquisition

Referees: Dr. Knud Jahnke
Prof. Hans-Walter Rix

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Entstehung von Scheibengalaxien, welche sich durch ihre geordnete Rotation von anderen Galaxienarten unterscheiden. Im Fokus steht die Überprüfung theoretischer Vorhersagen, welche zu erklären versuchen wie diese Scheibengalaxien zu ihrem Drehimpuls kommen. Ein wesentlicher Teil der Arbeit ist die Untersuchung wie sich Galaxienformen sinnvoll und korrekt quantifizieren lassen. Dies ist eine wesentliche Voraussetzung für die automatisierte und reproduzierbare Klassifikation von Scheibengalaxien in umfangreichen Datensätzen. Besonders vertieft wird der Ansatz Galaxienmorphologien in orthonormale Basisfunktionen zu zerlegen, welche auf einem realistischen Lichtprofil beruhen – dem Sérsicprofil. Der Hauptteil der Arbeit befasst sich mit der statistischen Analyse der Drehimpulse mehrerer Tausend Scheibengalaxien aus dem Sloan Digital Sky Survey. Hierbei wird besonderer Wert auf die korrekte Behandlung aller relevanten Fehlerquellen gelegt. Es wird detailliert aufgezeigt, dass bisherige Untersuchungen die wesentlichen Fehlerquellen unzureichend betrachtet haben und folglich eine unzutreffende Bestätigung theoretischer Vorhersagen lieferten. Bei Berücksichtigung dieser Fehlerquellen findet sich keine statistische Evidenz für die Korrektheit dieser Vorhersagen. Die hierbei entwickelten Methoden lassen sich leicht auf andere astrophysikalische Fragestellungen verallgemeinern. Die Arbeit schließt mit einer detaillierten Analyse möglicher Verbesserungen und einer Beobachtungsstrategie, welche die Datenlage hinreichend verbessern sollte, sodass die theoretischen Vorhersagen in absehbarer Zeit mit größerer Genauigkeit überprüft werden können.

Summary

This thesis is concerned with the formation of disc galaxies, a special type of galaxies characterised by ordered rotation. The main focus is on testing predictions of theories that try to explain the angular-momentum acquisition of these galaxies. A major issue concerns the proper quantification of galaxy morphologies. This is of vital importance for automated and reproducible classification of disc galaxies in large data samples. The most promising approach is to expand galaxy morphologies into orthonormal basis functions based on a well motivated light profile – the Sérsic profile. In the main part of this thesis, the angular momenta of several thousand disc galaxies from the Sloan Digital Sky Survey are analysed. It is demonstrated that previous investigations lacked a rigorous account for the relevant error sources, which lead to unjustified confirmations of theoretical predictions. If all relevant errors are considered, there is no statistical evidence confirming these predictions. The methods developed here are also relevant in a wider context, as they are applicable to other astrophysical investigations. This thesis concludes with a detailed discussion of potential improvements of measurements and an outline of an observational strategy that will allow for a more decisive empirical test of the theoretical predictions.

Contents

1	Introduction	1
I	Galaxy Morphology	5
2	Parametrising galaxy morphologies	7
2.1	Galaxy morphologies	7
2.2	Parametrisation schemes	9
2.2.1	Model-independent schemes I: CAS	9
2.2.2	Model-independent schemes II: M_{20} and Gini	10
2.2.3	Model-based schemes I: Sérsic profile	10
2.2.4	Model-based schemes II: Basis functions	14
2.2.5	Assumptions	16
2.3	Intertwinement of morphological observables	18
2.3.1	Example I: Sérsic profile vs. concentration index	18
2.3.2	Simulating ellipticity and lopsidedness	20
2.3.3	Example II: Steepness of light profile vs. ellipticity	21
2.3.4	Example III: Impact of lopsidedness on centroid estimation	22
2.3.5	Example IV: Impact of lopsidedness on ellipticity estimates	25
2.3.6	Reliability assessment	26
2.3.7	How to disentangle observables	26
2.4	Impact of PSF on the concentration index	27
2.4.1	Forward vs. backward PSF modelling	27
2.4.2	Impact on concentration index	28
2.5	Parametrisation and classification	29
2.5.1	Loss of discriminative information	29
2.5.2	Nonlinear parameter spaces	31
2.5.3	Discontinuous parameter spaces	34
2.5.4	High-dimensional parameter spaces	35

3	Polar sérsiclets and beyond	37
3.1	General formalism	37
3.1.1	Weight function and orthonormalisation	38
3.1.2	Cartesian basis functions	38
3.1.3	Polar basis functions	39
3.1.4	Elliptical basis functions	40
3.1.5	Basis expansion	40
3.2	Shapelets and sérsiclets	41
3.2.1	Pros and cons of the Gaussian profile	41
3.2.2	The first attempt to introduce sérsiclets	41
3.2.3	Mathematical justification	43
3.2.4	Mathematical derivation of radial parts	44
3.2.5	Definition of polar sérsiclets	45
3.2.6	Real-valued models	46
3.2.7	Interpretation of the Sérsic index	47
3.3	Measures and operations in sérsiclet space	47
3.3.1	Image flux and second moments	48
3.3.2	Rotations in model space	49
3.3.3	Parity flips in model space	49
3.3.4	Forward PSF modelling	50
3.3.5	Inferring spiral-arm handedness	50
3.4	Optimisation procedure for sérsiclet expansion	51
3.4.1	Maximum-likelihood solution	51
3.4.2	Optimising the nonlinear parameters for given N_{\max}	52
3.4.3	Estimating the maximum order	52
3.4.4	Error estimation	53
3.5	Testing sérsiclets	53
3.5.1	Completeness	53
3.5.2	Image decompositions	54
3.5.3	Orthonormality and sampling	55
3.5.4	Impact of undersampling	58
3.5.5	Analogy to Nyquist frequency	60
3.6	Application to weak-lensing data	61
3.7	Orthonormalising higher-order Taylor expansions	63
3.7.1	Third-order profiles	63
3.7.2	Numerical orthonormalisation	65
3.7.3	Revisiting the problems of sérsiclets	67
3.7.4	Computational feasibility	69
3.8	Assumptions	69

II Disc Alignment 71

4	Disc alignment in the Local Group	73
4.1	Angular-momentum-orientation vector of the Milky Way	73
4.2	Angular-momentum-orientation vectors of Andromeda, M33, and the LMC	75
4.2.1	Andromeda (M31)	75
4.2.2	Triangulum Galaxy (M33)	76

4.2.3	Large Magellanic Cloud (LMC)	79
4.3	Hypothesis test of random orientation	79
4.4	The next step	80
5	The data	83
5.1	Choice of data set	83
5.2	Sloan Digital Sky Survey	84
5.2.1	Telescope and instrument	84
5.2.2	Imaging data and meta-information	85
5.3	Galaxy Zoo	86
5.3.1	Project description	86
5.3.2	How to operate Galaxy Zoo	87
5.3.3	Criticism	87
5.4	Classifications by Huertas-Company et al. (2011)	88
5.5	Data selection	88
5.5.1	Handedness sample	89
5.5.2	Angular-momentum-orientation sample	90
5.6	Cosmological model	90
6	Measurements and results	93
6.1	Previous work and motivation	93
6.2	From axis ratio to angular-momentum orientation	94
6.2.1	Local Cartesian coordinates	94
6.2.2	Disc orientation	95
6.2.3	Global coordinate system	95
6.2.4	Thick-disc approximation	96
6.2.5	Geometric degeneracies	96
6.2.6	Leading spiral arms	97
6.3	Correlation estimators	98
6.3.1	General correlation estimator	98
6.3.2	Error estimation	98
6.3.3	Angular-momentum orientation	99
6.3.4	Handedness	100
6.4	The impact of errors	101
6.4.1	Conditional vs. marginal errors	101
6.4.2	Error sources	102
6.4.3	Propagating errors numerically	105
6.4.4	Impact on autocorrelation of handedness	106
6.4.5	Impact on autocorrelation of angular-momentum orientation	109
6.4.6	Consequences for parameter estimation	112
7	Improvements and potential of future surveys	117
7.1	Biased ellipticity estimates from second moments	117
7.1.1	Revealing the bias	118
7.1.2	Point-spread function	118
7.1.3	Galactic bulges	120
7.1.4	Simulating pairs of angular-momentum-orientation vectors	121
7.1.5	Bias simulation	125

7.1.6	Discussion	128
7.2	Improving number statistics	129
7.3	Improving redshift estimates	130
7.4	Morphological classification in future surveys	130
7.5	Front-edge estimation	131
7.5.1	Visual classification	131
7.5.2	Automated classification	132
8	Summary and outlook	135
8.1	Outlining an observational strategy	137
	Acknowledgements	141

List of Figures

1.1	Roadmap of this thesis.	4
2.1	Three examples of Sérsic profiles.	11
2.2	Definition of scale radii for Sérsic profiles.	12
2.3	Artificial correlations of Sérsic index and scale radius.	13
2.4	Initial guess for fitting Sérsic profiles.	14
2.5	Comparing concentration and Sérsic indices.	19
2.6	Impact of ellipticity on concentration estimates.	21
2.7	Gaussian profiles of different lopsidedness.	23
2.8	Impact of lopsidedness on various morphological quantities.	24
2.9	Impact of lopsidedness on ellipticity estimates.	25
2.10	Impact of PSF on estimation of concentration index.	28
2.11	Discriminative information of concentration index.	30
2.12	Profiles used for demonstration of nonlinearity.	32
2.13	Trajectories in CA-Gini subspaces revealing nonlinearities.	33
2.14	Discontinuity of concentration index and M_{20}	34
3.1	Cartesian ground states using different weight functions.	39
3.2	Radial basis functions used by Ngan et al. (2009).	42
3.3	Polar sérsiclet basis functions.	47
3.4	Rotation and parity flip in sérsiclet space.	49
3.5	Testing the completeness of the sérsiclet basis.	54
3.6	Comparing example fits of shapelets and sérsiclets.	55
3.7	Demonstrating undersampling for polar shapelets.	56
3.8	Demonstrating undersampling for general sérsiclets.	57
3.9	Impact of undersampling on parameter estimation.	59
3.10	Test results of applying sérsiclets to the GREAT08 data.	62
3.11	Third-order profile vs. Sérsic profile.	64
3.12	Fitting Sérsic and third-order profile to an elliptical galaxy.	64
3.13	Model comparison via distributions of normalised residuals.	65
3.14	Non-invertibility of substitution variable.	66
3.15	Performance of numerical orthonormalisation algorithms.	67
3.16	Undersampling test for third-order profiles.	68

4.1	Artist’s impression of the Milky Way.	74
4.2	GALEX Near-UV and false-colour image of Andromeda.	76
4.3	GALEX Near-UV image of M33.	77
4.4	Front-edge estimation of M33 via Cepheid distances.	78
4.5	KS-test of randomly oriented angular momenta in the Local Group.	80
4.6	KS-test of randomly oriented spin axes in the Local Group.	81
5.1	Distribution of galaxy redshifts in SDSS DR 7.	84
5.2	SDSS filter response functions.	85
5.3	Comoving distance and linear Hubble distance.	91
6.1	Local coordinate system for inferring angular-momentum orientation.	94
6.2	Geometric degeneracies in estimating the angular-momentum orientation.	97
6.3	A sketch of conditional and marginal errors.	101
6.4	Errors in estimates of comoving distance.	103
6.5	Errors in isophotal-ellipticity estimates.	104
6.6	Conditional estimates of the handedness autocorrelation.	107
6.7	Conditional estimate of the handedness autocorrelation.	108
6.8	Marginal estimate of the handedness autocorrelation.	108
6.9	Conditional angular-momentum-orientation autocorrelations.	110
6.10	Conditional and marginal angular-momentum-orientation autocorrelations.	111
6.11	Distribution of Monte-Carlo realisations in distance bins.	112
6.12	Marginal autocorrelation estimate vs. results of Lee (2011).	113
6.13	Marginal likelihoods of fitting the marginal angular-momentum-orientation autocorrelation.	114
7.1	Angular-momentum-orientation autocorrelation biased by second moments.	118
7.2	Comparing ellipticities estimated from second moments and isophotes for Scd galaxies.	119
7.3	PSF biasing ellipticities from second moments.	119
7.4	Bulge-disc decomposition of an example Scd galaxy.	120
7.5	Comparing ellipticities estimated from second moments and isophotes for Sab galaxies.	121
7.6	Angular autocorrelations of angular-momentum-orientation vectors.	125
7.7	Debiasing the autocorrelation function of angular-momentum-orientation vectors.	126
7.8	Self-consistency test of debiasing the autocorrelation function.	127
7.9	Impact of number statistics on the errors of handedness correlations.	129
8.1	6dF survey of spectroscopic redshifts on southern hemisphere.	138

List of Tables

2.1	Characteristics of parametrisation schemes.	17
3.1	Minimal scale radii avoiding undersampling.	61
3.2	Data sets chosen from the GREAT08 sample.	62

1

Introduction

Disc galaxies constitute a significant part of the galaxy population in the nearby universe. Bamford et al. (2009) find a number fraction as high as $\approx 40\%$ of galaxies with redshifts $z \leq 0.2$ to be disc galaxies in the spectroscopic sample of the Sloan Digital Sky Survey (York et al. 2000). This type of galaxies is supported against gravity by rotation of its baryonic matter. Considering the stars and gas of the galaxy as a fluid, it is then obvious that rotational flattening forms a galactic disc. Conversely, spheroidal galaxies – another significant part of the galaxy population – are supported by the velocity dispersion of their stellar constituents and are generally considered to form through hierarchical merging (e.g. Toomre & Toomre 1972; Dressler 1980). Consequently, ordered rotation is an essential and defining property of disc galaxies. Understanding angular-momentum acquisition therefore provides the key to learn how a significant part of the galaxy population has formed. The current theoretical model explaining the angular-momentum acquisition of disc galaxies – the tidal-torque theory – predicts that the angular-momentum vectors of disc galaxies should be aligned to a certain extent, i.e., the orientations and inclinations of disc galaxies at small separations are not supposed to be completely random (for a recent review see Schäfer 2009).

The basic theoretical picture is rather simple. In the current cosmological framework, galaxies reside in the centres of non-baryonic dark-matter host haloes. These dark-matter haloes evolved from primordial density fluctuations as observed in the cosmic microwave background and have formed under gravitational collapse. In contrast to collisionless dark matter, baryonic matter can cool through dissipative processes and sink deeper into the potential well of the assembling halo, eventually reaching densities large enough to form stars and galaxies. However, this formation process does not take place in isolation. The collapsing proto-halo is embedded into the large-scale structure of the universe which exerts tidal gravitational forces during the collapse. These tidal forces cause the collapsing proto-halo to acquire angular momentum. Numerical simulations (van den Bosch et al. 2002) have shown that the angular momentum of the dark-matter host halo is largely inherited by the baryonic matter, i.e., the angular momenta of the resulting disc galaxy

and its host halo are approximately aligned. This enables us to estimate the angular-momentum orientation of the dark-matter host halo from the angular-momentum orientation of the disc galaxy residing in that halo.

The alignment effects arise naturally in this theoretical framework. If two disc galaxies formed in each other's vicinities, their dark-matter haloes collapsed in the same large-scale environment, i.e., both haloes were subject to potentially similar tidal forces imprinting the angular momenta. Consequently, theory predicts that the angular momenta of the disc-galaxy population contain cosmological information inherited from the large-scale structure of the universe.

Detecting this alignment effect of disc galaxies is the principal goal of this project. Starting out with a simple setting, it will be tested whether the angular momenta of the four large disc galaxies in the Local Group – Milky Way, Andromeda (M31), M33 and the Large Magellanic Cloud – are consistent with purely random orientation or not. Sophisticating the simple test in the Local Group by using spatial two-point correlation functions, alignment effects are searched in a large sample of disc galaxies taken from the SDSS. A successful detection of such alignment effects would confirm our understanding of disc-galaxy formation, which places the angular-momentum acquisition of individual disc galaxies into a larger cosmological context. Conversely, a non-detection might enforce a critical reassessment and modification of our theoretical picture. Moreover, understanding alignment effects of disc galaxies is also important for investigations of gravitational weak lensing. Disc alignment can cause systematic effects in shear measurements of weak-lensing surveys (e.g. Crittenden et al. 2001; Blazek et al. 2011), thereby compromising the estimation of cosmological parameters. Therefore, a detailed understanding of disc-galaxy alignment may enable us to correct for this bias in weak-lensing studies.

In this thesis, it is necessary to touch on several important issues involved in the search of the theoretically predicted alignment effects. First and foremost, it is necessary to investigate how disc galaxies can be identified and separated from other types of galaxies in practice. This requires some discussion about the classification of galaxy morphologies. However, before galaxy morphologies can be classified, they have to be parametrised first. The parametrisation of galaxy morphologies is a highly non-trivial task, since the galaxy population exhibits a very rich diversity of morphologies. Therefore, different approaches to the parametrisation of galaxy morphologies are presented. Categorising these different approaches, the method most promising for the task of classifying disc galaxies is identified and elaborated in detail. The basic idea of this method is to decompose a given galaxy image into different components which are the modes of an orthonormal set of basis functions. This method by design excels in high flexibility which is necessary to provide a faithful parametrisation of the morphology typically exhibited by disc galaxies, which includes spiral-arm patterns, bulges, galactic bars and rings, as well as star-forming regions. It is investigated in great detail if this method indeed can provide reliable parametrisations. Another important aspect concerning galaxy morphologies is the seemingly trivial question of how to estimate a galaxy's ellipticity. Such an ellipticity estimate is required in order to infer the orientation of the angular-momentum vector of the rotating disc. Con-

sequently, the investigation of galaxy morphologies naturally forms the first major part of this thesis.

After disc galaxies have been identified and their angular-momentum orientations estimated, the task is to search for alignment effects by estimating two-point correlation functions. The second major part of this thesis is therefore dedicated to this estimation process. Here, it is necessary to rigorously assess and propagate all potentially important sources of errors, namely classification uncertainties and errors in estimates of redshift and ellipticity, in order to obtain correlation estimates with *marginal* errors which enable us to test theoretical prediction against data. Previous results (e.g. Slosar et al. 2009; Lee 2011) did not account for these error sources. Consequently, the reported errors necessarily underestimate the real errors and therefore lead to overly optimistic significance levels of the detected correlations. This evidently corrupts any tests of the presence of alignment effects. Therefore, a methodological objective of this thesis is to understand how all these errors can be propagated through the parameter-estimation process.

After these excursions into fundamental work this thesis has established the required tools for estimating marginal autocorrelation functions and will then apply this to real data. Evidence of autocorrelations of spiral-arm handedness is then searched in a sample of 30,154 spiral galaxies, while evidence of autocorrelations of angular-momentum-orientation vectors is searched in a sample of 4,216 Scd galaxies. The resulting correlation estimates can be used to try to constrain free parameters in the theoretical model (Pen et al. 2000; Lee & Pen 2008; Schäfer & Merkel 2011). However, as these parameter values are yet unknown, this investigation cannot be considered as an empirical test of the theory, since at this point in time the tidal-torque theory is still free to adopt to almost any experimental result. Hence, this study is only a first step to constrain the model parameters and a second future test is required for an empirical test. In particular, the hypothesis test in the Local Group using the three large disc galaxies is an important sanity check but it does not provide an astrophysically independent test of the theory. The thesis concludes by investigating the potential of future sky surveys to enhance these correlation estimates, thereby improving observational constraints on theoretical parameters.

In detail, this thesis is organised as shown in Fig. 1.1: The first part focusses on the parametrisation of galaxy morphologies with focus on the task of classification. Chapter 2 recapitulates several common classification schemes and scrutinises their reliability. Chapter 3 then elaborates in more detail on the method that has previously been identified as the most promising approach. The second part of this thesis is then dedicated to the actual investigation of disc alignment. As an initial test, the angular momenta of the disc galaxies in the Local Group are studied in Chapter 4. Chapter 5 presents the SDSS data used for the main analysis. The correct methodology and results are then discussed in Chapter 6. Chapter 7 discusses potential of future sky surveys to improve estimates of disc alignment. Finally, Chapter 8 summarises and discusses the results and provides an outlook.

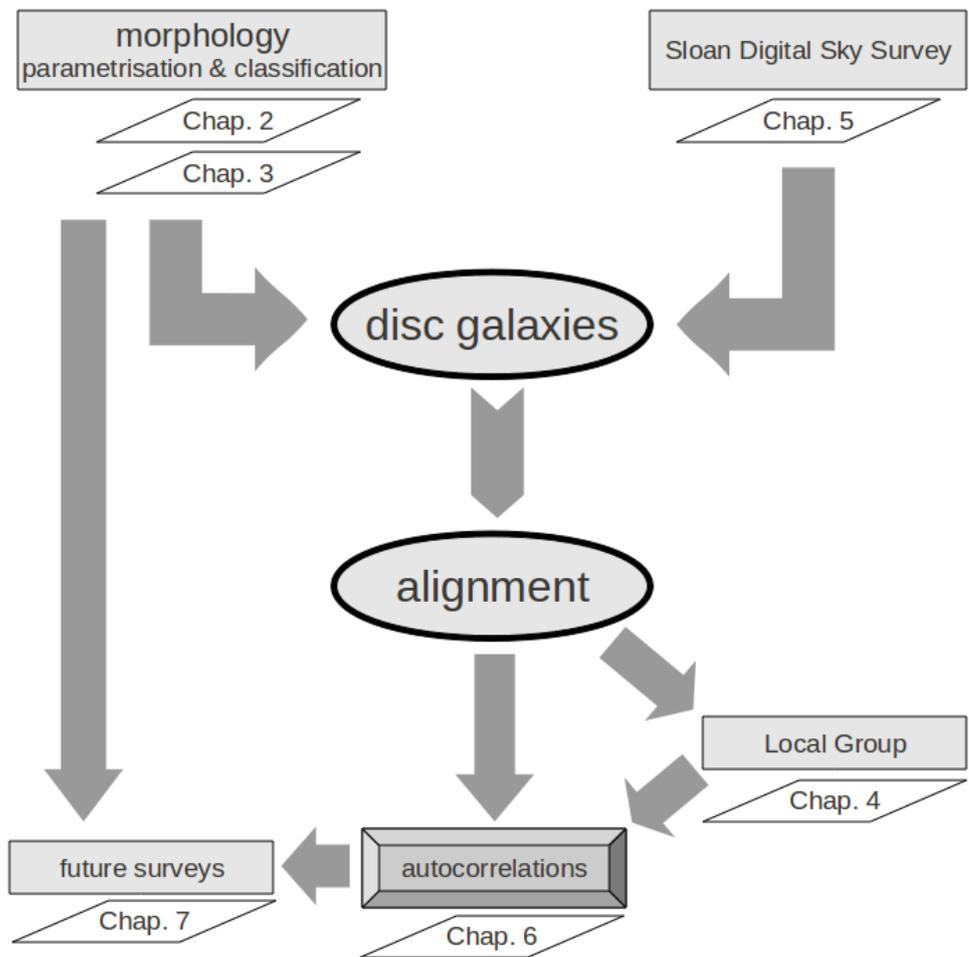


Figure 1.1: Roadmap of this thesis.

Part I

Galaxy Morphology

2

Parametrising galaxy morphologies

In this section, we discuss several methods for parametrising galaxy morphologies. This is necessary since modern sky surveys such as the Sloan Digital Sky Survey (SDSS) compile databases with millions of galaxies. Consequently, traditional classification of galaxy morphologies through visual inspection is becoming infeasible and at the same time, an objective or at least reproducible classification mechanism becomes desirable. However, an automated classification requires a parametrisation of the observed morphologies of galaxies. Given the rich variety of galaxy morphologies, parametrisation is a very challenging task and many different approaches have been attempted. Evidently, a parametrisation method has to be robust, flexible and reliable in order not to compromise morphological classification. We therefore rigorously test the most popular parametrisation methods in this section. Ideally, this may enable us to choose a suitable parametrisation method that automatically identifies disc galaxies in a given galaxy sample in order to investigate their potential alignment. If neither methods turns out to perform with satisfying reliability, we can at least expect to learn how to construct a better method than the existing ones.

2.1 Galaxy morphologies

There are two key diagnostics in modern astrophysical research concerning the formation and evolution of galaxies – morphology and star formation. Galaxy morphology by itself is not a fundamental physical parameter of a galaxy, though it correlates weakly with other more physical parameters such as star formation (e.g., Kennicutt 1998) or environment (e.g., van der Wel et al. 2010). However, morphology provides a direct observable, whereas, e.g., star-formation rates are derived quantities based on additional assumptions and extrapolations (e.g., Rosa-González et al. 2002). For instances, star-formation rates – which are used extensively as astrophysical diagnostics – rely on the

CHAPTER 2. PARAMETRISING GALAXY MORPHOLOGIES

following assumptions:

- An initial mass function (IMF) that is identical for all types of galaxies at all cosmological times, i.e., a universal IMF.
- Knowledge of the star-formation history of the galaxy under consideration.
- Knowledge of the mass-luminosity relation of stars in the observed part of the spectrum.

In particular, the assumption of a universal IMF appears to be a somewhat naïve oversimplification and recently has been challenged by observational data (van Dokkum & Conroy 2010, 2011).¹ Apparently, the estimation of star-formation rates is not the best manifestation of one of the basic rules of good scientific practice, namely the restriction to an absolute minimum of well-justified assumptions. The merit of this practice is obvious, since the fewer assumptions are involved, the less likely is a failure of the method. Therefore, studies of galaxy morphologies are an important complementary means for understanding the physics of galaxies. Furthermore, certain investigations are solely based on morphologies, e.g., weak-lensing measurements (e.g., Bernstein & Jarvis 2002) or studies of alignment of disc galaxies in this thesis.

¹As will be discussed by Andrae et al. (in prep.), common attempts to estimate the IMF are flawed to different extents, e.g., by inappropriate binning of data, a-priori unphysical models (e.g. Kroupa 2002) or omission of errors in stellar-mass estimates. Hence, the IMF is in fact unknown and its proclaimed universality may be inherited from poor methodology.

The morphology of a galaxy is specified by the characteristics of its two-dimensional light distribution, i.e., the shape of the galaxy projected onto the plane of the sky. Apart from the projection, an observed galaxy morphology is also influenced by the point-spread function of the observing instrument, pixellation on the CCD chip and the addition of Poissonian background noise which originates from the counting of photons. Some morphological observables are:

- steepness of the radial light profile,
- ellipticity, i.e., orientation angle and axis ratio,
- asymmetry, e.g., lopsidedness,
- substructures, e.g., spiral-arm patterns, bars, rings, star-forming regions, etc.,
- galaxy size.

The centroid position is an important morphological observable as well, since it is often required to derive other morphological estimators (cf. Table 2.1). For decades galaxy morphologies have been studied in the visual regime where all these observables are reasonably well defined. In the visual or optical regime – at least the red part – the light distribution is dominated by stellar emission and not by dust or gas. However, with increasing observational coverage of the electromagnetic spectrum, it became evident that morphology is a strongly varying function of wavelength. For instance, in the UV we observe mostly star-forming regions and dust extinction, such that galaxies can look patchy and highly irregular. On the other hand, in the far infra-red regime, there is almost no stellar but only dust emission. In the radio regime, morphologies may even be dominated by jets extending far beyond the actual galaxy (e.g. Lin

et al. 2010). Hence, radio morphologies do not correlate with the distribution of stars and are therefore very different from optical morphologies.

There has been substantial effort to define automated parametrisation schemes for galaxy morphologies (e.g., Sérsic 1968; Abraham et al. 1996, 2003; Bershady et al. 2000; Conselice 2003; Lotz et al. 2004, 2008, to name just a few). Unfortunately, these parametrisation schemes usually invoke rather restrictive assumptions, such that they lack the flexibility to describe the huge variety of different galaxy morphologies present in modern databases. Although we are well able to parametrise the morphologies of individual galaxies of certain types (e.g. Simmat et al. 2010), finding a parametrisation scheme that is flexible enough to describe arbitrary galaxy morphologies is a completely different task.

2.2 Parametrisation schemes

In order to assess the advantages and deficits of different parametrisation schemes we now briefly summarise the most common approaches. We divide them into model-independent and model-based methods. The most important difference is that the model-based approaches try to *model* the two-dimensional light distribution of an image and are thus mostly descriptive. Model-independent – or “non-parametric” – approaches more directly try to extract physical information, hence mixing description and inference. The current paradigm favours model-independent methods, hailing their seeming simplicity in comparison with model-based approaches which require fitting a model to data. We conclude this section by summarising the assumptions involved in the different parametrisation schemes.

2.2.1 Model-independent schemes I: CAS

A widely used set of morphological parameters is provided by the CAS system, which is based on the so-called concentration, asymmetry and clumpiness indices (Abraham et al. 1994, 1996; Bershady et al. 2000). The concentration index is defined as

$$C = 5 \log_{10} \left(\frac{r_{80}}{r_{20}} \right), \quad (2.1)$$

where r_{80} and r_{20} are the radii of circular (or elliptical) apertures containing 80% and 20% of the total image flux.² The concentration index is always strictly positive since $r_{80} > r_{20}$. Furthermore, the asymmetry index is defined as

$$A = \frac{\sum_{\text{pixels}} |I(\vec{x}) - I^{180^\circ}(\vec{x})|}{\sum_{\text{pixels}} I(\vec{x})}, \quad (2.2)$$

where I^{180° denotes the image I rotated by 180° . Technically, the asymmetry index is bound in the interval $[0, 2]$. In practice, A is bound to $[0, 1]$ because the observed photon fluxes cannot be negative. Finally, the clumpiness index is defined as

$$S = 10 \frac{\sum_{\text{pixels}} |I(\vec{x}) - I^\sigma(\vec{x})|}{\sum_{\text{pixels}} I(\vec{x})}, \quad (2.3)$$

²There are several variations of the concentration index: Sometimes it is based on the ratio of r_{90} and r_{50} . Some authors (e.g. Bershady et al. 2000) consider the whole image for estimating C , others (e.g. Scarlata et al. 2007) estimate C only within a region given by one Petrosian radius.

where I^σ has been convolved by a Gaussian of width σ . The specific choice of σ is somewhat arbitrary within a certain range, being sensitive to clumps of varying spatial extent. As far as the author knows, there is no systematic investigation of the impact of the choice of σ on the parametrisation results, i.e., it is in fact unknown what kind of information the clumpiness index actually is responding to.

2.2.2 Model-independent schemes II: M_{20} and Gini

Two further model-independent morphological parameters are M_{20} and the Gini coefficient. We define the second-order moment of pixel n with value I_n at position \vec{x}_n as (Lotz et al. 2004)

$$M_n = I_n (\vec{x}_n - \vec{x}_c)^2, \quad (2.4)$$

where \vec{x}_c denotes the reference position. Summation of the M_n over all pixels yields the total second moment M_{tot} with respect to \vec{x}_c . There is a theoretical preference to choose the reference position \vec{x}_c to be the centre of light, because this choice minimises M_{tot} . Then, M_{20} is defined as

$$M_{20} = \log_{10} \left(\frac{\sum_i M_i}{M_{\text{tot}}} \right), \quad (2.5)$$

where the summation $\sum_i M_i$ is over the pixels in descending order $I_1 \geq I_2 \geq \dots \geq I_N$ and stops as soon as $\sum_i I_i \geq 0.2 \sum_{n=1}^N I_n$, i.e., as soon as 20% of the total flux is reached. This design of M_{20} is meant to estimate the spatial distribution of the most luminous parts of a galaxy image.

The Gini coefficient was defined by Lotz et al. (2004, 2008) based on Glasser (1962) as

$$G = \frac{\sum_{n=1}^N (2n - N - 1) |I_n|}{(N - 1) \sum_{n=1}^N |I_n|}, \quad (2.6)$$

where N is the number of image pixels and $|I_1| \leq |I_2| \leq \dots \leq |I_N|$ are the absolute values of the pixel fluxes sorted in ascending order. In contrast to M_{20} , Gini does not require an estimate of the centroid position. The Gini coefficient is supposed to estimate the distribution of the pixel values over the image. However, as was shown by Lisker (2008), it strongly depends on the signal-to-noise distribution within a galaxy's image and is thus a highly unstable morphological estimator.³

2.2.3 Model-based schemes I: Sérsic profile

We now discuss the first model-based parametrisation scheme. The radial light profiles of many galaxies are found to be reasonably well described by the Sérsic profile (see Sérsic 1968; Graham & Driver 2005, for a compilation of relevant formulae),

$$I(R) = I_\beta \exp \left\{ -b_n \left[\left(\frac{R}{\beta} \right)^{1/n_s} - 1 \right] \right\}, \quad (2.7)$$

³In fact, Lisker (2008) concludes somewhat ironically that “the Gini coefficient could serve as a sophisticated measure of the S/N distribution within a galaxy's image, if it did not depend so much on the galaxy's morphology.”

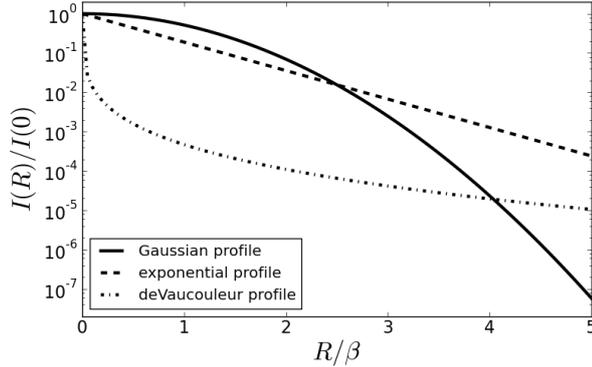


Figure 2.1: Three examples of Sérsic profiles.

We show a Gaussian profile ($n_S = 0.5$, solid line), an exponential profile ($n_S = 1$, dashed line), and a de Vaucouleur profile ($n_S = 4$, dashed-dotted line).

where n_S is the Sérsic index and β is the scale radius. The scale radius is expressed in units of pixels, i.e., β^{-1} is the pixel size relative to the object size. The constant b_n is usually chosen such that the radius β encloses half of the total light. I_β is the intensity at the half-light radius β . For fixed n_S , b_n is then given by (e.g. Graham & Driver 2005)

$$\Gamma(2n_S) = 2\gamma(2n_S, b_n) , \quad (2.8)$$

where Γ and γ denote the complete and incomplete gamma functions. For $n_S > 0.5$ one can approximate $b_n \approx 2n_S - \frac{1}{3}$ (Ciotti & Bertin 1999). The Sérsic profile corresponds to a Gaussian profile if $n_S = 0.5$, to an exponential profile if $n_S = 1$, and to a de Vaucouleur profile if $n_S = 4$. Figure 2.1 displays these example profiles. The exponential profile is of particular interest because it resembles the radial light profile of disc galaxies (e.g. MacArthur et al. 2003; Elmegreen et al. 2005). Similarly, the steeper de Vaucouleur profile is found to resemble the light profile of elliptical galaxies (de Vaucouleurs 1948). Hence, the Sérsic profile unifies these empirical laws. As is discussed in Sect. 3.2.3, the Sérsic profile is the first-order Taylor expansion of *any* (differentiable) light profile. Consequently, it is not surprising that it provides a good approximation to radial light profiles of real galaxies. Moreover, this provides a mathematical justification for the Sérsic profile, whereas the other methods discussed so far are mere ad-hoc constructions.

Figure 2.1 reveals that Sérsic profiles decline very slowly for large Sérsic indices, wherefore we need to introduce a truncation. Throughout this chapter we use the truncated Sérsic profile of the form

$$\tilde{I}(R) = \begin{cases} I(R) - I(5\beta) & \Leftrightarrow R \leq 5\beta \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

such that all profiles are 0 for $R > 5\beta$ but still continuous. A discontinuous cut-off would cause severe problems with Fourier transforms which are required for PSF convolutions in Sect. 2.4.2.

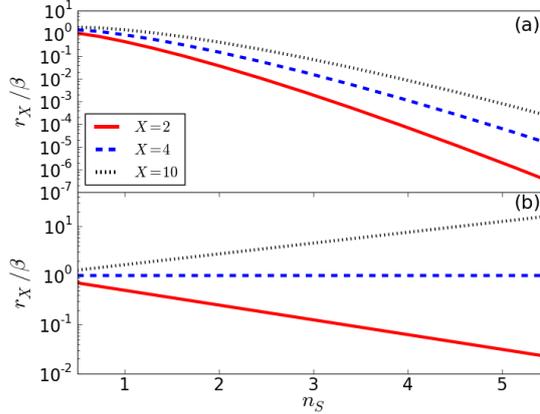


Figure 2.2: Definition of scale radii for Sérsic profiles.

Radii r_X where Sérsic profile takes values $I(r_X)/I(0) = 1/X$ for $X = 2, 4, 10$ and b_n given by Eq. (2.8) (panel a) and $b_n = \log 4$ (panel b).

2.2.3.1 Redefining b_n and β

It is important to note that b_n and β in Eq. (2.7) are completely degenerate. We are free to make any choice of b_n that is different from Eq. (2.8), thereby redefining the model and changing the meaning of β . There are two reasons why Eq. (2.8) potentially is not a good choice for b_n :

1. When fitting imaging data, half-light radius and Sérsic index are not independent of each other but strongly correlated (e.g. Caon et al. 1993; Trujillo et al. 2001). Due to this correlation, half-light radii β are *not* comparable for different values of n_S , i.e., the size of one galaxy relative to a second galaxy can be inferred from their scale radii if *and only if* both Sérsic models use identical n_S .⁴ However, in practice this is rarely a problem, since studies usually compare only sizes of galaxies of similar Hubble types, e.g., in studies of the size-evolution of disc galaxies. Figure 2.2 shows that the radii where the profile reduces to $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{10}$ of its value at $r = 0$ vary over several orders of magnitude for different n_S . For instances, panel (a) shows that a de Vaucouleur profile ($n_S = 4$) that drops to one tenth of its central value within 10 pixels distance from the centre has a half-light radius of only $\approx 10^{-3}$ pixels. Such small numbers seem to be counter-intuitive at first glance and may in fact cause investigators to falsely reject a correct fit result. This peculiar behaviour has to be kept in mind when using half-light radii, i.e., Eq. (2.8). The scale radius β could be defined more intuitively such that

$$\frac{I(\beta)}{I(0)} = 1/X \quad (2.10)$$

for some $X > 0$ *independent* of n_S . This can be achieved by setting $b_n = \log X$ for all n_S . Equation (2.10) enforces that the Sérsic profile drops to a certain value, which corresponds to the actual meaning of a

⁴Strictly, even this is not correct. A comparison of half-light radii inferred from Sérsic fits requires to marginalise out the Sérsic index.

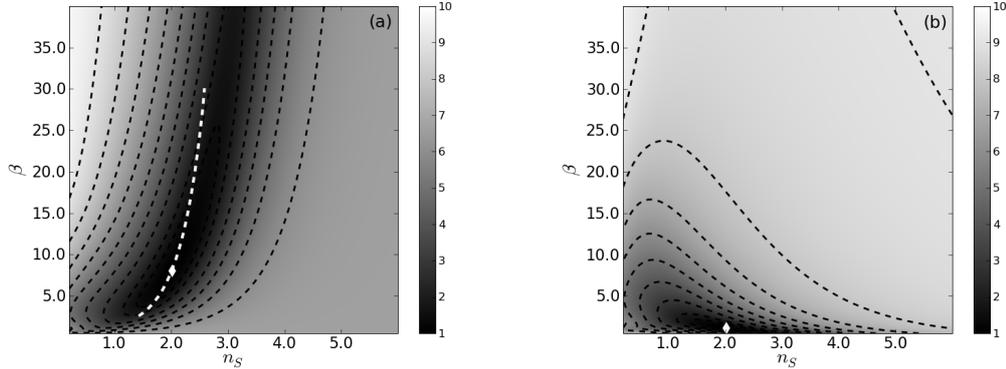


Figure 2.3: Artificial correlations of Sérsic index and scale radius.

The χ^2/dof -manifolds demonstrate how Eq. (2.8) induces the artificial correlation of n_S and β . Panel a: χ^2/dof manifold for b_n defined by Eq. (2.8). The white diamond indicates the optimum. The dashed white line is given by $b_n/\beta^{1/n_S} = \text{const}$ and approximately follows the valley, thereby illustrating that the correlation of n_S and β is artificial. Panel b: Same as in (a) but for $b_n = \log 4$ for all n_S . The valley is approximately parallel to the n_S -axis, i.e., the correlation is gone. Both panels use the same artificial light profile with low noise level to evaluate the χ^2/dof manifold. It is much easier to find the optimum in panel (b) than in panel (a). The optimal values of n_S are identical in (a) and (b), whereas the optimal values of β are different due to the different choice of b_n . χ^2/dof is not a simple quadratic form, because the Sérsic profile is a nonlinear model.

scale radius. Panel b of Fig. 2.2 shows that in this case the radii for different n_S change by less than two orders of magnitude and hence can be compared much better.

2. The strong correlation of n_S and β is problematic for many fit algorithms. This correlation of n_S and β is almost completely induced by Eq. (2.8), i.e., it is artificial. We can remove this correlation by setting $b_n = \log X$ for all n_S , thereby simplifying the fit problem. We demonstrate this in Fig. 2.3 showing χ^2 manifolds for fitting an artificial light profile once using Eq. (2.8) (panel a) and once using $b_n = \log X$ for all n_S (panel b). The noise level in this simulation is low (the signal-to-noise ratio of the central peak is 100). Higher noise levels do not change the curvatures of the χ^2 “valleys” in Fig. 2.3 but only broaden them and reduce their depth.

These issues are not fundamental and there is no theoretical preference for choosing between these approaches. In fact, it is possible to convert to and fro the definitions of Eqs. (2.8) and (2.10) via $b/\beta^{1/n_S} = \text{const}$. However, using Eq. (2.10), we encountered serious problems with fitting galaxy images, wherefore we do not further pursue this approach.

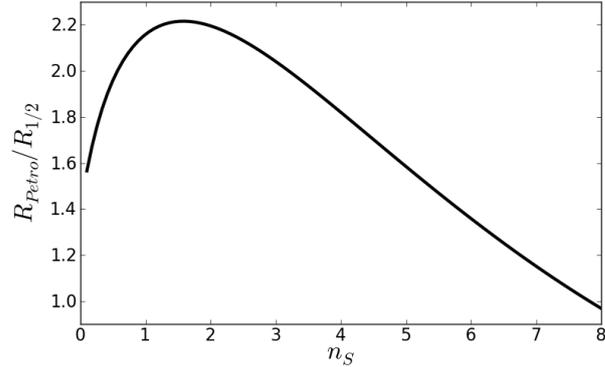


Figure 2.4: Initial guess for fitting Sérsic profiles.

Relation of Sérsic index and the ratio of Petrosian radius R_{Petro} over half-light radius $R_{1/2}$. If Eq. (2.8) is used, then $R_{1/2} = \beta$.

2.2.3.2 Initial guess for fitting Sérsic profiles

As the Sérsic profile defined in Eq. (2.7) is a nonlinear model, it cannot be fitted to data analytically but only through iterative optimisation. This requires an initial guess of the model parameters and a good initial guess is of obvious interest since it reduces the number of required iterations till convergence and hence reduces the computational effort. For general Sérsic profiles, the Petrosian radius (Petrosian 1976)⁵ can be computed through numerical integration (Graham & Driver 2005). The Sérsic index and the ratio of Petrosian radius over half-light radius are then related as is shown in Fig. 2.4. As Petrosian radius and half-light radius are both observables that can be estimated *model-independently* from a given galaxy image (Bertin & Arnouts 1996), this relation can be used in order to obtain an initial guess for the model parameters when fitting a Sérsic profile. However, there are two caveats: First, pixel noise in the given galaxy image may corrupt estimates of both radii such that their ratio exceeds the value ≈ 2.2 and strictly no solution for the Sérsic index exists. In this case, a fair initial guess would be $n_S \approx 1.5$, i.e., where the maximum of the relation is situated. Second, for several ratios there are two possible Sérsic indices. In this case, both solutions have to be tried and the one which produces a better fit is used as an initial guess. Hence, we can deduce an initial guess for the Sérsic index. The half-light radius directly provides an initial guess for the Sérsic scale radius β , if we employ Eq. (2.8).

⁵For a radial profile $f(r)$, the Petrosian radius R_{Petro} is defined via $f(R_{\text{Petro}}) = 0.2 \cdot \frac{2\pi}{\pi R_{\text{Petro}}^2} \int_0^{R_{\text{Petro}}} dr r f(r)$.

2.2.4 Model-based schemes II: Basis functions

An alternative model-based parametrisation approach is the expansion into basis functions. The most important advantage of this concept is that the parametrisation is more flexible, whereas all previous schemes are highly specialised for certain morphologies. For instance, the Sérsic profile is only a radial profile that ignores any azimuthal structures usually exhibited by disc galaxies. Conversely, a good set of basis functions should be able to fit almost anything, provided the signal-to-noise ratio of the given data is sufficiently high. Hence, this approach should in principle be favoured when the task at

hand is to parametrise arbitrary morphologies. Basis functions also by design can describe azimuthal structures such as spiral-arm patterns of disc galaxies. Therefore, these methods are of particular interest in the context of this thesis. In fact, Chapter 3 is dedicated to discuss basis-function expansions such that we do not elaborate in detail here. In the following, we only briefly comment on several example sets of basis functions, leaving all the details to the next chapter.

2.2.4.1 Shapelets

Shapelets were introduced by Réfrégier (2003). They are a scaled version of Gauss-Hermite polynomials, i.e.

$$B_n(x; \beta) = (2^n n! \sqrt{\pi} \beta)^{-1/2} H_n \left(\frac{x}{\beta} \right) \exp \left[-\frac{x^2}{2\beta^2} \right], \quad (2.11)$$

where H_n denotes the Hermite polynomial of order n and β is the shapelet scale size. A centroid can be introduced via $x \rightarrow x - x_0$. All basis functions take identical parameters $\vec{\theta}_n = \vec{\theta} = (x_0, \beta)$ in order to allow for orthogonality. From this definition, we can build two-dimensional basis functions, namely Cartesian shapelets and polar shapelets.

The Gaussian weight function of shapelets leads to very nice analytical properties. For instance, shapelets are nearly invariant under Fourier transformation, which makes any convolution or deconvolution a closed and analytic operation in shapelet space, as described in Melchior et al. (2009). However, in the presence of pixel noise we cannot fit an infinite number of basis functions but have to truncate the basis expansion at a certain maximum order. This limitation of basis functions due to pixel noise has a severe consequence: Shapelets employ a Gaussian weight function (cf. Eq. (2.11)), but real galaxies have typically much steeper profiles, e.g., see exponential and de Vaucouleur profiles in Fig. 2.1. This gives rise to characteristic modelling failures that typically manifest themselves in ring-like artefacts in the shapelet reconstructions of galaxies with exponential or steeper light profiles. This severely limits the diagnostic power of shapelets (cf. Melchior et al. 2010) and we therefore exclude them from our subsequent simulations.

Despite these fundamental problems, shapelets demonstrate a very important aspect of basis-function expansions: For highly resolved galaxies of high signal-to-noise ratios Sérsic profiles are incapable of providing excellent models as they are not flexible enough to account for azimuthal substructures such as spiral-arm patterns, i.e., their residuals do not always reach noise level. In the case of shapelets – as an example of basis functions – this is fundamentally different. They are highly flexible and reach noise level even for galaxies that are very large, highly resolved and bright (e.g. Andrae et al. 2010b).

2.2.4.2 Sérsiclets

Given the problematic impact of the Gaussian profile on shapelets, a set of basis functions based on the Sérsic profile is an obvious means to overcome the limitations of shapelets. The resulting basis functions are called sérsiclets. Ngan et al. (2009) were the first to realise the potential of this approach, which

CHAPTER 2. PARAMETRISING GALAXY MORPHOLOGIES

is capable of accounting for all morphological observables listed in Sect. 2.1. However, for technical reasons their implementation of sérsiclets was flawed (cf. Sect. 3.2.2). We postpone a detailed investigation of sérsiclets to Chapter 3.

2.2.4.3 Outlook: Template libraries

In practice, no basis set – apart from the pixel grid itself – is actually complete due to the limitations induced by pixel noise. Now, we want to briefly touch – without going into details – on a set of basis functions that is finite and thus incomplete from the beginning. The motivation is very simple: For both shapelets and sérsiclets the basis functions lack a physical interpretation. Why not use basis functions that directly correspond to spiral arms, galactic bars or rings? We can use a set of such *templates* – a template library – to form linear models and decompose the image, resulting in a set of coefficients that form a vector space. The individual templates do not even need to be orthogonal, but just as linearly independent as possible in order to avoid heavy degeneracies during the fitting procedure. Unfortunately, the direct physical motivation is also the major drawback of this approach, since we are strongly prejudiced and lack flexibility in this case. For instance, template libraries are likely to have severe problems in decomposing irregular galaxies, i.e., they are inappropriate for parametrising arbitrary morphologies. Moreover, the set of morphological features is very large, hence such a library has to contain numerous templates. Currently, such a template library is not within reach.

2.2.5 Assumptions

Using a certain method, it is crucial to be aware of all invoked assumptions, since if a method fails, it usually fails because one or more of its assumptions break down. Furthermore, the rules of good scientific practice clearly argue that the number of assumptions should be minimised and that those assumptions that are indispensable should be well justified. In the case of model-based approaches, the assumptions are usually rather obvious from the model design and therefore can be easily challenged. Conversely, the assumptions of model-independent approaches are implicit and often well hidden. This may lead to the misapprehension that model-independent schemes were superior since they required fewer or even no assumptions.

Table 2.1 shows our categorisation of parametrisation schemes. Based on this table and the definitions given in the previous sections, we now work out the assumptions of all schemes from a *theoretical* point of view. In practice, it is virtually impossible to satisfy all assumptions. Whether the violation of some assumption leads to a breakdown of a certain method depends on the specific question under consideration, the desired precision, the details of the method’s implementation, and the quality of the data. In detail, the assumptions are:

- Concentration index: There are no azimuthal structures such as spiral-arm patterns or galactic bars⁶ such that a curve of growth and radii containing certain percentages of the total galaxy flux can be defined meaningfully. The pixel noise is negligible and the object is not grossly

⁶This assumption is problematic since we want to parametrise disc galaxies.

CHAPTER 2. PARAMETRISING GALAXY
MORPHOLOGIES

characteristic	C	A	S	M_{20}	G	Sérsic profile	shapelets	sérsiclets
model-based	n	n	n	n	n	y	y	y
centroid estimate necessary	y	y	n	y	n	y	y	y
account for steepness of light profile	n	n	n	n	n	y	n	y
account for ellipticity	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$	n	n	y	$y^{(4)}$	y
account for substructures	n	y	y	n	n	n	y	y

Table 2.1: Characteristics of parametrisation schemes.

(1) We can employ elliptical isophotes to estimate C .

(2) A is invariant under all operations that are symmetric under rotations by 180° . Ellipticity is such an operation.

(3) It is possible to use an elliptical Gaussian for convolution.

(4) There is also a spherical shapelet formalism.

asymmetric such that a centroid is well defined (cf. Sect. 2.3.4). The scheme can be enhanced using elliptical apertures.

- Asymmetry index: A centre of rotation is well defined. The pixel noise is negligible. Both issues have been addressed by Conselice et al. (2000). The asymmetry of interest is visible under rotations of 180° .
- Clumpiness index: The functional type of the convolution kernel matches the galaxy profile. The width of the kernel is chosen such that the information of interest is extracted. The ellipticity of the kernel matches the ellipticity of the object.
- M_{20} : The pixel noise has negligible impact on the estimates of centroid and second moments.⁷ The centre of light and the object’s centre coincide, i.e., there is no substantial asymmetry. The structures dominating M_{20} are of circular shape with the centroid at their centres (see term $(\vec{x}_n - \vec{x}_0)^2$ in Eq. (2.4)).
- Gini coefficient: The pixel noise is negligible (see Lisker 2008).
- Sérsic profile: The Sérsic profile is a good match of the object’s light profile. In particular, this means that the object’s light profile is symmetric, monotonically decreasing and the steepness is correctly described by the model, and there are no azimuthal structures such as spiral-arm patterns, galactic bars or rings. As the Sérsic profile is the first-order Taylor expansion of any differentiable radial light profile (Sect. 3.2.3), this assumption is well justified in the limit of low resolution and low signal-to-noise ratios, where azimuthal structures are usually wiped out.
- Shapelets: The Gaussian weight function fits galaxy profiles.

⁷In fact, pixel noise can have a severe impact on estimates of centroid and second moments, in particular for disc galaxies with shallow light profiles.

We now clearly see that model-independent schemes implicitly make assumptions, too. Evidently, a general statement that model-independent or model-

based approaches rely on fewer assumptions is not possible. We also want to emphasise that shapelets – as an example of basis functions – can model asymmetries.

2.3 Intertwinement of morphological observables

The basic idea of model-independent schemes is to estimate the different morphological observables listed in Sect. 2.1 independently of each other, thereby simplifying the problem. However, in this section, we argue that these morphological observables are intertwined, which means that it is impossible to measure them independently of each other. Even if we try to measure only a single observable using a method unaware of the other observables, the mere presence of these other observable features influences the results. The notion of intertwinement should not be confused with redundancy. For instance, Sérsic index and concentration index are perfectly redundant (Sect. 2.3.1) but asymmetry index and concentration index are not (Sect. 2.3.4). Of course, for some observables the intertwinement may be stronger than for others. This intertwinement is not of physical origin but stems from the fact that usually all morphological observables are present simultaneously, such that the assumptions listed in Sect. 2.2.5 are *never* truly satisfied.

We carry out noise-free simulations of the different parametrisation schemes, revealing several systematic misestimations – in particular of the concentration index which is commonly used for morphological classification (e.g. Gauci et al. 2010; Huertas-Company et al. 2011). Pixel noise in real data may hide these biases to some extent, but they will still be present. All simulations invoke Sérsic profiles that are the first-order Taylor expansions of real light profiles (see Sect. 3.2.3) but we want to explicitly emphasise that it is *not* necessary for real galaxies to actually follow Sérsic profiles.⁸

⁸To be more precise, it is perfectly valid to use idealised simulations to discover and isolate these biases, but in order to correct for them more realistic simulations are necessary.

2.3.1 Example I: Sérsic profile vs. concentration index

We begin with comparing Sérsic profiles and the concentration index, establishing a relation between both schemes that allows us to assess systematic effects on the concentration. The Sérsic index estimates how steeply the radial light profile falls off, while the concentration index estimates the concentration of the radial light profile. Consequently, Sérsic index and concentration index are essentially two estimators for the same morphological feature, namely the steepness of the light profile. This is also evident from the fact that both schemes have almost identical assumptions (cf. Sect. 2.2.5), e.g., both ignore azimuthal structures. In fact, we can compute the concentration of a two-dimensional Sérsic profile using numerical integration, i.e., Sérsic index and concentration index are perfectly redundant (see also Trujillo et al. 2001). Numerically integrating the Sérsic profile to infinite radius yields the power law

$$C \approx 2.770 \cdot n_S^{0.466}, \quad (2.12)$$

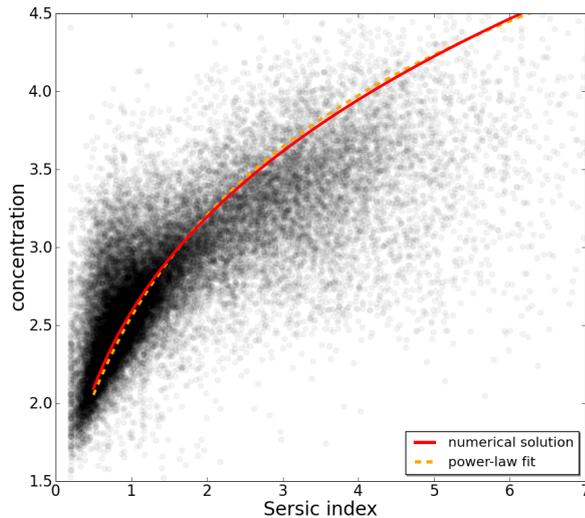


Figure 2.5: Comparing concentration and Sérsic indices.

Values of 31,288 COSMOS galaxies from the Zurich Structure & Morphology catalogue (Sargent et al. 2007) (points) with the numerical solution (solid curve) and power-law fit of Eq. (2.13) (dashed curve). Shown are unflagged galaxies with $I < 22.5$, and valid axis ratios ($0 < q \leq 1$). Concentration indices are predicted from analytic Sérsic profiles using numerical integration out to one Petrosian radius.

which provides a good approximation for the exact numerical solution for $0.5 \leq n_S \leq 7$. The resulting concentration values of $n_S = 0.5, 1$ and 4 are identical to those given by Bershadsky et al. (2000). Integrating the Sérsic profile to one Petrosian radius instead of infinity, the approximate solution is

$$C \approx 2.586 \cdot n_S^{0.305} . \quad (2.13)$$

Obviously, any declining radial profile can be mapped onto the concentration index this way, irrespective of whether or not it is a good description of a galaxy. Therefore, Fig. 2.5 also compares this theoretical expectation with the estimated concentration indices and Sérsic indices of 31,288 COSMOS galaxies from the Zurich Structure & Morphology catalogue (Scarlata et al. 2007; Sargent et al. 2007).⁹ Evidently, the *independent* estimates of concentration indices conducted by Scarlata et al. (2007) and of Sérsic indices conducted by Sargent et al. (2007) are in excellent agreement with the theoretical prediction of Eq. (2.13). This clearly demonstrates that concentration and Sérsic indices are not only theoretically equivalent parametrisations but, in the case of COSMOS galaxies, also providing largely unbiased estimates of similar quality. Nevertheless, this single example does *not* supersede a detailed study of potential biases that may occur in practice. In particular, the COSMOS data shown in Fig. 2.5 exhibits a large scatter that may hide biases.

⁹<http://irsa.ipac.caltech.edu/data/COSMOS/datasets.html>

2.3.2 Simulating ellipticity and lopsidedness

Presently, we want to simulate two-dimensional Sérsic profiles exhibiting ellipticity and a certain type of asymmetry, namely lopsidedness, which essentially describes an offset between peak of light and centre of light. In order to introduce ellipticity and lopsidedness analytically, we apply the shear and flexion transformations from gravitational weak lensing (Goldberg & Bacon 2005) which are briefly resumed now.

Given the complex ellipticity, $\epsilon = \epsilon_1 + i\epsilon_2$, the ratio of semi-minor axis b over semi-major axis a is

$$q = \frac{b}{a} = \frac{1 - |\epsilon|}{1 + |\epsilon|} \leq 1 \quad (2.14)$$

and the orientation angle θ is

$$\theta = \frac{1}{2} \arctan \left(\frac{\epsilon_2}{\epsilon_1} \right). \quad (2.15)$$

Then, the “sheared” coordinates, (x'_1, x'_2) , are given by

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 - \epsilon_1 & -\epsilon_2 \\ -\epsilon_2 & 1 + \epsilon_1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \quad (2.16)$$

For given pixel coordinates (x_1, x_2) , we then evaluate the model at (x'_1, x'_2) . Similarly, the flexion transformation is parametrised by the first flexion

$$F = F_1 + iF_2 \quad (2.17)$$

and the second flexion

$$G = G_1 + iG_2. \quad (2.18)$$

Given these parameters, we compute the derivatives of the gravitational shear $\gamma = (\gamma_1, \gamma_2)$,

$$\gamma_{1,1} = \frac{1}{2}(F_1 + G_1) \quad (2.19)$$

$$\gamma_{2,2} = \frac{1}{2}(F_1 - G_1) \quad (2.20)$$

$$\gamma_{1,2} = \frac{1}{2}(G_2 - F_2) \quad (2.21)$$

$$\gamma_{2,1} = \frac{1}{2}(G_2 + F_2). \quad (2.22)$$

Based on these derivatives, we compute the two matrices

$$D_{ij1} = \begin{pmatrix} -2\gamma_{1,1} - \gamma_{2,2} & -\gamma_{2,1} \\ -\gamma_{2,1} & -\gamma_{2,2} \end{pmatrix} \quad (2.23)$$

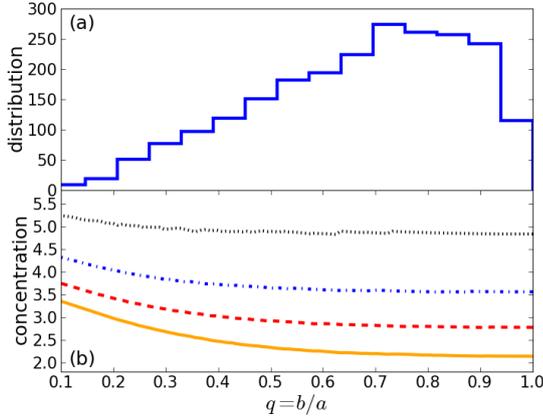


Figure 2.6: Impact of ellipticity on concentration estimates.

Panel a shows the distribution of axis ratios $q = b/a$ for 2,272 SDSS galaxies from the data sample of Fukugita et al. (2007). Panel b shows concentration estimates using circular isophotes for elliptical Sérsic profiles with $n_S = 0.5$ (solid line), $n_S = 1$ (dashed line), $n_S = 2$ (dotted-dashed line), and $n_S = 4$ (dotted line).

and

$$D_{ij2} = \begin{pmatrix} -\gamma_{2,1} & -\gamma_{2,2} \\ -\gamma_{2,2} & 2\gamma_{1,2} - \gamma_{2,1} \end{pmatrix}. \quad (2.24)$$

Using these matrices, we do not evaluate a flexed Sérsic profile at position $\vec{x} = (x_1, x_2)$, but rather at position

$$x'_i = x_i + \frac{1}{2} D_{ijk} x_j x_k. \quad (2.25)$$

The scaling of the coordinates by the scale radius β of the Sérsic profile is applied *prior* to this flexion transformation. In fact, the flexion transformation of Eq. (2.25) produces a second solution of $\vec{x}' = 0$, which corresponds to a secondary image in weak lensing. We only consider cutouts with just one image, where the other image resulting from the second solution to $\vec{x}' = 0$ is far away.

2.3.3 Example II: Steepness of light profile vs. ellipticity

Our second example is the intertwinement of the steepness of the radial light profile and the ellipticity. These two are certainly the most important morphological observables listed in Sect. 2.1, having the largest impact on parametrisation results.

It is obvious that estimates of the steepness of the radial light profile must take into account ellipticity. Therefore, it is necessary to use elliptical isophotes in the case of the concentration index or to fit a two-dimensional Sérsic profile that is enhanced by an ellipticity parameter. Unfortunately, in the case of the SDSS, the aperture radii containing 50% and 90% of the total image flux given

in the SDSS database are chosen as *circular* apertures (Strauss et al. 2002). This implies that estimates of the concentration index drawn from these values may be biased. In fact, this bias was already discussed by Bershady et al. (2000). They investigated how the concentration index changes with axis ratio for samples of real galaxies of similar morphological types. Bershady et al. (2000) claim that using circular apertures causes an overestimation of concentration indices of at most 3% and is therefore negligible. We investigate this effect in Fig. 2.6 for a realistic range of axis ratios, as is evident from panel a. Panel b shows how the concentration index is influenced by the axis ratio for Sérsic profiles with fixed Sérsic indices, corresponding to galaxy samples of similar morphologies as in Bershady et al. (2000).¹⁰ Evidently, for $q \gtrsim 0.5$ – which is the majority of galaxies in the given set – the bias is negligible. There are galaxies with $q < 0.5$, which are typically disc-like galaxies with shallow light profiles. For those objects, concentration estimates based on circular isophotes are substantially overestimated ($\approx 30\%$ for $n_S = 1$). This bias is *not* negligible. Moreover, this systematic overestimation of concentration indices would cause highly inclined disc galaxies to be confused with roundish galaxies that actually exhibit steeper light profiles, as is evident from Fig. 2.6. Consequently, this bias is problematic when using concentration indices in order to classify disc galaxies. Bershady et al. (2000) based their investigation on estimated concentration indices of *real* galaxies. Hence, the most likely origin of this discrepancy in our results is that the intrinsic scatter in the real data used by Bershady et al. (2000) hid this bias. Considering ellipticity and *circular* concentration index together – instead of using an elliptical concentration index – is *not* likely to solve this problem. The reason is that incorporating an ellipticity estimate may add information about the cause of the bias of the circular concentration index, but it does not provide information about the effect of this bias. Finally, we want to emphasise that Fig. 2.6 must not be used to calibrate the biased concentration estimates resulting from circular apertures. The reason is that such a correction would require that real galaxies are indeed Sérsic profiles. Moreover, also the study of Bershady et al. (2000) cannot be used for such a purpose, because the bias clearly depends on the intrinsic concentration index. This means that such a correction would require prior knowledge about the object’s true concentration.

Vice versa, Melchior et al. (2010) showed in the context of weak gravitational lensing that ellipticity measurements using shapelets are strongly biased in the case of steep profiles. In other words, shapelets fail to provide reliable ellipticity estimates because they do not properly account for the steepness of the radial light profile. This impressively demonstrates that these two observables are closely intertwined.

2.3.4 Example III: Impact of lopsidedness on centroid estimation

As a third example for the intertwinement of morphological observables, we consider the impact of lopsidedness on centroid estimates and the resulting parameter estimation using two-dimensional flexed Sérsic profiles. The strength of the flexion transformation is parametrised by F_1 , F_2 , G_1 , and G_2 . There is no pixel noise in this simulation. Figure 2.7 shows Gaussian profiles result-

¹⁰Obviously, Sérsic profiles are rather idealised and by far not as realistic as the sample used by Bershady et al. (2000). However, this does *not* hamper the validity of this test, but rather serves the purpose of isolating this bias. Apart from that, there is no difference in both studies.

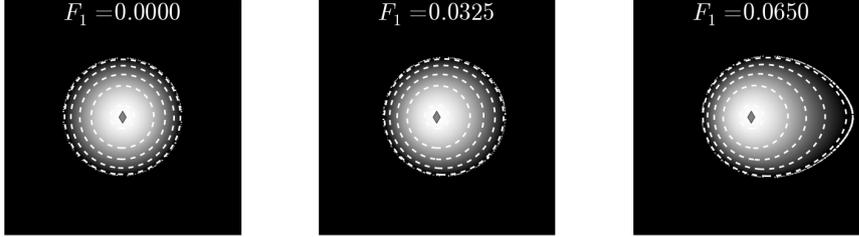


Figure 2.7: Gaussian profiles of different lopsidedness.

The applied flexions are $F_1 = 0.0$ (left panel), $F_1 = 0.0325$ (centre panel), and $F_1 = 0.065$ (right panel). The resulting profiles exhibit realistic lopsidedness. All profiles are evaluated on a 1000×1000 pixel grid using a scale radius of $\beta = 50$. White diamonds indicate the maximum position.

ing from this transformation. The resulting distortions are not unrealistically strong.

In Fig. 2.8 we investigate the impact of this type of asymmetry on the centroid, the asymmetry index and the concentration index. The first and foremost consequence is that in the presence of asymmetry the maximum position and the centre of light as given by

$$\hat{\vec{x}}_0 = \langle \vec{x} \rangle = \frac{\sum_n f_n \vec{x}_n}{\sum_n f_n}, \quad (2.26)$$

where \vec{x}_n and f_n denote the position vector and value of pixel n , do not coincide anymore. Hence, we call this special type of asymmetry “lopsidedness”. The centre of light $\vec{x}_{\text{col}} = \langle \vec{x} \rangle$ and the maximum position \vec{x}_{max} coincide if and only if the light distribution is symmetric. As is evident from Fig. 2.8, the lopsidedness is stronger for steeper profiles, where the maximum lopsidedness is $|\vec{x}_{\text{col}} - \vec{x}_{\text{max}}|/\beta \approx 0.25$. Moreover, Fig. 2.8 demonstrates that, especially for steep profiles, estimates of asymmetry and concentration strongly depend on the choice of centroid. Asymmetry indices estimated with respect to maximum and centre of light may differ substantially in the presence of lopsidedness considering the allowed parameter range.¹¹ Moreover, Fig. 2.8 reveals that the concentration estimated with respect to the maximum position is almost insensitive to lopsidedness, whereas the concentration estimated with respect to the centre of light can be biased low by up to 15%. In Fig. 2.5, the observed concentration indices were estimated with respect to the centre of light rather than the maximum position (cf. Scarlata et al. 2007). This may provide an explanation for the large scatter in Fig. 2.5.

We have demonstrated that the parametrisation results differ significantly depending on whether we use the centre of light or the maximum position as centroid. How do we resolve this ambiguity? And how do we estimate the maximum position in practice, when we suffer from pixel noise? If the parametrisation scheme was model-based, the model would define the centroid during the fit procedure – even in the presence of pixel noise. For instance, the Sérsic profile should use the maximum position as centroid, whereas shapelets

¹¹The steps in panel (c) are due to the computation of A_{col} , since \vec{x}_{col} is changing as F_1 increases. Whenever \vec{x}_{col} enters a new pixel, the set of pixels used for computing A_{col} changes. There are also steps in C_{col} , but they are very small.

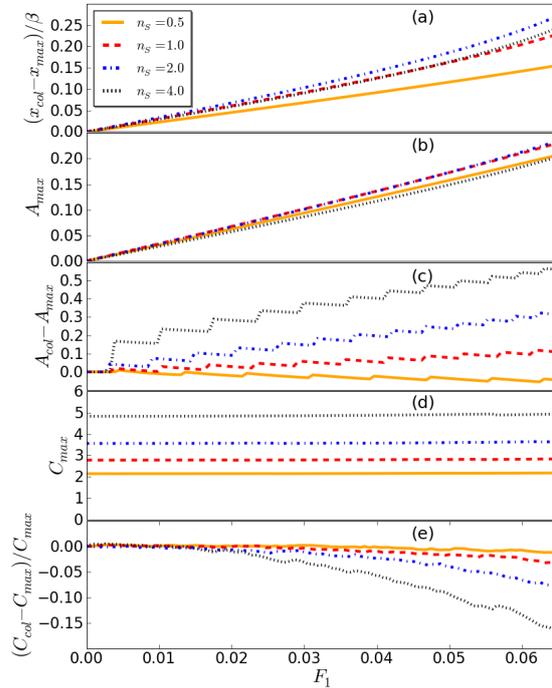


Figure 2.8: Impact of lopsidedness on various morphological quantities.

Impact on centroid (a), asymmetry with respect to maximum (b), absolute difference of asymmetries with respect to centre of light and maximum (c), concentration with respect to maximum (d), and relative difference of concentrations with respect to centre of light and maximum (e). Lopsidedness leads to a difference in maximum position and centre of light. Furthermore, lopsidedness creates asymmetry. Asymmetries evaluated with respect to the maximum or centre of light can differ substantially given that $A \in [0, 2]$. The concentration evaluated at the maximum position is almost insensitive to lopsidedness. However, the concentration with respect to centre of light is strongly underestimated. All Sérsic profiles are evaluated on a 1000×1000 pixel grid using $\beta = 50$. See footnote for explanation of the steps in panels (c) and (e).

can use both maximum position or centre of light. However, since C , A and M_{20} are not model-based, we have to resort to convention or ad-hoc solutions. Concerning the asymmetry index, Conselice et al. (2000) solved this problem by searching for the position that minimises the value of the asymmetry index, also considering resampling the image on a refined pixel grid. They were able to show that there are usually no local minima of asymmetry indices and hence that their method is stable. In the case of the concentration index, using the maximum position appears to be more plausible than the centre of light, since C_{max} appears to be robust against lopsidedness. Unfortunately, the concentration index does not provide us with model and residuals, hence we cannot estimate the most likely maximum position in the presence of noise. However, we can apply the same ad-hoc solution that Conselice et al. (2000) introduced for the asymmetry index, by searching the position that *maximises* the concentration estimate. Nevertheless, this method increases the computational

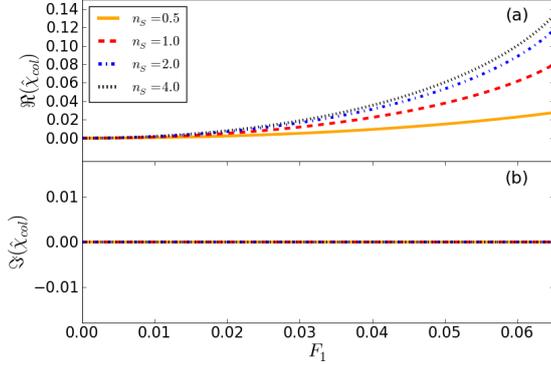


Figure 2.9: Impact of lopsidedness on ellipticity estimates.

The real parts are biased (a) while the imaginary parts are not (b). Considering $0 \leq |\hat{\chi}_{\text{col}}| < 1$, the real part is strongly biased by the lopsidedness. The imaginary part is unbiased due to the geometry of F_1 (cf. Fig. 2.7). All Sérsic profiles are evaluated on a 1000×1000 pixel grid using $\beta = 50$.

effort tremendously such that the required computation time is approximately of the same order as, e.g., fitting a Sérsic model. Evidently, concentration and asymmetry estimates are neither easy to implement nor computationally faster than model-based approaches.

2.3.5 Example IV: Impact of lopsidedness on ellipticity estimates

As our last example, we discuss the impact of asymmetry on estimators of ellipticity. Again, we simulate asymmetry as lopsidedness as in the previous section. We apply flexion transformations to two-dimensional Sérsic profiles without noise. However, we do *not* apply shear transformations, i.e., all profiles have no intrinsic ellipticity. From the pixellised images we then estimate the second moments of the light distribution,

$$Q_{ij} = \frac{\sum_n I_n (x_{n,i} - x_{0,i})(x_{n,j} - x_{0,j})}{\sum_n I_n}, \quad (2.27)$$

where \vec{x}_0 is the point of reference, e.g., centre of light or maximum position. Using the second moments, we compute the ellipticity estimator (e.g. Bartelmann & Schneider 2001)

$$\hat{\chi} = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}}. \quad (2.28)$$

This estimator is related to the axis ratio via $q = \frac{b}{a} = \sqrt{\frac{1-|\hat{\chi}|}{1+|\hat{\chi}|}} \leq 1$ and to the orientation angle θ via $\tan(2\theta) = \frac{\Im(\hat{\chi})}{\Re(\hat{\chi})}$. If this estimator detects any ellipticity, it is completely artificial, i.e., it indicates a bias.

Figure 2.9 shows results of this simulation. For perfectly symmetric profiles ($F_1 = 0$) the estimator indeed does not detect any ellipticity. However, if F_1

increases, the ellipticity estimator will be biased. The bias is stronger for steeper profiles. The maximum bias is $\mathfrak{R}(\hat{\chi}_{\text{col}}) \approx 0.13$, which corresponds to an axis ratio of $b/a \approx 0.877$ and is substantial.

We conclude from this simulation that asymmetries have a potentially strong impact on ellipticity estimates, i.e., asymmetry and ellipticity are intertwined. For instance, this is relevant in the case of using elliptical isophotes for estimating the concentration index.

2.3.6 Reliability assessment

In the previous sections we have demonstrated that some important morphological observables cannot be measured independently of each other. Given this insight, it cannot be guaranteed that estimates of an individual observable result in a parametrisation which is unbiased by the other observables. As all the parametrisation schemes mentioned in Sect. 2 are derived on rather restrictive assumptions (cf. Sect. 2.2.5), their flexibility in describing arbitrary galaxy morphologies is therefore limited. Consequently, it cannot be expected that these schemes provide accurate descriptions of *all* individual objects in a given data sample.

Can we assess the quality or reliability of the parametrisation results for *individual* objects, i.e., can we detect objects where the parametrisation failed in order to sort them out?¹² If we are using a model-based parametrisation scheme (e.g., shapelets or Sérsic profiles), the residuals of the resulting best fit provide us with an estimate of the goodness of fit. For instance, a residual map can reveal systematic mismodelling. However, if the parametrisation scheme is not model-based – as in the case of CAS, M_{20} and Gini – we have no residuals and hence we have no way of assessing the reliability for individual objects.¹³

2.3.7 How to disentangle observables

As we showed above, morphological observables are intertwined and cannot be measured independently. Nevertheless, there is a way to get disentangled estimates.

Let us consider two morphological observables A and B (e.g., Sérsic index and ellipticity). Intertwinement means that the joint probability of A and B does not factorise, i.e.

$$\text{prob}(A, B|\text{data}) \neq \text{prob}(A|\text{data}) \text{prob}(B|\text{data}) . \quad (2.29)$$

Using Bayes' theorem, we can rewrite the joint probability of A and B ,

$$\text{prob}(A, B|\text{data}) = \frac{\text{prob}(A, B) \text{prob}(\text{data}|A, B)}{\text{prob}(\text{data})} , \quad (2.30)$$

where $\text{prob}(A, B)$ denotes the joint prior probability of A and B , $\text{prob}(\text{data}|A, B)$ is the likelihood function and $\text{prob}(\text{data})$ a normalisation factor. A model that simultaneously measures A and B provides us with the likelihood function. For instance, in the case of Gaussian residuals, this likelihood is

$$\text{prob}(\text{data}|A, B) \propto e^{-\chi^2/2} . \quad (2.31)$$

¹²Note that this task is completely different from testing the reliability using simulations. Such simulations allow to assess and calibrate a parametrisation scheme in general, but they do *not* help in detecting parametrisation failures for *individual* objects.

¹³Note that reliability assessment and error estimation are two different tasks. Error estimation is possible for model-independent approaches, e.g., via bootstrapping the imaging data.

We then get independent estimates of A and B via marginalisation

$$\text{prob}(A|\text{data}) = \int dB \text{prob}(A, B|\text{data}) , \quad (2.32)$$

$$\text{prob}(B|\text{data}) = \int dA \text{prob}(A, B|\text{data}) . \quad (2.33)$$

Obviously, this only works for model-based parametrisation schemes, since otherwise we do not have residuals, i.e., the concept of a likelihood function is undefined. In other words, even if we found a model-independent parametrisation scheme that accounts for all observables simultaneously, we would not know how to disentangle the estimates. In addition to reliability assessment, this is another strong argument in favour of model-based approaches.

The marginalisation integrals of Eqs. (2.32) and (2.33) are usually very hard to evaluate, unless we use Markov-Chain Monte-Carlo methods (e.g. MacKay 2008). Using such methods for model fitting, we get the marginalisations for free, without any further effort.

2.4 Impact of PSF on the concentration index

In Sect. 3, we introduced the notion of intertwinement that may systematically influence morphological parameters. Another important origin of systematic effects is the point-spread function (PSF), as we illustrate in this section. The fact that parameters such as the concentration index may be influenced by the PSF is not new but has been long known. For instance, Scarlata et al. (2007) find that the PSF has a significant effect for objects with half-light radii smaller than two FWHM of the HST ACS PSF and with high Sérsic index, while the effect is negligible for larger objects. In an attempt to overcome this bias, Ferreras et al. (2009) applied an ad-hoc correction to the estimated concentration index, based on the half-light radius. The aim of this section is to reassess the impact of the PSF on estimates of the concentration index.

2.4.1 Forward vs. backward PSF modelling

In the case of model-based parametrisation schemes it is standard practice to account for the PSF by forward modelling, i.e., to fit a convolved model to the convolved data. In the case of parametrisation schemes that are not model-based this is impossible and we have to resort to backward PSF modelling, i.e., we deconvolve the data before the actual parametrisation is done. However, deconvolution in the presence of pixel noise is numerically unstable, so forward PSF modelling is to be favoured if possible. This is another practical disadvantage of parametrisation schemes that are not model-based, because they need to perform either an unstable backward modelling or they need to invoke another ad-hoc correction calibrated in simulations. Such simulation-based calibrations introduce a further assumption into the parametrisation process. Model-based schemes are much more rigorous in this respect, since they allow

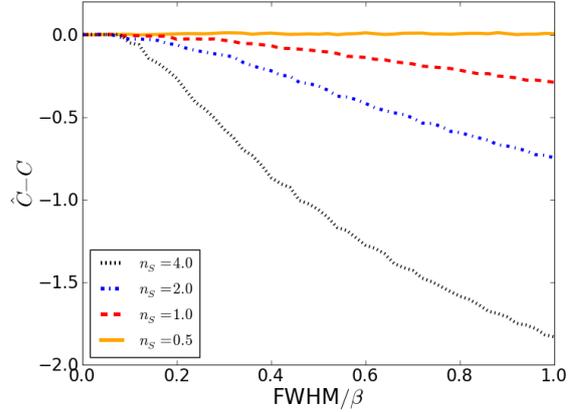


Figure 2.10: Impact of PSF on estimation of concentration index.

Bias $\hat{C} - C$ for different PSF sizes and Sérsic profiles. All Sérsic profiles are evaluated on a 1000×1000 pixel grid using $\beta = 50$ and $b_n = 2n_S - 1/3$. With increasing PSF size with respect to the object size the concentration index estimated from the convolved image is increasingly underestimated.

for a mathematically well-defined PSF treatment that does not introduce any further assumption.

2.4.2 Impact on concentration index

In the case of the ZEST, Sargent et al. (2007) accounted for the PSF by forward modelling when estimating the Sérsic index, while Scarlata et al. (2007) neglected the PSF when estimating the concentration index. The fact that the results shown in Fig. 2.5 are in agreement with theoretical predictions suggests that in the case of the COSMOS data the PSF can indeed be neglected for estimates of the concentration index. Therefore, the theoretical prediction supports the claim by Scarlata et al. (2007). Nevertheless, this single example should not mislead us to generalise this conclusion. It is *not* guaranteed that the PSF will have no impact on the concentration index for data sets other than COSMOS that exhibit different signal-to-noise ratios, PSF, and resolution.

In order to test the impact of the PSF on the concentration index, we generate two-dimensional Sérsic profiles with $n_S = 0.5, 1, 2, 4$ and convolve these profiles with a Gaussian kernel of increasing FWHM.¹⁴ We expect that the concentration indices of very steep Sérsic profiles are severely underestimated, since the PSF washes out the sharp peak. For lower Sérsic indices this effect becomes smaller. For $n_S = 0.5$, the concentration index should not be affected at all, since convolution of a Gaussian with a Gaussian yields a Gaussian, i.e., the steepness of the profile does not change. Figure 2.10 confirms our expectation. If we ignore the PSF, we can significantly underestimate the concentration index in particular for very compact objects such as elliptical galaxies. Hence, this bias can be problematic when using the concentration index for morphological classification, since it can cause elliptical galaxies to

¹⁴We are aware that the COSMOS PSF is not a Gaussian. This test is meant to demonstrate the principle of this effect.

look more similar to disc galaxies than they actually are.

We conclude from this test that although the PSF is indeed negligible in the case of the ZEST, this cannot be generalised to other data sets. Consequently, a PSF treatment is always necessary at least when using the concentration index. In particular concerning ground-based telescopes, the PSF is usually *not* small compared to the peak exhibited by highly concentrated objects.

2.5 Parametrisation and classification

We now discuss the parametrisation of galaxy morphologies in the context of classification, since our primary interest in this subject is to separate disc galaxies from other types of galaxies. First, we show that if we do not account for all morphological observables simultaneously, the effects discussed in the previous sections can dilute discriminative information. Second, we show that all parametrisation schemes discussed here form nonlinear or even discontinuous parameter spaces. Third, we comment on the problem of high-dimensional parameter spaces.

2.5.1 Loss of discriminative information

The conclusion from our investigation of the intertwinement was: If a parametrisation scheme does not account for all morphological observables simultaneously, the results will be systematically altered, i.e., biased. How does this influence classification results? For a large sample of objects, the origins of these systematic effects have random strength. Consequently, we have to expect an increase in the scatter of the resulting parameters. The sample distributions of the parameters are broadened due to the additional scatter, i.e., peaks in the distributions are reduced and troughs between different peaks are washed out. In other words, we are losing discriminative information.

We now demonstrate this broadening of parameter distributions: We generate samples of two-dimensional Sérsic profiles with fixed Sérsic indices of $n_S = 1, 2, 3, 4$. We then add a random ellipticity and a random lopsidedness via the flexion transformation of Eq. (2.25). The flexion parameter F_1 is drawn from a uniform distribution on the interval $[-0.065, 0.065]$. The ellipticity is drawn from the joint distribution of Sérsic indices and axis ratios of 2,000 COSMOS galaxies randomly drawn from the Zurich Structure & Morphology catalogue. We then sample the Sérsic profiles on a $1,000 \times 1,000$ pixel grid using a scale radius of $\beta = 50$. We convolve the resulting image with a Gaussian PSF of FWHM= 37.5 chosen such that the effects of Fig. 2.10 are present but moderate. There is no pixel noise in this simulation. From the pixellised image we then estimate the concentration index with respect to the peak of light and the centre of light, since Sérsic index and concentration are two different estimators for the same morphological feature. Concentration estimates also take into account elliptical isophotes, where the ellipticity is estimated via Eq. (2.27) with respect to the maximum position and the centre of light, respectively.

Figure 2.11 shows the results of this simulation. The distributions of concentration indices have a finite width, in contrast to the distribution of the

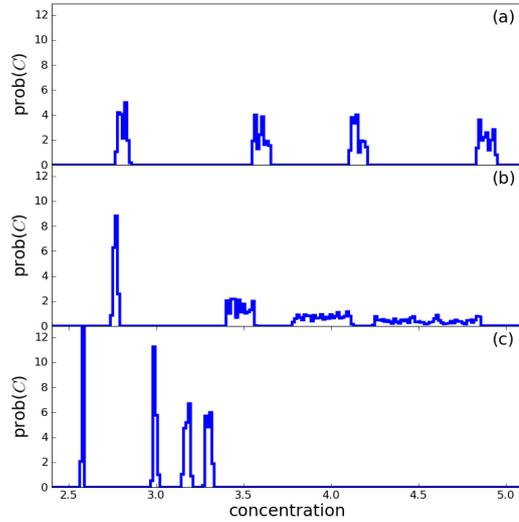


Figure 2.11: Discriminative information of concentration index.

Normalised sample distributions of concentration indices estimated with respect to (a) the maximum position of unconvolved image, (b) the centre of light of unconvolved image, and (c) the centre of light of convolved images. The modes in the distributions correspond to samples of 10,000 profiles each with fixed Sérsic indices of exactly $n_S = 1, 2, 3, 4$ (from left to right). The finite widths of all modes in all distributions indicate the loss of discriminative information. This is particularly evident in panel (b), where the modes of very compact objects are substantially broadened. All Sérsic profiles were evaluated on a 1000×1000 pixel grid using a scale radius of $\beta = 50$. The Gaussian convolution kernel for panel (c) was evaluate on the same pixel grid with FWHM= 37.5.

Sérsic indices, which are infinitely thin δ -peaks. Consequently, we are indeed loosing discriminative information. In reality this loss may be even more severe, since the distribution of Sérsic indices has itself a finite width. Moreover, Fig. 2.11 reveals that the loss of discriminative information is stronger for the concentration index evaluated at the centre of light. Especially for large Sérsic indices the peaks are lowered and broadened. This is a strong argument to evaluate the concentration at the maximum position (if it were accessible), since we conserve more discriminative information. In the presence of an unconsidered PSF, the parameter space is substantially biased. This has the advantage of reducing the width of the distributions, but it also shifts the different modes closer together. If the distribution of Sérsic indices had a finite width, this would wash out the troughs separating the peaks.

This simulation demonstrates that an incautious use of the concentration index (ignoring asymmetries and the PSF) can lead to a substantial loss of discriminative information. In practice, this loss causes sample distributions of the concentration index to be of low modality, despite the diversity of the galaxy population – a problem already mentioned by Faber et al. (2007). Consequently, the concentration index can only provide a lower bound on the

number of classes in a given data sample. If the sample distribution of the concentration is unimodal, this does *not* imply that all objects are of the same type.

2.5.2 Nonlinear parameter spaces

This section highlights an additional problem, which is independent of the previous considerations. It is based on the fact that all parametrisation schemes discussed here are nonlinear in the data. Consequently, the resulting morphological parameters inevitably form nonlinear spaces. For instance, Fig. 2.5 demonstrates that COSMOS galaxies populate a nonlinear submanifold in the two-dimensional parameter space spanned by concentration index and Sérsic index. In fact, the nonlinear submanifold is described by Eq. (2.13), i.e., it is known which is a rare situation. If the parameter space is nonlinear, the distance metric will be nonlinear, too. Although this fact may be known, it is typically ignored in practice. Usually, the Euclidean metric is employed whenever a distance-based algorithm is used, e.g., a principal components analysis (Scarlata et al. 2007) or classification algorithms (e.g. Gauci et al. 2010). The crucial question is: Does ignoring the nonlinearity and employing the Euclidean distance leads us to misestimate the true distances between galaxy morphologies in the parameter space? If so, galaxies may seem to be more similar or less similar than they actually are and hence distance-based classification algorithms may face serious problems. There are only few classification algorithms that do not rely on distances (e.g. Fraix-Burnet et al. 2009).

2.5.2.1 Definition of nonlinearity

Let us consider a parametrisation $P(I)$ of an image I . This parametrisation is said to be *linear* in the image data, if

$$P(\alpha I_A + \beta I_B) = \alpha P(I_A) + \beta P(I_B) \quad (2.34)$$

for any two images I_A and I_B and any real-valued α and β . Otherwise P is nonlinear.

We begin by considering CAS (Eqs. (2.1)–(2.3)). Apart from the obvious nonlinearities in C due to the logarithm and the ratio of radii, the computation of the radii containing 20% and 80% of the total flux itself is highly nonlinear. The nonlinearities in A and S are caused by the fractions and absolute values in the numerators. Gini coefficient (Eq. (2.6)) and M_{20} (Eq. (2.5)) are both nonlinear in the data, too. For both of them the major nonlinearity is hidden in the sorting of the pixel values. The Sérsic model given by Eq. (2.7) contains the Sérsic index and the scale radius as nonlinear parameters.

The nonlinearity of (spherically symmetric) shapelets is due to the scale radius β and the centroid \vec{x}_0 . Both enter the basis functions nonlinearly, as is evident from Eq. (2.11). The nonlinearity of shapelets has been investigated in detail by Melchior et al. (2007), so we do not need to elaborate on this here. In the case of sérsiclets, the Sérsic index is another nonlinear model parameter in addition to the scale radius.

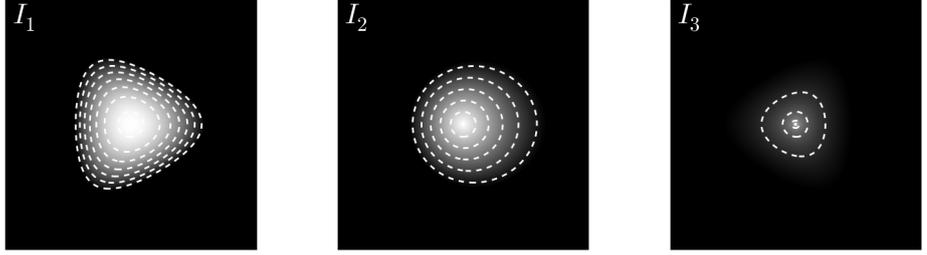


Figure 2.12: Profiles used for demonstration of nonlinearity.

All objects are lopsided and have been evaluated on a $1,000 \times 1,000$ pixel grid with scale radius $\beta = 50$. No intrinsic ellipticity was applied. All maximum positions are identical. Profile I_1 (left) has flexion $G_1 = 0.1$ and $n_S = 0.5$. Profile I_2 (centre) has flexion $F_1 = 0.05$ and $n_S = 1$, i.e., it resembles a disc galaxy. Profile I_3 (right) has flexion $G_1 = -0.1$ and $n_S = 4$, i.e., it resembles an elliptical galaxy.

2.5.2.2 Demonstration of nonlinearity of C , A & Gini

As emphasised above, CAS, Gini, M_{20} and the Sérsic index are nonlinear in the data. The crucial question is: Is the nonlinearity severe or can we assume local flatness in the parameter space and use the Euclidean metric as an approximation? In order to answer this question, we now show a demonstration using three Sérsic profiles with different Sérsic indices and different flexion values as shown in Fig. 2.12. There is no pixel noise in this simulation. We perform a linear transformation in the image space such that two images I_A and I_B linearly transform into each other, i.e.

$$I(\alpha) = (1 - \alpha)I_A + \alpha I_B, \quad (2.35)$$

where $\alpha \in [0, 1]$ parametrises this linear transformation. In reality, the superpositions of this linear transformation may not represent viable galaxy morphologies, e.g., $\alpha = 0.5$ for $I_1 \leftrightarrow I_3$. A proper trajectory should be a geodesic on the submanifold of viable morphologies. If this submanifold is linear, the trajectory defined by Eq. (2.35) will pass through viable morphologies only. If it is nonlinear, it will add additional nonlinearity to this test. This means that even though Eq. (2.35) passes through unrealistic morphologies in this setup, it provides a lower limit to the nonlinearity. For 100 equidistant values of $\alpha \in [0, 1]$ we evaluate the mixed image $I(\alpha)$ in pixel space and then estimate the concentration and asymmetry indices with respect to the maximum position. We also estimate the Gini coefficient. Figure 2.13 shows the trajectories in the subspaces of C , A and Gini. Examples I_1 and I_2 have very similar Sérsic indices and flexion parameters, hence their transition produces trajectories that are only moderately nonlinear. However, example I_3 is very different from I_1 and I_2 and thus its transitions produce trajectories that exhibit substantial nonlinearities. As example I_3 uses a de Vaucouleur profile and thus resembles an elliptical galaxy, while example I_2 uses an exponential profile and thus resembles a disc galaxy, this nonlinearity leads us to underestimate the difference between both galaxy types when using the Euclidean

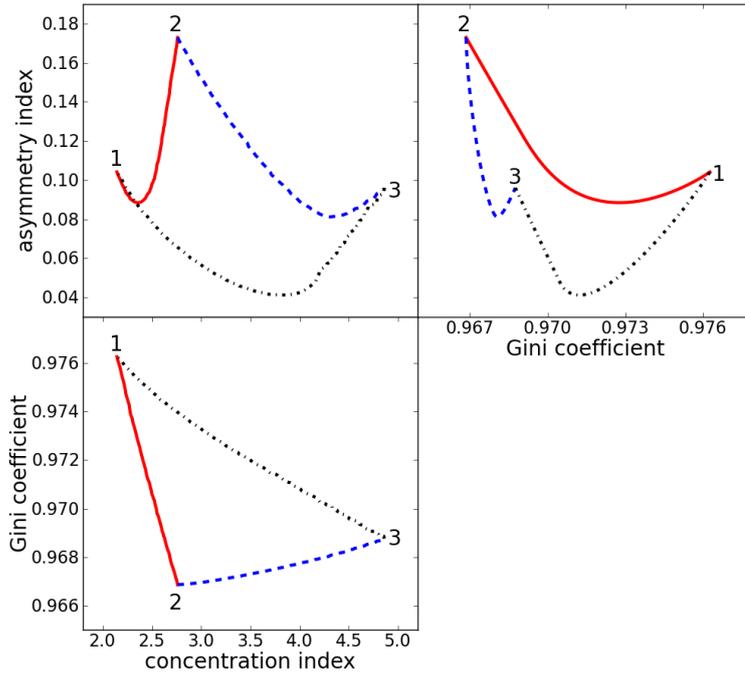


Figure 2.13: Trajectories in CA-Gini subspaces revealing nonlinearities.

Top left: Trajectories in CA space. Top right panel: Trajectories in Gini-A space. Bottom left: Trajectories in C-Gini space. In this simulation the nonlinearity is induced by the different lopsidedness of all objects (cf. Fig. 2.12). The asymmetry is evaluated with respect to the maximum position, whereas the concentration is evaluated with respect to the centre of light.

metric. Note that the nonlinearities in Fig. 2.13 are primarily induced by the lopsidedness via the asymmetry parameter, as is evident from the centre panel where A is not shown and virtually all nonlinearity is gone.

2.5.2.3 Discussion

We conclude from this simulation that for galaxy morphologies exhibiting realistic asymmetries the Euclidean distance is a very poor approximation to distances in parameter space. Consequently, any algorithm based on Euclidean distances would severely underestimate the true distances, i.e., objects would appear more similar than they actually are. This may be an explanation why the drop in the spectrum of eigenvalues of the principal components analysis of Scarlata et al. (2007) (their Fig. 2) – which justifies the reduction of dimensionality – is not very decisive. This is no particular drawback of C , A and Gini, but applies to all other parametrisation schemes discussed here. It is highly questionable whether a “calibration” of the Euclidean distance in order to account for the nonlinearity is possible. The reason for this is that, due to nonlinearity, the distance is an unknown function of the positions of both objects in parameter space, i.e., the distance depends on the morphology.¹⁵ One possible solution is to try to estimate the true distance via a linear transformation as given by Eq. (2.35), although that is computationally very expensive.

¹⁵If galaxies were Sérsic profiles, we could deduce the metric of the parameter space spanned by Sérsic and concentration index from Eq. (2.13).

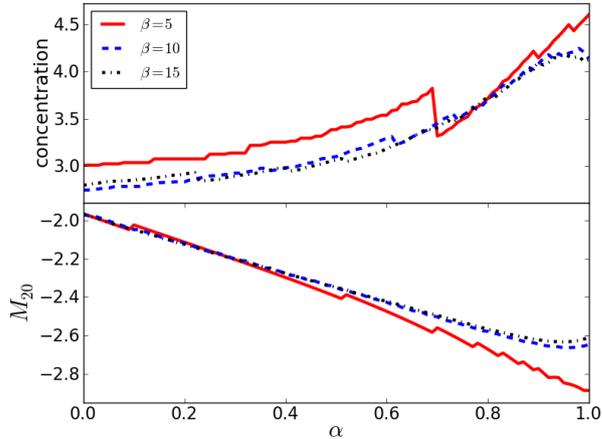


Figure 2.14: Discontinuity of concentration index and M_{20} .

For poor sampling (small β), concentration index (top) and M_{20} (bottom) exhibit substantial discontinuities. For better sampling (larger β) the discontinuities decrease. The transition was between two Sérsic profiles with $n_S = 1$ and $n_S = 3$ and no intrinsic ellipticity or lopsidedness. The scale radii were $\beta = 5$, $\beta = 10$ and $\beta = 15$, respectively, for the exponential profile. The scale radii for $n_S = 3$ are a factor of 3 larger as the profile is more compact. We used Eq. (2.8) in this test, i.e., the scale radii quoted here are half-light radii. The profiles were evaluated on a 500×500 pixel grid.

Another option is to employ a method called “diffusion distance” (Richards et al. 2009) in order to estimate the true nonlinear distances.

2.5.3 Discontinuous parameter spaces

In Fig. 2.13 we used images simulated on a $1,000 \times 1,000$ pixel grid. What happens if we reduce the object size? In Fig. 2.14 we investigate the behaviour of concentration index and M_{20} under a linear transformation between two Sérsic profiles with $n_S = 1$ and $n_S = 3$, i.e., a profile resembling a disc galaxy and steeper profile resembling an elliptical galaxy. In fact, the linear transformation between both profiles corresponds to varying bulge-disc compositions and thus can be interpreted as a morphological sequence from bulge-less disc galaxies (Scd) over bulge-dominated discs (Sa) towards elliptical galaxies without any disc. C and M_{20} exhibit substantial discontinuities due to pixellation effects. These effects increase for decreasing resolution, i.e., decreasing β in Fig. 2.14.

In the case of C , the discontinuities occur because the radii containing 20% and 80% of the total image flux can only change in discrete steps. With increasing resolution, the pixel size decreases and the discontinuities of R_{20} and R_{80} become smaller. Hence, this is not a problem for well resolved galaxies as in Fig. 2.13. However, it is a problem for poorly sampled galaxies. In this case, we can overcome this problem by interpolating the pixellised image and integrating numerically. Unfortunately, this would drastically increase the computational effort. In fact, the discontinuity of the concentration index has

already been observed by Lotz et al. (2006) but they used real galaxy images in their investigation, wherefore they could neither isolate this effect and nor infer resolution limits.

In the case of M_{20} , the origin of the discontinuity is the sum over the second-order moments in the numerator of Eq. (2.5), which stops as soon as 20% of the total flux are reached. This threshold is the problem, as it causes the set of pixels fulfilling this criterion to change abruptly during the linear transformation. Again, the discontinuities of M_{20} decrease with increasing resolution. However, for poorly sampled galaxies we cannot overcome these discontinuities by interpolation, since the definition of M_{20} only makes sense for pixellised images.

A parametrisation scheme forming discontinuous parameter spaces is problematic because it is not guaranteed that objects with similar morphologies end up in neighbouring regions of the parameter space. This implies that distances in the space formed e.g., by M_{20} do not necessarily correlate with the similarity of galaxy morphologies. *We need similar morphologies to have smaller distances than dissimilar morphologies*, but this is not guaranteed for C and M_{20} if the resolution is poor. Figure 2.14 suggests that such discontinuities become important when disc galaxies are smaller than 10 pixels in half-light radius and elliptical galaxies smaller than 30 pixels in half-light radius, maybe even earlier depending on the precise morphology. This applies to $\approx 35\%$ of all COSMOS galaxies with $0.5 \leq n_S \leq 1.5$ and $\approx 97.7\%$ of all COSMOS galaxies with $2.5 \leq n_S \leq 3.5$.¹⁶ Evidently, the problem of discontinuity is substantial in this case and neither concentration index nor M_{20} should be employed for morphological classification of the COSMOS galaxy sample. In this case, we even cannot rely on hard-cut classifications and it is highly questionable whether meaningful classification is possible at all in these parameter spaces.

¹⁶The impact on elliptical galaxies is so large because these galaxies are very compact in comparison to disc galaxies.

2.5.4 High-dimensional parameter spaces

Concerning classification, the current paradigm appears to favour low-dimensional parameter spaces (e.g. Scarlata et al. 2007) that simplify the analysis or even allow a visual representation. However, we have to keep in mind that a high-dimensional parameter space may be necessary in order to differentiate between different groups of galaxy morphologies. There is no astrophysical reason to expect that a two- or even three-dimensional parameter space should be able to host such groups without washing out their differences. This solely depends on the complexity of the physics governing galaxy morphologies.

On the other hand, basis-function expansions typically form parameter spaces of high dimensionality. For instance, the morphological parameter space used by Kelly & McKay (2005) had 455 dimensions. Apart from problems with visualisation, we suffer from what is commonly called the *curse of dimensionality* (Bellman 1961): The hypervolume of a (parameter) space grows exponentially with its number of dimensions.¹⁷ Consequently, the density of data points in this parameter space is suppressed exponentially. In fact, a parameter space with a hundred or more dimensions is virtually empty, regardless of how many data is populating this space. Therefore, it is impossible to reliably

¹⁷Consider a hypercube of edge length L in d dimensions. Its hypervolume L^d grows exponentially with d .

model a data distribution in a parameter space of several hundred dimensions, no matter how much data is available. Nevertheless, it is preferable to employ a parametrisation scheme that produces a high-dimensional parameter space. Loosely speaking, it is better to start with too much information than with too little. We can overcome the curse of dimensionality, if we compress the parameter space, i.e., if we reduce its number of dimensions by identifying and discarding unimportant or redundant information. For instance, Kelly & McKay (2004, 2005) applied a principal component analysis in order to reduce the dimensionality of their parameter space. Unfortunately, principal component analysis is inappropriate in the context of classification because it diagonalises the sample covariance matrix, i.e., it assumes that the whole data comes from a *single* Gaussian distribution. This assumption obviously contradicts the goal of assigning objects to *different* classes. An alternative approach to overcome the curse of dimensionality is to employ a kernel approach by describing the data using a similarity matrix. We demonstrated in Andrae et al. (2010b) that this yields excellent results, e.g., allowing us to classify 84 galaxies populating a 153-dimensional parameter space into three classes.

The bottom line

- Morphological observables are intertwined and cannot be estimated independently without introducing potentially serious biases, questioning the conception of most model-independent parametrisation schemes. Intertwinement can wash out discriminative information in the classification context.
- Model-based approaches allow for forward PSF modelling, disentangling observables through marginalisation, and reliability assessment via model residuals. Model-independent approaches do not and are therefore questionable.
- All parametrisation schemes form nonlinear parameter spaces with a potentially highly nonlinear and unknown metric. Classification algorithms which employ the Euclidean metric underestimate distances and therefore suffer from a loss of discriminative information.
- For poorly resolved galaxies, concentration index and M_{20} form discontinuous parameter spaces that do not conserve neighbourhood relations of morphologies and may therefore fool classification algorithms. This problem applies to $\approx 38\%$ of disc galaxies and $\approx 98\%$ of elliptical galaxies in the COSMOS sample.
- Concerning methods for parametrisation and classification of galaxy morphologies, there is a clear tendency towards *oversimplification*. Methodological sophistication is necessary in order to optimally exploit the large databases accumulated in the recent past and by future surveys.
- A model-based parametrisation scheme is required which accounts for all relevant morphological observables and is highly flexible. In particular, it has to account for azimuthal substructures such as spiral-arm patterns, since we are interested in the classification of disc galaxies.

3

Polar sérsiclets and beyond

As we have seen in the previous chapter, there has been substantial effort to define automated parametrisation schemes for galaxy morphologies. Unfortunately, we have also seen that these parametrisation schemes usually invoke rather restrictive assumptions such that they lack the flexibility to describe the huge variety of different galaxy morphologies present in modern databases. In particular, none of these schemes is capable of representing the rich azimuthal structure of disc galaxies, whereof spiral-arm patterns are of particular interest for our investigations of angular-momentum correlations because through the handedness of their winding they enable us to estimate the rotation sense of the disc. Conversely, basis-function expansions are explicitly designed to model this rich azimuthal structure. However, the shapelet basis functions suffer from severe modelling failures since their Gaussian weight function is not a good match for the radial light profiles of real galaxies, as was shown by Melchior et al. (2010). Consequently, it is obvious to set up a set of basis functions which are an orthogonalisation of the Sérsic profile. This set of basis functions is called sérsiclets. It was first introduced by Ngan et al. (2009) but they faced severe implementation problems that prevented them to get this method to work. Given the conclusions from the previous chapter, this appears to be the most promising approach to faithfully parametrise and hence classify disc galaxies.

3.1 General formalism

In this section, we describe the general recipe how to build a set of basis functions, starting from the orthonormalisation of a chosen weight function. We then briefly discuss the pros and cons of Cartesian and polar basis functions, before touching on the issues of incorporating ellipticity and basis expansion. This discussion is wholly general and we do not specialise on any basis functions yet.

3.1.1 Weight function and orthonormalisation

The first step for building a set of basis functions is to choose a weight function. For the sake of simplicity, let us consider the one-dimensional case where we want to build basis functions on the domain $x \in D$. In this case, the one-dimensional basis function, B_n , of order n takes the generic form

$$B_n(x) = p_n(x)w(x), \quad (3.1)$$

where p_n denotes a polynomial of order n in x and $w(x)$ is the weight function. Therefore, the different modes of a set of basis functions only differ in their polynomial parts, whereas the weight function is identical for all modes.

The second step is to orthonormalise these polynomials with respect to the chosen weight function. Employing the Dirac notation from quantum mechanics, we define

$$\langle B_n | B_m \rangle = \int_D dx p_n^*(x)w^*(x)p_m(x)w(x), \quad (3.2)$$

where $*$ denotes complex conjugate. It is straightforward to show that this integral is a complex-valued scalar product, i.e., it is a positive-definite, Hermitian, sesquilinear form on the linear space of complex-valued functions defined on the domain $x \in D$. The squared absolute value of the weight function $w^*(x)w(x) = |w(x)|^2$ acts as a kernel or metric in this scalar product. The desired result of the orthonormalisation is

$$\langle B_n | B_m \rangle = \delta_{mn}. \quad (3.3)$$

For certain weight functions, the orthonormal polynomials are known analytically, e.g. Hermite polynomials for a Gaussian weight function. For other weight functions an analytic solution often does not exist and the orthonormal polynomials have to be constructed numerically, e.g., using the Gram-Schmidt procedure (see Sect. 3.7.2).

3.1.2 Cartesian basis functions

As galaxy images are two-dimensional, two-dimensional extensions of basis functions are required. The first intuitive choice for building two-dimensional basis functions are those that are separable in Cartesian coordinates (x, y) , since they are real-valued and naturally fit the rectangular pixel grid. Let us consider the Gaussian weight function, which satisfies

$$e^{-x^2/2} \cdot e^{-y^2/2} = e^{-(x^2+y^2)/2} = e^{-r^2/2}, \quad (3.4)$$

i.e., this Cartesian weight function is also spherically symmetric. In fact, the Gaussian profile is the only possible weight function satisfying

$$w(x)w(y) = w(r). \quad (3.5)$$

All other Cartesian weight functions *cannot* be spherically symmetric. As an example, Fig. 3.1 compares the Cartesian ground states (lowest order basis function or 0-0 mode) when using the Gaussian profile and the Sérsic profile

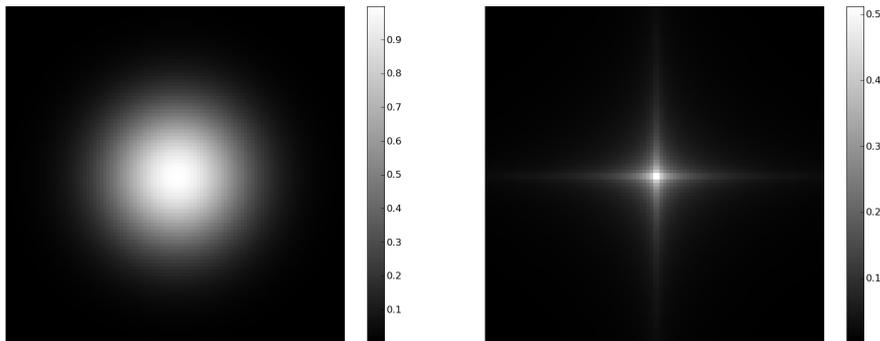


Figure 3.1: Cartesian ground states using different weight functions.

In the left panel the weight function is a Gaussian profile and it is a Sérsic profile with $n_S = 2$ in the right panel. Both ground states have a scale radius $\beta = 15$ and are evaluated on a 101×101 pixel grid. Unless we employ the Gaussian profile, the Cartesian weight function is not spherically symmetric.

as weight function. While the former provides a fair initial guess of some galaxy's morphology and therefore acts as useful ground state, the latter does not resemble any known galactic shape. From these considerations we have to conclude that Cartesian basis functions – apart from Gaussian-weighted – are inappropriate for describing galaxy morphologies.

3.1.3 Polar basis functions

As we have seen, general Cartesian basis functions are ruled out by their unphysical ground states. The alternative is to build basis functions that are separable in polar coordinates (r, φ) . Employing the language of quantum mechanics again, polar basis functions are the eigenfunctions of the angular-momentum operator and are specified by the eigenstate quantum numbers l and m . These polar basis functions take the generic form

$$B_{lm}(\vec{x}) = B_{lm}(r, \varphi) = (2\pi)^{-1/2} R_l(r) e^{im\varphi} . \quad (3.6)$$

We explain in Sect. 3.2.6 how to build real-valued models from these complex-valued basis functions. The radial functions $R_l(r)$ are real-valued and orthonormal, i.e.,

$$\int_0^\infty dr r R_l(r) R_{l'}(r) = \delta_{l,l'} . \quad (3.7)$$

Note the functional determinant $dr r$ which arises from integration in polar coordinates. The R_l depend on the choice of the weight function,

$$R_l(r) = p_l(r) w(r) , \quad (3.8)$$

where $p_l(r)$ is a polynomial of order l in the radial coordinate r . The azimuthal part does not depend on the weight function and is orthonormal due to

$$\int_0^{2\pi} d\varphi e^{i(m-m')\varphi} = 2\pi \delta_{mm'} . \quad (3.9)$$

In fact, this is nothing else than the orthonormality of Fourier modes. Obviously, basis functions with $\pm m$ are their complex conjugates and thus linearly independent of each other.

For a given maximum order N_{\max} , the radial index ranges from $l = 0, 1, \dots, N_{\max}$ and the azimuthal index from $m = -l, -l + 2, \dots, l - 2, l$, i.e., m changes in steps of $\Delta m = \pm 2$. The reason is that the polar basis functions are simultaneous eigenstates of the angular-momentum *and* the Hamiltonian operator. Consequently, the creation of a right-handed quantum is always accompanied by the destruction of a left-handed quantum and vice versa. Moreover, if we allowed for $\Delta m = \pm 1$, the resulting azimuthal basis functions would exhibit non-orthogonalities and the resulting coefficients would be strongly correlated.

3.1.4 Elliptical basis functions

Due to the presence of pixel noise in real data and the finite resolution of objects, we cannot decompose a given galaxy image into infinitely many basis functions. Instead, we have to truncate the basis expansion at a certain maximum order. However, Melchior et al. (2010) pointed out that truncated series of spherical basis functions give rise to a systematic underestimation of intrinsic ellipticities, i.e., ellipticity estimates are biased low. Consequently, we seek to incorporate ellipticity into the design of our basis functions. The ellipticity, ϵ , is introduced as a complex-valued quantity as described in Sect. 2.3.2 via the shear transformation of Eq. (2.16). Axis ratio and the orientation angle of the ellipse can then be computed from the complex ellipticity via Eqs. (2.14) and (2.15).

3.1.5 Basis expansion

For the moment, let us assume that the set of basis functions is complete, which means we can decompose an arbitrary image. Using Dirac notation, completeness is expressed as

$$\sum_{l=0}^{\infty} \sum_{m=-l, -l+2, \dots}^l |B_{lm}\rangle \langle B_{lm}| = \delta^2(\vec{x} - \vec{x}'), \quad (3.10)$$

where $\delta^2(\vec{x})$ denotes the two-dimensional Dirac delta function. Let us consider a given image $I(\vec{x})$, which is a scalar-valued function of the two-dimensional pixel position vector \vec{x} . Then we can use the completeness relation of Eq. (3.10) to derive

$$c_{lm} = \langle B_{lm} | I \rangle = \int d^2x B_{lm}^*(\vec{x}) I(\vec{x}), \quad (3.11)$$

where c_{lm} denotes the coefficient of basis function B_{lm} in the linear expansion

$$I(\vec{x}) = \sum_{l=0}^{\infty} \sum_{m=-l, -l+2, \dots}^l c_{lm} B_{lm}(\vec{x}). \quad (3.12)$$

This identity of $I(\vec{x})$ and its basis expansion will only hold strictly, if either we can indeed fit infinitely many basis functions or $I(\vec{x})$ is the superposition of a finite number of basis functions. In practice, neither is the case and Eq. (3.13) is only an approximation

$$I(\vec{x}) \approx \sum_{l=0}^{N_{\max}} \sum_{m=-l, -l+2, \dots}^l c_{lm} B_{lm}(\vec{x}) \quad (3.13)$$

that converges to $I(\vec{x})$ with increasing maximum order N_{\max} . Testing this convergence enables us to assess whether a given set of basis functions indeed satisfies the completeness relation defined by Eq. (3.10). Furthermore, the integral in Eq. (3.11) turns into a summation in practice, due to the pixellation of the imaging data.

3.2 Shapelets and sérsiclets

After introducing the general formalism, we now discuss the pros and cons of specific choices for the weight function. We briefly comment on shapelets and a prior attempt to introduce sérsiclets, also justifying the choice of the Sérsic profile for orthonormalisation from a mathematical point of view. Finally, we give our definition of polar sérsiclets.

3.2.1 Pros and cons of the Gaussian profile

When choosing the Gaussian profile as the weight function, we end up with shapelets (Réfrégier 2003; Massey & Réfrégier 2005). This choice has some important advantages: First, as discussed in Sect. 3.1.2, Cartesian shapelets are the only reasonable Cartesian basis functions in the context of galaxy morphologies. It is easier to fit and handle Cartesian basis functions than polar ones. Second, as already mentioned in Sect. 2.2.4.1, shapelets are nearly invariant under Fourier transformation, which renders convolution and especially deconvolution from the point-spread function an analytic operation in shapelet space (Melchior et al. 2009). Third, due to the Gaussian, shapelets also excel in further nice analytic properties other than analytic convolution.

However, as was shown by Melchior et al. (2010), shapelets can suffer from strong biases that originate from the radial light profiles of galaxies potentially being much steeper than a Gaussian profile, in particular for early-type galaxies. These biases manifest themselves as ring-like artefacts in the models and the residual maps (cf. Fig. 3.6). They are able to wash out virtually all the information about ellipticity of an object and are likely to affect other morphological quantities in a similar way.

3.2.2 The first attempt to introduce sérsiclets

Given these findings about shapelets, it was obvious to employ the Sérsic profile as weight function which is a much better match to the radial light profiles of real galaxies. Consequently, this is likely to remove the strong biases exhibited by shapelets. The resulting basis functions are called sérsiclets and were first investigated by Ngan et al. (2009).

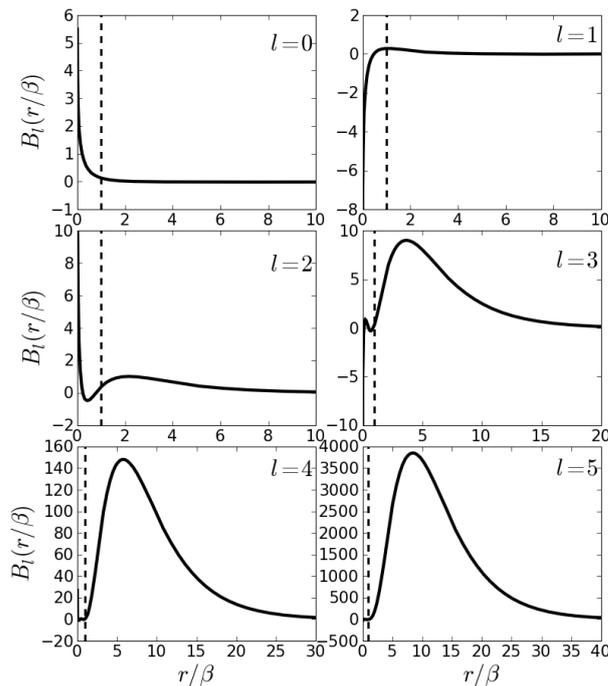


Figure 3.2: Radial basis functions used by Ngan et al. (2009).

Radial basis functions orthonormalised on integration interval $r \in [0, \beta]$ using a Sérsic profile with $n_S = 2$ as weight function. The vertical dashed lines indicate $r = \beta$, where the orthonormalisation interval ends. The Sérsic profile is not able to suppress the polynomials outside the orthonormalisation interval.

However, Ngan et al. (2009) faced a severe problem with sérsiclets. In a simple test case, Ngan et al. (2009) observed that the polar sérsiclets were incapable of fitting a given object. Even when increasing the number of basis functions used for the decomposition, the model did not converge to the given image data. This directly implies that the basis functions constructed by Ngan et al. (2009) are not complete. This problem originates from the fact that Ngan et al. (2009) orthonormalised their basis functions out to one half-light radius – not on the interval $r \in [0, \infty[$ – but then extended their basis functions beyond this range. Figure 3.2 shows what happens in this case. In the inner region, the basis functions may be orthonormal, but for larger radii the leading order monomial in the radial polynomials dominates over the Sérsic profile. This creates an artificial bump outside the region of orthonormality, which appears in all modes of order $l > 0$ and becomes the dominant feature. Consequently, these basis functions lose their linear independence, which explains the non-convergence observed by Ngan et al. (2009).

Ngan et al. (2009) suggested to solve this problem by discarding all modes with $m \neq 0$, i.e., all modes with azimuthal structure, and only maintain the radial modes. While this approach may be viable for weak-lensing applications, where azimuthal structures are rarely visible, it does not solve the actual problem, which is the extension beyond the orthonormalisation interval. Moreover, the resulting basis functions may still represent an expansion into radial modes, but it is not only radial structure of galaxies that is inter-

esting. Rather, a description of azimuthal structures is required in order to faithfully parametrise and classify disc galaxies. In fact, it is the ability to describe azimuthal structure in a (theoretically) well defined way that makes basis-function expansions of galaxy morphologies such a compelling approach.

3.2.3 Mathematical justification

The empirical relation found by Sérsic (1968) is a surprisingly good match to the radial light profiles of real galaxies. This is the motivation to orthonormalise the Sérsic profile. We are now going to demonstrate that the Sérsic profile is the first-order Taylor expansion of *any* light profile, which is a new insight.¹⁸ First, this naturally explains why this profile is such a good fit to real galaxies. Second, it provides a mathematical justification for an orthonormalisation of the Sérsic profile. Later in Sect. 3.7, we are also going to consider higher-order Taylor expansions, going beyond the first-order expansion.

Let us consider the real, two-dimensional light profile of a galaxy $I(\vec{x})$ projected onto the sky, which may exhibit arbitrary radial and azimuthal structures. Due to observational effects, such as PSF, background noise and pixellation, the observed light profile $I_{\text{obs}}(\vec{x})$ is degraded. For observations of low quality, we cannot identify azimuthal structures anymore and only the radial decline is left, i.e., $I_{\text{obs}}(\vec{x}) \approx I_{\text{obs}}(r)$. Let us further consider a rescaling of the observed light profile $p(r) = I_{\text{obs}}(r)/I_{\text{obs}}(0)$, such that $0 < p(r) \leq 1$. Then we can take the logarithm of $p(r)$ and due to $\log p(r) \leq 0$, for $r > 0$ we can also take the logarithm a second time, introducing

$$\tilde{p}(r) = \log(-\log p(r)) . \quad (3.14)$$

We now Taylor expand this function $\tilde{p}(r)$ in $\log r$ at a characteristic radius $\log \beta$ to first order, i.e., a constant plus the first nontrivial term,

$$\tilde{p}(r) \approx A + B(\log r - \log \beta) = A + \log(r/\beta)^B . \quad (3.15)$$

Let us transform backwards now,

$$\log p(r) = -e^{\tilde{p}(r)} \approx -e^A (r/\beta)^B \quad (3.16)$$

and hence

$$p(r) \approx \exp[-e^A (r/\beta)^B] . \quad (3.17)$$

All we need to do now is to identify the coefficients A and B of the Taylor expansion in Eq. (3.15). If we simply rename these constants by $B = 1/n_S$ and $A = \log(b_n)$, then we have arrived at the definition of the Sérsic profile

$$p_S(r) = \exp \left[-b_n \left(\frac{r}{\beta} \right)^{1/n_S} \right] . \quad (3.18)$$

This differs from Eq. (2.7) only by a constant factor of $I_\beta e^{b_n}$. Evidently, the Sérsic profile is the first-order Taylor expansion at radius $\beta > 0$. In other words, in the limit of low resolution and low signal-to-noise ratios (where substructures are negligible) *any* radial light profile is approximately a Sérsic profile.¹⁹

¹⁸Ciotti & Bertin (1999) also investigate a Taylor expansion of the Sérsic profile in their Appendix A. However, they expand the profile in powers of Sérsic index, not in powers of radius.

¹⁹This is not true for radial profiles that are not differentiable, i.e., where no Taylor expansion exists. However, such profiles are unphysical.

3.2.4 Mathematical derivation of radial parts

The crucial idea to overcome the problems faced by Ngan et al. (2009) is to realise that we *cannot* expect a galaxy’s structure being well captured within one half-light radius (see discussion in Sect. 2.2.3.1). For instance, disc galaxies may exhibit spiral-arm patterns in their outskirts. Therefore, we demand orthonormality on the full interval $r \in [0, \infty[$. Apart from solving the problems reported by Ngan et al. (2009), this also has the advantage that the orthonormal polynomials exist analytically as we are going to show now. Our starting point is the scalar product of two radial modes of orders l and l' ,

$$\langle l|l' \rangle = \int_0^\infty dr r R_l(r/\beta) R_{l'}(r/\beta) \exp \left[-b \left(\frac{r}{\beta} \right)^{1/n_S} \right] \quad (3.19)$$

where we have adopted Dirac notation from quantum mechanics. $R_l(r)$ denotes the radial polynomials we are looking for and the Sérsic profile acts as the weight function or “metric” of this scalar product. Note the functional determinant $dr r$, which is due to our integration in polar coordinates. We now change variables according to

$$u(r) = b \left(\frac{r}{\beta} \right)^{1/n_S}, \quad (3.20)$$

such that

$$dr r = \beta^2 \frac{n_S}{b^{2n_S}} u^{2n_S-1} du. \quad (3.21)$$

The limits of integration do not change under this transformation. Then Eq. (3.19) reads

$$\langle l|l' \rangle = \beta^2 \frac{n_S}{b^{2n_S}} \int_0^\infty du \tilde{R}_l(u) \tilde{R}_{l'}(u) u^{2n_S-1} e^{-u}. \quad (3.22)$$

The “new” weight function of this transformed scalar product is now of the form $u^k e^{-u}$, where $k = 2n_S - 1$, and the corresponding set of orthogonal polynomials are the associated Laguerre polynomials,

$$L_l^k(u) = \frac{e^u u^{-k}}{l!} \frac{d^l}{du^l} \left(e^{-u} u^{l+k} \right). \quad (3.23)$$

The L_l^k exist if and only if $k > -1$. This is guaranteed, since $k = 2n_S - 1$ and the Sérsic index satisfies $n_S > 0$. The normalisation factor is

$$\int_0^\infty du L_l^k(u) L_l^k(u) u^k e^{-u} = \frac{\Gamma(l+k+1)}{l!}, \quad (3.24)$$

where Γ denotes the Gamma function. Consequently, the radial parts of the sérsiclets read

$$R_l(r) = \frac{1}{N_l} L_l^{2n_S-1} \left[b (r/\beta)^{1/n_S} \right] \exp \left[-\frac{b}{2} (r/\beta)^{1/n_S} \right] \quad (3.25)$$

with normalisation factor

$$N_l = \sqrt{\frac{\beta^{2n_S} \Gamma(l+2n_S)}{b^{2n_S} l!}}. \quad (3.26)$$

Mind the factor of $1/2$ in the exponent of Eq. (3.25), which arises from our definition that the Sérsic profile is the weight function in Eq. (3.19). We could also have defined Eq. (3.25) with a pure Sérsic profile, resulting in a factor 2 arising in Eq. (3.19). Fit results obtained from both definitions will not differ, since these factors are absorbed in the parameter b . In the special case of polar shapelets these expressions simplify. First, for $n_S = 0.5$, we note that $k = 2n_S - 1 = 0$, which implies $\Gamma(l + k + 1) = \Gamma(l + 1) = l!$ and simplifies the associated Laguerre polynomial L_l^k to the “normal” Laguerre polynomial $L_l^0 = L_l$. Second, in order to obtain shapelets, we need to set $b = 1$. Therefore, the radial parts of polar shapelets read

$$R_l(r) = \frac{1}{\sqrt{\beta^2/2}} L_l \left[(r/\beta)^2 \right] \exp \left[-\frac{r^2}{2\beta^2} \right]. \quad (3.27)$$

The basis functions introduced by Massey & Réfrégier (2005) as “polar shapelets” (their Eq. (8)) differ from Eq. (3.25) by using $x^{|m|} L_{(l-|m|)/2}^{|m|}(x^2)$ instead of $L_l(x^2)$. These two expressions are not identical, e.g., consider any state with $l > 0$ and $m = 0$. Consequently, the “polar shapelets” defined by Massey & Réfrégier (2005) are incorrect and their basis functions cannot be orthogonal. For instance, consider the scalar product

$$\langle l = 1, m = 1 | l = 2, m = 0 \rangle \propto \int_0^\infty dr r e^{-r^2} r^{|1|} L_0^1(r^2) r^{|0|} L_1^0(r^2), \quad (3.28)$$

which equals $-\sqrt{\pi}/8 \neq 0$, i.e., the two states $|l = 1, m = 1\rangle$ and $|l = 2, m = 0\rangle$ are not orthogonal. However, the linear independence is preserved because no two states with identical values of m (otherwise the azimuthal parts $e^{im\varphi}$ would differ) and different values of l can have identical $x^{|m|} L_{(l-|m|)/2}^{|m|}(x^2)$.

3.2.5 Definition of polar sérsiclets

Given the radial parts derived in the previous section, the basis functions of polar sérsiclets read

$$B_{l,m}(r, \varphi) = \frac{1}{\sqrt{2\pi}} R_l(r) e^{im\varphi}. \quad (3.29)$$

Figure 3.3 displays an example of sérsiclet basis functions. Apart from the linear expansion coefficients, sérsiclets have the following nonlinear model parameters:

- The maximum order N_{\max} of the expansion.
- The Sérsic index n_S .
- The scale radius β .
- The centroid position \vec{x}_0 .
- The complex ellipticity ϵ .

It is advantageous to use the real and imaginary parts of the complex ellipticity as model parameters, rather than orientation angle and axis ratio. Using the orientation angle as parameter would cause severe problems with convergence in the case of nearly spherical objects.

In the case of $n_S = 0.5$ and $b = 1$ the polar sérsiclets reduce to the special case of polar shapelets. Furthermore, for $n_S = 1$ and $b = 1$, we get a set of basis function that could be called “disclets”, since they have an exponential profile as weight function. Similarly, we could define “de Vaucouleurlets”, though elliptical galaxies usually do not exhibit azimuthal structures and a basis expansion is not necessary. All these sets are special cases of Eq. (3.29). Finally, we emphasise that when fitting a sérsiclet model, we always also fit a Sérsic model, which is the ground state of the sérsiclets.

3.2.6 Real-valued models

Given the definition of polar sérsiclets in Eq. (3.29), these basis functions are generally complex-valued. This is unphysical because observed galaxy images always have real-valued pixel. Therefore, we have to impose a constraint that removes the complex values. In the case of polar basis functions such as sérsiclets, the model $f(\vec{x})$ of the data $I(\vec{x})$ takes the generic form

$$f(\vec{x}) = \sum_{l,m} c_{lm} B_{lm}(\vec{x}; \beta), \quad (3.30)$$

where the expansion coefficients

$$c_{lm} = a_{lm} + i b_{lm} \quad (3.31)$$

are in general complex-valued, too, and $a_{lm}, b_{lm} \in \mathbb{R}$. As the data is real-valued, we impose the constraint

$$c_{l,-m} = c_{l,+m}^* \quad (3.32)$$

onto the coefficients. It is easily shown that Eq. (3.30) becomes real-valued under this constraint. The constraint of Eq. (3.32) limits the degrees of freedom in our model as follows: For fixed l there are $l + 1$ possible values of m , i.e., there are $l + 1$ coefficients c_{lm} of this radial order l . Equation (3.31) suggests that there are $2(l + 1)$ free parameters, namely the real and imaginary parts of every coefficient. Now, if l is an odd integer, there will be no coefficient with $m = 0$ and thus Eq. (3.32) will half the number of free parameters down to $l + 1$. If l is an even integer, the same argument will apply to those coefficients with $m \neq 0$. For $m = 0$, Eq. (3.32) implies that $c_{l,pm0}$ is real-valued, i.e., the number of free parameters is halved, too. Consequently, Eq. (3.32) ensures that there are only $l + 1$ free parameters for each l . Therefore, the polar model has

$$P = \sum_{l=0}^{N_{\max}} (l + 1) = \frac{1}{2} (N_{\max} + 1)(N_{\max} + 2) \quad (3.33)$$

free parameters, which is the same number as in the case of a Cartesian model of maximum order N_{\max} .

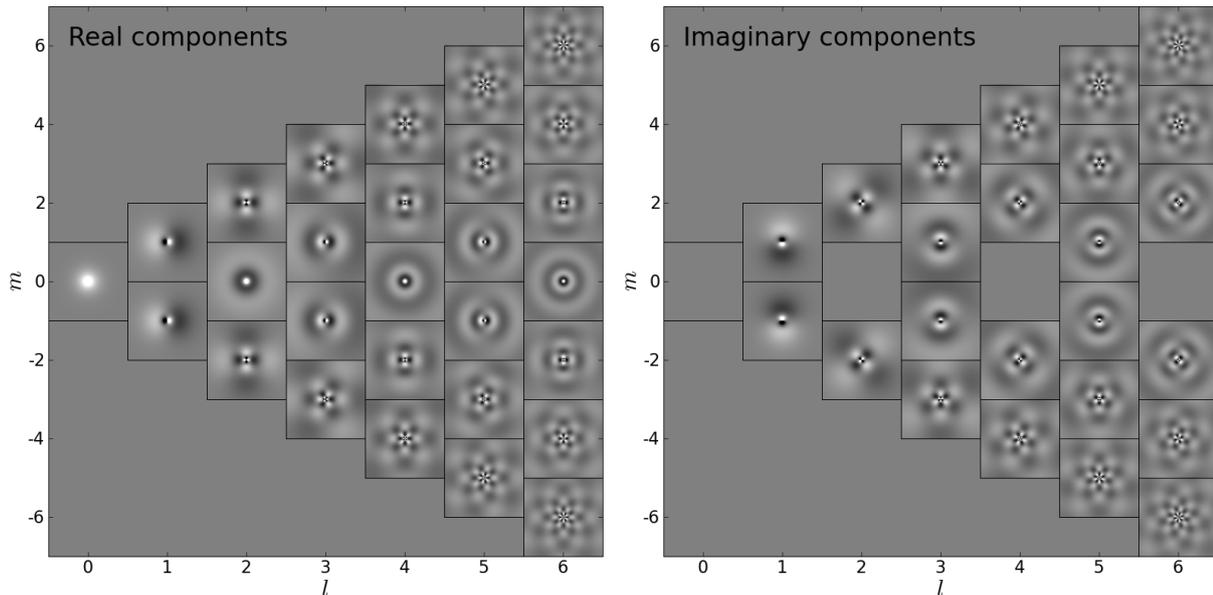


Figure 3.3: Polar sérsiclet basis functions.

The model parameters are $n_S = 1$, $b = 1$ and $\epsilon = 0$. The real components of the complex-valued basis functions are shown in the left panel, the imaginary components in the right panel. The basis functions with $m = 0$ are wholly real. We note that the polar basis functions exhibit lots of substructure in the central region which becomes very small with increasing radial order l .

3.2.7 Interpretation of the Sérsic index

In the case of sérsiclet basis functions, the Sérsic index n_S changes its interpretation. First, n_S regulates the steepness of the weight function, which is a normal Sérsic profile. Second, via the steepness of the weight function, n_S also regulates the spatial scale on which the associated Laguerre polynomials oscillate as is evident from Eq. (3.25). In simple words, there is a fixed relation between steepness of the weight function and oscillation scale of the polynomials. However, real galaxy morphologies do not necessarily obey such a relation. For instance, viewing the same disc galaxy under different inclination angles, the (apparent) steepness of the light profile varies while the scale of visible substructures along the (apparent) semi-major axis remains constant. Furthermore, the steepness of a disc galaxy’s bulge does not generally correlate with the size and distribution of star-formation knots or spiral-arm patterns. In practice, this can lead to modelling problems, if the steepness of the profile and the range of oscillation scales required for a faithful description of a given galaxy morphology do not match.

3.3 Measures and operations in sérsiclet space

One advantage of modelling data is that certain operations and part of the scientific inference can then be conducted in model space, rather than data

space. We now briefly discuss several important operations in the model space of sérsiclets.

3.3.1 Image flux and second moments

We now explain how to estimate flux and second moments from sérsiclet models. The flux F and the second moments Q_{ij} of a model image $I(\vec{x})$ are defined as

$$F = \int d^2x I(\vec{x}) \quad (3.34)$$

and

$$Q_{ij} = \frac{1}{F} \int d^2x x_i x_j I(\vec{x}), \quad (3.35)$$

respectively. Given the analytic form of polar sérsiclets, it is possible to derive analytic expressions for F and Q_{ij} . However, for reasons that we are going to explain in Sect. 3.5.4, we do not recommend using these analytic expressions. Instead, we recommend estimating F and Q_{ij} from the best-fitting sérsiclet model by sampling the model on the pixel grid of the data. These estimates are then given by summations over the pixel grid,

$$\hat{F} = \sum_n I(\vec{x}_n) \quad (3.36)$$

and

$$\hat{Q}_{ij} = \frac{1}{\hat{F}} \sum_n x_{i,n} x_{j,n} I(\vec{x}_n). \quad (3.37)$$

From the second moments we can then estimate the following morphological parameters:

- The complex ellipticity of the overall model,

$$\hat{\epsilon} = \frac{\hat{Q}_{11} - \hat{Q}_{22} + 2i\hat{Q}_{12}}{\hat{Q}_{11} + \hat{Q}_{22} + 2\sqrt{\hat{Q}_{11}\hat{Q}_{22}} - \hat{Q}_{12}^2}. \quad (3.38)$$

- The ratio $q = b/a$ of semi-minor axis b to semi-major axis a according to Eq. (2.14).
- The orientation angle θ according to Eq. (2.15).
- The root-mean-squared (RMS) radius,

$$\hat{R}_{\text{RMS}} = \sqrt{\hat{Q}_{11} + \hat{Q}_{22}}. \quad (3.39)$$

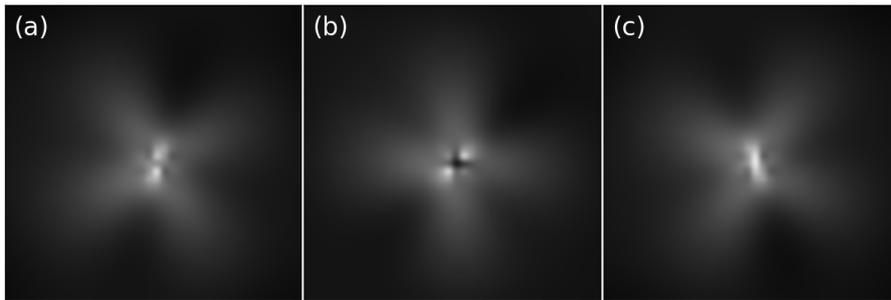


Figure 3.4: Rotation and parity flip in sérsiclet space.

Panel a: A random sérsiclet model. Panel b: The original model of panel a rotated by 30° . Panel c: The original model of panel a flipped.

3.3.2 Rotations in model space

One of the major advantages of polar basis functions is that rotations are easily performed in model space, where we do not suffer from pixellation. Essentially, rotating a model counter-clock wise by some angle α amounts to replacing the azimuthal angle φ by $\varphi' = \varphi + \alpha$. From Eq. (3.6) it is evident that the transformed coefficients are then given by

$$c'_{lm} = c_{lm} e^{im\alpha}. \quad (3.40)$$

The rotated coefficients still satisfy Eq. (3.32), which ensures that also the rotated model is purely real-valued. Moreover, we also need to transform the complex ellipticity, ϵ . Rewriting $\epsilon = |\epsilon| e^{2i\theta}$, where θ is the orientation angle ($\theta \neq \varphi$), it is obvious that the rotated ellipticity is given by

$$\epsilon' = |\epsilon| e^{2i(\theta+\alpha)} = \epsilon e^{2i\alpha}. \quad (3.41)$$

This transformation only changes the orientation but not the axis ratio, since $|e^{2i\alpha}| = 1$. However, Fig. 3.4 displays a sérsiclet model and a rotated version of it. Evidently, the rotated model looks strikingly different from the original model. We need to understand the origin of this problem.

3.3.3 Parity flips in model space

Another advantage of polar basis functions is that parity flips are easy to perform. For instance, parity flips are useful in order to estimate the similarity of galaxy morphologies (Andrae et al. 2010b). Here we consider only parity flips along the x -axis. A flip along any other axis is performed by combining model rotations and the flip technique we are describing now.

A parity flip along the x -axis is equivalent to replacing y by $-y$. In polar coordinates this implies $\varphi \rightarrow -\varphi$. From Eq. (3.6) it is evident that the coefficients transform via complex conjugation. In other words, we can easily perform a parity flip by changing the sign of the imaginary parts of all coefficients, while leaving the real parts unchanged. Again, we also need to transform the complex ellipticity. Its orientation angle undergoes the same transformation, $\theta \rightarrow -\theta$, i.e., the complex ellipticity also transforms via complex conjugation. Figure 3.4 also shows an example of flipping a sérsiclet model. Evidently, there is a discrepancy here, too, which needs to be understood.

3.3.4 Forward PSF modelling

As discussed in Sects. 2.4 and 2.5.1, a PSF treatment is always necessary when dealing with steep light profiles as exhibited by elliptical galaxies. In the case of shapelets, the PSF treatment can be performed analytically in shapelet space, as was shown by Refregier & Bacon (2003). This is possible because Gauss-Hermite polynomials are invariant under Fourier transform and convolution in real space is equivalent to multiplication in Fourier space. For general sérsiclets with $n_S \neq 0.5$, this invariance under Fourier transform is lost and hence an analytic PSF treatment is impossible. We account for the PSF by forward modelling, i.e., we convolve the sérsiclet model with the PSF and fit the convolved model to the given image data. The advantage of this approach is that the convolution of the model does not suffer from pixel noise, since there is no noise in model space.

3.3.5 Inferring spiral-arm handedness

As we shall see in Chapter 6, it is possible to test alignment effects of disc galaxies via the statistical distribution of spiral-arm handedness. This requires a handedness estimate for numerous spiral galaxies. The Galaxy Zoo project (Land et al. 2008) obtained such handedness estimates via visual classification. However, this approach is very time-consuming. Therefore, we now discuss how spiral-arm handedness can be inferred from basis-function expansions, which does not require human interaction, is trivial to parallelise, and is objective and reproducible. We discuss this for general polar basis functions and do *not* specialise on sérsiclets.

A first naïve approach is to exploit the analogy to quantum mechanics, where polar basis functions are eigenfunctions of the angular-momentum operator. In particular, a polar state $|l, m\rangle$ will yield $m|l, m\rangle$, if the operator \hat{L}_z acts on it. The expectation value of \hat{L}_z is exactly what we are interested in. If this expectation value is positive, the spiral arms are right-handed and they are left-handed, if the expectation value is negative.²⁰ Given a galaxy image decomposed into polar basis functions,

$$|I\rangle = \sum_{l,m} c_{lm} |l, m\rangle, \quad (3.42)$$

we can easily evaluate the expectation value of \hat{L}_z . The result is

$$\langle I | \hat{L}_z | I \rangle = \sum_{l,m} m |c_{lm}|^2. \quad (3.43)$$

Unfortunately, the constraint of Eq. (3.32) that enforces real-valued models also enforces that this expectation value is always zero. Consequently, we cannot estimate spiral-arm handedness this way.

A more promising approach is to employ parity flips, which have been introduced in Sect. 3.3.3. Unfortunately, the parity flip cannot be cast into a simple operator acting on the polar basis, whose expectation value could then be computed analytically. In order to estimate the spiral-arm handedness of a given spiral galaxy, its model can be compared to the actual and flipped model of a galaxy with *known* handedness. A useful metric for this comparison has

²⁰This assumes that spiral arms are always leading. This assumption is discussed in Sect. 6.2.6.

been introduced by Andrae et al. (2010b), where the scalar product of the expansion-coefficient vectors of two galaxies is taken as a similarity measure. Spiral-arm handedness is a purely azimuthal feature, i.e., this information should be solely encoded in the expansion coefficients. For two decomposed galaxies with coefficient vectors \vec{c}_1 and \vec{c}_2 , a distance measure is given by,

$$d = \arccos\left(\frac{\vec{c}_1 \cdot \vec{c}_2}{|\vec{c}_1| |\vec{c}_2|}\right). \quad (3.44)$$

As discussed in Andrae et al. (2010b), this distance metric is invariant against size and brightness of both galaxies. In the case of sérsiclets, it is also invariant against ellipticity and steepness of the radial profiles, as these two features are – at least theoretically – encoded in other model parameters but not in the expansion coefficients. By evaluating this distance for a galaxy of unknown handedness and galaxies or templates of known handedness, one can then construct a handedness estimator. In fact, this is a traditional classification problem, where a classifier is calibrated on a training sample. However, also this approach may be severely hampered by the problems with parity flips shown in Fig. 3.4, whose origin is still to be determined.

3.4 Optimisation procedure for sérsiclet expansion

We now discuss how to fit polar sérsiclets to a given galaxy image. First, we discuss the maximum-likelihood solution in the case of given nonlinear model parameters. Second, we explain how to fit the nonlinear model parameters. Finally, we discuss how to estimate errors of the model parameters.

3.4.1 Maximum-likelihood solution

For the moment, we assume that we are given estimates of the nonlinear model parameters $(N_{\max}, n_S, \beta, \vec{x}_0, \epsilon)$. Let us rewrite Eq. (3.30) as $f(\vec{x}) = \sum_{l,m} f_{lm}(\vec{x})$, where the $f_{lm}(\vec{x}) = c_{lm} B_{lm}(\vec{x})$ denote partial models. Due to Eq. (3.32), we can rewrite this summation where for $m > 0$ we get terms like

$$f_{l,+m}(\vec{x}) + f_{l,-m}(\vec{x}) = 2(a_{lm} \Re(B_{lm}(\vec{x})) - b_{lm} \Im(B_{lm}(\vec{x}))) \quad (3.45)$$

and for $m = 0$

$$f_{l,0}(\vec{x}) = a_{l,0} \Re(B_{l,0}(\vec{x})). \quad (3.46)$$

These two equations now define the “partial” models of each of the P fit parameters. We now proceed as usual: We define a vector \vec{c} of free parameters that form the linear expansion coefficients. Furthermore, we write the image as a vector \vec{y} of size N , where N denotes the number of pixels in the given image. Finally, we define the so-called $N \times P$ design matrix X , such that X_{np} is the p -th partial model²¹ (cf. Eqs. (3.45) and (3.46)) evaluated at the n -th image pixel (after forward convolution with the PSF). If the basis functions depend on certain parameters θ , e.g., scale radius β , centroid position \vec{x}_c , etc.,

²¹The index p is given by the “quantum numbers” l and m according to a sorting convention. Different sorting conventions are possible and do not change the result.

then X will depend on θ , too. With these definitions, we can now write the residual vector

$$\vec{R}(\theta) = \vec{y} - X(\theta) \cdot \vec{c}, \quad (3.47)$$

which is an vector of size N itself and also depends on θ . Here “ \cdot ” denotes matrix multiplication. As the pixel noise of modern imaging data is usually Gaussian in excellent approximation, we are allowed to define

$$\chi^2(\theta) = \vec{R}^T(\theta) \cdot \Sigma^{-1} \cdot \vec{R}(\theta), \quad (3.48)$$

where Σ denotes the $N \times N$ pixel covariance matrix. In the case of uncorrelated Gaussian noise with constant variance σ^2 in all pixels, Σ takes the simple form $\Sigma = \text{diag}(\sigma^2, \dots, \sigma^2) = \sigma^2 I$. It is straightforward to show that the maximum-likelihood estimate that minimises χ^2 is given by

$$\hat{c} = (X^T \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X^T \cdot \Sigma^{-1} \cdot \vec{y}. \quad (3.49)$$

The matrix $X^T \cdot \Sigma^{-1} \cdot X$ has a special meaning: It is the $P \times P$ covariance matrix of the coefficients \vec{c} , which would follow from a Fisher matrix analysis. In this formalism, we get this information for free. The advantage of expressing the maximum-likelihood solution in terms of matrix operations is that we can employ fast and efficient linear-algebra algorithms.

3.4.2 Optimising the nonlinear parameters for given N_{\max}

We now assume that we are given an estimate of the maximum order, N_{\max} , and need to get estimates of $(n_S, \beta, \vec{x}_0, \epsilon)$ by minimising Eq. (3.48). These estimates are derived from a Simplex algorithm (Nelder & Mead 1965) that we incorporate from the GNU Scientific Library (GSL)²². The Simplex algorithm is a very robust but not very efficient optimisation algorithm that does not employ derivatives but only evaluations of χ^2 . We also employ priors to restrict the model parameters to reasonable values. These priors in detail ensure that:

- The Sérsic index is in the range $0.1 \leq n_S \leq 8$.
- The scale radius satisfies $\beta > 0$.
- The object centroid \vec{x}_0 is within the pixel grid.
- The complex ellipticity satisfies $|\epsilon| < 1$.

Within these allowed parameter ranges, we employ a non-informative flat prior which assigns equal prior probability to all values.

3.4.3 Estimating the maximum order

In the case of shapelets, the maximum order, N_{\max} , is estimated via a reduced χ^2 . However, demanding that χ^2 equals the number of degrees of freedom is justified if *and only if* the model is purely linear (e.g., Barlow 1993; Andrae et al. 2010c). In the case of sérsiclets, this assumption is definitely violated,

²²<http://www.gnu.org/software/gsl/manual/>

since sérsiclets contain many nonlinear fit parameters. We rather recommend to estimate N_{\max} by comparing the normalised residuals for models of different maximum orders. The optimal N_{\max} is then defined by the model whose normalised residuals are closest to a Gaussian with mean zero and variance one (Andrae et al. 2010c).

Concerning large samples of galaxy images, one can also decompose all objects using identical N_{\max} . This approach may be favourable because many techniques to analyse the resulting catalogue of models require that all coefficient vectors have the same dimensionality (e.g. clustering analysis, cf. Kelly & McKay 2004, 2005; Andrae et al. 2010b).

3.4.4 Error estimation

As we are going to point out in Sect. 3.5.4, error estimation for the best-fit parameters is crucial. Given the best-fit parameters resulting from the optimisation using the Simplex algorithm, we employ a Markov-chain Monte-Carlo (MCMC) algorithm to derive error estimates. MCMC algorithms are iterative algorithms. Loosely speaking, we can interpret an MCMC algorithm as a random walk inside a potential, where the potential is given by the χ^2 -manifold. The number fraction of iteration steps spent at a certain place in parameter space is directly proportional to the likelihood of these parameter values. Consequently, the covariance of the Markov chain is identical to the covariance of the χ^2 -manifold.

A technical remark: The Metropolis-Hastings algorithm – a special kind of MCMC algorithm – is inappropriate in this context because it involves stepsizes in parameter space that need to be fine-tuned in order to allow an efficient sampling of the χ^2 -manifold. This fine-tuning of stepsizes is very hard to automatise in many dimensions and for fitting numerous objects (e.g., Miller et al. 2000). Therefore, we employ a slice-sampling algorithm (Neal 2003) – another kind of MCMC algorithm – that does not involve pre-defined step sizes.

3.5 Testing sérsiclets

Before we can apply sérsiclets to scientific questions, we need to investigate the performance of sérsiclets in some detail.

3.5.1 Completeness

First, we need to test the completeness of sérsiclets in order to proof that we indeed overcome the problems reported by Ngan et al. (2009). We decompose three real galaxy images from the Sloan Digital Sky Survey into sérsiclets with increasing maximum order. If the χ^2 -values do not decrease with increasing maximum order, then our basis functions cannot be linearly independent, i.e., they cannot be complete. Figure 3.5 shows the results. First and foremost, the χ^2 -values of sérsiclets are decreasing with increasing maximum order, i.e., we indeed set up a set of basis functions that are linearly independent. Furthermore, it is interesting to compare the residuals of sérsiclets with those of

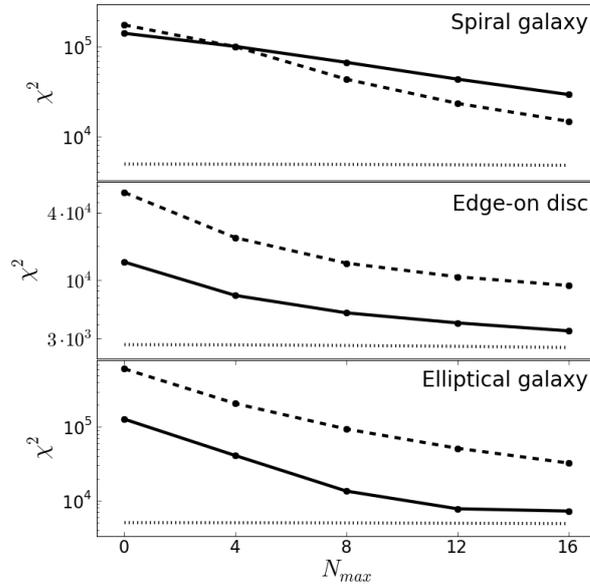


Figure 3.5: Testing the completeness of the sérsiclet basis.

Dependence of χ^2 of sérsiclet decomposition (solid lines) and shapelet decomposition (dashed lines) on maximum order for the spiral galaxy, the edge-on disc, and the elliptical galaxy shown in Fig. 3.6. The horizontal dotted line indicates the number of pixels.

circular shapelets. In the case of the spiral galaxy, sérsiclets yield residuals comparable to shapelets. This is not surprising, because shapelets excel in modelling extended objects with lots of substructure, such as a face-on spiral galaxy. However, in the case of the edge-on disc and the elliptical galaxy, the χ^2 -values of sérsiclets are substantially lower. Evidently, sérsiclets outperform shapelets in modelling galaxies of these types.

3.5.2 Image decompositions

We have seen in the previous section that sérsiclet decompositions produce substantially lower residuals than (circular) shapelet decompositions when it comes to modelling galaxies that exhibit steep profiles, such as elliptical galaxies or edge-on discs. However, we still have to check whether sérsiclets indeed overcome the ring-like artefacts produced by shapelets.

Figure 3.6 compares the best-fitting models and residuals of (circular) shapelets and (elliptical) sérsiclets using $N_{max} = 8$ for the three test galaxies, namely a face-on spiral galaxy, an edge-on disc, and an elliptical galaxy. As expected from the similar χ^2 -values, in the case of the spiral galaxy, both shapelets and sérsiclets perform well in modelling the spiral-arm patterns. Sérsiclets fit the central region better than shapelets, whereas shapelets tend to describe the outskirts better. As expected from a sérsiclet model with $n_S > 0.5$, the polynomial oscillations appear on smaller scales than for the shapelet model with $n_S = 0.5$. Concerning the edge-on disc, sérsiclets are clearly superior. First, the ring-like artefacts of shapelets are gone. Second,

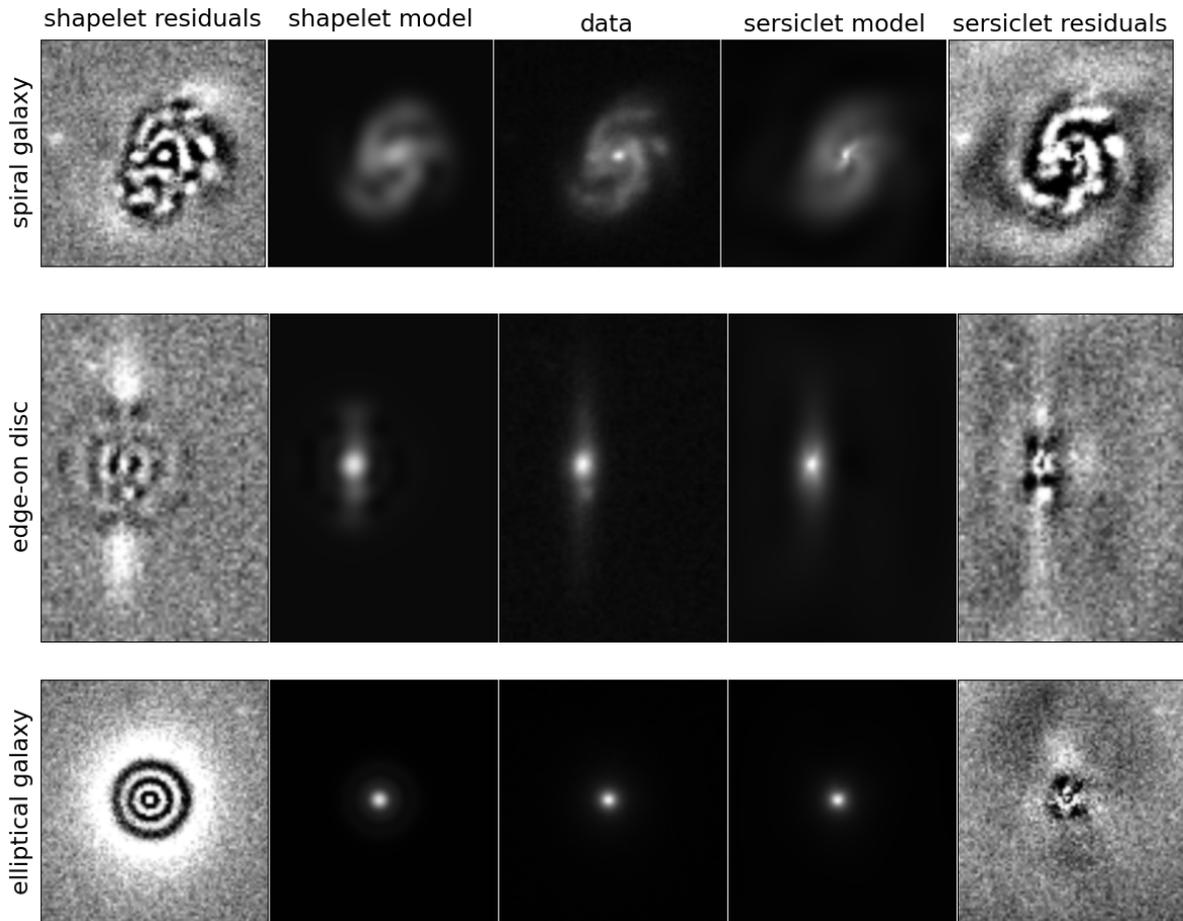


Figure 3.6: Comparing example fits of shapelets and sérsiclets.

Models and residual maps resulting from sérsiclets and shapelets for the spiral galaxy (top row), the edge-on disc (central row), and the elliptical galaxy (bottom row). All models used $N_{\max} = 8$. For the sake of visualisation, the colour code of the model and data maps is nonlinear and ranges from -4.5σ to the maximum value. The colour code of the residual maps is linear and ranges from -4.5σ to $+4.5\sigma$, where σ denotes the standard deviation of the background noise. The *peak* significance of the imaging data is 132σ for the spiral galaxy, 101σ for the edge-on disc, and 610σ for the elliptical galaxy.

the intrinsic ellipticity of the sérsiclet model allows the basis functions to also describe the outermost regions of the disc, whereas these regions go almost unfitted by circular shapelets. In the case of the elliptical galaxy, the ring-like artefacts are very prominent in the shapelet residuals. Conversely, the sérsiclet residuals do not exhibit any artefacts of this kind, i.e., the steepness of the light profile is indeed described properly.

3.5.3 Orthonormality and sampling

In order to get the maximum-likelihood estimate of the expansion coefficients from Eq. (3.49), the matrix $X^T \cdot \Sigma^{-1} \cdot X$ – or $X^T \cdot X$ for uncorrelated and

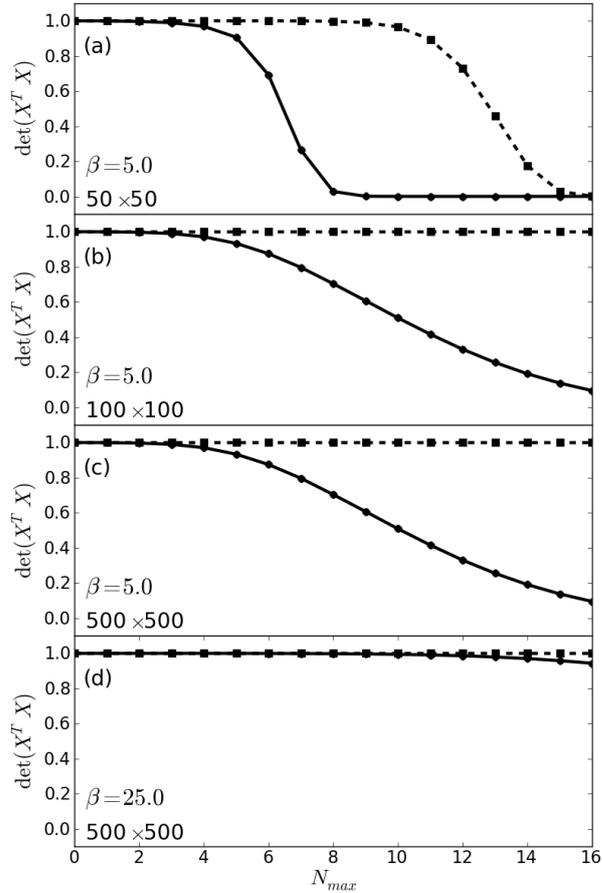


Figure 3.7: Demonstrating undersampling for polar shapelets.

Determinant of $X^T \cdot X$ for Cartesian (dashed lines with squares) and polar (solid lines with dots) shapelets for increasing maximum order N_{\max} and different scale radii β and grid sizes. Panel a: At large N_{\max} both Cartesian and polar shapelets deviate from $\det(X^T \cdot X) = 1$. Panel b: Cartesian shapelets maintain $\det(X^T \cdot X) = 1$ for all N_{\max} , proving that boundary effects are now negligible. Panel c: Polar shapelets still differ from $\det(X^T \cdot X) = 1$ at large N_{\max} on this enlarged grid, i.e., this is not a boundary effect. Panel d: Polar shapelets now also maintain $\det(X^T \cdot X) = 1$, i.e., the remaining effect was due to undersampling.

constant pixel noise – needs to be invertible. If it were not for pixellation, finite image grids, and the PSF, $X^T \cdot X$ should be an identity matrix due to the orthonormality of the basis functions. We now investigate the orthonormality of sérsiclets in two tests.

First, we compare the orthonormality of *polar* shapelets (sérsiclets with $n_S = 0.5$ and $b = 1$) and *Cartesian* shapelets of Melchior et al. (2007). We sample polar and Cartesian shapelets of constant scale radius $\beta = 5$ on pixel grids of sizes 50×50 , 100×100 , and 500×500 and compute the determinant of $X^T \cdot X$ for maximum orders $0 \leq N_{\max} \leq 16$, where $\det(X^T \cdot X) < 1$ indicates a violation of orthonormality. Figure 3.7 shows the results of this test. In the case of the 50×50 grid (panel a) both Cartesian and polar shapelets

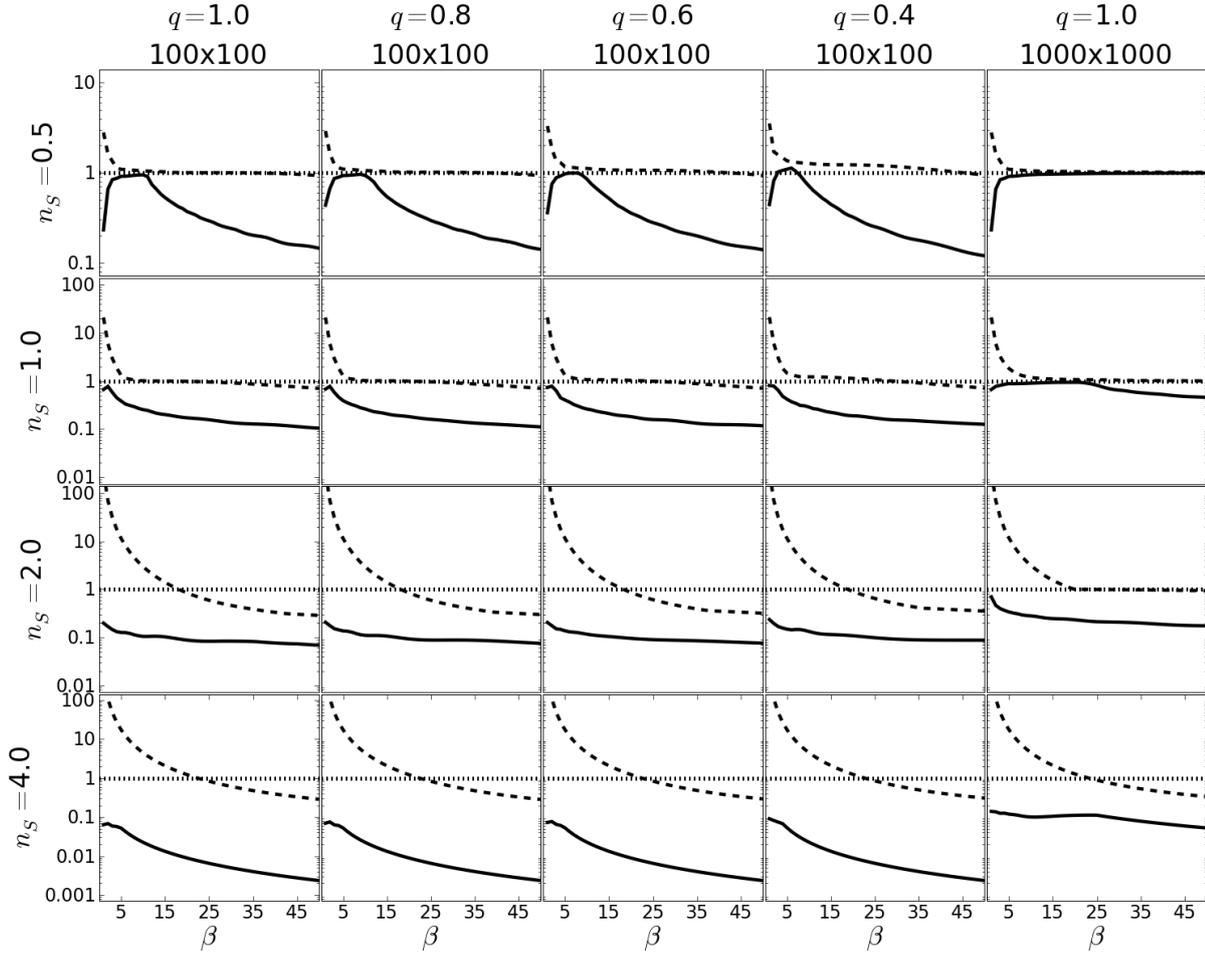


Figure 3.8: Demonstrating undersampling for general sérsiclets.

Smallest (solid line) and largest (dashed line) diagonal elements of matrix $X^T \cdot X$ as a function of β at $N_{\max} = 12$. Deviation from unity (dotted line) indicates the loss of orthonormality. The image size was 100×100 pixels (left columns) and 1000×1000 pixels (right column). We used $b = 2 \log 3$, such that at one scale radius the weight function drops to one third of its central value.

suffer from non-orthonormality for increasing N_{\max} . In particular, for polar shapelets and $N_{\max} \geq 8$ the determinant of $X^T \cdot X$ is zero, i.e., $X^T \cdot X$ is not invertible and the estimator of the expansion coefficients given by Eq. (3.49) breaks down. This problem can be caused either by boundary truncation or undersampling of higher-order modes (with rapidly oscillating polynomials). Therefore, in panel b we increase the pixel grid from 50×50 to 100×100 . The non-orthonormality of Cartesian shapelets is now cured, i.e., it has indeed been caused by boundary truncation. However, polar shapelets still exhibit non-orthonormality. Therefore, in panel c we increase the pixel grid again now from 100×100 to 500×500 . The behaviour of polar shapelets is unchanged, i.e., the non-orthonormality is not caused by boundary truncation. In order to demonstrate that the non-orthonormality is caused by undersampling, we increase in panel d the object size from $\beta = 5$ to $\beta = 25$ while keeping the 500×500 pixel grid. The non-orthonormality of polar shapelets is now almost cured,

which verifies that this was indeed an undersampling effect. Evidently, undersampling has a larger impact on polar shapelets than on Cartesian shapelets. The reason is that polar basis functions exhibit most of their structure in the central region (cf. Fig. 3.3) and therefore suffer strongly from pixellation. Loosely speaking, *polar* basis functions do not appreciate being sampled on a *Cartesian* pixel grid.

Second, having attested problems of polar shapelets with orthonormality due to undersampling, we now investigate the orthonormality of (polar) sérsiclets in general. We compare the value of the largest and smallest diagonal element of $X^T \cdot X$, respectively, repeating the same test as Berry et al. (2004) did for shapelets. We evaluate sérsiclet models of maximum order $N_{\max} = 12$, Sérsic indices $n_S = 0.5, 1, 2, 4$ and axis ratios $q = 1, 0.8, 0.6, 0.4$ on a 100×100 pixel grid for varying scale radii β . There is no PSF in this test. Figure 3.8 shows test results. For the moment, we only consider the first row in Fig. 3.8, which corresponds to polar shapelets. The results agree with Fig. 3.7, revealing undersampling effects for small β and boundary truncation for large β (cf. Melchior et al. 2007, and discussion therein). Furthermore, we can see that the axis ratio has only a mild impact on the orthonormality. However, inspecting the other rows corresponding to steeper profiles ($n_S = 1, 2, 4$), Fig. 3.8 reveals serious violations of orthonormality. Obviously, for $n_S \geq 1$ the undersampling regime is not overcome before the boundary effects set in. This is confirmed by the right-most column in Fig. 3.8, which used an enlarged pixel grid while keeping the resolution (scale radius β) constant. In this case, sérsiclets with $n_S = 1$ exhibit decent orthonormality on this larger grid, while the cases with $n_S > 1$ still do not reach an acceptable level of orthonormality. The reason for this peculiar behaviour is the argument $b(r/\beta)^{1/n_S}$ of the associated Laguerre polynomials in Eq. (3.25). It implies that the polynomials are only slowly varying with r for large values of n_S . Consequently, polar sérsiclets have a serious problem with orthonormality, especially for large Sérsic indices.

3.5.4 Impact of undersampling

We have seen that undersampling is a fundamental problem of polar sérsiclets. The only way to avoid undersampling is to ensure critical sampling of the basis functions, i.e., to ensure that all galaxy images that are to be analysed using sérsiclets have high resolution. This is not always possible to guarantee (especially in the case of weak-lensing studies), so we now need to understand the impact of undersampling. There are four major consequences of undersampling, ordered by importance:

1. As we have seen in Fig. 3.7, undersampling compromises the invertibility of $X^T \cdot X$. If $\det(X^T \cdot X)$ is too close to zero, numerical inaccuracies will dominate its inversion and the estimate of the expansion coefficients using Eq. (3.49) will catch up random biases. Consequently, this increases the uncertainty in the model parameters. We demonstrate this in Fig. 3.9, where we compare the dependence of χ^2 on the scale radius β for Cartesian and polar shapelets fitting the same galaxy. Evidently, the minimum of χ^2 is much sharper for Cartesian shapelets than for polar shapelets. We conclude that an error estimate as described in Sect. 3.4.4 is crucial for any sérsiclet analysis. In the worst case, if $\det(X^T \cdot X)$

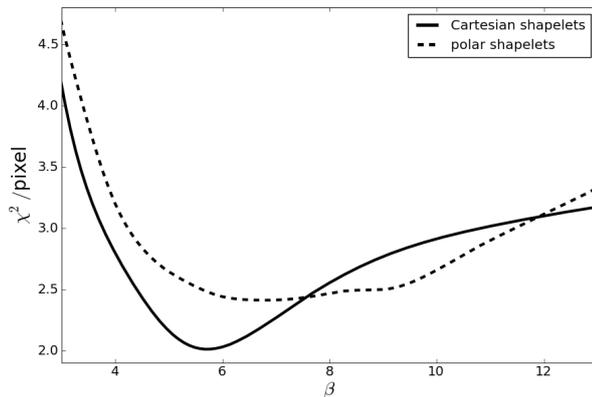


Figure 3.9: Impact of undersampling on parameter estimation.

We show the dependence of χ^2 per pixel on the scale radius β for Cartesian (solid curve) and polar (dashed curve) shapelets for a given face-on disc galaxy. Both types of shapelets used identical $N_{\max} = 12$, identical centroids and zero intrinsic ellipticity.

approaches zero, not only the orthonormality is seriously violated but also the linear independence of the basis functions breaks down.

2. Figure 3.9 shows that Cartesian shapelets also reach lower values of χ^2 than polar shapelets. The reason is that the computation of χ^2 for polar shapelets only considers pixel noise, but not the additional uncertainty introduced by undersampling (random sampling).
3. Estimates of flux and second moments should *not* be derived mathematically from the abstract models, since they can vary substantially because of the sampling. We rather recommend to sample the best-fitting model on the same pixel grid where the fit was performed and then estimate flux and second moments from the sampled model. These estimates still do *not* suffer from pixel noise.
4. As described in Sects. 3.3.2 and 3.3.3, we can perform rotations and parity flips in model space. However, because of undersampling rotated and flipped models may look dramatically different from the original model when sampled on the same pixel grid, as demonstrated by Fig. 3.4. This effect also compromises estimates of spiral-arm handedness as discussed in Sect. 3.3.5, i.e., we cannot estimate spiral-arm handedness from sérsiclet models.

The impact of undersampling could be alleviated by sampling the model several times within each pixel.²³ However, in the limit of infinitely fine sampling this amounts to a convolution of the model with the pixel-response function. As the undersampling problem stems from a key feature of the employed sérsiclet model, namely rapid polynomial oscillations, an oversampled model may better approximate the image data. But the information about the galaxy morphology contained in the expansion coefficients degrades as the subpixel oscillations are washed out. As we require meaningful information from the

²³This corresponds to a Riemann integration of the model within one pixel.

expansion coefficients, we will not further pursue this approach. If one wants to employ this approach, one has to bear in mind, that the additional convolution will lead to covariances among the coefficients and therefore render the fitting procedure more complicated.

3.5.5 Analogy to Nyquist frequency

In order to avoid the undersampling problem, we want to derive a lower limit to the scale radius β of a sérsiclet model. This limit is given by comparing the pixel size to the scale on which the radial components of sérsiclets – essentially the Laguerre polynomials – vary. In fact, this is an analogy to the Nyquist frequency in the case of Fourier transform.

The key to set this lower limit is to identify the shortest scale on which a Laguerre polynomial varies. This scale can be inferred from the roots of the Laguerre polynomial. An associated Laguerre polynomial $L_l^k(x)$ of order l has l real-valued, positive roots within the interval $(0, l + k + (l - 1)\sqrt{l + k}]$. The smallest scale that can be resolved by an associated Laguerre polynomial is given by the distance between $x = 0$ (the peak of the radial component) and the first root $x_1 > 0$. Unfortunately, the roots of associated Laguerre polynomials are not known analytically, so the first root x_1 has to be inferred numerically. As the argument of the associated Laguerre polynomial is $x = b(r/\beta)^{1/n_s}$, given x_1 , we can infer the radius of the first root to be $r_1 = \beta(x_1/b)^{n_s}$. Hence, the distance between $r = 0$ and the first root is

$$\Delta r = r_1 - 0 = r_1 = \beta(x_1/b)^{n_s} . \quad (3.50)$$

This scale Δr should be well resolved by the pixel grid in order to avoid undersampling. The absolute lower limit is $\Delta r = 1$, i.e., the radial component drops from its central peak to its first root within a single pixel.²⁴ Setting $\Delta r = 1$ and solving for the minimal β , we obtain,

$$\beta \geq \beta_{\min} = (b/x_1)^{n_s} , \quad (3.51)$$

where β and hence β_{\min} are given in units of pixels. Table 3.1 provides values of β_{\min} for some realistic values of n_s and polynomial order l . Considering this table, we also have to keep in mind that the image grid has to be large enough such that several scale radii fit into it. Evidently, as n_s increases, the limit becomes larger and larger. This can also be inferred qualitatively from Fig. 3.8, but Eq. (3.51) provides a quantitative result. Obviously, for $n_s \geq 2$ the lower limit becomes extraordinarily large. In particular, $n_s = 4$ would require an extremely well resolved object with radius of several thousand pixels.

In the case of shapelets, it is standard practice to choose a maximum order N_{\max} for a decomposition according to signal-to-noise ratio and resolution. For sérsiclets, the resolution constraint plays a much more important role and restricts models with large n_s to low N_{\max} , with which complicated morphological features could receive improper description. As elliptical galaxies rarely exhibit such features, this restriction may not be overly problematic.

²⁴Actually, the region where the profile drops from its maximum to its first root should be resolved by several pixels.

l	$n_S = 0.5$	$n_S = 1.0$	$n_S = 2.0$	$n_S = 3.0$	$n_S = 4.0$
1	0.8165	0.8333	0.8403	0.8424	0.8435
2	1.0668	1.3145	1.7599	2.2042	2.6658
3	0.3256	1.7810	2.9324	4.3311	6.0286
4	1.4377	2.2423	4.3625	7.3782	11.537
5	1.5904	2.7011	6.0513	11.499	19.888
6	1.7296	3.1586	7.9993	16.848	31.873
7	1.8584	3.6151	10.207	23.577	48.375
8	1.9787	4.0712	12.674	31.841	70.372
9	2.0921	4.5268	15.400	41.792	98.931
10	2.1996	4.9821	18.386	53.584	135.22
11	2.3021	5.4373	21.632	67.371	180.49
12	2.4002	5.8922	25.137	83.305	236.08
13	2.4944	6.3471	28.902	101.54	303.46
14	2.5853	6.8018	32.926	122.23	384.13
15	2.6730	7.2565	37.210	145.53	479.75
16	2.7579	7.7110	41.754	171.59	592.02
17	2.8403	8.1656	46.557	200.57	722.76
18	2.9204	8.6201	51.620	232.61	873.87
19	2.9983	9.0745	56.943	267.88	1047.4
20	3.0742	9.5289	62.525	306.52	1245.3

Table 3.1: Minimal scale radii avoiding undersampling.

Lower limits of scale radii β in units of pixels for different radial orders l according to Eq. (3.51). Here we used $b = 2n_S - 1/3$.

3.6 Application to weak-lensing data

Given the fact that shapelets were also intended for shear measurements in weak-gravitational lensing (e.g., Refregier & Bacon 2003; Chang et al. 2004), we now study the potential of sérsiclets in this field of research. From the simulated GREAT08 data (Bridle et al. 2009) we selected ten sets of artificial galaxies, whose shear values are given in Table 3.2.²⁵ The data sets were taken from the GREAT08 RealNoiseKnown branch, for which the galaxies are modelled as either of bulge or disc type, i.e., having Sérsic index of $n_S = 1$ or $n_S = 4$, respectively. Every set contains 10,000 objects sampled on a 40×40 pixel grid. Within such a set the applied shear is constant and the aim is to retrieve the correct value. Furthermore, all objects have been convolved with a PSF that is a truncated Moffat profile,

$$I_{\text{Moffat}}(R) = \begin{cases} (1 + (R/R_d)^2)^{-\beta} & \text{if } R < R_c \\ 0 & \text{else} \end{cases}, \quad (3.52)$$

with $\beta = 3.5$, R_d chosen such that FWHM = 2.85 pixels, cut-off radius R_c of twice the FWHM, and intrinsic ellipticity $\epsilon = 0.019 - 0.007i$ (cf. Bridle et al. 2010).

We decompose all objects into sérsiclets with maximum orders $N_{\text{max}} =$

²⁵This restriction on only ten sets was necessary, because analysing the whole GREAT08 data set would have been too time-consuming.

Set	g_1	g_2
0007	0.0005405	0.0069236
0026	-0.0527875	-0.0090224
0035	0.0166067	-0.0045223
0048	0.0708862	-0.0377040
0056	0.0325078	0.0978346
0091	-0.0246346	-0.0488837
0126	0.0170977	-0.1383142
0135	0.0596913	0.0416342
0268	-0.0653126	-0.0883511
0281	-0.0431769	0.0462176

Table 3.2: Data sets chosen from the GREAT08 sample.

We also quote the complex shears $g = g_1 + ig_2$ of each set. The selected sets have shear values spanning the diversity of GREAT08 samples.

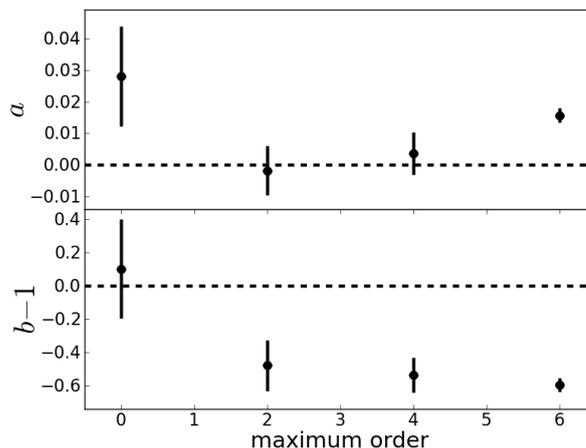


Figure 3.10: Test results of applying sérsiclets to the GREAT08 data.

Offset a (top) and deviation from unity slope $b - 1$ (bottom) of GREAT08 data sets given in Table 3.2 for different maximum orders, N_{\max} , of sérsiclet decomposition. Horizontal dashed lines indicate the ideal case of $a = 0$ and $b = 1$. Errors of a and b were estimated by bootstrap fitting of $\hat{g} = a + b \cdot g_{\text{input}}$ to the estimated and input shear values.

0, 2, 4, 6, taking into account the PSF as described in Sect. 3.3.4. For each set, we estimate the mean ellipticity $\hat{\epsilon}$ from the 10,000 artificial galaxies via the ellipticity parameter of the sérsiclet model, which provides us with an unbiased estimator of the gravitational shear (cf. Bartelmann & Schneider 2001). Consequently, the error estimate is the error of the mean value (see Sect. 6.3.2). The goodness of the shear estimate \hat{g} with regard to the known input shear g_{input} is parametrised by a straight-line model

$$\hat{g} = a + b \cdot g_{\text{input}} . \quad (3.53)$$

A perfect shear estimator yields $a = 0$ and $b = 1$. If a real shear estimator

yields an offset $a \neq 0$, i.e., a shear is detected although the input shear was zero, this typically implies that the PSF is not properly corrected for. If $b < 1$, the true shears are underestimated.

Figure 3.10 shows the test results for offset a and deviation from unit slope $b - 1$. Evidently, the sérsiclets perform best for $N_{\max} = 0$, where the offset a is consistent with zero, while the slope b is consistent with 1. This is not surprising, because for $N_{\max} = 0$ sérsiclets reduce to pure Sérsic profiles, which were used to simulate this subset of the GREAT08 data. However, the real lesson from Fig. 3.10 is that for $N_{\max} > 0$ the shear estimates correlate less strongly with the input shears, i.e., the slopes are significantly below 1. Obviously, the higher-order modes of sérsiclets do more harm than good in this case. The likely explanation is that using higher-order modes, we are suffering from the undersampling problem. As we have seen comparing Cartesian and polar shapelets in Fig. 3.9, undersampling washes out the optimum of the χ^2 manifold and may therefore lead to misestimations of parameters. Low resolution is a generic feature of weak-lensing data, i.e., shear estimates based on sérsiclet decompositions will always suffer from substantial undersampling effects. Oversampling of the model within each pixel could cure this, but is – at least in our implementation – computationally infeasible.

3.7 Orthonormalising higher-order Taylor expansions

We have seen in Sect. 3.2.3 that the Sérsic profile is the first-order Taylor expansion of any light profile. This naturally leads us to the expectation that with improving imaging quality the Sérsic profile may not provide a good match anymore because it is “only” a *first-order* expansion. Consequently, an obvious strategy to enhance the Sérsic profile is to allow for higher orders in the Taylor expansion of Eq. (3.15). Such higher-order radial profiles of course can also be used for orthogonalisation in order to describe azimuthal structures.

3.7.1 Third-order profiles

A realistic profile has to be unity at $r = 0$ and it has to approach zero for $r \rightarrow \infty$. As the Taylor expansion is in $\log r$, only expansions where the leading order term is of odd power in $\log r$ and has a positive expansion coefficient can satisfy these constraints. Therefore, the next useful higher-order expansion beyond the Sérsic profile is a third-order expansion,

$$\tilde{p}(r) \approx A + B \log(r/\beta) + C \log^2(r/\beta) + D \log^3(r/\beta) \quad (3.54)$$

or rather

$$p(r) \approx \exp \left[-e^{A+B \log(r/\beta)+C \log^2(r/\beta)+D \log^3(r/\beta)} \right], \quad (3.55)$$

where $D > 0$ and A , B , and C are arbitrary. The coefficient A and the scale radius β are fully degenerate and we can freely set A to some value, very much

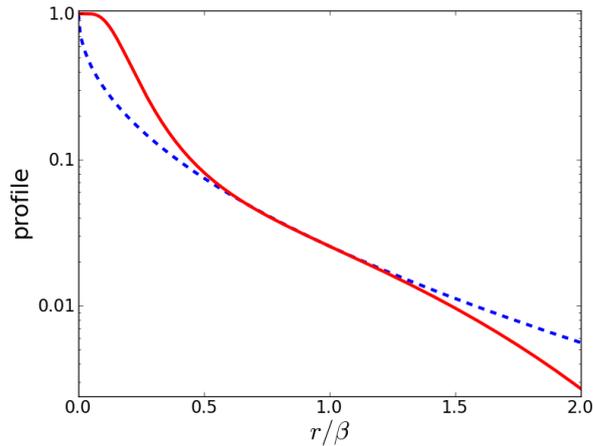


Figure 3.11: Third-order profile vs. Sérsic profile.

We compare a Sérsic profile with $n_S = 2$ (dashed line) to a third-order profile with $B = 1/n_S = 0.5$, $C = 0.1$ and $D = 0.25$ (solid line).

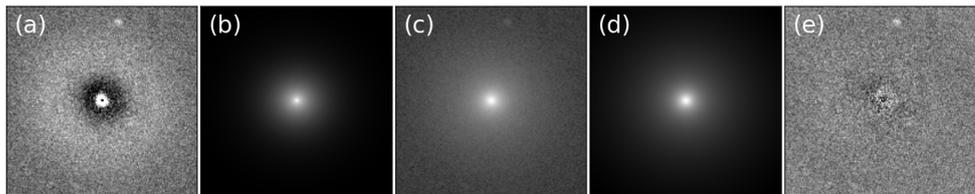


Figure 3.12: Fitting Sérsic and third-order profile to an elliptical galaxy.

The panels denote residual map of Sérsic fit (a), Sérsic model (b), original data (c), best-fitting third-order profile (d), and residual map from third-order profile (e). Panels, b, c, and d have identical scaling. The residual maps (panels a and e) both use plot ranges from -5σ to 5σ .

like scale radius and b_n in the case of a Sérsic profile (see Sect. 2.2.3). Figure 3.11 shows an example of such a third-order profile in comparison to an exponential disc profile. Evidently, the third-order profile can overcome two essential problems of pure Sérsic profiles, since it (a) exhibits a central cusp, which is also observed in real galaxies, and, (b) approaches zero for increasing radii faster than the pure Sérsic profile. Therefore, we may speculate that such higher-order Taylor expansions could provide a reasonable generalisation, if deviations from the normal Sérsic profile are observed while azimuthal structures are still absent. For instance, this may help to describe the light profiles of elliptical galaxies or unbarred S0 galaxies.

In Fig. 3.12 we compare the performances of the third-order profile and the Sérsic profile by fitting an elliptical galaxy from the SDSS database. Clearly, the residual map of the third-order profile is almost perfectly random noise whereas the residual map of the Sérsic profile reveals systematic mismodelling. Correspondingly, the ratio of χ^2 -values of the third-order profile over Sérsic fit is $\chi_3^2/\chi_S^2 \approx 0.48$. In fact, the third-order profile fits the data so well that we need to check that it is not an overfit. For this purpose, Fig. 3.13 displays

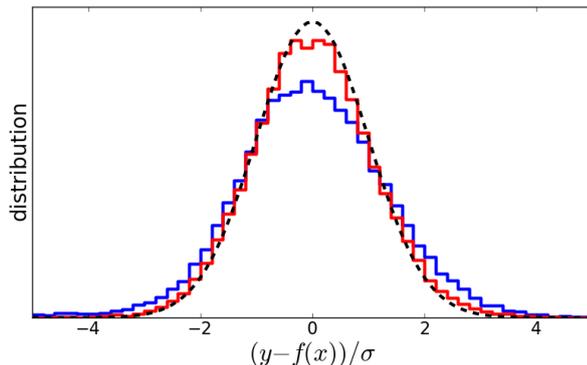


Figure 3.13: Model comparison via distributions of normalised residuals.

The blue histogram shows the distribution of normalised residuals from the Sérsic fit. The red histogram are the normalised residuals from the third-order profile. The dashed line is a unit Gaussian.

the distributions of normalised residuals for both models in comparison to the unit Gaussian. Evidently, the normalised residuals of the Sérsic profile have a broader distribution than the unit Gaussian, i.e., the Sérsic fit is an underfit. The normalised residuals of the third-order profile are closer to the unit Gaussian, i.e., they are closer to the truth. However, their distribution does *not* peak sharper than the unit Gaussian, i.e., the third-order profile is not an overfit.

3.7.2 Numerical orthonormalisation

Third-order profiles such as Eq. (3.55) can be orthonormalised, too. Unfortunately, it is not possible to do this orthonormalisation analytically. The reason is that the substitution step in Sect. 3.2.4 does not work anymore. In order to understand this, let us consider the scalar product

$$\begin{aligned} \langle l|k \rangle &= \int_0^\infty dr r R_l(r) R_k(r) \exp \left[-2e^{A+B \log r + C \log^2 r + D \log^3 r} \right] \\ &= \int_{-\infty}^\infty du R_l(e^u) R_k(e^u) \exp \left[2u - 2e^{A+Bu+Cu^2+Du^3} \right], \end{aligned} \quad (3.56)$$

where we already substituted $u(r) = \log r$. In order to bring the weight function into the generic form e^{-v} , we have to substitute

$$v(u) = 2e^{A+Bu+Cu^2+Du^3} - 2u. \quad (3.57)$$

However, substitution now requires to solve *analytically* for u as a function of v , which is plainly impossible in this case as shown in Fig. 3.14. Therefore, the orthonormalisation has to be performed numerically.

The brute-force Gram-Schmidt algorithm starts from a set of linearly independent but not orthogonal vectors. In our case, these are the radial monomials $(r^0, r^1, r^2, r^3, \dots)$. Using the definition of the scalar product according to Eq. (3.19), the algorithm then reads:

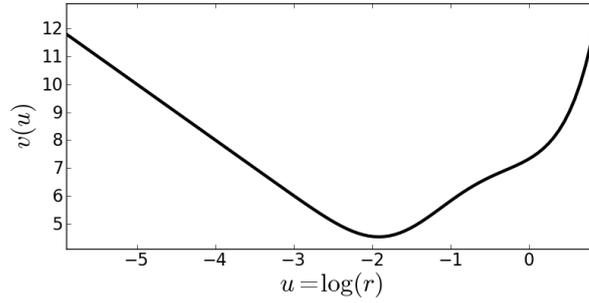


Figure 3.14: Non-invertibility of substitution variable.

We plot $v(u)$ as given by Eq. (3.57) for the profile parameters of Fig. 3.11. Clearly, $v(u)$ is not injective, i.e., there is no unique solution of u for a given v . Consequently, $v(u)$ is not bijective, i.e., it is not invertible.

Algorithm 3.1 *Brute-force Gram-Schmidt.*

- Step 1: Initialise the ground state $|0\rangle = \frac{|r^0\rangle}{\langle r^0|r^0\rangle}$.
 Step 2: Iterate from $l = 1$ to ∞ ,

$$|l\rangle = |r^l\rangle - \sum_{k=0}^{l-1} \langle k|r^l\rangle |k\rangle \quad (3.58)$$

and normalise such that $\langle l|l\rangle = 1$.

This algorithm creates a set of orthonormal basis functions. However, simple to understand as this algorithm may be, it is highly inefficient from a computational point of view. We therefore chose the following three-term recurrence relation:

Algorithm 3.2 *Three-term recurrence relation.*

- Step 1: Initialise the ground state $|0\rangle = \frac{|r^0\rangle}{\langle r^0|r^0\rangle}$.
 Step 2: Compute $|1\rangle = (\hat{r} - \langle 0|\hat{r}|0\rangle)|0\rangle$ and normalise.
 Step 3: Iterate from $l = 2$ to ∞ ,

$$|l\rangle = (\hat{r} - \langle l-1|\hat{r}|l-1\rangle)|l-1\rangle - \langle l-1|\hat{r}|l-2\rangle|l-2\rangle. \quad (3.59)$$

and normalise such that $\langle l|l\rangle = 1$.

Here, we used the expectation of the operator \hat{r} over the radial basis functions B_k and B_l

$$\langle k|\hat{r}|l\rangle = \int_0^\infty dr r B_k(r) r B_l(r). \quad (3.60)$$

In order to compute a basis set up to maximum order L , the first algorithm requires the evaluation of $\frac{1}{2}L(L+1)$ projection integrals and $L+1$ normalisation integrals. Conversely, the second algorithm also requires the evaluation of $L+1$ normalisation integrals but only $2L-1$ projection integrals. First, this has the trivial advantage of being computationally faster.²⁶ Second, and more importantly, the three-term recurrence relation is numerically more stable than

²⁶For example, let us consider $L = 12$ where the first algorithm requires the evaluation of 91 integrals, whereas the second algorithm only involves 36 integrations.

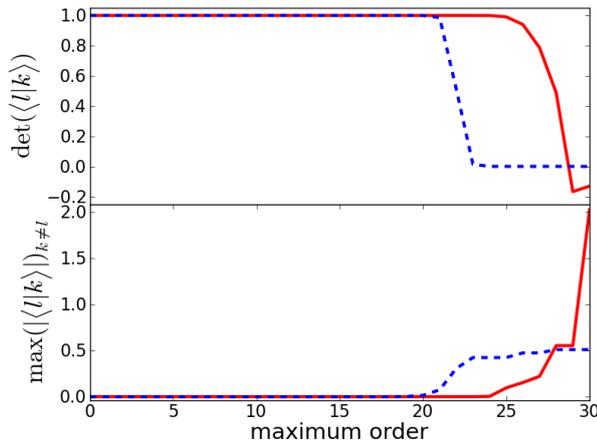


Figure 3.15: Performance of numerical orthonormalisation algorithms.

We compare the brute-force Gram-Schmidt algorithm (dashed line) and the three-term recurrence relation (solid line). In the top panel, we show the determinant of the projection matrix, which should be unity. In the bottom panel, we show the maximum absolute value of the off-diagonal elements, which should be zero.

the first algorithm. In the first algorithm, the computation of a single basis function requires numerous integrations such that numerical errors quickly accumulate. We demonstrate this by investigating the projection matrix

$$p_{lk} = \langle l|k \rangle \quad (3.61)$$

for orthonormalisations of different maximum orders using both algorithms. If the basis set is indeed orthonormal, this projection matrix should be an identity matrix. For the profile parameters of Fig. 3.11 and maximum radial orders up to $L = 30$, Fig. 3.15 compares the performances of both algorithms. Evidently, for maximum orders larger than 20, the brute-force Gram-Schmidt algorithm breaks down whereas the three-term recurrence relation is still numerically stable up to maximum order of 25. When exactly both algorithms break down depends on the parameters of the weight function. Nevertheless, the three-term recurrence relation is generally more stable and also computationally faster, so it is the preferred method.

3.7.3 Revisiting the problems of sérsiclets

The conceptual problem of sérsiclets was the postulated fixed relation of steepness of the weight function and scale of polynomial oscillations that is not obeyed by real galaxies. This fixed relation is loosened by the third-order profiles. Of course, the Sérsic index – or rather $B = \frac{1}{n_S}$ – still has an influence on the oscillation scale of the radial polynomials. However, now there are two further model parameters C and D which also influence the polynomials. Consequently, basis functions based on third-order profiles are much more flexible and do not impose such a rigid relation as sérsiclets do.

The practical problem of sérsiclets was undersampling, in particular for Sérsic indices $n_S \geq 1$. In fact, Fig. 3.11 already suggests that third-order

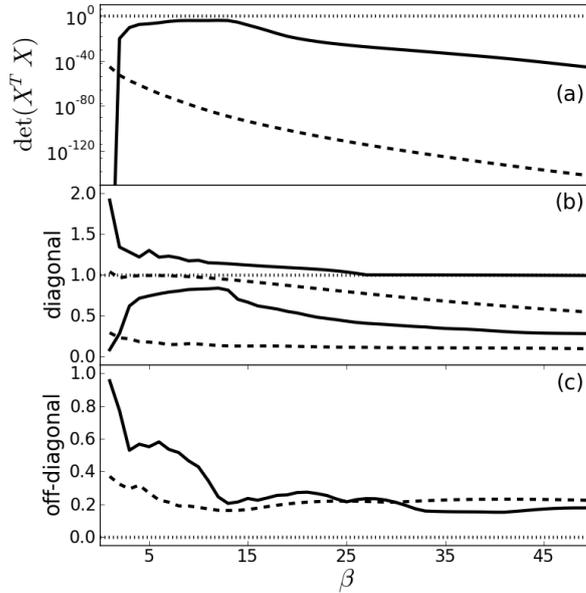


Figure 3.16: Undersampling test for third-order profiles.

We compare basis sets of the Sérsic (dashed lines) and third-order profile (solid lines) shown in Fig. 3.11. The test is performed on a 100×100 pixel grid and the weight function has no intrinsic ellipticity. In contrast to Fig. 3.8, we have chosen $b = 4 - 1/3$ in this case. Panel a: Determinant of projection matrix $X^T \cdot X$ which should be 1. Panel b: Largest and smallest diagonal elements of $X^T \cdot X$ which should be 1. Panel c: Largest absolute value of off-diagonal elements of $X^T \cdot X$ which should be 0.

profiles can overcome this problem. The undersampling problem of sérsiclets stems from the fact that the weight function is strongly peaked in the centre for $n_S \geq 1$, thereby concentrating the weight and squeezing the polynomials to rapid oscillations. Conversely, third-order profiles can exhibit central cusps instead of peaks. Therefore, the weights are not as highly concentrated and the polynomials do not oscillate as rapidly. We demonstrate this by repeating the orthonormality test of Fig. 3.8 for the third-order profile with parameters given in Fig. 3.11 and its set of basis functions. In direct comparison to sérsiclets with $n_S = \frac{1}{B} = 2$, the orthonormalised third-order profile indeed suffers less strongly from undersampling, which is most obvious for the determinant of $X^T \cdot X$. In particular, we can see that there appears to be a regime $5 \leq \beta \leq 15$ where undersampling is roughly overcome and boundary truncation has not yet set in. Such a regime does not exist for the corresponding set of sérsiclets shown here for comparison. Consequently, third-order profile can indeed overcome the undersampling problem of sérsiclets. Nevertheless, there are parameter choices for third-order profiles where the undersampling problem is as bad as for sérsiclets, e.g., when $C = D = 0$ and the third-order profile reduces to a Sérsic profile. However, by choosing appropriate priors it might be possible to exclude such parameter values in order to avoid undersampling.

3.7.4 Computational feasibility

We have shown that orthonormalisations of third-order profiles can indeed overcome the limitations of sérsiclets while preserving their mathematical justification. However, we have not yet mentioned a serious practical limitation of this novel approach: The numerical orthonormalisation of third-order profiles is computationally highly expensive. Fitting a galaxy on a, say, 100×100 pixel grid while freely adjusting all model parameters is completely infeasible on a standard computer. A problematic work-around would be to first fit a simple third-order profile to the galaxy and then only orthonormalising the best-fit profile in order to model the data’s deviation from the mean profile. However, this step-wise approach by construction is very unlikely to find the best-fitting basis-function expansion. Fitting a third-order profile does *not* provide a “mean” profile whereof the basis functions could account for the “deviations” from the mean profile. Instead, it provides a *biased* profile. Another work-around would be to set up a library of basis functions for all realistic parameter values of third-order profiles, such that numerical orthonormalisation has not to be conducted “on-the-fly” during the fit. This approach is hampered by the fact that third-order profiles have three free parameters such that a decently detailed sampling of this three-dimensional parameter space – which is necessary in order to provide reliable fits – would produce an extensive library. The best solution is certainly to maintain the numerical orthonormalisation during the fit and to employ Graphics Processing Units (GPUs) instead of normal CPUs. For such pure “number crunching” like numerical orthonormalisation, GPUs impressively outperform CPUs (Fluke et al. 2011).

3.8 Assumptions

Resuming our critical assessment of assumptions in Sect. 2.2.5, how do sérsiclets and orthonormalised higher-order profiles fare in comparison to other parametrisation schemes? Both approaches are basis-function expansions, i.e., they can handle azimuthal structures and do not need to assume a radial profile like, e.g., the concentration index. As discussed above, sérsiclets assume that the steepness of the radial profile is tightly correlated with the scales of substructures and that the galaxy is well resolved. This first assumption is absolutely unrealistic, while the second is certainly problematic in practice. Orthonormalised third-order profiles cure both of these problems and effectively make no assumptions here. However, both methods assume that their radial weight functions are realistic. In contrast to other basis-function approaches such as shapelets (Réfrégier 2003) or Chebyshev rational functions (Jiménez-Teja & Benítez 2011), this assumption is well justified since these radial profiles are the Taylor expansions of the true yet unknown light profiles of galaxies (Sects. 3.2.3 and 3.7.1). Consequently, orthonormalised third-order profiles effectively involve only a single assumption that is backed up by mathematics. In fact, they would present an excellent example of good scientific methodology, if they did not also involve the “pseudo-assumption” that excessive computational power is available.

— The bottom line —

- Basis-function expansions are designed to faithfully represent the azimuthal structure of galaxies, e.g., enabling us to infer the handedness of spiral-arm patterns of disc galaxies in order to estimate their sense of rotation.
- Being the first-order Taylor expansion of *any* radial profile, the Sérsic profile is a better match to real galaxies than the Gaussian profile, i.e., sérsiclets by design overcome the limitations of shapelets. Furthermore, this provides a mathematical justifications to orthonormalise the Sérsic profile, whereas shapelets are a pure ad-hoc construction.
- From general Sérsic profiles we can only build *polar* basis functions because the ground states of Cartesian basis functions do not exhibit the azimuthal symmetry required in the context of galaxy morphologies.
- The polar sérsiclet basis functions can be computed analytically.
- Polar sérsiclets postulate a fixed relation between steepness of the weight function and spatial scale of polynomial oscillations that is not obeyed by real galaxy morphologies. This will give rise to modelling biases which are not generally predictable.
- Polar sérsiclets suffer strongly from undersampling effects on Cartesian pixel grids. For steep weight functions, polar sérsiclets require incredibly high resolution in order to maintain orthogonality and linear independence. In particular, this problem corrupts estimates of spiral-arm handedness based on sérsiclet expansions of observed spiral galaxies.
- Orthonormalisations of higher-order Taylor expansions overcome the undersampling problem and loosen the tight coupling of steepness of the weight function and scale of polynomial oscillations. However, their numerical orthonormalisation is computationally expensive. This renders this method as computationally infeasible.

Part II

Disc Alignment

4

Disc alignment in the Local Group

As we want to investigate disc alignment, the Local Group provides a natural first testbed. The Milky Way Galaxy is not an isolated field galaxy but part of a galaxy group – the Local Group. Apart from numerous dwarf galaxies, the Local Group consists of four disc galaxies, namely the Milky Way, Andromeda (M31), M33, and the Large Magellanic Cloud (LMC), all with pairwise distances of less than 1Mpc. Therefore, if theory predicts disc alignment on spatial scales of 1Mpc, we might well use the Local Group as a simple test case. However, given the limited diagnostic power of only four disc galaxies, this is meant more as an introduction and motivation for our later analysis of much larger sample of disc galaxies.

4.1 Angular-momentum-orientation vector of the Milky Way

Figure 4.1 shows an artist’s impression of the Milky Way. In order to estimate the angular-momentum-orientation vector of the Milky Way, we need two ingredients:

1. The unit vector \vec{r}_{\odot} pointing from the Galactic centre to the position of the Sun.
2. The unit vector \vec{v}_{\odot} of the Sun’s velocity on its trajectory around the Galactic centre.

From these two ingredients, and the well justified assumption that the Sun’s orbit lies inside the Galactic disc, we can then compute the Milky Way’s angular-momentum-orientation vector via

$$\vec{L}_{\text{MW}} = \vec{r}_{\odot} \times \vec{v}_{\odot} . \quad (4.1)$$

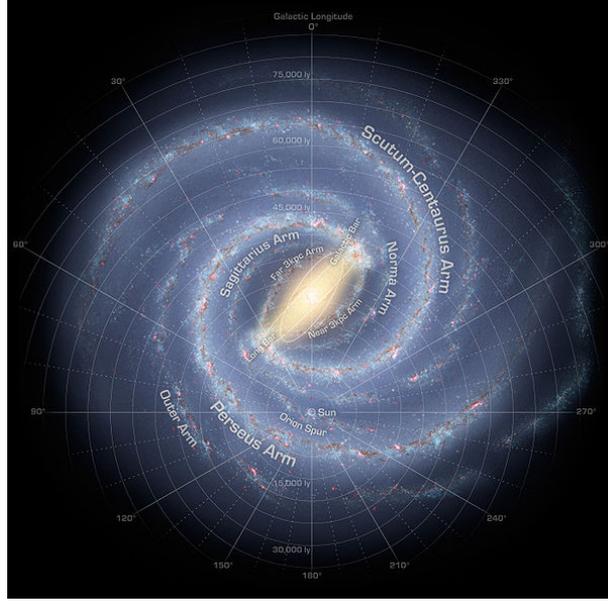


Figure 4.1: Artist's impression of the Milky Way.

The Hubble type of the Milky Way is SBc, i.e., it is a barred spiral with a less prominent bulge in its centre and rather open spiral arms. Source: NASA public domain.

We can infer \vec{r}_\odot from the equatorial coordinates of the Galactic centre, $\alpha_{\text{GC}} \approx 266.416833^\circ$ and $\delta_{\text{GC}} \approx -29.007806^\circ$. These are the coordinates of the radio source Sagittarius A, which marks the position of the central supermassive black hole of the Milky Way (e.g. Shen et al. 2005). Here, we have to keep in mind that these coordinates are pointing from the Sun towards the Galactic centre, i.e., \vec{r}_\odot is the inverted direction,

$$\vec{r}_\odot = - \begin{pmatrix} \cos \alpha_{\text{GC}} \sin(90^\circ - \delta_{\text{GC}}) \\ \sin \alpha_{\text{GC}} \sin(90^\circ - \delta_{\text{GC}}) \\ \cos(90^\circ - \delta_{\text{GC}}) \end{pmatrix}. \quad (4.2)$$

Note that this vector is defined in a global coordinate system (Sect. 6.2.3). This coordinate system is centred at the Earth's position and – apart from its radial coordinate – its two spherical angles are $\varphi = \alpha$ and $\vartheta = 90^\circ - \delta$.²⁷ Such a global coordinate system is necessary in order to compare angular-momentum-orientation vectors of different galaxies. The unit vector \vec{v}_\odot has to be inferred from the rotation of the Galactic disc. B. Lindblad and J. H. Oort investigated the rotation of stars in the Galactic disc in 1926/27 and their results can be found in any astronomy textbook. By definition, \vec{v}_\odot points into the direction specified by Galactic longitude $\ell = 90^\circ$ and Galactic latitude $b = 0^\circ$ (e.g. Brunthaler et al. 2005). In equatorial coordinates, this direction is given by $\alpha_v \approx 318.004563^\circ$ and $\delta_v \approx 48.329553^\circ$, such that

$$\vec{v}_\odot = \begin{pmatrix} \cos \alpha_v \sin(90^\circ - \delta_v) \\ \sin \alpha_v \sin(90^\circ - \delta_v) \\ \cos(90^\circ - \delta_v) \end{pmatrix}. \quad (4.3)$$

²⁷This means the xy -plane coincides with the equatorial plane and the x -direction points towards a right ascension of $\alpha = 0^\circ$. The z -direction points towards the North Pole where the declination angle is $\delta = 90^\circ$.

Inserting these values into Eq. (4.1), we obtain the following estimate of the angular-momentum-orientation vector of the Milky Way,

$$\vec{L}_{\text{MW}} \approx \begin{pmatrix} 0.86771 \\ 0.19878 \\ -0.45560 \end{pmatrix}. \quad (4.4)$$

We conduct two cross-checks: First, the two unit vectors \vec{r}_{\odot} and \vec{v}_{\odot} should be orthogonal and indeed their scalar product is $\vec{r}_{\odot} \cdot \vec{v}_{\odot} \approx 0.00097$. Second, \vec{L}_{MW} by construction should be normal to the plane of the Galactic disc, i.e., it should be parallel or antiparallel to the unit vector pointing into the direction of the Galactic North pole

$$\vec{u}_{\text{NP}} = \begin{pmatrix} \cos \alpha_{\text{NP}} \sin(90^\circ - \delta_{\text{NP}}) \\ \sin \alpha_{\text{NP}} \sin(90^\circ - \delta_{\text{NP}}) \\ \cos(90^\circ - \delta_{\text{NP}}) \end{pmatrix}, \quad (4.5)$$

whose equatorial coordinates are given by $\alpha_{\text{NP}} \approx 192.859508^\circ$ and $\delta_{\text{NP}} \approx 27.128336^\circ$. Indeed, the scalar product is $\vec{L}_{\text{MW}} \cdot \vec{u}_{\text{NP}} \approx -0.9999992$, i.e., both vectors are almost perfectly antiparallel. If the Sun's orbit did not lie in the plane of the Galactic disc, this projection value would be a highly unlikely coincidence.

4.2 Angular-momentum-orientation vectors of Andromeda, M33, and the LMC

In order to estimate the angular-momentum orientations of Andromeda, M33, and the LMC, we use the formalism described in Lee (2011) which is based on ellipticity estimates and the assumption of intrinsically circular galactic discs. The details of this formalism are discussed later in Sect. 6.2.

4.2.1 Andromeda (M31)

We now estimate the angular-momentum-orientation vector of Andromeda (M31). Figure 4.2 displays a GALEX Near-UV image of the Andromeda galaxy. We adopt an inclination angle of 77° (Walterbos & Kennicutt 1988) and an orientation angle of 38° (Walterbos & Kennicutt 1987). Furthermore, dust lanes in the foreground of stellar emission enable us to identify the front-edge of Andromeda's galactic disc, as indicated in Figure 4.2. Our front-edge estimate agrees with the result of Iye & Ozawa (1999) who investigated the reddening of globular clusters as a function of height above the major axis. Given its equatorial coordinates $\alpha_{\text{M31}} \approx 10.684708^\circ$ and $\delta_{\text{M31}} \approx 41.268750^\circ$, the angular-momentum-orientation vector of Andromeda then reads

$$\vec{L}_{\pm} \approx \pm \begin{pmatrix} -0.08031 \\ -0.79651 \\ 0.59926 \end{pmatrix}. \quad (4.6)$$

As Andromeda does not exhibit spiral arms, we cannot break the \pm -degeneracy this way. However, Chemin et al. (2009) published spatially

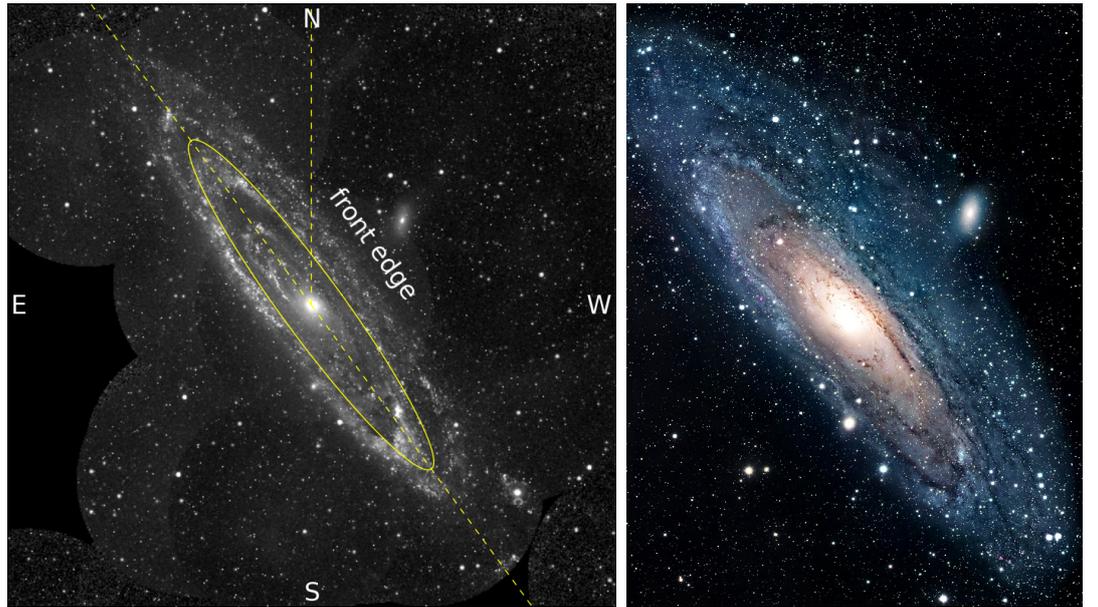


Figure 4.2: GALEX Near-UV and false-colour image of Andromeda. Left panel: The estimated orientation angle is $\theta \approx 38^\circ$ and the inclination angle is $\approx 77^\circ$, as indicated by the yellow dashed lines. Source: SkyView, NASA. Right panel: The false-colour image enables us to estimate the front-edge via identification of dust lanes. Source: NASA.

resolved HI spectra of M31, which enable us to infer the disc rotation directly. Their Figs. 5 and 6 and in particular their velocity field of Fig. 8 clearly show that the North-Eastern part of the disc has larger radial velocity, whereas the South-Western part of the disc has smaller radial velocities. This directly implies that the North-Eastern part is receding from us, whereas the South-Western part is rotating towards us. Consequently, the angular-momentum-orientation vector of Andromeda points South-East in Fig. 4.2 and into the plane of the paper away from the reader. Therefore, if we project \vec{L}_{M31} onto the unit direction vector pointing from the Milky Way towards Andromeda, this projection must be positive. We can understand this by considering the rotation near the front-edge. There, the velocity vector points towards North-East, such that $\vec{L} = \vec{r} \times \vec{v}$ points into the plane of the paper and thus away from the Milky Way. Given the position vector of Andromeda, his condition selects the “-” in Eq. (4.6) such that the angular-momentum-orientation vector of Andromeda is given by

$$\vec{L}_{M31} \approx \begin{pmatrix} -0.08031 \\ -0.79651 \\ 0.59926 \end{pmatrix}. \quad (4.7)$$

Hence, the angular-momentum-orientation vector of Andromeda is fully known like in the case of the Milky Way.

4.2.2 Triangulum Galaxy (M33)

Concerning M33, we adopt an inclination angle of 49° and an orientation angle of 21° (Corbelli & Schneider 1997). M33 clearly is a right-handed (Z-wise)

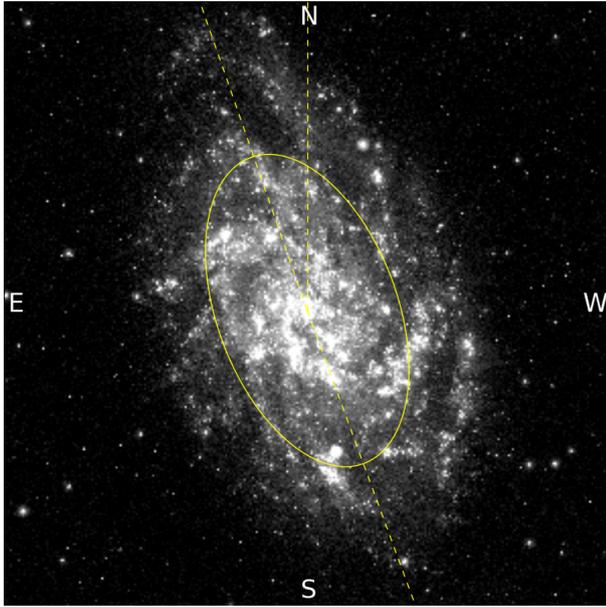


Figure 4.3: GALEX Near-UV image of M33.

The estimated orientation angle is $\theta \approx 21^\circ$ and the inclination angle is $\approx 49^\circ$, as indicated by the yellow dashed lines. There are no obvious dust lanes such that we cannot estimate the front-edge. Source: SkyView, NASA.

spiral. This rotational sense agrees with the results of Brunthaler et al. (2005) who observed the proper motion of two H_2O masers in M33. It also agrees with the results of Putman et al. (2009), who measured the radial-velocity field of HI gas in M33. Again, this implies that the projections of both possible front-edge configurations of \vec{L}_{M33} onto the unit direction vector pointing from the Milky Way towards M33 have to be positive. Unfortunately, M33 does not exhibit dust lanes, such that the front-edge cannot be identified this way. This is not surprising since M33 is not as highly inclined as Andromeda such that we are less likely to observe a dust lane. From dust reddening of C-rich AGB stars Cioni et al. (2008) concluded that there is weak evidence that the North-Western side of M33 is the front-edge. Assuming this front-edge and given its equatorial coordinates, $\alpha_{\text{M33}} \approx 23.46^\circ$ and $\delta_{\text{M33}} \approx 30.66^\circ$, the angular-momentum-orientation vector of M33 then reads

$$\vec{L}_{\text{M33}} \approx \begin{pmatrix} 0.67170 \\ -0.47655 \\ 0.56721 \end{pmatrix}. \quad (4.8)$$

However, the front-edge estimate of Cioni et al. (2008) is still rather uncertain (see their Fig. 9).

4.2.2.1 M33 front-edge estimation using parallaxes

Given the distance of M33 of $\approx 0.8\text{Mpc}$ to the Milky Way, what is the required angular resolution to measure parallax distances of individual stars? Given a baseline of 2 astronomical units and the distance of M33, the parallax angle

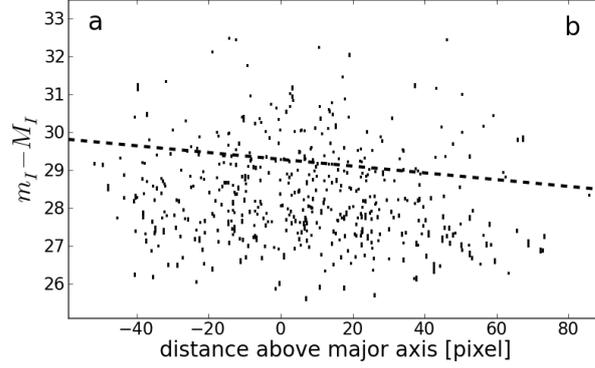


Figure 4.4: Front-edge estimation of M33 via Cepheid distances.

We plot distance modulus $m_I - M_I$ over height above or below the major axis of M33 for the 563 Cepheids of Pellerin & Macri (2011). Error bars indicate errors in apparent I -band magnitudes, assuming that Cepheid periods and therefore absolute I -band magnitudes have negligible errors. The two front-edge configurations a and b are indicated. The dashed line corresponds to a linear fit whose slope is $b = -(8.94 \pm 8.92) \cdot 10^{-3} \text{pixel}^{-1}$ marginalised over 100 bootstrap samplings.

is given by

$$\delta \approx \sin \delta = \frac{1 \text{AU}}{0.8 \text{Mpc}} \approx 1.3 \cdot 10^{-6} \text{arcsec}. \quad (4.9)$$

Clearly, this is way beyond the resolution limits of existing instruments such as the HUBBLE Space telescope ($\approx 80 \cdot 10^{-3} \text{arcsec}$) or HIPPARCOS ($\approx 2 \cdot 10^{-3} \text{arcsec}$). It is also beyond the planned resolution limit of the GAIA satellite ($\approx 20 \cdot 10^{-6} \text{arcsec}$). Consequently, it is currently impossible to estimate the front-edge of M33 directly via parallaxes.

4.2.2.2 M33 front-edge estimation using Cepheid distances

As parallax estimates are limited by resolution, we try to estimate the front-edge of M33 using Cepheid distances from the DIRECT project (Macri et al. 2001) and its recent extension by Pellerin & Macri (2011). This database contains 563 Cepheids in M33 with equatorial coordinates, period estimates P , and estimates of the apparent HST I -band magnitude m_I . In the absence of metallicity dependence, Freedman et al. (2001) defined the following relation between absolute I -band magnitude and Cepheid period in units of days

$$M_I = -(2.962 \pm 0.020)(\log P - 1) - (4.904 \pm 0.010). \quad (4.10)$$

Given the absolute magnitude and an observed apparent magnitude, we can then employ the distance modulus

$$m_I - M_I = -2.5 \log_{10} \left(\frac{F(d)}{F(10 \text{pc})} \right) = 5 \log_{10} \left(\frac{d}{10 \text{pc}} \right) \quad (4.11)$$

which increases with distance because m_I increases. There is no need to correct the distance modulus for *foreground* dust extinction in this case, since all

Cepheids are affected equally due to their very similar equatorial coordinates. If we plot the distance modulus $m_I - M_I$ vs. the height above or below the major axis of M33, we should observe a slope which enables us to estimate the front-edge. The side with larger distance modulus is further away from the Milky Way. Figure 4.4 shows that the distance moduli exhibit a very large scatter, though the cloud of data points exhibits a weak “tail” in the lower right regime. Nevertheless, there is very weak evidence – barely 1σ – for a negative slope, favouring the front-edge configuration b , i.e., North-West in Fig. 4.3. Although this is not a reliable result, it demonstrates that a front-edge estimation via Cepheids is indeed feasible. However, more Cepheid variables would be required in order to reduce the uncertainties. Due to the geometry of the front-edge degeneracy, observing a small number of Cepheids at very large heights above and below the major axis may indeed suffice to place decisive constraints. Moreover, our results using Cepheid distances point into the same direction as results obtained by Cioni et al. (2008) from carbon-rich AGB stars. As aforementioned, also Cioni et al. (2008) find only weak evidence for a North-Western front-edge (see their Fig. 9). However, the agreement of our and their results is sufficient for this exercise.

4.2.3 Large Magellanic Cloud (LMC)

Concerning the LMC, we adopt an inclination angle of 35° and an orientation angle of 123° (van der Marel & Cioni 2001). Furthermore, van der Marel & Cioni (2001) find clear evidence that the North-Eastern side of the disc is the front-edge (their Fig. 5). The rotational sense of the LMC is right-handed as is evident from observed velocity fields (e.g. Olsen & Massey 2007). Given its equatorial coordinates $\alpha_{\text{LMC}} \approx 80.8938^\circ$ and $\delta_{\text{LMC}} \approx -69.7561^\circ$, the angular-momentum-orientation vector of the LMC then reads

$$\vec{L}_{\text{LMC}} \approx \begin{pmatrix} -0.29699 \\ -0.46945 \\ -0.83152 \end{pmatrix}. \quad (4.12)$$

4.3 Hypothesis test of random orientation

Are the angular-momentum-orientation vectors in the Local Group compatible with the null hypothesis of random orientation? In order to test this, we investigate the distribution of projection values using a KS-test (e.g. Press et al. 2002). For the four disc galaxies, there can only derive three statistically independent projections. We choose the projections onto the Milky Way:

- $\vec{L}_{\text{MW}} \cdot \vec{L}_{\text{M31}} \approx -0.5010$
- $\vec{L}_{\text{MW}} \cdot \vec{L}_{\text{M33}} \approx +0.2297$
- $\vec{L}_{\text{MW}} \cdot \vec{L}_{\text{LMC}} \approx +0.0278$

Adding further projection values, e.g., $\vec{L}_{\text{M31}} \cdot \vec{L}_{\text{M33}}$, would introduce correlations, thereby compromising the KS-test. These three projection values provide an unbiased estimate of the *cumulative* distribution of projection values, as shown in Fig. 4.6. The null hypothesis of random orientation predicts that

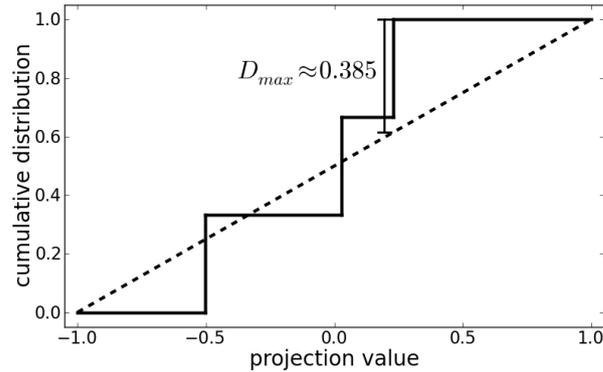


Figure 4.5: KS-test of randomly oriented angular momenta in the Local Group. Step function: Empirical (unbiased) estimate of cumulative distribution for the Local Group. Dashed line: Cumulative distribution of null hypothesis of random orientations. The maximum KS-distance is $D_{\max} \approx 0.385$.

all projection values are equally likely, i.e., its cumulative distribution is a straight line also shown in Fig. 4.6. The maximum vertical distance between empirical and predicted cumulative distribution is then $D_{\max} \approx 0.385$. For three projection values, this yields a p -value of ≈ 0.648 (Press et al. 2002). Consequently, with 64.8% probability we make a mistake if we reject the null hypothesis of randomly oriented angular-momentum-orientation vectors in the Local Group.

How do these results change if we only consider the spin *axes*, ignoring the rotational sense? In this case, we have to investigate the cumulative distribution of absolute values of the projections given above. Again, the null hypothesis of random orientation predicts a uniform distribution of projected spin axes, i.e., its cumulative distribution is again a straight line. Both cumulative distributions are shown in Fig. 4.6. The maximum vertical distance is now $D_{\max} \approx 0.499$, which yields a p -value of ≈ 0.32 . This still cannot be considered as evidence against the null hypothesis.

4.4 The next step

This simple test case showed no evidence for disc alignment in the Local Group. Our results are consistent with randomly oriented angular momenta. However, we analysed the data of only four disc galaxies. We could only have detected such alignment effects, if they had been very strong. Consequently, the next step is to investigate a larger sample of disc galaxies. Such a sample also enables us to refine the search for alignment effects. Instead of just comparing pairwise angular-momentum projections, we can estimate angular-momentum correlations and their dependence on galaxy-galaxy separation. In other words, using a large sample of disc galaxies, we can test disc alignment via the autocorrelation function of angular-momentum-orientation vectors. This will be the major part of this thesis and is described in Chapter 6.

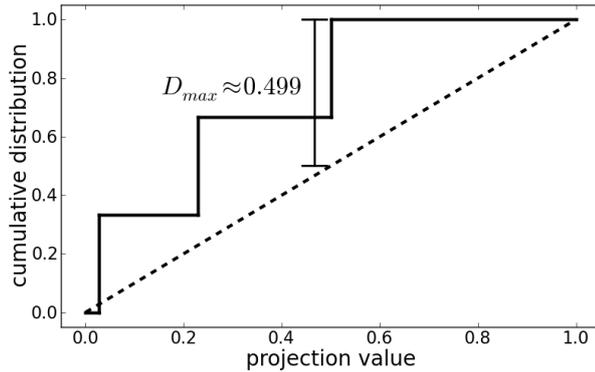


Figure 4.6: KS-test of randomly oriented spin axes in the Local Group. Step function: Empirical (unbiased) estimate of cumulative distribution for the Local Group. Dashed line: Cumulative distribution of null hypothesis of random orientations. The maximum KS-distance is $D_{\max} \approx 0.499$.

The bottom line

- For Milky Way, Andromeda (M31) and the Large Magellanic Cloud, the angular-momentum-orientation vectors are fully known. For M33, the angular-momentum-orientation vector is only known up to the two-fold front-edge degeneracy, which cannot be resolved decisively.
- Given the projection values of angular-momentum-orientation vectors, the null hypothesis of random orientations has a p -value of 0.65. If we consider only spin axes, neglecting rotational sense, the null hypothesis of random orientations has a p -value of 0.32.
- Concerning the Local Group, there is no evidence that the hypothesis of random orientation is incorrect. However, this does not rule out the presence of disc alignment. A more elaborate investigation is necessary, using a large sample of disc galaxies.

CHAPTER 4. DISC ALIGNMENT IN THE LOCAL GROUP

5

The data

In the previous chapter, we have seen that a large sample of disc galaxies is required in order to detect potential alignment effects. This means, a large database of galaxies including a morphological classification is needed. In this chapter, we present the data we are using for our investigation of disc alignment. As we want to investigate alignment effects in two ways – for general disc galaxies and specialised for spiral-arm handedness – corresponding morphological classifications have to be accessible. We justify our choice of data and present its details. Finally, we briefly explain the details of the cosmological model used to infer distances from cosmological redshifts.

5.1 Choice of data set

In order to investigate disc alignment in the galaxy population, we require a large data sample that contains a sufficient number of disc galaxies. Furthermore, theory predicts that angular-momentum vectors should be correlated on distance scales of approximately 1Mpc (Schäfer & Merkel 2011). Therefore, the galaxy sample has to cover a decently large volume in space. Finally, investigations of disc alignment are compromised by weak gravitational lensing at redshifts $z > 0.3$ or even earlier (Crittenden et al. 2001), i.e., the galaxy sample has to be restricted to the local universe. These three requirements have to be satisfied.

We have already met the Cosmic Evolution Survey (COSMOS) galaxy sample in Sect. 2.3.1. Is the COSMOS database suitable for such an investigation? First, the COSMOS survey contains 31,288 galaxies with valid morphological parametrisations but not all of them are disc galaxies and many lie beyond $z = 0.3$. Second, the survey area is a square of only 2 square degrees. At a maximum allowed redshift of $z = 0.3$ having an angular-diameter distance of $\approx 900\text{Mpc}/h$, this survey area corresponds to a square with edge length of only $\approx 20\text{Mpc}/h$. In comparison to the theoretically predicted correlation length of $1\text{Mpc}/h$, this box size is rather small and allows only for a little amount of disc galaxies. We clearly prefer a survey with much larger field of

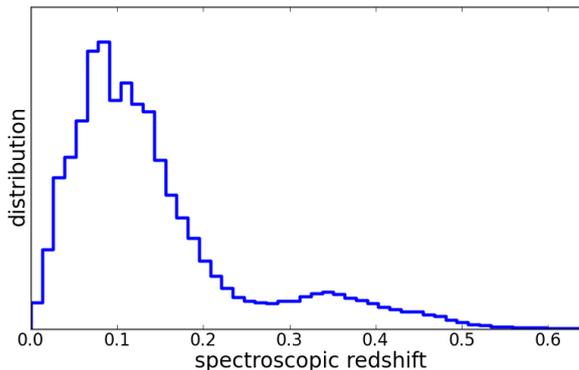


Figure 5.1: Distribution of galaxy redshifts in SDSS DR 7.

We show the distribution of 929,483 galaxies from the spectroscopic sample. These objects are taken from the `SPECOBJALL` table and have been classified as galaxies.

view. We conclude that the COSMOS survey is not suitable for investigations of disc alignment.

The Sloan Digital Sky Survey (SDSS) covers roughly one fourth of the full sky and contains approximately one million galaxies with spectroscopic redshifts (Data Release 7, Abazajian et al. 2009). In contrast to the COSMOS survey which is deep but small, the SDSS survey is wide and “shallow”. The distribution of galaxy redshifts is shown in Fig. 5.1. The SDSS database contains 818,722 galaxies with spectroscopic redshift $z < 0.3$. Consequently, the SDSS database satisfies all our requirements and we therefore use this catalogue for our investigation of disc alignment in the local universe.

Initially, we envisioned to use the PanSTARRS survey for our investigation. While SDSS covers only one quarter of the sky, the PanSTARRS survey is planned to be a fully-sky survey. Furthermore, PanSTARRS at its final depth will be able to detect somewhat fainter objects than SDSS. All in all, we can expect that the PanSTARRS catalogue will contain at least three times more galaxies than the SDSS catalogue.²⁸ Unfortunately, the PanSTARRS survey was delayed and the first test measurements suffered from severe problems in the data-reduction pipeline. Therefore, we decided not to wait for PanSTARRS but to opt for SDSS. Moreover, in the course of this work it turned out that PanSTARRS cannot be used for investigations of disc alignment because it only provides photometric redshift estimates. Photometric redshifts are not accurate enough for this task (see Sect. 7.3).

²⁸The Milky Way covers roughly one quarter of the sky such that PanSTARRS cannot detect galaxies in this area.

5.2 Sloan Digital Sky Survey

Using SDSS data, we have to acquaint ourselves with the relevant technical characteristics of the SDSS instruments to some extent.

5.2.1 Telescope and instrument

The Sloan Digital Sky Survey (SDSS; York et al. 2000; Abazajian et al. 2009, DR 7) was one of the first large sky surveys, commenced in the year 2000 and

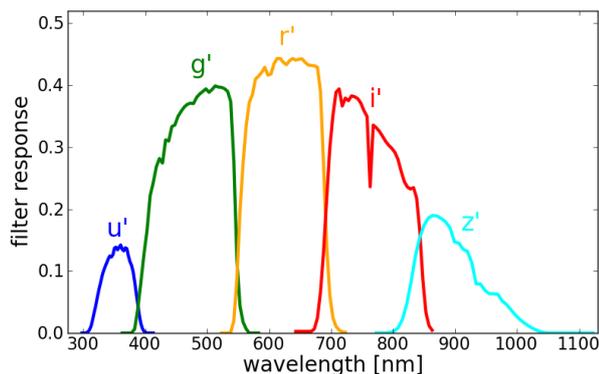


Figure 5.2: SDSS filter response functions.

finished in 2008 (SDSS-II). The SDSS covered an area of 8,400 square degrees and contains imaging and spectra of 930,000 galaxies, 120,000 quasars, and 225,000 stars. The survey used the 2.5m telescope at the Apache Point Observatory, New Mexico, USA, to observe the Northern Galactic Cap and along the equator. The telescope uses a 1m secondary mirror and its instruments are set up in Cassegrain focus which allows a short and compact design. However, in contrast to a standard Cassegrain telescope with concave-parabolic primary mirror and convex-hyperbolic secondary mirror, primary and secondary mirrors are shaped differently in order to enlarge the field of view and provide in-focus imaging.²⁹ The 120-megapixel CCD camera was able to detect objects to a flux limit of $r \approx 22.5\text{mag}$, and spectroscopic observations of galaxies were possible down to $r \approx 17.77\text{mag}$. The camera has five bands (u' , g' , r' , i' , and z'), whose filter response functions are shown in Fig. 5.2. We note that the u' and z' bands have very poor response, i.e., these bands are very noisy compared to the other three bands. The camera is operated in drift mode, i.e., the telescope shutter remains open and the telescope in fixed positioning such that the observed objects drift over the CCD from West towards East during the night. The CCD is continuously read out synchronised such that pixel columns are read out continuously while the object drifts from one to the next pixel column.

²⁹<http://cas.sdss.org/dr7/en/sdss/telescope/telescope.asp>

5.2.2 Imaging data and meta-information

SDSS imaging data is publicly accessible.³⁰ However, the imaging data is provided in a way that is far from convenient or ready-to-use. One possibility is to download whole imaging plates. The tasks of locating the desired target object, e.g., a disc galaxy, and making a cutout is “outsourced” to the user. Downloading the full imaging plates is very time-consuming and requires a fast internet connection and lots of hard-disk memory, in particular since we are interested in imaging data of numerous galaxies. Conversely, the NASA/IPAC infrared science archive³¹ offers a web interface to download object cutouts for given coordinates. However, this tool is limited to 20 objects per query and is therefore also inappropriate to download imaging data of thousands of galaxies, which is necessary for our investigation. Nevertheless, we used this web

³⁰<http://www.sdss.org/>

³¹<http://irsa.ipac.caltech.edu/Missions/sdss.html>

interface to download imaging data of the 500 largest galaxies in our sample.

The SDSS database also provides a wealth of additional information apart from pure imaging data. The access to these additional information is public, too, but this time it is conveniently provided by a MySQL-database interface. In particular, the SDSS table `GALAXY` provides us with Petrosian radii and ellipticity estimates based on isophotes and second moments of the galaxy’s light distribution. Furthermore, the SDSS table `SPECOBJALL` contains spectroscopic redshift estimates and its errors for many SDSS galaxies.

5.3 Galaxy Zoo

For reasons discussed in Chapters 2 and 3, we are unsatisfied with the parametrisation of galaxy morphologies for the task of automated classification. Therefore, we resort to visual classification by human beings in order to separate disc galaxies from other galaxy types. Such a catalogue of visual classifications of galaxy morphologies is provided by the Galaxy Zoo project.

5.3.1 Project description

³²<http://zoo1.galaxyzoo.org/>

The Galaxy Zoo project³² provides visual classifications of approximately 900,000 galaxies from the SDSS spectroscopic sample (Lintott et al. 2008, 2011), whose redshift distribution is shown in Fig. 5.1. This enormous effort has been accomplished with the help of the internet community. The crucial idea of the Galaxy Zoo project was to show galaxy images to non-expert internet users and let them classify these objects according to a simple pre-defined classification scheme. Every object has been classified by more than a single person, reaching 39 votes on average for all 900,000 galaxies (Land et al. 2008). In fact, 98% of the sample have at least 20 votes per galaxy (Bamford et al. 2009). These multiple votes per galaxy provide a direct estimate of the galaxy’s class probabilities. In detail, the Galaxy Zoo database provides probabilistic assignments to the following morphological classes:

- elliptical galaxy, $p_{\text{ell}}^{\text{GZ}}$, including S0 galaxies,
- disc galaxy, $p_{\text{disc}}^{\text{GZ}}$,
- edge-on disc galaxy, $p_{\text{edge}}^{\text{GZ}}$,
- clock-wise/Z-wise spiral galaxy, p_{Z}^{GZ} ,
- anti-wise/S-wise spiral galaxy, p_{S}^{GZ} ,
- merger, $p_{\text{mg}}^{\text{GZ}}$.

Of particular interest are the handedness classifications of spiral-arm patterns, p_{Z}^{GZ} and p_{S}^{GZ} . All probabilities that are taken from Galaxy Zoo are denoted with a “GZ” superscript. The normalisation is given by

$$p_{\text{ell}}^{\text{GZ}} + p_{\text{disc}}^{\text{GZ}} + p_{\text{edge}}^{\text{GZ}} + p_{\text{Z}}^{\text{GZ}} + p_{\text{S}}^{\text{GZ}} + p_{\text{mg}}^{\text{GZ}} = 1. \quad (5.1)$$

5.3.2 How to operate Galaxy Zoo

The classification provided by Galaxy Zoo offers probabilistic object-to-class assignments which naturally account for the uncertainty in the classification. As we argued in Andrae et al. (2010b), this is the only meaningful and clean way to classify galaxy morphologies. Our line of argument was that hard “either–or” object-to-class assignments are inappropriate because, first, observational effects wash out differences between classes creating class overlap, and, second, that also theoretically galaxy evolution should produce overlapping morphological classes.³³ In simple words, galaxy morphologies *cannot* be clearly assigned to morphological types in general – apart from singular prototypical examples of very obvious morphology. Nevertheless, hard assignments are easier to carry out and interpret, wherefore many astronomers have a natural affinity to this approach. Sometimes, astronomers are so used to hard-cut classifications that even the probabilistic assignments of Galaxy Zoo have been converted into hard cuts, which is inappropriate for above reasons (e.g. Slosar et al. 2009; Lee 2011). Even Bamford et al. (2009) themselves do not fully understand their own work when they claim that the probabilities “cannot provide classifications for individual objects”, instead recommending to threshold or cut the probabilities for this purpose.³⁴ A probabilistic object-to-class assignment is *all we can get*. As we demonstrated in Andrae et al. (2010b), cutting probabilistic classifications in the presence of overlap can lead to substantial biases. We address this problem in Sects. 6.4.4.2 and 6.4.5.2.

5.3.3 Criticism

The major point of criticism usually raised against Galaxy Zoo is the fact that the visual classification has been performed by non-experts. However, that is really not an argument at all for two reasons: First, Lintott et al. (2008, 2011) made an effort to weight the contributions of every internet user according to hers/his performance in comparison to all other users. This reduces the influence of individual persons misclassifying objects. Second, a single expert classifier might indeed produce better results than a single non-expert, but a group of non-expert users eventually also produces highly accurate results. This combination of many “weak learners” to form a “strong learner” is well known from *ensemble learning* (e.g. Hastie et al. 2009), a subfield of machine learning which has created several modern and highly successful learning algorithms.³⁵

A real criticism of Galaxy Zoo is that the project is neither reproducible nor scalable. The project was an enormous success because it quickly gathered the interest of the internet community. However, as quickly as the internet community can gain interest in something, it can also lose this interest very rapidly. Therefore, it appears highly questionable whether it would be possible to repeat such an exercise, e.g., with a different or refined classification scheme. This brings us to the second problem: The success of Galaxy Zoo is not scalable, meaning it is impossible to do something similar with a galaxy sample 10 times as large as SDSS. Although the internet community is very powerful, classifying 900,000 galaxies took already more than two years to complete. We dare say that the Galaxy Zoo project is an impressive mark placed right at

³³For instances, consider bulged disc galaxies or galaxy mergers.

³⁴Such a cut would only be justified, if the classification probabilities were biased and the nature of this bias was understood. However, if such a bias is understood, it will be better to correct the probabilities instead of cutting them (see Sect. 5.5.1).

³⁵This principle is also known as “wisdom of crowds”. Some quiz shows even exploit it via “audience jokers”.

the limit where visual classification becomes infeasible.

5.4 Classifications by Huertas-Company et al. (2011)

Similar to the Galaxy Zoo project, Huertas-Company et al. (2011) performed a morphological classification on the SDSS spectroscopic galaxy sample.³⁶ There are two important differences with respect to Galaxy Zoo. First, instead of visual inspection, a support-vector machine, i.e., an automated classification algorithm, has been used in order to classify the galaxies. The classification was based on several morphological and non-morphological parameters, namely colours, isophotal axis ratio, goodness of de Vaucouleur fit, and concentration index. The classification algorithm was trained on a sample of 2,253 SDSS galaxies visually classified by Fukugita et al. (2007). Although based on a problematic parameter like concentration index (see Sects. 2.3, 2.4, and 2.5), Huertas-Company et al. (2011) can show that the resulting probabilistic galaxy-to-class assignments are robust and reliable. In particular, their classifications are in good agreement with the Galaxy Zoo results, as far as both classification schemes are comparable. Second, the classification scheme is slightly different from Galaxy Zoo. The morphological classes are:

- elliptical, $p_{\text{ell}}^{\text{HC}}$,
- S0 galaxy, $p_{\text{S0}}^{\text{HC}}$,
- Sab disc galaxy, $p_{\text{Sab}}^{\text{HC}}$,
- Scd disc galaxy, $p_{\text{Scd}}^{\text{HC}}$, including irregulars.

Of particular interest is the classification of Scd disc galaxies, $p_{\text{Scd}}^{\text{HC}}$. All probabilities taken from the catalogue of Huertas-Company et al. (2011) are denoted with a superscript “HC”. The normalisation reads

$$p_{\text{ell}}^{\text{HC}} + p_{\text{S0}}^{\text{HC}} + p_{\text{Sab}}^{\text{HC}} + p_{\text{Scd}}^{\text{HC}} = 1. \quad (5.2)$$

As mentioned in Huertas-Company et al. (2011), Sect. 3.1 therein, the “Scd” class not only contains Scd galaxies but also irregular galaxies, since there were not enough irregulars in the training sample to make up their own class.

5.5 Data selection

Not all objects in the two catalogues presented above can be used for our analysis. Some objects have to be removed for various reasons. We now describe the data selection for this investigation of disc alignment. Our investigation essentially aims at two different approaches, namely looking for statistically significant autocorrelations of spiral-arm handedness and angular-momentum-vector orientations, respectively. Separate galaxy samples are required for both tasks.

³⁶http://gepicom04.obspm.fr/sdss_morphology/Morphology_2010.html

5.5.1 Handedness sample

A fully automated handedness estimation as outlined in Sect. 3.3.5 is not possible, due to the severe limitations of sérsiclets as discussed in Sect. 3.5.4. Fortunately, Galaxy Zoo provides us with a classification of the winding sense of spiral-arm patterns in disc galaxies. However, Land et al. (2008) reported a bias in the handedness classifications, p_Z^{GZ} and p_S^{GZ} , where more spiral galaxies are classified as S-wise than as Z-wise.³⁷ This bias is corrected in an asymmetric, additive fashion by Land et al. (2008) and Slosar et al. (2009) in order to enforce that the proportions of Z-wise and S-wise spirals are equal with regard to the whole sample. In contrast to this, we employ a *symmetric*, additive bias correction of the form

$$p_Z = p_Z^{\text{GZ}} + b \quad \text{and} \quad p_S = p_S^{\text{GZ}} - b, \quad (5.3)$$

where b is chosen such that the numbers of Z-wise and S-wise spirals are approximately identical. There are two reasons:

1. The symmetric correction preserves the normalisation of Eq. (5.1). This is important because in contrast to Slosar et al. (2009) we are handling the Galaxy Zoo results fully probabilistically in our analysis (cf. Sect. 6.4.2).
2. Demanding that the proportions of Z-wise and S-wise spirals are equal only provides a single condition, such that an asymmetric correction with two biases, b_Z and b_S , is not fully constrained and therefore arbitrary.

Slosar et al. (2009) argued correctly that such a bias can only lead to a constant offset in the handedness autocorrelation function, but it cannot feign a distance-dependent autocorrelation, which is the desired astrophysical signal.

First, starting from the complete Galaxy Zoo sample, we select a subsample galaxies with either $p_Z^{\text{GZ}} \geq 0.778$ or $p_S^{\text{GZ}} \geq 0.8$, which results in 36,999 galaxies. This preselection is necessary in order to minimise the number of galaxies for which we have to retrieve meta-information. We only *select* these objects but we do *not* cut their class probabilities. The probability thresholds are chosen this way in order to allow for some flexibility in the correction of the handedness bias. Like Slosar et al. (2009), we do not restrict the redshift in our selection.

Second, for all galaxies we tried to download the r -band Petrosian radii from the SDSS GALAXY table, the spectroscopic redshift estimate and its error estimate from the SDSS SPECOBJALL table. Actually, all objects in the Galaxy Zoo sample have been selected from the SDSS spectroscopic sample. Strangely though, we could not find 103 objects in the GALAXY table and another 5,106 objects were untraceable in the SPECOBJALL table.³⁸ This leaves us with 31,790 objects with r -band Petrosian radius and estimates of spectroscopic redshift and its error.

Third, we remove multiple objects from the sample, i.e., very extended galaxies that have been shredded by the SDSS pipeline producing multiple entries of a single object. Evidently, such multiple entries would compromise the handedness-correlation estimate by causing artificial positive correlations on

³⁷The exact origin of this bias remains unknown, though one option considered by Land et al. (2008) is a psychological effect.

³⁸Cross-matching was done by retrieving the SPECOBJID from the GALAXY table or – if this label was unavailable – by matching the given OBJID with the BESTOBJID from the SPECOBJALL table.

short distances. We automatically removed galaxy pairs whose angular separations were less than 1.5 times the maximum r -band Petrosian radius of both galaxies, by randomly dropping a single of the two partners. Furthermore, Slosar et al. (2009) removed another 69 objects through visual inspection. This list has been kindly provided by Anže Slosar such that we are capable of removing these objects, too. This leaves us with 31,621 galaxies.

Finally, we apply the additive and symmetric bias correction of the handedness classifications given by Eq. (5.3) to the 36,999 galaxies from the preselected sample. Naïvely interpreting any galaxy with $p_Z \geq 0.8$ as likely Z-wise spiral and any galaxy with $p_S \geq 0.8$ as likely S-wise spiral, we end up with 15,083 Z-wise and 15,071 S-wise spirals for a bias correction of $b = 0.0105$.³⁹ Therefore, our sample is slightly smaller than the one used by Slosar et al. (2009) (18,074 Z-wise spirals and 18,052 S-wise spirals, Land et al. 2008) for reasons unknown to us. The most likely explanation for this discrepancy is that Slosar et al. (2009) used a slightly different version of the SDSS database.

³⁹Note that the preselection $p_Z^{GZ} \geq 0.778$ or $p_S^{GZ} \geq 0.8$ was only intended to minimise the amount of objects requiring meta-information. The preselection was *not* our actual data selection.

5.5.2 Angular-momentum-orientation sample

Concerning correlations of angular-momentum-orientation vectors, one goal is to reproduce the results reported by Lee (2011). Therefore, we have to apply the same data selection criteria. Based on the catalogue of morphological classifications by Huertas-Company et al. (2011), we select those galaxies with spectroscopic redshifts $0 < z \leq 0.02$ and probability $p_{\text{Scd}}^{\text{HC}} > 0.5$ to be a galaxy of type Sc or Sd. The redshift cut ensures completeness, i.e., our sample contains virtually all Scd galaxies out to redshift $z \leq 0.02$.⁴⁰ This leaves us with 4,236 galaxies satisfying these criteria, the same number of objects as reported by Lee (2011). For 25 of these objects we could not find any information in the SDSS database, i.e., estimates of r -band Petrosian radii, Stokes parameters, their errors, and error estimates of spectroscopic redshift are missing. For these objects, we set the spectroscopic redshift error to 10^{-4} , which is a typical value for this sample. Petrosian radii are set to zero. Using the automated method described above, we find 20 rogue pairs of shredded galaxy parts in this sample. For each rogue pair, we randomly discard one of the two galaxies, such that we are left with a sample of 4,216 Scd galaxies.

⁴⁰The limiting r -band magnitude is approximately $r_{\text{max}} \approx 18$.

5.6 Cosmological model

For the analyses carried out in the next chapters, we have to infer cosmological distances from estimates of cosmological redshift. The sample of Scd galaxies with $z \leq 0.02$ is well in the linear regime of the simple Hubble law. However, the handedness sample contains galaxies from the whole SDSS spectroscopic sample whose redshifts also explore the nonlinear regime as is evident from Fig. 5.1. This requires the specification of a cosmological model because different cosmological models exhibit different space-time curvatures and therefore lead to different relations of redshift and distance.

We employ a flat Λ CDM cosmology with Dark-Energy density $\Omega_\Lambda = 0.734$ and Dark-Matter density $\Omega_m = 1 - \Omega_\Lambda$ (Larson et al. 2011). Furthermore, we use the standard parametrisation of Hubble's constant

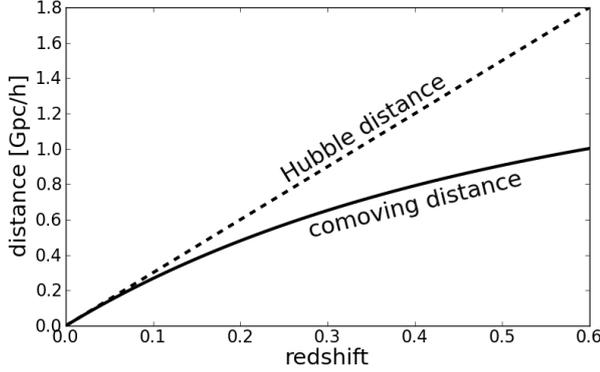


Figure 5.3: Comoving distance and linear Hubble distance.

The redshift range is the same as for the SDSS spectroscopic sample (Fig. 5.1).

$H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$. For a given spectroscopic redshift estimate, \hat{z} , we then estimate the comoving distance,

$$d(\hat{z}) = \frac{c}{H_0} \int_0^{\hat{z}} \frac{dz}{(1+z)^2 \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}}. \quad (5.4)$$

We use the comoving distance because it is the spatial distance at fixed time of two objects comoving with the cosmic expansion.⁴¹ Figure 5.3 shows the comoving distance as a function of redshift for the relevant redshift regime.

⁴¹As the concept of simultaneity does not exist in relativity, the fixed time is specified by the observer.

The bottom line

- For investigating disc alignment, a galaxy sample with numerous galaxies at cosmological redshift $z < 0.3$ is required. The Sloan Digital Sky Survey (SDSS) satisfies these requirements.
- Classifications of galaxy morphologies in SDSS are available through the Galaxy Zoo project and the work of Huertas-Company et al. (2011).
- We select a sample of 15,083 Z-wise and 15,071 S-wise spiral galaxies from Galaxy Zoo. Furthermore, we select a sample of 4,216 Scd galaxies from the catalogue of Huertas-Company et al. (2011).
- From spectroscopic redshifts we estimate comoving distances using a Λ CDM cosmology with parameters $\Omega_\Lambda = 0.734$, $\Omega_m = 1 - \Omega_\Lambda$ and $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$.

6

Measurements and results

In this chapter, the central results of testing disc alignment via autocorrelations of spiral-arm handedness and angular-momentum-orientation vectors are presented. We start with a brief review of previous work, motivating our own approach. Then, the details of inferring angular-momentum-orientation vectors from the data are explained. We proceed by rigorously investigating the impact of various sources of errors.

6.1 Previous work and motivation

Disc alignment explains angular-momentum acquisition of disc galaxies partially by the influence of environment on the collapsing dark-matter halo. Therefore, galaxies close to each other reside in haloes which were influenced by the same environment, thereby giving rise to autocorrelation effects on short distance scales. The search for such alignment effects is not new. For instances, Sugai & Iye (1995) and Land et al. (2008) investigated the distribution of spiral-arm handedness over the sky via multipole expansion, searching for signatures of disc alignment. More recently, Slosar et al. (2009) looked for autocorrelations of spiral-arm handedness in the Galaxy Zoo sample and reported the discovery of positive autocorrelations on distance scales of $0.5\text{Mpc}/h$ at a statistical significance level of $\approx 3\sigma$. Furthermore, positive autocorrelations of angular-momentum orientations at similarly low statistical significance levels have been reported, e.g., by Lee (2011). Unfortunately, the recent work by Slosar et al. (2009) and Lee (2011) did not take into account several important sources of errors, namely errors in redshift estimates, morphological classification and – in the case of angular-momentum orientations – ellipticity estimates. Consequently, these investigations systematically underestimate the errors of their correlation estimates, leading to overly optimistic estimates of statistical significance. In this chapter, we investigate the impact of these errors on the correlation estimates. The crucial question is whether we can still find statistical evidence for the existence of such correlations.

Why is it worth looking for such autocorrelations? Indeed, their discovery of such correlations could be regarded as important confirmation of disc

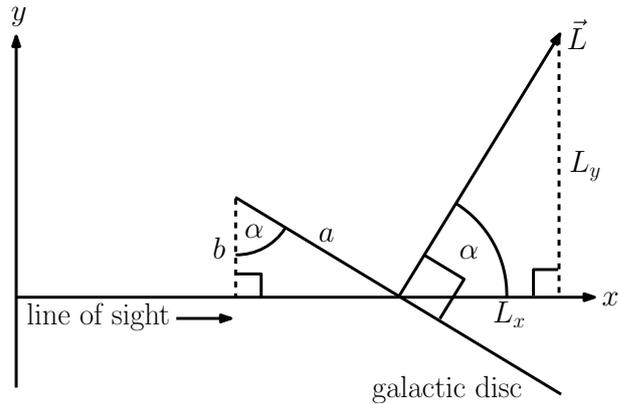


Figure 6.1: Local coordinate system for inferring angular-momentum orientation. Line of sight and x -direction coincide. Shown is a cross-section through the inclined galactic disc with apparent axis ratio $q = b/a$. The inclination angle α appears in both dashed triangles.

alignment. However, a non-detection could be explained by suitable values of theoretical parameters producing correlations which are too weak to be detectable with current data. Therefore, the only valid scientific approach is to use this experiment to constrain the theoretical parameters from the data, and to devise a second future test of the constrained theory. Consequently, the task is well defined: Estimate these autocorrelations in order to constrain theoretical parameters which enable us to devise a future more decisive empirical test.

6.2 From axis ratio to angular-momentum orientation

The orientation of the angular-momentum vector of a disc galaxy is not an observable. It has to be inferred from the observed galactic disc by invoking several assumptions. We now explain how to estimate the angular-momentum orientation from observables, as described in Lee & Erdogdu (2007) or Lee (2011). In fact, we used this formalism already earlier in Sect. 4.2. Let us emphasise that we can only infer the *orientation* of the angular-momentum vector but not its absolute value. The latter would require knowledge of the density distribution and precise dynamics of the galaxies in our sample.

6.2.1 Local Cartesian coordinates

Assuming a circular and infinitely thin disc, the observed axis ratio $q = b/a \leq 1$ enables us to directly estimate the disc's inclination angle α (cf. Fig. 6.1),

$$\cos \alpha = q. \quad (6.1)$$

We choose a local Cartesian coordinate system – “local” because it is different for every individual galaxy – where the line-of-sight (from us towards

the galaxy) points along the x -direction, while the semi-minor axis of the observed inclined disc points along the y -direction (cf. Fig. 6.1). The angular-momentum-orientation vector then reads

$$\vec{L}_{\text{loc}} = \begin{pmatrix} \pm \cos \alpha \\ \sin \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \pm q \\ \sqrt{1-q^2} \\ 0 \end{pmatrix}. \quad (6.2)$$

The \pm stems from the geometric degeneracy caused by our ignorance of which edge of the inclined disc is pointing towards us (the front edge). If we assign probabilities p_a and p_b to both possible configurations of the front edge, we can write

$$\vec{L}_{\text{loc}} = \begin{pmatrix} (p_a - p_b) \cos \alpha \\ \sin \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} (p_a - p_b)q \\ \sqrt{1-q^2} \\ 0 \end{pmatrix}, \quad (6.3)$$

where $p_a + p_b = 1$. If no estimate of the front edge is available, we set $p_a = p_b = 1/2$, such that \vec{L}_{loc} has no x -component on average.

6.2.2 Disc orientation

Apart from the inclination angle $\alpha \approx \arccos(q)$ the observed disc galaxy also has an orientation angle θ that can be estimated from the imaging data along with the axis ratio q . Therefore, we need to rotate Eq. (6.3) by θ around the x -axis, which yields

$$\begin{aligned} \vec{L}'_{\text{loc}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} (p_a - p_b) \cos \alpha \\ \sin \alpha \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} (p_a - p_b) \cos \alpha \\ \cos \theta \sin \alpha \\ \sin \theta \sin \alpha \end{pmatrix} = \begin{pmatrix} L'_{\text{loc},x} \\ L'_{\text{loc},y} \\ L'_{\text{loc},z} \end{pmatrix}. \end{aligned} \quad (6.4)$$

This means for an orientation angle of $\theta = 0$ the semi-major axis of the disc points into y -direction, whereas it points into z -direction for $\theta = +90^\circ$.

6.2.3 Global coordinate system

Next, we need to transform from the local coordinate system – where the line-of-sight points along the x -axis and which therefore differs from galaxy to galaxy as the line-of-sights are different – to a global coordinate system which is identical for all galaxies. This global coordinate system is chosen best in spherical coordinates (r, φ, ϑ) , where the azimuthal/longitude angle φ is identified with the right ascension, whereas the latitude angle ϑ is identified with 90° minus the declination. Then $L'_{\text{loc},x}$ always points into r -direction, while $L'_{\text{loc},y}$ points into φ -direction and $L'_{\text{loc},z}$ into ϑ -direction. Consequently, the angular-momentum vector in global coordinates is given by

$$\vec{L} = L'_{\text{loc},x} \vec{e}_r + L'_{\text{loc},y} \vec{e}_\varphi + L'_{\text{loc},z} \vec{e}_\vartheta, \quad (6.5)$$

where we have used the spherical unit vectors,

$$\vec{e}_r = \begin{pmatrix} \cos \varphi \sin \vartheta \\ \sin \varphi \sin \vartheta \\ \cos \vartheta \end{pmatrix}, \quad \vec{e}_\varphi = \begin{pmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{pmatrix} \quad \text{and}$$

$$\vec{e}_\vartheta = \begin{pmatrix} \cos \varphi \cos \vartheta \\ \sin \varphi \cos \vartheta \\ -\sin \vartheta \end{pmatrix}.$$

Following the notation of Lee (2011), we denote the two possible orientations of the front edge by \vec{L}_a and \vec{L}_b ,

$$\vec{L}_a = +L'_{\text{loc},x}\vec{e}_r + L'_{\text{loc},y}\vec{e}_\varphi + L'_{\text{loc},z}\vec{e}_\vartheta, \quad (6.6)$$

and

$$\vec{L}_b = -L'_{\text{loc},x}\vec{e}_r + L'_{\text{loc},y}\vec{e}_\varphi + L'_{\text{loc},z}\vec{e}_\vartheta = \vec{L}_a - 2(\vec{L}_a \cdot \vec{e}_r)\vec{e}_r, \quad (6.7)$$

which only differ in the sign of their radial components.

6.2.4 Thick-disc approximation

Equation (6.1) assumes circular and infinitely thin discs. In reality, however, disc galaxies are not infinitely thin. First, galactic discs have a non-zero scale height. Second, disc galaxies typically exhibit a central bulge. Therefore, disc galaxies are not infinitely thin even when viewed edge-on. If we want to make this inference more physical by allowing for circular thick discs, all we need to do is to change Eq. (6.1). Let p denote the thickness parameter of the disc, we can redefine (e.g. Lee & Erdogdu 2007; Lee 2011)

$$\cos^2 \alpha = \begin{cases} \frac{q^2 - p^2}{1 - p^2} & \text{if } q > p \\ 0 & \text{else} \end{cases}. \quad (6.8)$$

The value of p depends on morphological type. Haynes & Giovanelli (1984) quote values of p that continuously decrease from 0.23 down to 0.1 over the morphological sequence Sa, Sab, Sb, Sbc, Sc, Scd. In fact, Lee (2011) decided to restrict the investigations of angular-momentum-orientation correlations on Scd galaxies only, since those galaxies by definition do not exhibit prominent bulges. Consequently, Scd galaxies can provide an unbiased estimate of the disc ellipticity, depending on the ellipticity estimator as discussed in Sect. 7.1. For later purposes, we note that Heidmann et al. (1972) compared different estimates of the intrinsic axial ratios and find values between $p = 0.083$ and 0.145 for Scd galaxies.

6.2.5 Geometric degeneracies

We have already met the geometric degeneracy originating from the ambiguity of the front edge (cf. Fig. 6.2). This degeneracy can in principle be broken by searching for an asymmetric dust extinction, which would indicate the front edge (see Sect. 7.5).

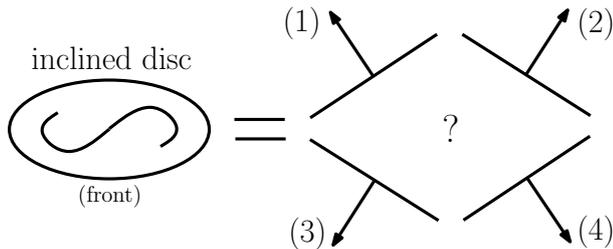


Figure 6.2: Geometric degeneracies in estimating the angular-momentum orientation. Observing only the inclined disc allows four different configurations. If we knew the lower edge of the inclined disc were the front edge pointing towards us, this would eliminate options (2) and (3). If we observed spiral arms with the sketched handedness, this would eliminate options (1) and (2) (“right-hand rule”, \vec{L} points along the right thumb). Knowledge of the front edge and handedness together would enable us to identify (4) as the only possible configuration in this case.

There is yet a second geometric degeneracy originating from our ignorance of the rotational sense. In fact, Eq. (6.5) only tells us the plane of rotation, such that $\pm\vec{L}$ are both possible options (cf. Fig. 6.2). This degeneracy can be broken by observing the rotational sense of the disc. One possibility to do this is to employ spatially resolved spectroscopy and study the distribution of radial velocities within the disc. However, obtaining spatially resolved spectroscopy for hundreds of thousands of disc galaxies is infeasible. In fact, the ambitious CALIFA project aims at assembling spatially resolved spectra of ≈ 600 galaxies (Sánchez et al. 2010). A less expensive approach is to look for spiral-arm patterns. We can infer the rotational sense of the disc from the winding sense of spiral arms, assuming that spiral arms are always trailing. We discuss this issue in further detail in the following section.

6.2.6 Leading spiral arms

As mentioned in the previous section, inferring the rotational sense of a disc galaxy from the handedness explicitly assumes that all spiral arms are trailing arms. In fact, Pasha (1985) found that 4 out of 190 spiral galaxies have leading arms.⁴² As counting the number of leading spiral arms is a Bernoulli trial and its outcome is given by a binomial distribution, this yields a maximum-likelihood estimate of $\hat{f} = 4/190 \approx 2.1\%$ for the fraction of leading arms. The 99% cumulative confidence is given by $f_{99} \approx 6.0\%$, i.e., there is a 1% chance that the true fraction of leading spiral arms is larger than 6.0%. Similarly, the 1% cumulative confidence is $f_1 \approx 0.7\%$, i.e., there is a 1% chance that the true fraction of leading spiral arms is smaller than 0.7%.⁴³

What is the impact of a small fraction of leading spiral arms on the estimator given by Eq. (6.24)? Slosar et al. (2009) showed that it leads to a constant multiplicative bias of

$$b = 4 \left(f - \frac{1}{2} \right)^2, \quad (6.9)$$

which stems from the fact that for a pair of galaxies a bias in the handedness-

⁴²In fact, all four objects with leading arms were found to dynamically interact with companions, wherefore Pasha (1985) argued that leading arms may be induced by interactions.

⁴³These confidence levels have been estimated by numerical integration of the beta distribution without any approximation (Cameron 2010).

correlation estimate only appears if exactly one spin was randomly reversed, but not if both were reversed. Consequently, this effect cannot feign a distance-dependent correlation. Furthermore, a multiplicative bias does *not* interfere with an assessment of statistical significance in a detected signal because the multiplicative bias also applies to the error of the correlation estimate and therefore cancels out. As $b < 1$, the possibility of leading spiral arms causes us to underestimate the correlation while the true correlation is actually larger. In order to assess the impact of leading arms, we may use the two extremes f_1 and f_{99} , which yield $b_1 \approx 0.97$ and $b_{99} \approx 0.77$, respectively.

6.3 Correlation estimators

We now introduce the general correlation estimators for angular-momentum-orientation vectors and handedness. We also explain how to estimate errors in general. We start by explaining the general formalism and then specialise on both angular-momentum orientations and handedness.

6.3.1 General correlation estimator

Given two random variates X and Y , we want to estimate their correlation ξ_{XY} and its error. If N pairs $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ have been drawn from (X, Y) , a general correlation estimator is given by,

$$\hat{\xi}_{XY} = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle, \quad (6.10)$$

where the hat on ξ_{XY} indicates an estimator and the expectation values are approximated by sample means,⁴⁴

$$\langle XY \rangle = \frac{1}{N} \sum_{n=1}^N x_n y_n, \quad (6.11)$$

$$\langle X \rangle = \frac{1}{N} \sum_{n=1}^N x_n \quad \text{and} \quad \langle Y \rangle = \frac{1}{N} \sum_{n=1}^N y_n. \quad (6.12)$$

Strictly, Eq. (6.10) defines a covariance estimator instead of a correlation estimator. In order to obtain a true correlation estimator, we actually need to normalise Eq. (6.10) by the factor $\sqrt{(\langle X^2 \rangle - \langle X \rangle^2)(\langle Y^2 \rangle - \langle Y \rangle^2)}$. However, this is usually not done in practice and Eq. (6.10) is nevertheless referred to as an estimate of the correlation of X and Y .

6.3.2 Error estimation

Merely obtaining a value of the correlation $\hat{\xi}_{XY}$ via Eq. (6.10) alone is not informative in any way. A meaningful result also requires an error estimate of $\hat{\xi}_{XY}$, e.g., in order to assess the confidence of the correlation value. $\hat{\xi}_{XY}$ is the *mean* of the random variate

$$Z = (X - \langle X \rangle)(Y - \langle Y \rangle). \quad (6.13)$$

⁴⁴Expectation values are integrals over probability distributions. Mean values are averages over samples drawn from probability distributions. Their equality only holds for large data samples.

If we were interested in estimating the random variate Z , we would employ its mean $\langle Z \rangle$ and its variance $\langle Z^2 \rangle - \langle Z \rangle^2$. However, in this case we are not interested in estimating Z but in estimating the *mean* of Z and the variance of $\hat{\xi}_{XY} = \langle Z \rangle$ equals the variance of Z divided by the number of samples drawn from Z . We can understand this very easily assuming that Z were subject to a Gaussian distribution – for the sake of argument. If we are given N samples z_1, z_2, \dots, z_N drawn from Z we would estimate the mean μ of Z via minimising χ^2 ,

$$\frac{\partial \chi^2}{\partial \mu} = \frac{\partial}{\partial \mu} \sum_{n=1}^N \frac{(z_n - \mu)^2}{\sigma_Z^2} = -2 \sum_{n=1}^N \frac{z_n - \mu}{\sigma_Z^2} = 0, \quad (6.14)$$

where σ_Z denotes the Gaussian standard deviation of Z . Solving the last equation for μ yields the maximum-likelihood estimate of the mean of Z ,

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N z_n, \quad (6.15)$$

which corresponds to our correlation estimate of Eq. (6.10). An error estimate $\hat{\sigma}_\mu$ of $\hat{\mu}$ is then given by a Fisher analysis (e.g. Heavens 2009). Taking the second derivative of χ^2 with respect to μ , we obtain

$$\frac{1}{\hat{\sigma}_\mu^2} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \mu^2} = \sum_{n=1}^N \frac{1}{\sigma_Z^2} = \frac{N}{\sigma_Z^2}, \quad (6.16)$$

and therefore

$$\hat{\sigma}_\mu^2 = \sigma_Z^2 / N \quad (6.17)$$

or rather $\hat{\sigma}_\mu = \sigma_Z / \sqrt{N}$.⁴⁵ In this particular case, we assumed that Z were subject to a Gaussian distribution such that for fitting the linear mean value there is no approximation involved. In practice, Z is usually not Gaussian. However, the central-limit theorem ensures that the likelihood of the mean $\hat{\xi}_{XY} = \langle Z \rangle$ is approximately Gaussian if the number N of samples drawn from Z is large enough. This guarantees that Eq. (6.17) is a valid approximation also in the more general case. Consequently, for the correlation estimate of Eq. (6.10) we obtain an error estimate of

$$\hat{\sigma}(\hat{\xi}_{XY}) = \frac{\hat{\sigma}((X - \langle X \rangle)(Y - \langle Y \rangle))}{\sqrt{N}}. \quad (6.18)$$

Here, we assume that N is large enough such that the likelihood function of the mean $\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$ is approximately Gaussian and we are allowed to take the square-root of the variance in order to obtain a standard deviation “ σ ”.

6.3.3 Angular-momentum orientation

Our aim is to estimate the scalar two-point autocorrelation function of angular-momentum orientations, $\hat{\xi}_{LL}(r)$. Here, we assume spherical symmetry such

⁴⁵Loosely speaking, if we draw more samples from Z , the distribution of Z does not change, in particular its width (variance) stays constant. However, drawing more samples from Z enables us to estimate the mean of the distribution more accurately.

that $\hat{\xi}(\vec{r}) = \hat{\xi}(r)$. This is a first-order approximation because the spatial distribution of galaxies in the universe is not isotropic on short scales (“Cosmic Web”). Usually, the following estimator is employed (e.g. Pen et al. 2000; Lee 2011),

$$\begin{aligned} \hat{\xi}_{\text{LL}}(r) = & \langle p_a p'_a |\vec{L}_a \cdot \vec{L}'_a|^2 \rangle + \langle p_a p'_b |\vec{L}_a \cdot \vec{L}'_b|^2 \rangle \\ & + \langle p_b p'_a |\vec{L}_b \cdot \vec{L}'_a|^2 \rangle + \langle p_b p'_b |\vec{L}_b \cdot \vec{L}'_b|^2 \rangle - \frac{1}{3}, \end{aligned} \quad (6.19)$$

where primes indicate the data of the second galaxy in the pair and subscripts a, b denote the two possible orientations of the disc’s front edge with probabilities p_a and p_b . The constant $\frac{1}{3}$ is subtracted because $\vec{L}_{a/b} \cdot \vec{L}'_{a/b} = \cos \alpha$ where α is a random angle. Hence, $\langle (\vec{L}_{a/b} \cdot \vec{L}'_{a/b})^2 \rangle$ equals

$$\langle \cos^2 \alpha \rangle = \frac{\int_{-1}^1 \cos^2 \alpha \, d \cos \alpha}{\int_{-1}^1 d \cos \alpha} = \frac{1}{3}. \quad (6.20)$$

Introducing the abbreviation $Z = (p_a p'_a |\vec{L}_a \cdot \vec{L}'_a|^2 + p_a p'_b |\vec{L}_a \cdot \vec{L}'_b|^2 + p_b p'_a |\vec{L}_b \cdot \vec{L}'_a|^2 + p_b p'_b |\vec{L}_b \cdot \vec{L}'_b|^2)$, an error estimate of $\hat{\xi}_{\text{LL}}(r)$ is given by,

$$\hat{\sigma}(\hat{\xi}_{\text{LL}}) = \frac{\hat{\sigma}(Z)}{\sqrt{N}}, \quad (6.21)$$

where N denotes the number of galaxy pairs in the relevant distance bin. The division by \sqrt{N} stems from the fact that the correlation is the mean of the random variate Z and not Z itself.⁴⁶

⁴⁶The subtracted constant $\frac{1}{3}$ has no impact on the error estimate.

6.3.4 Handedness

We also want to estimate the two-point autocorrelation function of handedness $\hat{\xi}_{\text{HH}}(r)$. Again assuming spherical symmetry, a general estimator is given by,

$$\hat{\xi}_{\text{HH}}(r) = \langle h h' \rangle, \quad (6.22)$$

where we have defined the handedness

$$h = p_Z - p_S, \quad (6.23)$$

based on Eq. (5.3). As explained in Sect. 5.5.1, the mean handedness of galaxies is zero in the whole sample, i.e., $\langle h \rangle = \langle h' \rangle = 0$. Handedness alignments cannot change this in individual distance bins, if the number of galaxy pairs is large enough.

For the moment, let us assume that the spiral-arm handedness is known precisely, i.e., $h = \pm 1$. In every distance bin, n_+ denotes the number of galaxy pairs with $h h' = +1$ and n_- the number of galaxy pairs with $h h' = -1$. We can then rewrite Eq. (6.22) to read

$$\hat{\xi}_{\text{HH}}(r) = \frac{n_+ - n_-}{n_+ + n_-} = f_+ - f_- = 2f_+ - 1, \quad (6.24)$$

where $f_{\pm} = n_{\pm}/(n_+ + n_-)$ denotes the fraction of galaxy pairs with positive or negative handedness products, respectively. An error estimate of $\hat{\xi}_{\text{HH}}(r)$ is obtained from the fact that counting positive handedness products is a Bernoulli trial, i.e., n_{\pm} are subject to the binomial distribution while f_{\pm} are subject to the beta distribution (e.g. Cameron 2010).

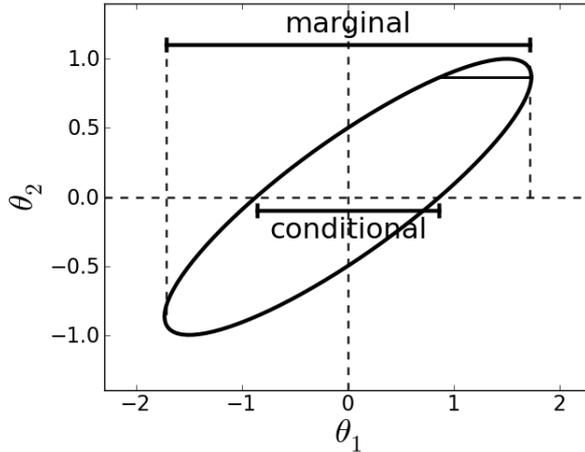


Figure 6.3: A sketch of conditional and marginal errors.

Shown is a Gaussian error ellipse for two correlated parameters θ_1 and θ_2 . Slicing at $\theta_2 = 0$, we obtain a conditional error for θ_1 of ≈ 0.86 . Furthermore, slicing at $\theta_2 = 0.87$, we obtain another conditional error for θ_1 of only ≈ 0.42 . Projecting the Gaussian ellipse onto the θ_1 -axis, we obtain a marginal error for θ_1 of ≈ 1.718 . This sketch has been inspired by a similar figure in Press et al. (2002).

6.4 The impact of errors

As we are going to demonstrate in this section, previous estimates of correlations in handedness and angular-momentum orientations neglected the most important sources of errors. Consequently, their results are overly optimistic concerning the statistical significance of the detections. First, we introduce some important statistical terminology. Second, we discuss the additional sources of errors that have to be taken into account. Finally, we investigate their impact on estimates of handedness correlations and angular-momentum-orientation correlations.

6.4.1 Conditional vs. marginal errors

Our criticism of earlier investigations of alignment effects via correlations functions (e.g. Slosar et al. 2009; Lee 2011) can be formulated as: Previous estimates are only conditional instead of marginal estimates. Consequently, we now need to briefly explain the conceptual difference between conditional and marginal errors.

For the sake of simplicity, let us consider fitting data D with Gaussian noise using a model with two linear parameters θ_1 and θ_2 . In this case, the likelihood function $\mathcal{L}(D|\theta_1, \theta_2)$ is a bivariate Gaussian whose 2×2 covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (6.25)$$

can be estimated from a Fisher analysis (e.g. Heavens 2009). Here, σ_1 is the standard deviation of this Gaussian if sliced at the optimal value of θ_2 and

vice versa, whereas $-1 \leq \rho_{12} \leq 1$ is the correlation coefficient. Therefore, σ_1 is the *conditional* error of θ_1 , “conditional” because it depends on where the Gaussian has been sliced, i.e., the optimal value of θ_2 . Conversely, the *marginal* error of θ_1 is independent of the value of θ_2 . This marginal error is obtained by projecting the bivariate Gaussian onto the θ_1 -axis, instead of slicing it. Marginal errors are never smaller than conditional errors. Figure 6.3 shows a simple example.

In practice, one is always interested in either marginal errors or Σ as a whole. Statements like “assuming the value of θ_2 is ..., we obtain a conditional error for θ_1 of σ_1 ” are interesting, if and only if θ_2 has zero or negligible error.⁴⁷ Otherwise, the conditional error σ_1 underestimates the true error on θ_1 , which, for instances, compromises any estimate of statistical significance.

⁴⁷In fact, if θ_2 has zero or negligible error, this means that the conditional error of θ_1 actually is a “quasi-marginal” error.

6.4.2 Error sources

We now briefly discuss various error contributions that have not been taken into account in previous studies (e.g. Slosar et al. 2009; Lee 2011) and assess their importance and effects.

6.4.2.1 Uncertainties in classifications

The morphological classifications of Galaxy Zoo and Huertas-Company et al. (2011) are probabilistic, i.e., every object is assigned a probability to belong to either of the possible morphological types. Conversely, non-probabilistic or “hard” assignments clearly assign a certain type to every object. As argued in Sect. 5.3.2, the morphological types are overlapping and hard classification schemes become inappropriate. For instances, a galaxy with $p_Z = 0.8$ still has a 20% chance not to be a Z-spiral – or a disc galaxy at all. In fact, Slosar et al. (2009) turned the probabilistic assignments of Galaxy Zoo into hard assignments by introducing a hard cut: For the clean sample, every galaxy with $p_Z \geq 0.8$ is considered as Z-wise spiral and every galaxy with $p_S \geq 0.8$ is considered as S-wise spiral, while all other galaxies are discarded. Similarly, Lee (2011) considers every galaxy with $p_{\text{Scd}}^{\text{HC}} > 0.5$ as Scd galaxy. Discarding the classification uncertainty by introducing a hard cut pretends that the data is more accurate than it actually is. This inevitably leads us to underestimate the errors, thereby compromising estimates of statistical significance. We explain in Sects. 6.4.4 and 6.4.5 how to account for these classification uncertainties in estimating the correlation functions of handedness and angular-momentum orientations.

6.4.2.2 Errors in spectroscopic redshift estimates

Both autocorrelation functions require estimates of distances of galaxy pairs and these distances are uncertain due to errors in the redshift estimates. In order to assess the impact of redshift errors, we randomly select a single galaxy and draw 10,000 random samples from its redshift-error distribution. For every sampled value of redshift, we compute the comoving distance and monitor its distribution. As is evident from Fig. 6.4, the errors in the comoving distances are of the same order of magnitude as the typical distance scale of the

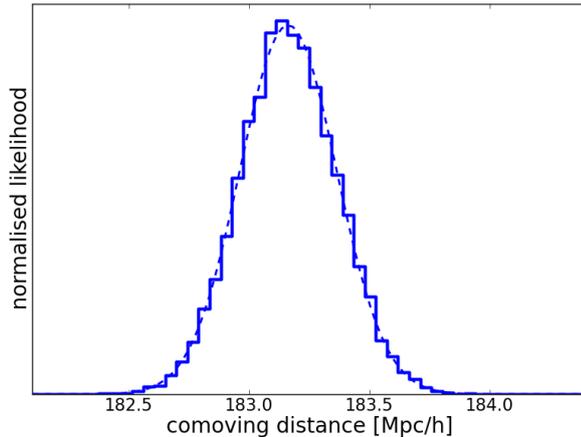


Figure 6.4: Errors in estimates of comoving distance.

We show the likelihood function of comoving distance for a galaxy with spectroscopic redshift of $z = (6.5993 \pm 0.0078) \cdot 10^{-2}$. The likelihood has been estimated by drawing 10,000 random samples from the error distribution of the spectroscopic redshift and is approximately Gaussian with mean $(183.16 \pm 0.20)\text{Mpc/h}$.

correlations reported in the literature. Consequently, these errors are important and must not be neglected. We explain in Sect. 6.4.3 how to propagate redshift errors by Monte-Carlo sampling.

6.4.2.3 Errors in ellipticity estimates

Errors in ellipticity estimates clearly have an impact on the estimation of angular-momentum-orientation vectors and their correlation function. We now try to estimate these errors. We explain in Sect. 6.4.3 how to propagate ellipticity errors by Monte-Carlo sampling.

First, considering the isophotal ellipticities used by Lee (2011), the SDSS database unfortunately does not offer error estimates.⁴⁸ Consequently, employing isophotal ellipticities, the SDSS database actually does not enable us to estimate a marginal autocorrelation function. In order to get a rough estimate of the errors in isophotal ellipticities, we make use of the rogue pairs in the SDSS database, i.e., multiple entries of identical galaxies. Starting out from 698,420 galaxies in the classification table provided by Huertas-Company et al. (2011), we identify rogue pairs as galaxy pairs whose angular separation is less than 0.4arcsec, which equals one pixel size. We find 1,596 such pairs. We then monitor the difference in axis ratios and orientation angles of every pair. The resulting distributions are shown in Fig. 6.5. As rough error estimate for the isophotal axis ratio, we obtain a standard deviation of

$$\hat{\sigma}(q_{\text{iso}}) \approx 0.0562 . \quad (6.26)$$

The distribution of differences in orientation angles is not Gaussian but has more prominent wings. We therefore model the likelihood function of orientation angles with mean angle θ_0 as a mixture of two Gaussians of different

⁴⁸In fact, the table GALAXY contains columns for the errors of the isophotal ellipticities. However, for the relevant objects these columns are only filled with invalid default values.

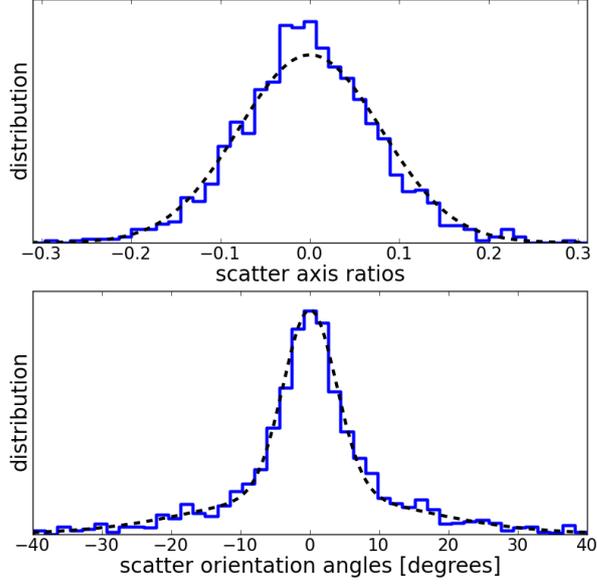


Figure 6.5: Errors in isophotal-ellipticity estimates.

Distributions of differences in isophotal axis ratios (top) and isophotal orientation angles (bottom) for the 1,596 rogue pairs in the catalogue of Huertas-Company et al. (2011). The top panel is well approximated by a Gaussian with mean zero and standard deviation of ≈ 0.0795 (dashed line), which yields an error estimate of $\hat{\sigma}(q_{\text{iso}}) = \frac{0.0795}{\sqrt{2}} \approx 0.0562$. The distribution of differences in orientation angles (bottom panel) is not Gaussian, but described by the ad-hoc model of Eq. (6.28) (dashed line) based on Eq. (6.27) with manually adjusted parameters $\hat{\alpha} \approx 0.73$, $\hat{\sigma}_1 \approx 2.7^\circ$ and $\hat{\sigma}_2 \approx 15.0^\circ$.

width,

$$\mathcal{L}(\theta|\theta_0, \sigma_1, \sigma_2, \alpha) = \alpha N(\theta|\theta_0, \sigma_1) + (1 - \alpha)N(\theta|\theta_0, \sigma_2). \quad (6.27)$$

The bottom panel of Fig. 6.5 displays the distribution of differences of two values drawn from Eq. (6.27), whose likelihood is obtained by convolving $\mathcal{L}(\theta|\theta_0, \sigma_1, \sigma_2, \alpha)$ with itself. The resulting likelihood function then reads

$$\begin{aligned} \mathcal{L}(\Delta\theta|\sigma_1, \sigma_2, \alpha) &= \alpha^2 N(\Delta\theta|0, \sqrt{2}\sigma_1) \\ &+ 2\alpha(1 - \alpha)N(\Delta\theta|0, \sqrt{\sigma_1^2 + \sigma_2^2}) + (1 - \alpha)^2 N(\Delta\theta|0, \sqrt{2}\sigma_2). \end{aligned} \quad (6.28)$$

Manually adjusting the model parameters of Eq. (6.27), we obtain a rough error estimate for the isophotal orientation angle with parameters $\hat{\alpha} \approx 0.73$, $\hat{\sigma}_1 \approx 2.7^\circ$ and $\hat{\sigma}_2 \approx 15.0^\circ$. If we required an angular separation of 0, i.e., identical coordinates, we would still end up with 17 pairs exhibiting similar scatter in both parameters. Clearly, this error estimate is only a work-around. However, given the information provided by the SDSS database, this is all we can do.

Second, the correction for intrinsic disc thickness of Scd galaxies is subject to uncertainties, too. As mentioned in Sect. 6.2.4, Heidmann et al. (1972) find a scatter of intrinsic disc thickness for Scd galaxies in the range from 0.083 to 0.145. Wherever we neglect ellipticity errors, we also neglect errors in intrinsic disc thickness and simply adopt the value 0.1 for p in Eq. (6.8). Conversely, if we take into account ellipticity errors, we will automatically also take into account errors in the intrinsic disc thickness. In this case, we assume that p is drawn from a uniform distribution over the interval [0.083, 0.145].

6.4.2.4 Truly negligible error sources

There are further sources of errors which could be taken into account but are not relevant in our case.

For instances, uncertainties in the cosmological parameters have an impact on the comoving distances (see Sect. 5.6). In our case, this is irrelevant because all galaxies are affected the same way. However, if the task is to use marginal autocorrelation functions in order to do cosmological inference, it may in fact be mandatory to also incorporate uncertainties of cosmological parameters into the Monte-Carlo sampling described in Sect. 6.4.3. We experienced that this increases the error in comoving distances by approximately a factor of two.⁴⁹

Another negligible error source is the position estimate of a galaxy in equatorial coordinates. Given the pixel size of ≈ 0.4 arcsec of SDSS, at a redshift of $z = 0.066$ and comoving distance of $d = 183$ Mpc/h one pixel misestimation corresponds to a transversal error of 0.35kpc/h. This is several orders of magnitude below the theoretically expected correlation length of roughly 1Mpc/h (Schäfer & Merkel 2011). Consequently, this error source is completely negligible.

6.4.3 Propagating errors numerically

We now explain how to incorporate errors in redshift estimates and ellipticity estimates. The crucial problem is that both errors cannot be propagated analytically because it is infeasible to express the final correlation estimate as an analytic function of all input redshifts and ellipticities.

We propagate the measurement errors of spectroscopic redshift and ellipticity by drawing 1,000 Monte-Carlo realisations from the error distributions of both parameters and averaging the results over all Monte-Carlo realisations.⁵⁰ A value for the intrinsic disc thickness is drawn from the uniform interval [0.083, 0.145] once for every Monte-Carlo realisation, i.e., in each realisation all galaxies have the same correction for intrinsic disc thickness. This Monte-Carlo sampling is in fact a marginalisation over the errors of both observables, spectroscopic redshift and ellipticity. Propagating these errors is inevitable in order to obtain correlation estimates with *marginal* errors. Typically, both error sources are neglected (e.g. Slosar et al. 2009; Lee 2011), which yields correlation estimates with *conditional* errors – conditional because they assume, e.g., the observed redshifts were the true ones.

A final remark on the correlation estimation in this case: We monitor the distribution of the correlation values $\hat{\xi}$ resulting from the 1,000 Monte-

⁴⁹At first glance, it may be surprising that an error $\Omega_\Lambda = 0.734 \pm 0.030$ only leads to an error of $0.4/183.2 \approx 0.2\%$ in comoving distance. However, at a redshift of $z \approx 0.066$ we are in a regime where nonlinear effects and thus the impact of Ω_Λ are still small (see also Fig. 5.3).

⁵⁰Sampling from the error distribution centred at the observed instead of the true value, we convolve the error distribution with itself. Consequently, this technique slightly overestimates the errors. However, given the rules of scientific methodology, overestimated errors are an unfortunate necessity where analytic error propagation is impossible, whereas underestimated errors are out of the question.

Carlo realisations. However, a fundamental difference to Eq. (6.18) is that now $\hat{\xi}$ itself is a random variate. Consequently, we are now interested in the variance of $\hat{\xi}$ but not in the variance of the mean of $\hat{\xi}$. The difference is a factor of 1,000 in the variances. Otherwise, we could make the resulting errors arbitrarily small by increasing the number of Monte-Carlo realisations.

6.4.4 Impact on autocorrelation of handedness

Estimates of handedness autocorrelation do not involve ellipticity information. Consequently, we only need to investigate the impact of uncertainties of handedness classification and redshifts. Our ultimate goal is a marginal estimate of the handedness autocorrelation function, where errors of morphological classification and redshift estimates have been marginalised out.

6.4.4.1 Reproducing former results

First, we use the hard estimator from Slosar et al. (2009), which neglects uncertainties in classification and redshift. This estimator cuts the probabilities for handedness h at a certain level and thereby unambiguously assigns a clear handedness to every galaxy, whose two possible values we denote by $h = \pm 1$. We then look for pairs of galaxies, bin them according to their distance in real space and finally evaluate the correlation estimator of Eq. (6.24). Figure 6.6 (blue circles) shows our estimate of this autocorrelation function for the Galaxy Zoo clean sample. Qualitatively, our results agree with the results of Slosar et al. (2009). We observe positive correlations, i.e., an alignment of handedness, on short distances, too. However, we also observe an oscillatory pattern of the correlation function with distance, which is not present in the result of Slosar et al. (2009). First, Slosar et al. (2009) only show their correlation function for distances smaller than 3Mpc/h, while we show the correlation for distances of up to 10Mpc/h. Second, we use equidistant bins, whereas Slosar et al. (2009) use fewer and non-equidistant bins without giving details or providing a justification. As we are going to demonstrate later in this section, most of the substructure of the correlation functions are likely to be noise features enhanced by the binning. Furthermore, our maximum correlation on short scales is smaller than that in Slosar et al. (2009), while our error bars are larger. This may also be an effect of the different binning.

6.4.4.2 Taking into account classification uncertainties

Second, we now take into account uncertainties in the handedness classifications, but still ignore the errors in redshifts. In every distance bin, we compute the handedness products

$$h h' = (p_Z - p_S)(p'_Z - p'_S) = p_Z p'_Z + p_S p'_S - p_Z p'_S - p_S p'_Z, \quad (6.29)$$

which can now take any value in the interval $[-1, 1]$. The correlation estimator of Eq. (6.24) is unchanged. However, n_{\pm} now are *not* the number of pairs where $h h' = \pm 1$, but are rather defined by

$$n_+ = \sum_{\text{pairs}} (p_Z p'_Z + p_S p'_S) \quad \text{and} \quad n_- = \sum_{\text{pairs}} (p_Z p'_S + p_S p'_Z). \quad (6.30)$$

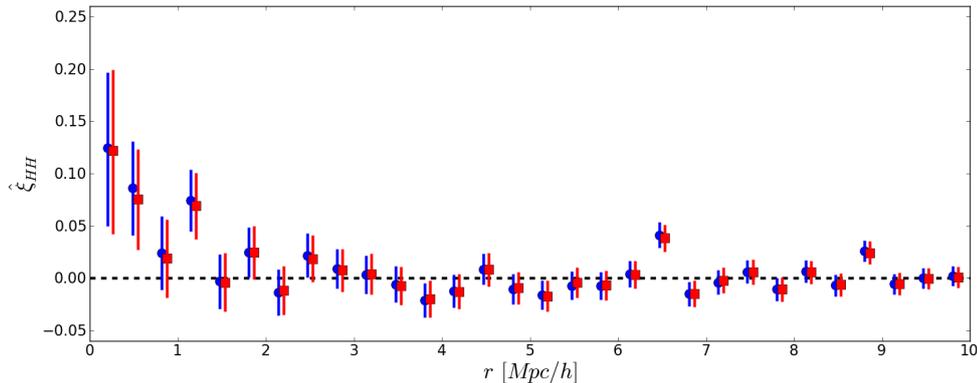


Figure 6.6: Conditional estimates of the handedness autocorrelation.

Circles: Conditional estimate accounting for number statistics but neglecting classification uncertainties and redshift errors. Squares: Conditional estimate accounting for number statistics and classification uncertainties but neglecting redshift errors. Circles and squares are slightly offset with respect to each other for the sake of visualisation but in fact mark identical distance bins.

Note that if N denotes the number of galaxy pairs in a given distance bin, then $n_+ + n_- \leq N$. Consequently, this reduces the “effective” number of galaxy pairs in a given distance bin because the contribution of every galaxy pair is downweighted by the probability that either galaxy is not a spiral with handedness. Furthermore, reducing the effective number of galaxy pairs also increases the error of the correlation estimate through the beta distribution (e.g. Cameron 2010). Results of this estimator are shown in Fig. 6.6 (red squares). The error bars are indeed slightly larger. However, ignoring uncertainties in the handedness classifications has negligible impact in this particular case. This is not surprising since the handedness probabilities were cut at 80%.

6.4.4.3 Taking into account redshift errors

Third, we consider the uncertainty in the redshift estimates. As described in Sect. 6.4.3, we draw 1,000 Monte-Carlo realisation from the error distributions of the spectroscopic redshift estimates. The redshift errors are given in the SDSS database and we interpret these values as one-sigma intervals of a Gaussian distribution. Every Monte-Carlo realisation then represents a database on its own. For every realisation and in every distance bin, we then evaluate the expectation value of the estimator of Eq. (6.24). The handedness autocorrelation is then estimated by the distribution of values of Eq. (6.24), which is found to be Gaussian in excellent approximation. This greatly simplifies the error estimation. Figure 6.7 shows the resulting estimate of the handedness autocorrelation. In comparison to Fig. 6.6, the autocorrelation function now looks remarkably smooth. The explanation is very simple: Errors in redshift cause uncertainties in the distances, i.e., the data points are “smeared out” along the x-axis by galaxy pairs ending up in different distance bins in different realisations. Consequently, the most likely explanation for all the small sub-

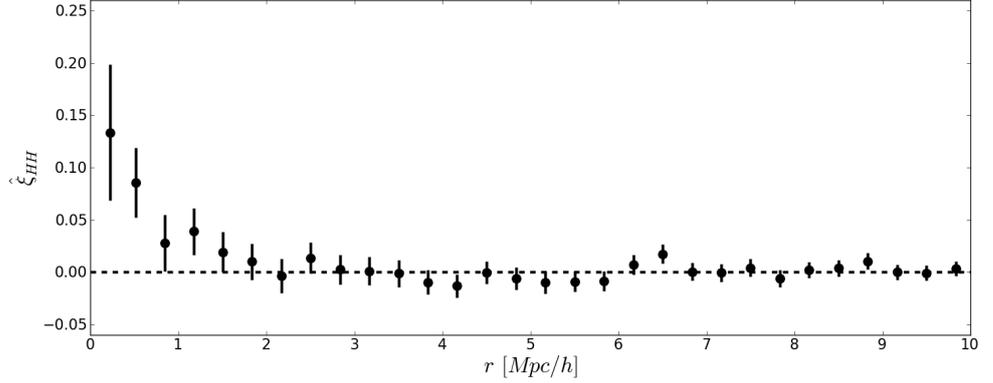


Figure 6.7: Conditional estimate of the handedness autocorrelation.

This estimate accounts for number statistics and redshift errors but not for classification uncertainties. Results have been averaged over 1,000 Monte-Carlo samples drawn from the error distribution of spectroscopic redshifts. The dots indicate mean values and the error bars correspond to one Gaussian standard deviation.

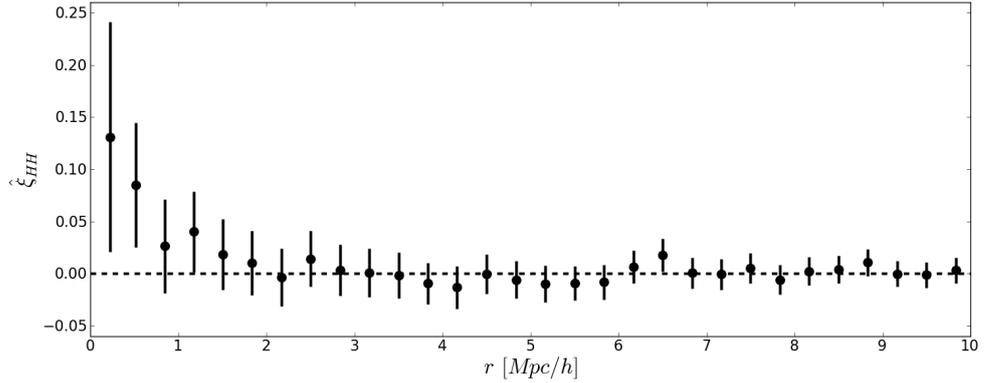


Figure 6.8: Marginal estimate of the handedness autocorrelation.

This final estimate takes into account number statistics and uncertainties in redshift and handedness classification. Results have been averaged over 1,000 Monte-Carlo samples drawn from the error distribution of spectroscopic redshifts. The dots indicate mean values and the error bars correspond to one Gaussian standard deviation. Although the error bar of the innermost distance bin is not small compared to the allowed interval $[-1, 1]$, the distribution of Monte-Carlo samples is still well approximated by a Gaussian.

structures in Fig. 6.6 is that they are noise features that have been enhanced by binning.

6.4.4.4 Final result: Marginal handedness autocorrelation

Unfortunately, Fig. 6.7 is still not our desired marginal estimate of the handedness autocorrelation. The reason is very simple: We have taken into account redshift errors. However, we did not take into account the uncertainty in-

duced by the number statistics of galaxy pairs in every distance bin, because we averaged the realisations over the maximum-likelihood values of Eq. (6.24), ignoring the fact that these maximum-likelihood values have errors themselves. This effect caused the errors in Fig. 6.6. For instance, let us consider the case where the redshift errors are zero, i.e., the true redshifts are known. In this case, all 1,000 Monte-Carlo realisations are identical, the pairs in all distance bins are always the same and hence the error bars in Fig. 6.6 would be zero. For every realisation and in every distance bin, we therefore draw a random sample from the beta distribution defined by n_{\pm} instead of merely taking the maximum-likelihood value of Eq. (6.24). Finally, this results in the desired estimate of the marginal autocorrelation function, which takes into account all important sources of uncertainty. This final marginal autocorrelation function is shown in Fig. 6.8. As a major result of this thesis, it is obvious that the error bars are so large that no statistically significant positive correlation of handedness can be detected, though we can see an indication for this effect. Therefore, we have to conclude that the only reason why Slosar et al. (2009) were capable of detecting such a positive correlation is that they did not account for the uncertainties in the redshift estimates. Given the present SDSS data, there is no statistically significant evidence that spiral-arm handedness is correlated for nearby spiral galaxies. Consequently, this result does not confirm this prediction of the tidal-torque theory.

6.4.5 Impact on autocorrelation of angular-momentum orientation

We again start by reproducing former results in order to validate our method. We then proceed to take into account the important error contributions.

6.4.5.1 Reproducing former results

First, we start by reproducing the autocorrelation estimate of Lee (2011). The only difference is that we have removed 20 objects from the galaxy sample in order to eliminate rogue pairs. Figure 6.9 (blue circles) shows our resulting estimate of the autocorrelation via Eq. (6.19). Our result is perfectly identical to the one of Lee (2011). This implies that, first, our method is working correctly, and, second, that rogue pairs have negligible impact on the results of Lee (2011).

6.4.5.2 Taking into account classification uncertainties

We now study the impact of uncertainties of morphological classification. Formally, the estimator defined in Eq. (6.19) does not change. The impact of classification uncertainties is again a reduction of the effective number of galaxy pairs in all redshift bins as discussed in the case of handedness correlations in Sect. 6.4.4. Considering a single of the four terms in Eq. (6.19), we change the definition

$$\langle p_q p'_a | \vec{L}_a \cdot \vec{L}'_a |^2 \rangle = \frac{\sum_{\text{pairs}} p_{\text{Scd}}^{\text{HC}} p_{\text{Scd}}^{\text{HC}'} p_q p'_a | \vec{L}_a \cdot \vec{L}'_a |^2}{\sum_{\text{pairs}} p_{\text{Scd}}^{\text{HC}} p_{\text{Scd}}^{\text{HC}'}}. \quad (6.31)$$

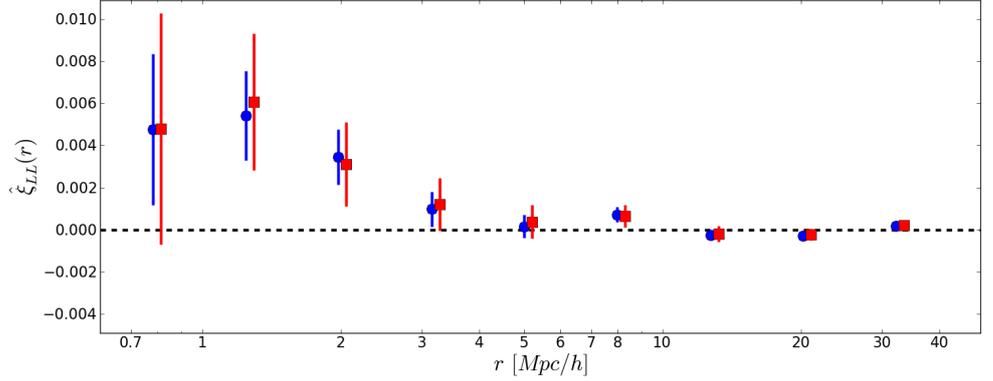


Figure 6.9: Conditional angular-momentum-orientation autocorrelations. Circles: Hard estimator from Lee (2011) which accounts for number statistics only but neglects errors in classification, redshifts and ellipticities. Squares: Probabilistic estimator which accounts for number statistics and classification uncertainties but neglects error in redshifts and ellipticities. Estimates are based on the sample of 4,216 Scd galaxies. Error bars correspond to one Gaussian standard deviation in good approximation. The dashed horizontal line indicates zero correlation. Circles and squares are slightly offset with respect to each other for the sake of visualisation but in fact mark identical distance bins.

This weights the contribution of every pair by the probability $p_{\text{Scd}}^{\text{HC}} p_{\text{Scd}}^{\text{HC}'}$ that *both* galaxies are Scd galaxies. Furthermore, the number N of pairs in the distance bin are replaced by the sum of weights $\sum_{\text{pairs}} p_{\text{Scd}}^{\text{HC}} p_{\text{Scd}}^{\text{HC}'} \leq N$. Obviously, this weighting also affects the error estimate of Eq. (6.21). The red squares in Fig. 6.9 show the probabilistic correlation estimate. Evidently, the hard estimator used by Lee (2011) substantially underestimates the errors, thereby compromising estimates of statistical significance. It is not surprising that the impact is larger than in Sect. 6.4.5.2, since here, class probabilities were cut at $p_{\text{Scd}}^{\text{HC}} > 0.5$ whereas Slosar et al. (2009) cut the handedness probabilities at 0.8.

6.4.5.3 Taking into account redshift errors

Now, we “switch on” the error in spectroscopic redshift by drawing 1,000 Monte-Carlo realisations from the redshift’s error distribution. For every single Monte-Carlo realisation, we then estimate the correlation and its error in every distance bin. Finally, we estimate the correlation over all 1,000 Monte-Carlo samples by drawing randomly a single value from the error distribution of each realisation’s correlation. This ensures that the error bars did not approach zero, if the spectroscopic redshifts were known precisely. Consequently, the resulting correlation estimate accounts for uncertainties in classification, number statistics and redshift errors. Nonetheless, it is still only a conditional estimate, as it does not yet include errors in ellipticity estimates. Furthermore, we now restrict the correlation to maximum distances of 10Mpc/h – instead of 40Mpc/h as in Fig. 6.9 – in order to reduce the required hard-disk memory. The resulting conditional estimate is shown in Fig. 6.10 (blue cir-

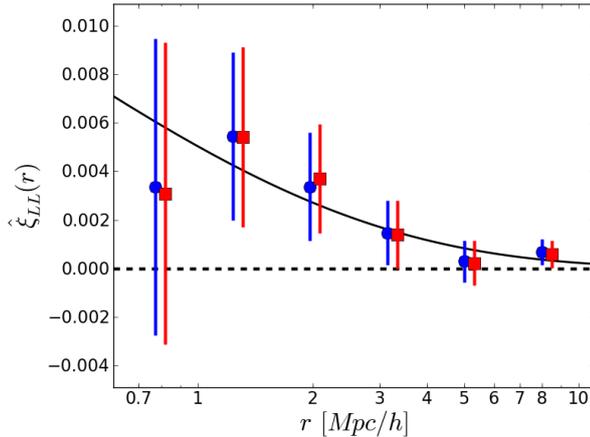


Figure 6.10: Conditional and marginal angular-momentum-orientation autocorrelations.

Circles: Conditional estimate that takes into account number statistics, classification uncertainties and redshift errors but still neglects ellipticity errors. Squares: Marginal estimate that accounts for all error sources. Results have been averaged over 1,000 Monte-Carlo samples drawn from the error distribution of spectroscopic redshifts. The dots indicate mean values and the error bars correspond to one Gaussian standard deviation. Again, circles and squares mark identical distance bins but are slightly offset with respect to each other for the sake of visualisation. The solid line indicates a fit to the binned marginal correlation function (see Sect. 6.4.6).

cles). Apparently, the impact of redshift errors on the correlation estimate of angular-momentum orientations is not as severe as in the case of handedness (cf. marginal estimate of Fig. 6.8). There are two reasons: First, the conditional correlation signals have larger statistical significance in this case. Second, the binsize in Fig. 6.10 is much larger than in Fig. 6.8, because here we are studying a smaller galaxy sample and therefore have fewer galaxy pairs. In Fig. 6.8), the equidistant binsize is $0.3Mpc/h$, while the smallest bin in Fig. 6.10 has a width of $0.36Mpc/h$ and thus is already slightly larger. Nonetheless, the estimated errors have indeed increased, which is obvious for the first distance bin. As the binning is logarithmic in distance, this is not surprising because the first distance bin has the smallest binsize and is thereby strongest affected by redshift errors “smearing out” galaxy pairs along the horizontal axis. We would not expect distance errors of the order of $0.2Mpc/h$ (cf. Fig. 6.4) to have a large impact on a distance bin of $1Mpc/h$ binsize. However, such large binsizes are clearly inappropriate to resolve a theoretically predicted correlation length of $1Mpc/h$ (Schäfer & Merkel 2011).

6.4.5.4 Taking into account ellipticity errors

As mentioned in Sect. 6.4.2, the SDSS database actually does not provide error estimates for the isophotal ellipticities. Hence, we need to proceed using the work-around described in Sect. 6.4.2. We adopt the rough error estimates of Eqs. (6.26) and (6.27) as well as the uniform error in intrinsic disc thickness

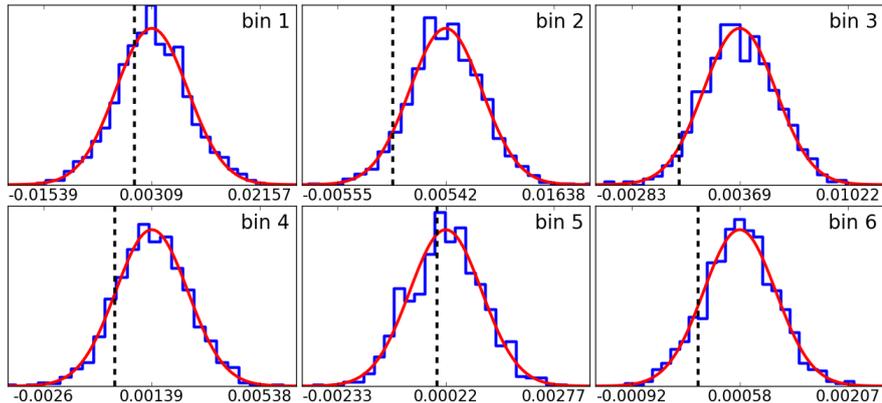


Figure 6.11: Distribution of Monte-Carlo realisations in distance bins.

For the six distance bins of the marginal autocorrelation estimate of Fig. 6.10, the distributions of 1,000 Monte-Carlo realisations are well approximated by Gaussians (solid lines). In every panel, the ticks indicate mean μ and $\mu \pm 3\sigma$. The vertical dashed lines indicate zero correlation.

for all galaxies in the data sample and proceed by Monte-Carlo sampling as described in Sect. 6.4.3 from the catalogue of 4,216 Scd galaxies. This enables us to estimate a marginal autocorrelation function. Figure 6.10 (red squares) shows the resulting marginal estimate of the angular-momentum-orientation autocorrelation function. We show the distribution of correlation values for the 1,000 Monte-Carlo realisations in Fig. 6.11, which confirms that the errors are indeed Gaussian in excellent approximation. In comparison to the conditional estimate also shown in Fig. 6.10, there is only a minor increase in the error bars. However, we would not put too much faith into the marginal estimate shown in Fig. 6.10 because the error estimate of ellipticities is rather handwavy. As the final result, we directly compare our marginal estimate to the conditional estimate of Lee (2011) in Fig. 6.12. Evidently, the marginal estimate differs substantially from the conditional estimate and there are no statistically significant autocorrelations. Again, we have to conclude that Lee (2011) obtained a seemingly significant detection because important error contributions were not taken into account. Given the present SDSS data, there is no statistically significant evidence that angular-momentum-orientation vectors are correlated for nearby disc galaxies. Consequently, this result also does not confirm this prediction of the tidal-torque theory.

6.4.6 Consequences for parameter estimation

The autocorrelation of angular-momentum orientations can be used to estimate free parameters in the tidal-torque theory (e.g. Lee & Pen 2008). As discussed in Sect. 6.1, such a parameter estimate is necessary in order to put the constrained theory to a second future empirical test. Let $\xi(r, R)$ denote the two-point correlation function of Scd galaxies, smoothed over scale R . In this case, one can derive a model prediction for the linear regime (e.g. Pen

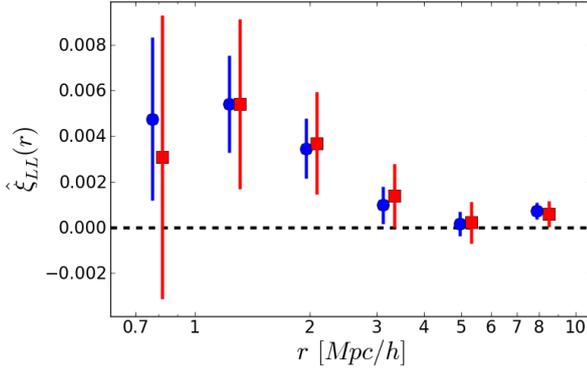


Figure 6.12: Marginal autocorrelation estimate vs. results of Lee (2011).

Blue circles show the conditional autocorrelation estimate of angular-momentum-orientation vectors published by Lee (2011). Red squares show our marginal estimate that takes into account all relevant error contributions. This figure brings together results shown in Figs. 6.9 and 6.10 and is solely meant to facilitate a direct comparison.

et al. 2000)

$$\xi_{LL}(r) \approx \frac{a^2 \xi^2(r, R)}{6 \xi^2(r, 0)}, \quad (6.32)$$

where a is a free model parameter. For the nonlinear regime, Lee & Pen (2008) derived the following model prediction

$$\xi_{LL}(r) \approx \frac{a_L^2 \xi^2(r, R)}{6 \xi^2(r, 0)} + \varepsilon_{NL} \frac{\xi(r, R)}{\xi(r, 0)}, \quad (6.33)$$

where a_L and ε_{NL} are free model parameters describing the linear and nonlinear contributions. Estimating values for these model parameters is important in order to constrain the tidal-torque theory.⁵¹ The impact of the additional error sources on this parameter estimation is devastating. First, the marginal estimate of $\xi_{LL}(r)$ has large errors. Second, errors in redshift estimates and morphological classification also affect the estimation of the two-point correlation function $\xi(r, R)$ of Scd galaxies. Given these considerations, we have to conclude that it is impossible to place decisive constraints on the parameters in Eqs. (6.32) or (6.33).

The same argument applies to the generic autocorrelation model proposed by Schäfer & Merkel (2011),

$$\xi_{LL}(r) = A \exp \left[- \left(\frac{r}{R} \right)^C \right], \quad (6.34)$$

which contains a linear amplitude A and two nonlinear model parameters R and C that cannot be constrained properly. Figure 6.13 demonstrates this by showing the marginal likelihoods of fitting Eq. (6.34) to the data of Fig. 6.10. Evidently, the (marginal) uncertainties in all model parameters are so large that they cover almost the complete allowed parameter ranges. Hence, this

⁵¹In fact, this is the reason why Lee (2011) restricts the sample to galaxies with $z \leq 0.02$ in order to obtain a volume-limited sample. Otherwise, the density field of galaxies cannot be meaningfully defined and $\xi(r, R)$ cannot be estimated.

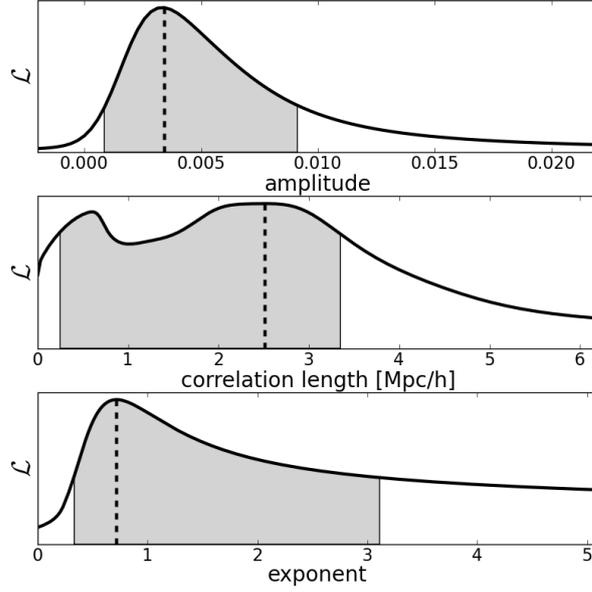


Figure 6.13: Marginal likelihoods of fitting the marginal angular-momentum-orientation autocorrelation.

The model given by Eq. (6.34) is fitted to the binned version of the marginal autocorrelation of Fig. 6.10. Top panel: Marginal likelihood of amplitude with maximum at $A = 0.0034^{+0.0057}_{-0.0027}$. Centre panel: Marginal likelihood of correlation length with maximum at $R = 2.5^{+0.8}_{-2.3}$. Bottom panel: Marginal likelihood of exponent with maximum at $C = 0.71^{+2.40}_{-0.38}$. The asymmetric errors denote 68% confidence intervals. The parameter estimation has been conducted on a three-dimensional brute-force grid. As the distributions of Monte-Carlo realisations in every distance bin are Gaussian in excellent approximation as shown in Fig. 6.11, the fit is done via χ^2 -minimisation. The likelihood is then given by $\mathcal{L} \propto e^{-\chi^2/2}$.

fit poorly constrains the model parameters. Nevertheless, the marginal likelihood of the correlation length is compatible with the theoretical prediction of 1Mpc/h from Schäfer & Merkel (2011). For later purposes, we note the best fitting model,⁵²

⁵²Note that the maximum of the *joint* likelihood does not coincide with the maxima of the *marginalised* likelihoods shown in Fig. 6.13.

$$\xi_{\text{LL}}(r) \approx 0.026 \cdot \exp \left[- \left(\frac{r}{0.34 \text{Mpc}/h} \right)^{0.46} \right]. \quad (6.35)$$

This fit is also shown in Fig. 6.10. We explicitly emphasise that we do *not* claim that this were by any means a model of the true correlation function. This fit is solely meant to provide us with *some* model that is *compatible* with the SDSS data. Such a model is later required in order to conduct simulations. This is also the reason why we do not need to estimate errors for the fit given by Eq. (6.35).

— The bottom line —

- The SDSS data contains no statistically significant evidence for theoretically predicted autocorrelations, neither in spiral-arm handedness nor in angular-momentum-orientation vectors. However, given the data, it appears *plausible* that such autocorrelations could exist and exhibit a correlation length of the order of 1Mpc/h as theoretically predicted.
- Marginal autocorrelation estimates of spiral-arm handedness and angular-momentum-orientation vectors require to account for errors in morphological classification, redshift estimates and estimates of ellipticities. If these errors are not taken into account, the resulting conditional autocorrelation function exhibits errors that are substantially underestimated. This leads to an overly optimistic assessment of statistical significance.
- The evidence for correlations in the SDSS data is so weak that theoretical parameters cannot be constrained decisively from this experiment.

7

Improvements and potential of future surveys

As an outlook, we discuss possible improvements of ellipticity estimates and briefly elaborate on the potential of future sky surveys in this chapter. Namely, we discuss the potentials of PanSTARRS, LSST and EUCLID to improve the estimates of handedness and angular-momentum-orientation autocorrelations. We discuss the estimation of ellipticities and the impact of number statistics. Furthermore, we discuss the impact of possible improvements of redshift estimates and the potential of front-edge estimation.

7.1 Biased ellipticity estimates from second moments

Given the marginal autocorrelation estimate of angular-momentum-orientation vectors shown in Fig. 6.12, it is evident that a reduction of noise in any of the error sources would be of great interest. Isophotal ellipticity estimates have the disadvantage that they strongly depend on the choice of the particular isophote. Typically, this isophote is chosen in the outskirts of the imaged galaxy where the signal-to-noise ratio is low. Conversely, ellipticity estimates based on the second moments of the galaxy's light distribution, e.g., as given by Eqs. (2.27) and (2.28), seem to be more promising, since no isophote is required and the complete data enters the estimate. Consequently, we may expect that ellipticity estimates based on second moments are more robust against pixel noise than isophotal ellipticities. In this section, we demonstrate that ellipticity estimates based on second moments of the light distribution are so strongly biased that they cannot be used for investigations of disc alignment. In particular, this bias would cause us to overestimate the correlation due to alignment such that, e.g., we would overestimate its impact on weak-lensing studies. However, this bias does not impair investigations of gravitational weak lensing.

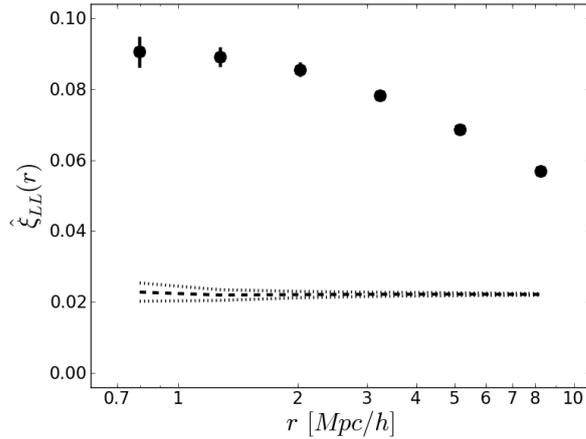


Figure 7.1: Angular-momentum-orientation autocorrelation biased by second moments.

The marginal correlation estimate like in Fig. 6.10 but now using ellipticity estimates based on second moments of the galaxies’ light distribution instead of isophotal ellipticities.

7.1.1 Revealing the bias

We also downloaded the ellipticity estimates based on unweighted second moments of the galaxies’ light distributions from the SDSS database. Furthermore, SDSS offers error estimates for these parameters. Figure 7.1 shows the resulting autocorrelation estimate of angular-momentum-orientation vectors. The most striking difference to Fig. 6.12 is that Fig. 7.1 exhibits correlations that are substantially larger. This difference stems from systematic differences in the axis ratios resulting from second moments and isophotal contours, which is shown in Fig. 7.2. Evidently, axis ratios estimated from second moments are systematically larger than isophotal axis ratios. This implies that in Fig. 7.1 galaxies are generally considered to be rounder and therefore their angular-momentum-orientation vectors are bent into the line-of-sight, thereby feigning these strong correlations. Figure 7.2 also reveals that second moments tend to prefer axis-parallel orientation angles, which might be caused from pixellation effects of poorly sampled disc galaxies. Our scepticism is further raised by the enormous statistical significance of the correlations, which still seems to hold at separations as large as 10Mpc/h substantially exceeding the theoretical prediction. Finally, we note that the background correlation estimated from randomly shuffling the galaxy positions in the sample (cf. Lee 2011) in Fig. 7.1 is not zero. This suggests the presence of a strong bias, corrupting the correlation estimate of Fig. 7.1.

7.1.2 Point-spread function

Is this bias an effect of the point-spread function (PSF) which makes galaxies look rounder than they actually are? This is unlikely because all our objects are large compared to the size of the PSF. The median r -band Petrosian radius of the 4,211 Scd galaxies with SDSS data is 15.8 pixel, whereas the r -band

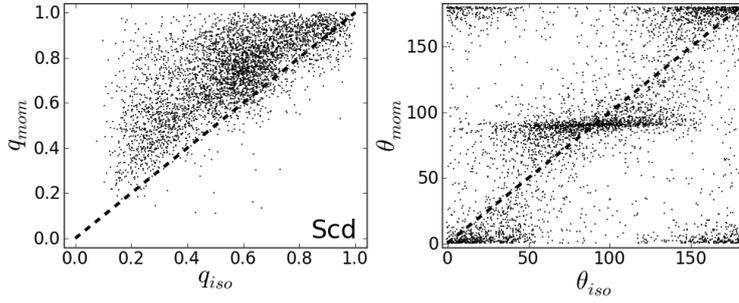


Figure 7.2: Comparing ellipticities estimated from second moments and isophotes for Scd galaxies.

Left panel: Axis ratios estimated from (unweighted) second moments are systematically larger than those estimated from isophotal contours. As we defined axis ratios as semi-minor over semi-major axis, this means that second moments find the disc galaxies to be rounder. Right panel: Orientation angles for second moments prefer axis-parallel values ($0^\circ, 90^\circ, 180^\circ$) due to the seeming roundness.

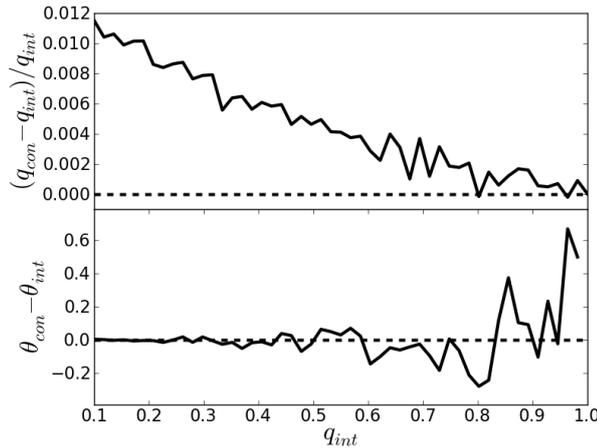


Figure 7.3: PSF biasing ellipticities from second moments.

Impact of circular Gaussian PSF with Petrosian radius of 1.3 pixel onto convolved axis ratios q_{con} and orientation angles θ_{con} of exponential-disc profiles with Petrosian radii of 15.8 pixels and intrinsic axis ratios $0.1 \leq q_{\text{int}} \leq 1$ and orientation angles $\theta_{\text{int}} = 30^\circ$. All profiles have been truncated at five scale radii. There was no noise in this simulation. The PSF leads to an overestimation of the axis ratios by at most 1.2% for highly elongated objects. As the PSF was circular in this test, orientation angles are not affected.

Petrosian radius of the SDSS PSF is approximately 1.3 pixel.⁵³ Consequently, the impact of the PSF should be small. This prediction is supported by Fig. 7.3, where we simulate the impact of a Gaussian PSF with Petrosian radius 1.3 pixel onto exponential-disc profiles with Petrosian radii of 15.8 pixel and different intrinsic axis ratios. We find a maximum overestimation of axis ratios of only 1.2%, which is not enough to explain the strong bias in Fig. 7.1 or the discrepancy in Fig. 7.2.

⁵³The r -band Petrosian radius of the SDSS PSF has been estimated as the median r -band Petrosian radius of 100,000 stars downloaded from the SDSS database.

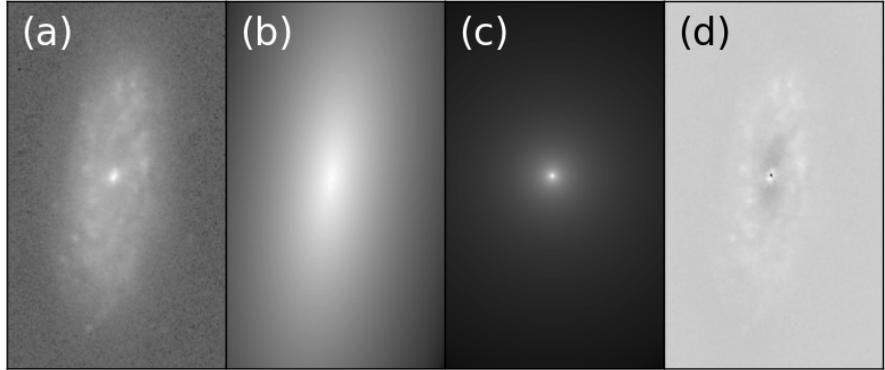


Figure 7.4: Bulge-disc decomposition of an example Scd galaxy.

We used g -band imaging data. The bulge is a circular de Vaucouleur profile, while the disc component is an exponential profile with ellipticity. The bulge is pinned to the pixel of the peak-of-light whereas the centroid of the disc component is free. Panel (a) shows the original galaxy. Panel (b) is the disc component, while panel (c) is the bulge component. Panel (d) displays the fit residuals. The fit was performed by χ^2 -minimisation using a Simplex algorithm (Nelder & Mead 1965) and reached a minimum value of 3.18 per pixel.

7.1.3 Galactic bulges

We now argue that the heavily biased correlation estimate of Fig. 7.1 stems from the galactic bulges biasing the second moments and thereby the ellipticity estimates. At first glance, this may seem to be a rather unlikely explanation, since we explicitly selected only Scd galaxies in order to minimise the impact of galactic bulges. However, this hypothesis can explain the substantial discrepancy between isophotal axis ratios and axis ratios based on second moments revealed by Fig. 7.2. If bulges were an issue, they would affect the second moments and would lead us to overestimate axis ratios, since bulges are typically of spheroidal shape or their ellipticity is misaligned with the disc ellipticity. On the other hand, isophotal ellipticity estimates should be almost unaffected by the presence of bulges as long as the isophote used is situated inside the disc component.

We demonstrate that the presence of a bulge can bias the estimate of axis ratio based on second moments. For this purpose, we perform a bulge-disc decomposition of a prototypical Scd galaxy from our data sample, which is shown in Fig. 7.4. Indeed, the axis ratio estimated from the second moments of the complete model (including bulge) is $q_{b+d} \approx 0.48$, whereas the axis ratio used by the disc model is only $q_{\text{disc}} \approx 0.38$. The g -band axis ratio noted in the SDSS database for this example galaxy is $q_{\text{iso}} \approx 0.41$ estimated from isophotes and $q_{\text{mom}} \approx 0.63$ estimated from second moments (Stokes parameters).⁵⁴ We conclude that the bulge is well capable of biasing the ellipticity estimate substantially, even in the case of Scd galaxies.

As a final test for our hypothesis to pass, we compare the axis ratios based on isophotes and second moments for Sab galaxies from the catalogue of Huertas-Company et al. (2011). As Sab galaxies have more prominent bulges

⁵⁴The discrepancy of the bulge-disc decomposition is the consequence of non-optimal modelling. We used a de Vaucouleur profile for the bulge and an exponential profile for the disc.

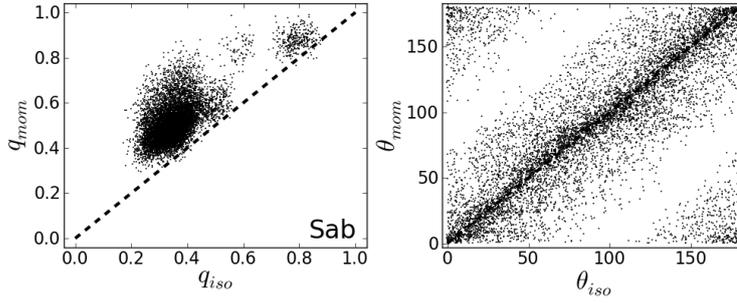


Figure 7.5: Comparing ellipticities estimated from second moments and isophotes for Sab galaxies.

We selected 8,496 galaxies with $p_{\text{Sab}} \geq 0.8$ from the catalogue of Huertas-Company et al. (2011). Left panel: Misestimation of axis ratios. The bimodal distribution of axis ratios agrees well with their Fig. 2. The bias of axis ratios estimated from second moments is stronger than for Scd galaxies in Fig. 7.2. Right panel: Orientation angles are unbiased.

than Scd galaxies, we would expect a stronger bias than in Fig. 7.2. We select all galaxies with $p_{\text{Sab}} \geq 0.8$ and download the r -band ellipticities (Stokes parameters) from the SDSS database, if available. For the resulting 8,496 Sab galaxies, Fig. 7.5 shows the comparison of axis ratios estimated from isophotes and second moments. Evidently, the second moments are biased, too, and the bias is also more pronounced than in Fig. 7.2, meaning there is less scatter. This confirms our expectation.

7.1.4 Simulating pairs of angular-momentum-orientation vectors

Presently, we want to simulate the bias that second moments create in correlation estimates of angular-momentum orientations. Therefore, we now explain how to simulate pairs of angular-momentum-orientation vectors which exhibit a given input correlation. This requires some cumbersome arithmetic.

7.1.4.1 Pairs of uncorrelated orientation vectors

As the orientation vectors indicate directions, the samples are drawn from the uniform distributions $\varphi \in [0, 2\pi)$ and $\cos \vartheta \in [-1, 1]$ of the two polar angles φ and ϑ . A random orientation vector is then given by

$$\vec{\ell}_1 = \begin{pmatrix} \cos \varphi \sin \vartheta \\ \sin \varphi \sin \vartheta \\ \cos \vartheta \end{pmatrix}. \quad (7.1)$$

This vector is normalised, i.e., $\vec{\ell}_1 \cdot \vec{\ell}_1 = 1$. Sampling a uniform angle $\phi \in [0, 2\pi)$, a second random orientation vector is

$$\vec{\ell}_2 = \sin \phi \begin{pmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{pmatrix} + \cos \phi \begin{pmatrix} \cos \varphi \cos \vartheta \\ \sin \varphi \cos \vartheta \\ -\sin \vartheta \end{pmatrix}. \quad (7.2)$$

This vector is again normalised, i.e., $\vec{\ell}_2 \cdot \vec{\ell}_2 = 1$, and also orthogonal to the first, i.e., $\vec{\ell}_1 \cdot \vec{\ell}_2 = 0$.

7.1.4.2 Pairs of correlated orientation vectors

In the first step, we sample a pair of uncorrelated angular-momentum-orientation vectors $\vec{\ell}_1$ and $\vec{\ell}_2$ as described in the previous section. In the second step, we mix these two uncorrelated vectors such that we obtain two correlated vectors,

$$\vec{L}_a = \cos \alpha \vec{\ell}_1 + \sin \alpha \vec{\ell}_2, \quad (7.3)$$

$$\vec{L}'_a = \cos \beta \vec{\ell}_1 + \sin \beta \vec{\ell}_2, \quad (7.4)$$

and their counter-parts due to the front-edge degeneracy,

$$\vec{L}_b = \cos \alpha \left[\vec{\ell}_1 - 2(\vec{e}_r \cdot \vec{\ell}_1) \vec{e}_r \right] + \sin \alpha \left[\vec{\ell}_2 - 2(\vec{e}_r \cdot \vec{\ell}_2) \vec{e}_r \right], \quad (7.5)$$

$$\vec{L}'_b = \cos \beta \left[\vec{\ell}_1 - 2(\vec{e}'_r \cdot \vec{\ell}_1) \vec{e}'_r \right] + \sin \beta \left[\vec{\ell}_2 - 2(\vec{e}'_r \cdot \vec{\ell}_2) \vec{e}'_r \right], \quad (7.6)$$

where \vec{e}_r and \vec{e}'_r are unit vectors pointing from the coordinate origin towards the positions of both galaxies. Due to the orthonormality of $\vec{\ell}_1$ and $\vec{\ell}_2$, all these vectors are unit vectors. The two mixing angles α and β have to be chosen such that the desired input correlation

$$\begin{aligned} \xi_{\text{input}} = & \frac{1}{4} \left(\langle (\vec{L}_a \cdot \vec{L}'_a)^2 \rangle + \langle (\vec{L}_a \cdot \vec{L}'_b)^2 \rangle + \langle (\vec{L}_b \cdot \vec{L}'_a)^2 \rangle \right. \\ & \left. + \langle (\vec{L}_b \cdot \vec{L}'_b)^2 \rangle \right) - \frac{1}{3} \end{aligned} \quad (7.7)$$

is exhibited by the sampled pairs of orientation vectors. This provides only a single constraint, i.e., we are allowed to freely choose one mixing angle. For convenience, we choose $\alpha = 0$ such that $\vec{L}_a = \vec{\ell}_1$, which simplifies the calculations. We now need to compute the four expectation values.

7.1.4.3 Computing the first term

We start by computing $\langle (\vec{L}_a \cdot \vec{L}'_a)^2 \rangle$, which is the simplest term and also presents the basic arithmetic steps. Evidently,

$$\vec{L}_a \cdot \vec{L}'_a = \cos \beta \vec{\ell}_1 \cdot \vec{\ell}_1 + \sin \beta \vec{\ell}_1 \cdot \vec{\ell}_2. \quad (7.8)$$

Using $\vec{\ell}_1 \cdot \vec{\ell}_1 = 1$ and $\vec{\ell}_1 \cdot \vec{\ell}_2 = 0$, this expression simplifies to

$$\vec{L}_a \cdot \vec{L}'_a = \cos \beta. \quad (7.9)$$

The autocorrelation is then given by

$$\langle (\vec{L}_a \cdot \vec{L}'_a)^2 \rangle = \cos^2 \beta. \quad (7.10)$$

7.1.4.4 Computing the other terms

The other three terms in Eq. (7.7) are computed in precisely the same way. Using

$$\langle (\vec{\ell}_1 \cdot \vec{e}'_r)^2 \rangle = \frac{1}{3}, \quad \langle (\vec{\ell}_1 \cdot \vec{e}'_r)^4 \rangle = \frac{1}{5}, \quad (7.11)$$

$$\langle (\vec{\ell}_1 \cdot \vec{\ell}_2)(\vec{\ell}_1 \cdot \vec{e}'_r)(\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0, \quad (7.12)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}'_r)^2 (\vec{\ell}_2 \cdot \vec{e}'_r)^2 \rangle = \frac{1}{15}, \quad (7.13)$$

and

$$\langle (\vec{\ell}_1 \cdot \vec{e}'_r)(\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = \langle (\vec{\ell}_1 \cdot \vec{\ell}_2)(\vec{\ell}_1 \cdot \vec{e}'_r)^2 \rangle = \langle (\vec{\ell}_1 \cdot \vec{e}'_r)^3 (\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0, \quad (7.14)$$

we obtain

$$\langle (\vec{L}_a \cdot \vec{L}'_b)^2 \rangle = \frac{7}{15} \cos^2 \beta + \frac{4}{15} \sin^2 \beta. \quad (7.15)$$

As the correlation estimate is invariant under exchanging the pair, we can directly conclude that

$$\langle (\vec{L}_b \cdot \vec{L}'_a)^2 \rangle = \frac{7}{15} \cos^2 \beta + \frac{4}{15} \sin^2 \beta, \quad (7.16)$$

as well. In order to compute the last term, $\langle (\vec{L}_b \cdot \vec{L}'_b)^2 \rangle$, we need the following expectation values:

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)(\vec{\ell}_1 \cdot \vec{e}'_r) \rangle = \frac{1}{3} \vec{e}_r \cdot \vec{e}'_r \quad (7.17)$$

$$\langle (\vec{\ell}_1 \cdot \vec{\ell}_2)^2 \rangle = 0 \quad (7.18)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)(\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0 \quad (7.19)$$

$$\langle (\vec{\ell}_1 \cdot \vec{\ell}_2)(\vec{\ell}_1 \cdot \vec{e}_r)(\vec{\ell}_1 \cdot \vec{e}'_r) \rangle = \langle (\vec{\ell}_1 \cdot \vec{\ell}_2)(\vec{\ell}_1 \cdot \vec{e}_r)(\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0 \quad (7.20)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^2 (\vec{\ell}_1 \cdot \vec{e}'_r)^2 \rangle = \frac{2}{15} (\vec{e}_r \cdot \vec{e}'_r)^2 + \frac{1}{15} \quad (7.21)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^2 (\vec{\ell}_1 \cdot \vec{e}'_r)(\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0 \quad (7.22)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^2 (\vec{\ell}_2 \cdot \vec{e}'_r)^2 \rangle = \frac{2}{15} - \frac{1}{15} (\vec{e}_r \cdot \vec{e}'_r)^2 \quad (7.23)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^3 (\vec{\ell}_1 \cdot \vec{e}'_r) \rangle = \langle (\vec{\ell}_1 \cdot \vec{e}_r) (\vec{\ell}_1 \cdot \vec{e}'_r)^3 \rangle = \frac{1}{5} \vec{e}_r \cdot \vec{e}'_r \quad (7.24)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r) (\vec{\ell}_1 \cdot \vec{e}'_r)^2 (\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0 \quad (7.25)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^3 (\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = 0 \quad (7.26)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r)^2 (\vec{\ell}_1 \cdot \vec{e}'_r) (\vec{\ell}_2 \cdot \vec{e}_r) \rangle = 0 \quad (7.27)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r) (\vec{\ell}_1 \cdot \vec{e}'_r) (\vec{\ell}_2 \cdot \vec{e}'_r)^2 \rangle = \langle (\vec{\ell}_1 \cdot \vec{e}_r)^2 (\vec{\ell}_2 \cdot \vec{e}_r) (\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = \frac{1}{15} \vec{e}_r \cdot \vec{e}'_r \quad (7.28)$$

$$\langle (\vec{\ell}_1 \cdot \vec{e}_r) (\vec{\ell}_1 \cdot \vec{e}'_r) (\vec{\ell}_2 \cdot \vec{e}_r) (\vec{\ell}_2 \cdot \vec{e}'_r) \rangle = \frac{1}{10} (\vec{e}_r \cdot \vec{e}'_r)^2 - \frac{1}{30} \quad (7.29)$$

Given these expectation values, we can finally compute

$$\begin{aligned} \langle (\vec{L}_b \cdot \vec{L}'_b)^2 \rangle &= \left(\frac{7}{15} - \frac{8}{5} (\vec{e}_r \cdot \vec{e}'_r)^2 + \frac{32}{15} (\vec{e}_r \cdot \vec{e}'_r)^4 \right) \cos^2 \beta \\ &\quad + \left(\frac{4}{15} + \frac{4}{5} (\vec{e}_r \cdot \vec{e}'_r)^2 - \frac{16}{15} (\vec{e}_r \cdot \vec{e}'_r)^4 \right) \sin^2 \beta, \end{aligned} \quad (7.30)$$

which depends on the angular separation $\vec{e}_r \cdot \vec{e}'_r$ of the galaxy pair that is simulated. This dependence is inherited from flipping the radial component of *both* angular-momentum-orientation vectors due to an unknown front edge.

7.1.4.5 Determining the mixing angles

We now have all ingredients to derive the mixing angle β from the imposed correlation of Eq. (7.7). Inserting Eqs. (7.10), (7.15), (7.16) and (7.30), we obtain

$$\cos \beta = \sqrt{\frac{1}{3} + \frac{20\xi_{\text{input}}}{16(\vec{e}_r \cdot \vec{e}'_r)^4 - 12(\vec{e}_r \cdot \vec{e}'_r)^2 + 8}}. \quad (7.31)$$

where $\alpha = 0$ has been chosen previously. The denominator is always strictly positive because $0 \leq (\vec{e}_r \cdot \vec{e}'_r)^2 \leq 1$. Furthermore, there is a maximum input

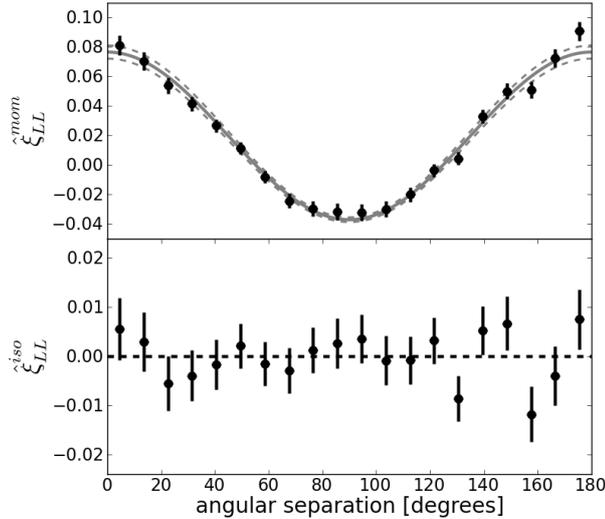


Figure 7.6: Angular autocorrelations of angular-momentum-orientation vectors. The autocorrelation estimate is now in angular separation not in real-space distance. Top panel: Angular autocorrelation for ellipticity estimates based on second moments. The bias model of Eq. (7.34) with 1σ errors is shown. Bottom panel: Angular autocorrelation for ellipticity estimates based on isophotes.

correlation where $\cos \beta = 1$. Given the minimal value of the denominator of $\frac{3}{8}$, the maximum input correlation is given by

$$\xi_{\max} = \frac{1}{80} = 0.0125 . \quad (7.32)$$

The existence of such an upper limit is obvious, since with unknown front edges it is impossible to reach full correlation. Similarly, there is also a minimal input correlation where the square-root becomes zero. Given the maximal value of the denominator for $(\vec{e}_r \cdot \vec{e}'_r)^2 = 1$ which is 12, this minimal input correlation is

$$\xi_{\min} = -\frac{1}{5} = -0.2 . \quad (7.33)$$

The existence of a lower limit is obvious, too, since the correlation is the squared projection of orientation vectors, i.e., it is impossible to reach full anticorrelation. Wherever we use this simulation method, we explicitly check that the result indeed exhibits the input correlation correctly.

7.1.5 Bias simulation

From our hypothesis of bulges biasing second moments, we can deduce the following prediction: If galactic bulges indeed bias second moments such that angular-momentum-orientation vectors are bent into the line of sight, the *angular* correlation function should exhibit a bias of the form

$$b(\theta) = A + B \cos^2 \theta , \quad (7.34)$$

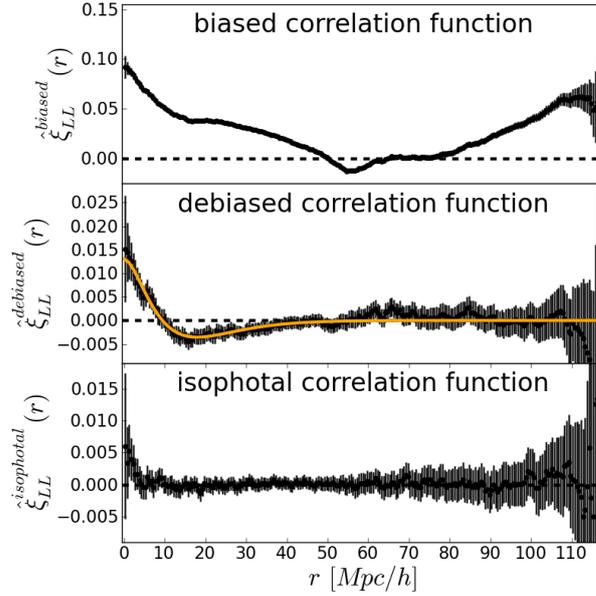


Figure 7.7: Debiasing the autocorrelation function of angular-momentum-orientation vectors.

Top panel: The biased autocorrelation function based on ellipticity estimates from second moments. Middle panel: “Debiased” correlation function where Eq. (7.34) has been subtracted from all pairwise projections. The solid line is the fit given by Eq. (7.35). Bottom panel: Autocorrelation function based on isophotal ellipticities.

where θ now denotes the angular separation of two galaxies. The parameters A and B depend on the details of the bias caused by the galactic bulges and are not generally predictable. This mathematical form stems from the bending of orientation vectors, such that the scalar product $\vec{L} \cdot \vec{L}'$ is *on average* equal to the cosine of the two galaxies’ separation angle. This prediction is confirmed by Fig. 7.6 which strongly suggests that $\hat{\xi}_{LL}(\theta)$ is dominated by this bias. Moreover, it would be difficult to find an alternative astrophysical explanation why the correlation increases for separation angles larger than 90° . This suspect behaviour is also exhibited by the autocorrelation function in real space, as shown in the top panel of Fig. 7.7. Concerning isophotal ellipticities, Fig. 7.6 also shows that $\hat{\xi}_{LL}(\theta)$ does not exhibit such a bias.⁵⁵

Is it possible to debias the autocorrelation function by subtracting Eq. (7.34) from all pairwise projections of angular-momentum-orientation vectors? We investigate this question in Fig. 7.7, where we show the biased and debiased autocorrelation function. Indeed, the debiased autocorrelation function looks very promising. For later modelling purposes, we parametrise the debiased autocorrelation function by

$$\xi_{LL}(r) \approx (0.013 + 0.002r - 0.00036r^2) \exp\left[-\frac{r}{6.1\text{Mpc}/h}\right], \quad (7.35)$$

where no error estimate is required since we only use this fit as input in simulations.

Is the debiased autocorrelation function trustworthy? For comparison,

⁵⁵Note that the angular correlation estimate in Fig. 7.6 looks worse than the spatial correlation estimate of Fig. 6.10. This is due to the fact that the angular correlation function does not use distance information.

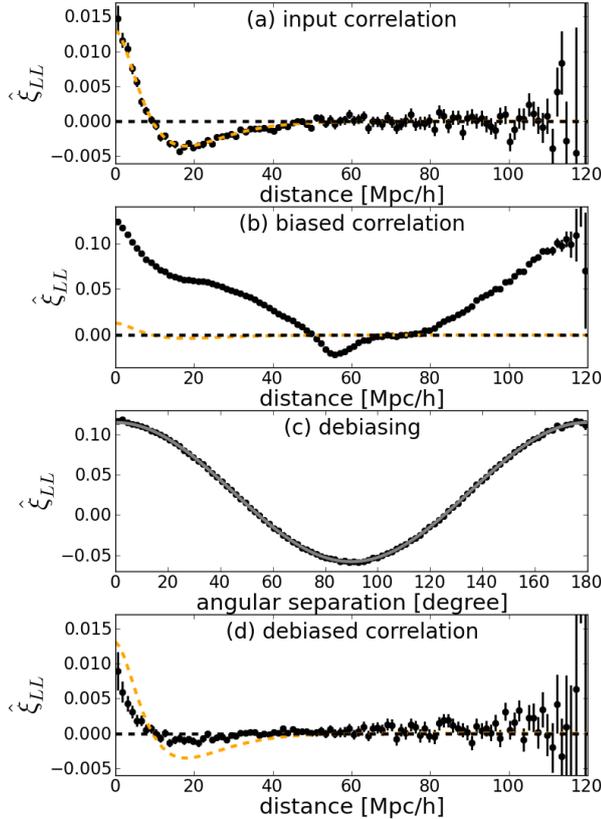


Figure 7.8: Self-consistency test of debiasing the autocorrelation function.

Panel (a): The input autocorrelation function as given by Eq. (7.35), validating our simulation technique. Panel (b): The biased autocorrelation function. Panel (c): The debiasing of the autocorrelation function in angular space. Panel (d): The debaised autocorrelation function, which exhibits significant deviations from the input.

Fig. 7.7 also shows the unbiased autocorrelation function based on isophotal ellipticities. Evidently, the debaised and isophotal autocorrelation functions do not agree. However, this does not necessarily rule out the debaised autocorrelation function. We actually expect that ellipticity estimates based on second moments are less noisy than isophotal ellipticity estimates since they use the whole light distribution instead of a single isophote. Hence, it is not a-priori implausible that the debaised autocorrelation function exhibits more information than the isophotal autocorrelation function. However, we did not account for the biases in orientation angles which are evident from Fig. 7.2. This bias favoured axis-parallel orientation angles, i.e., it may implant an orthogonality for pairs of angular-momentum-orientation vectors. In fact, the negative values observed in the debaised autocorrelation function of Fig. 7.7 require such an orthogonality.

In order to assess the trustworthiness of the debaised autocorrelation estimate, we conduct the following self-consistency test: We take the original galaxies as in Fig. 7.7, maintaining their true spatial positions, but when estimating the autocorrelation function, we replace the actual angular-momentum-orientation vectors by simulated vectors which exhibit the correlation function

given by Eq. (7.35). This simulation is described in Sect. 7.1.4. Panel (a) of Fig. 7.8 validates our simulation method. We then simulate the bias of second moments. For every galaxy, we take the simulated angular-momentum-orientation vector and infer the actual axis ratio q_{true} from it. Motivated by the left panel of Fig. 7.2, we then replace the true axis ratio by an “overestimate” drawn from the uniform distribution over the interval $[q_{\text{true}}, 1]$. Using this biased axis ratio, we recompute the angular-momentum-orientation vector and estimate the correlations. As shown in panel (b) of Fig. 7.8, the resulting biased autocorrelation function closely resembles the observation from Fig. 7.7. For debiasing, we then also estimate the autocorrelation in angular space, as shown in panel (c) of Fig. 7.8. Indeed, the estimate is dominated by a bias of the form of Eq. (7.35), i.e., our bias simulation is realistic. We then estimate the debiased autocorrelation function, which is shown in panel (d). Evidently, the debiased result exhibits systematic and significant deviations from the input autocorrelation function. We emphasise that the debiased result is *not* an obscured version of the input correlation function. Neither their difference nor their ratio is a constant, i.e., the debiasing was not successful. In simple words, the debiased result is *not* “weaker” than the input, it is “different”. Consequently, the debiasing is not self-consistent and the debiased autocorrelation estimate shown in Fig. 7.7 is *not* trustworthy.

7.1.6 Discussion

We demonstrated that ellipticity estimates based on second moments are strongly biased by galactic bulges even for Scd galaxies. In fact, Fig. 7.6 suggests that correlation estimates based on second moments are completely dominated by this bias which swamps the desired astrophysical signal. Therefore, we conclude that ellipticity estimates based on second moments overestimate axis ratios and thereby corrupt estimates of angular-momentum-orientation autocorrelation. This bias also corrupts similar correlation estimates, such as ellipticity autocorrelations (e.g. Blazek et al. 2011), leading us to overestimate the impact of disc alignment on weak-lensing studies. What are alternative ellipticity estimators? The bias also applies to adaptive moments (Bernstein & Jarvis 2002; Hirata & Seljak 2003) in this context. Furthermore, model-based ellipticity estimates are problematic, since nearby disc galaxies usually exhibit rich azimuthal structures, which are virtually impossible to model faithfully. The only kind of model designed to describe such rich azimuthal structure are basis-function expansions (e.g. Massey & Réfrégier 2005; Ngan et al. 2009), which unfortunately suffer from other severe conceptual problems (Melchior et al. 2010, Sects. 3.2.7 and 3.5.4). We have to conclude that isophotal ellipticities – though relying on a somewhat arbitrarily chosen isophote⁵⁶ – are the only useful ellipticity estimates for investigations of angular-momentum-orientation autocorrelation, since they are closest to the desired disc ellipticity.

There is yet another serious conceptual issue we have to face. In the weak-lensing context galaxies are usually rather small with radii of a few pixels only. In our case, however, we are considering large extended disc galaxies. These galaxies usually exhibit substructures such as galactic bars, rings or star-forming regions. In particular, the Scd galaxies considered by Lee (2011) and in this work typically exhibit very open spiral-arm patterns and other

⁵⁶The SDSS pipeline uses the 25 magnitudes per square arcsec isophote.
http://www.sdss.org/dr6/algorithms/classify.html#photo_stokes

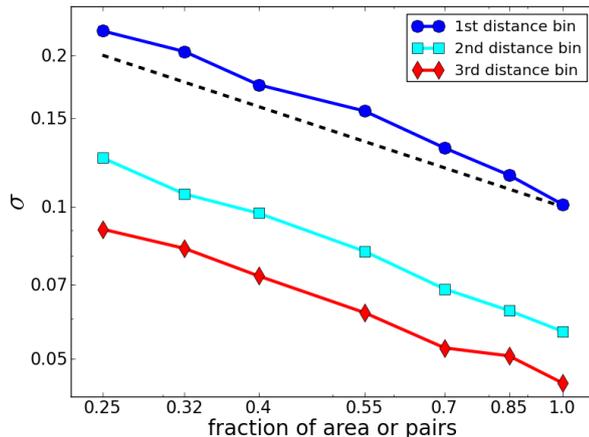


Figure 7.9: Impact of number statistics on the errors of handedness correlations.

We show how the errors of the innermost three distance bins in the marginal handedness autocorrelation function changes with number of galaxies. The x -axis shows the fraction of galaxy pairs selected from all pairs, which is equivalent to a survey covering the same fraction of the total survey area. Both axes are in logarithmic scale, i.e., the dependence of the errors is approximately a power law for all three bins. The dashed line indicates a power law according to $N^{-1/2}$, where N is the number of pairs in every bin.

prominent azimuthal structures such as bars or star-forming regions. For such objects, azimuthal structures are *not* a negligible second-order effect and “disc ellipticity” is not a well defined concept anymore.

7.2 Improving number statistics

An obvious strategy to improve estimates of handedness or angular-momentum-orientation autocorrelations is to increase the number of galaxies in the data sample. For instances, SDSS and thereby Galaxy Zoo cover approximately one quarter of the full sky. How would an extension to an all-sky survey improve the autocorrelation estimates? If we assume identical depth, this areal extension leaves the galaxy density unchanged, it only increases the number of galaxy pairs in all distance bins.

In order to study the improvement of a survey with larger coverage, we draw subsamples from the Galaxy Zoo database and estimate their handedness autocorrelations. A larger database is not available, so we use smaller databases to demonstrate the impact of number statistics. In fact, we do not draw the subsamples from the database itself, which would correspond to reducing the galaxy density. Instead, we randomly draw the subsamples from the list of galaxy *pairs*.⁵⁷ Figure 7.9 clearly shows that the errors in the handedness autocorrelation function are indeed dominated by number statistics, since the errors depend on sample size with a power law of exponent $-\frac{1}{2}$. Consequently, an extension from SDSS to full-sky coverage with SDSS quality would increase the database approximately threefold (the Milky Way obscures

⁵⁷Actually, we should select the galaxies by equatorial coordinates, defining subregions. However, our approach is equivalent and requires no new evaluation of the Galaxy Zoo data.

roughly one quarter of the sky) and thereby would decrease the errors by a factor of $\sqrt{3} \approx 1.7$. Given the results of Fig. 6.8, this would clearly be a major break-through in the measurability of potential handedness autocorrelation.

7.3 Improving redshift estimates

Substantial errors in redshift estimates clearly have an impact on the errors in the autocorrelation functions. However, for instances, the redshift error of $\sigma_z = 7.8 \cdot 10^{-5}$ at $z = 6.5993 \cdot 10^{-2}$ quoted in Fig. 6.4 corresponds to an error in the radial-velocity estimate of $\sigma_v = \frac{c\sigma_z}{(1+z)^2} \approx 20.6 \text{ km/s}$. Given the typical velocity dispersion of galaxies in small groups of $(202 \pm 10) \text{ km/s}$ and in large clusters of $(854 \pm 102) \text{ km/s}$ (Becker et al. 2007), the spectroscopic redshift estimates of SDSS are already picking up peculiar motions of individual galaxies instead of cosmological expansion. Consequently, further improving the accuracy of spectroscopic redshift estimates cannot improve estimates of, e.g., the handedness autocorrelation function.

Given the impact of uncertainties in spectroscopic redshift estimates on autocorrelation estimates (Sects. 6.4.4.3 and 6.4.5.3), it is obvious that larger surveys with *photometric* redshift estimates cannot help to improve the situation. Typically, uncertainties in photometric redshift estimates are two orders of magnitudes larger than uncertainties in spectroscopic redshift estimates (e.g. Csabai et al. 2003). Considering Fig. 6.4, this would lead to an error in the comoving distance of several tens of Mpc/h. Moreover, though there are many more galaxies with photometric redshift estimates than galaxies with spectroscopic redshift estimates (typically at least one order or magnitude more objects), these additional objects are typically also much fainter because selection for spectroscopic observations is usually triggered by the galaxy's brightness. The faintness of these additional objects would therefore also complicate the morphological classification. For a disc galaxy, the fainter the object, the more difficult it is to identify the disc. Consequently, surveys that offer only photometric but no spectroscopic redshift estimates are of no use to estimate handedness autocorrelation functions. This essentially rules out PanSTARRS⁵⁸ and LSST⁵⁹ because these surveys rely exclusively on photometric redshifts. Conversely, the EUCLID mission⁶⁰ will gather of the order of 100 million spectroscopic redshifts of galaxies. Unfortunately, the galaxy sample observed by EUCLID will have redshifts between 0.5 and 2. As was shown by Crittenden et al. (2001), estimates of handedness and angular-momentum-orientation correlations are compromised by weak-lensing signals for $z > 0.3$.

⁵⁸<http://pan-starrs.ifa.hawaii.edu/>

⁵⁹<http://www.lsst.org>

⁶⁰<http://sci.esa.int/euclid/>

7.4 Morphological classification in future surveys

Evidently, autocorrelation estimates of handedness and angular-momentum orientation require some morphological classification in future surveys. As we cannot probe high-redshift galaxies for this purpose, the morphological classes used by Galaxy Zoo or Huertas-Company et al. (2011) will be

sufficient and no further diversification is necessary. In particular, this implies that we can build on these two morphological catalogues to classify galaxies in future surveys: First, we match for the galaxies of known morphological types in a new survey. Second, we use a new survey’s imaging or spectroscopic data to estimate those galaxy’s parameters. Finally, using these parameters and the galaxies of known morphological types as a training sample, we can set up a probabilistic classification algorithm to extend this classification scheme to a new survey catalogue. In fact, this is precisely the same exercise as Huertas-Company et al. (2011) did, but on much larger scale. In particular, the Galaxy Zoo sample with approximately 900,000 visually classified galaxies would provide an extremely valuable training sample. Gauci et al. (2010) demonstrated that modern classification algorithms perform excellently in reproducing the visual classifications of the Galaxy Zoo sample. This strategy has several advantages: It is easily conductible, it does not require much computational time, and it is highly accurate and objective.

7.5 Front-edge estimation

With so little information in the data, using additional information can be very helpful. Such additional information is provided by an estimate of the disc’s front edge, i.e., which side of the semi-minor axis is pointing towards us. If we can estimate the front-edge, we can use the results as weights p_a and p_b in the correlation estimator of Eq. (6.19). Evidently, if we knew the front edge of every galaxy in our data sample, this would break the geometric degeneracy in the angular-momentum-orientation vector and thereby would improve the correlation estimate.

7.5.1 Visual classification

We estimate the front-edge by looking for dust extinction, in particular dust lanes. We visually inspect g -band images, since of all five SDSS bands this band is most strongly affected by dust extinction while still being of decent depth. The outcome of such a visual inspection is as follows:

- Equal weights $p_a = p_b = \frac{1}{2}$ if we are uncertain.
- Weight of 0.6 to indicate a somewhat uncertain trend.
- Weight of 0.9 if we believe to be certain.

We do not assign a weight of 1 in the last case, since there is always some uncertainty. By construction, this method works best for strongly inclined discs, since face-on discs may display dust lanes but a front-edge does not exist. Unfortunately, knowing the front-edge would have a larger impact for nearly face-on discs than for edge-on discs (see definitions in Lee 2011). We visually inspected g -band images of the 500 largest galaxies, sorted by their Petrosian radii.⁶¹ For smaller galaxies, the resolution is not good enough to identify dust lanes. This yields a sample of 40 disc galaxies with certain front-edge classifications (weight 0.9) and 39 disc galaxies with somewhat uncertain results (weight 0.6). Given these numbers of only very few decisive front-edge

⁶¹We do not find evidence for psychological preference in visual front-edge classifications similar to the bias in handedness classifications reported by Land et al. (2008).

classifications, we find no substantial improvement of the marginal correlation estimate. Nevertheless, future sky surveys may have an improved imaging quality, such that a visual front-edge classification is possible for more objects.

7.5.2 Automated classification

It would definitely be beneficial to obtain a front-edge classification for galaxies with intermediate inclinations, since the rounder the object the larger the information gain. Unfortunately, visual classification via dust lanes is restricted to highly inclined discs. Therefore, the front edge needs to be inferred in a different way, which should ideally be fully automated in order to ensure objectiveness. One potential approach is front-edge classification via colour gradients from dust extinction. However, this requires highly accurate photometric positions. In simple tests, we experienced that already coordinate offsets between the different bands of a hundredth of a pixel along the semi-minor axis can compromise such estimates, due to the rapidly falling radial light profiles of galaxies. Another approach is front-edge classification via dust extinction in single-band photometry. In the case of SDSS, this would ideally be the g -band, where the impact of dust extinction is larger than in r, i, z whereas the g -band is not as shallow as the u -band. This front-edge estimator would compare the fluxes above and below the major axis, which could differ due to unequal dust extinction. In contrast to colour-based methods, this approach does not rely on accurate photometric positions. We tested a number of simple implementations but the results were very poor. Apparently, the desired signal was not large enough, such that other effects, e.g., star-forming regions in the galaxy or foreground stars, compromise colour gradients and flux differences. These effects are the major obstacles which have to be overcome in order to set up a reliable front-edge classification algorithm.

— The bottom line —

- Second moments of the light distribution are strongly biased by galactic bulges even for Scd galaxies. More precisely, galaxies seem to be less inclined using ellipticity estimates based on second moments than using estimates based on isophotes. This bends the angular-momentum-orientation vectors into the line-of-sight, thereby creating artificial correlations. For instances, this leads us to overestimate the impact of disc alignment as a systematic effect in gravitational weak lensing.
- It is possible to reduce the errors in the correlation estimates by increasing the number of galaxies in the data sample. Already going from SDSS to a full-sky survey of SDSS quality or better would reduce the errors to an extent that autocorrelations in handedness and angular-momentum-orientation vectors could become statistically significant.
- Photometric redshift estimates have too large uncertainties. Although typically there are many more galaxies with photometric redshift estimates than with spectroscopic redshift estimates, these number statistics do not outweigh the increase in errors. Unless the accuracy of photometric redshift estimates is substantially improved in comparison to SDSS, spectroscopic redshift estimates are required. This rules out the use of PanSTARRS and LSST.
- Future surveys will compile galaxy catalogues surpassing the SDSS database in sky coverage and depth. Recent work successfully demonstrated that visual classifications of galaxy morphologies can be used to train fully automated classification algorithms which can then be used to classify more galaxies. In particular, the Galaxy Zoo sample provides a powerful training sample.
- Front-edge classification of inclined disc galaxies would help to optimally exploit the given data because it breaks one geometric degeneracy involved in the inference of angular-momentum-orientation vectors. For highly inclined disc galaxies, front edges can be detected using dust lanes. For less inclined disc galaxies, the overall dust extinction may provide a proxy. However, this turned out to be a challenge that definitely requires a sophisticated solution.

CHAPTER 7. IMPROVEMENTS AND POTENTIAL OF
FUTURE SURVEYS

8

Summary and outlook

As we have seen in Chapter 4, the angular momenta of the four large disc galaxies in the Local Group – Milky Way, Andromeda (M31), M33, and the Large Magellanic Cloud – are consistent with the null hypothesis of random orientation. Motivated from these results, we expanded our test of disc alignment to the SDSS spectroscopic galaxy sample using autocorrelation functions of spiral-arm handedness and angular-momentum-orientation vectors, respectively. We have seen that there are several important error sources that have to be taken into account, namely uncertainties in morphological classification, errors in redshift estimates and errors in ellipticity estimates. From our detailed investigations of the impact of these errors, we conclude that it is inevitable to incorporate them into the estimation process. Propagating all these error sources, we obtain marginal autocorrelation functions which do not exhibit statistically significant autocorrelations. Our results therefore falsify previous reports of such findings (Slosar et al. 2009; Lee 2011). Nevertheless, we could test the theoretically predicted autocorrelation length of 1Mpc/h (Schäfer & Merkel 2011), which is consistent with our data. However, due to the large errors in the autocorrelation estimates, the constraint on the autocorrelation length is subject to large uncertainties. We have to conclude that the SDSS data – which is the currently best database for such investigations – does not enable us to place decisive constraints on theoretical parameters of disc alignment or even to confirm the presence of alignment effects.

Do our results falsify the prediction of the tidal-torque theory? Not necessarily. The prediction concerns the alignment for angular momenta of dark-matter haloes and *not* for the disc galaxies residing inside these haloes. For instances, van den Bosch et al. (2002) find a median misalignment of angular momenta of disc galaxies and their host haloes of $\approx 30^\circ$. Furthermore, even minor mergers can significantly disturb the angular momenta of disc galaxies by transferring orbital angular momentum (e.g. Moster et al. 2010). Hence, numerical simulations suggest that an initial alignment is perturbed and may not survive. Conversely, we could speculate whether there is some relaxation mechanism, e.g., compensating for perturbations by mergers, which has not been taken into account by numerical simulations. However, we do not want

to push this discussion too far because we are wary of turning the tidal-torque theory from an empirical into a “vampirical” hypothesis (Gelman & Weakliem 2009), where virtually any observational result can be explained such that an empirical falsification becomes impossible. The best strategy is certainly to enlarge the database of disc galaxies in order to reduce the errors through number statistics.

Furthermore, we have revealed a systematic misestimation of ellipticities based on second moments of the light distribution which is induced by an ellipticity gradient between galactic disc and galactic bulge (Bernstein 2010). This leads to a substantial bias in the autocorrelation function of angular-momentum-orientation vectors which overwrites a potential astrophysical signal by roughly an order of magnitude.

Our results concerning the relevance of errors sources for estimates of autocorrelation functions also applies to further astrophysical and cosmological investigations, other than disc alignment. This is very typical for methodological investigations which are to some extent independent of the astrophysical or cosmological question. For instances, auto- and cross-correlation functions of real and imaginary parts of complex ellipticities in investigations of weak gravitational lensing need to take into account these error sources (e.g. Blazek et al. 2011). Furthermore, our findings also apply to two-point autocorrelation functions of galaxies in investigations of baryonic acoustic oscillations (e.g. Blake et al. 2011). Such autocorrelation functions need to account for uncertainties in redshift estimates, too. It may also be mandatory to account for uncertainties in the star-galaxy classification.

In preparation of data selection for estimates of autocorrelation functions, we were also concerned with the parametrisation of galaxy morphologies. Such a parametrisation is necessary in order to automatically identify disc galaxies in large data samples containing numerous galaxies such as the SDSS catalogue. In Chapter 2, we studied the performance of several very popular morphological parameters, e.g., the concentration index. However, we revealed the existence of an intertwinement of morphological observables such as light concentration, ellipticity and asymmetry. This intertwinement questions the conceptual setup of most of these parametrisation schemes and severely limits their practical application. Our findings challenge the current paradigm which favours model-independent parametrisation schemes due to their seeming simplicity. Based on these results, we identified basis-function expansions as a very promising approach due to its high flexibility in describing azimuthal structures such as spiral-arm patterns which are exhibited by disc galaxies. Unfortunately, the astrophysical interpretation of the expansion coefficients of such basis functions is not obvious and more work is needed here as soon as a reliable set of basis functions has been identified. We continued by orthonormalising the Sérsic radial profile in Chapter 3. We investigated this promising approach in great detail but revealed two severe problems, namely that these “sérsiclets” are prone to undersampling and that they impose an unphysical relation between steepness of the radial bulge profile and spatial scale of disc patterns such as spiral arms. Therefore, we were forced to enhance the simple Sérsic radial profile and orthonormalise these profiles. Indeed, the resulting set of basis functions overcomes both problems of sérsiclets but turns out to be computationally infeasible. Another promising approach was recently pub-

lished by Jiménez-Teja & Benítez (2011) who expand galaxy morphologies into Chebyshev rational functions. However, their handling of the scale radius is a work-around and the authors also did not investigate how their basis functions perform in the regime of low resolution. Further investigations are necessary in order to assess the reliability of this approach.

Moreover, we demonstrated in Chapter 2 that all these parametrisation schemes form nonlinear parameter spaces. This is a severe problem because virtually all classification algorithms require the definition of a distance metric (e.g. Fraix-Burnet et al. 2009). We have shown that the naïve Euclidean metric is a very poor approximation, while the true nonlinear metric remains unknown. Even worse, we have seen that several morphological parameters form discontinuous parameter spaces, i.e., similar galaxy morphologies are not guaranteed to have similar parameter values. Evidently, this can corrupt any classification attempt. Dissatisfied with these problems of morphological parametrisation schemes in the classification context, we resorted to visual classifications of galaxy morphologies, which are provided by the Galaxy Zoo project. In general, we have to conclude that the field of parametrising galaxy morphologies is still open. So far, there appear to be no trustworthy parametrisation schemes, although many (over-)simplified workarounds are often used uncritically. More work and sophistication is required in order to fully exploit the information content of galaxy morphologies, which provide an important diagnostic and are a direct observable.

8.1 Outlining an observational strategy

As we have discussed earlier, there are no planned or upcoming sky surveys that provide spectroscopic redshift estimates of sufficiently large numbers of galaxies in the relevant redshift regime. This would imply that testing the existence of disc alignment through autocorrelation functions would remain an unsolved problem for the unforeseeable future. This situation is of course unacceptable, wherefore we now outline an observational strategy. This strategy aims at assembling a database which enables us to test disc alignment with autocorrelation functions.

The strategy builds on upcoming photometric surveys such as PanSTARRS or LSST. As mentioned before, these surveys only provide photometric redshift estimates and imaging data but no spectroscopic redshifts. However, we can use these surveys to build a sample of galaxies which can in the future be targeted specifically in order to take spectra and estimate spectroscopic redshifts. Photometry can be used for morphological classification in order to identify Scd galaxies, as described in Sect. 7.4. Furthermore, photometry already enables us to estimate ellipticities of identified Scd galaxies. Photometric redshifts can then be used to identify Scd galaxies in the relevant redshift regime. Here, we have to account for the completeness of the Scd-galaxy sample. In the case of SDSS, Scd galaxies are complete up to the redshift $z \approx 0.02$ and Lee (2011) found $\approx 4,000$ Scd galaxies in this regime. Consequently, a full sky survey of SDSS depth would contain $\approx 12,000$ Scd galaxies in this redshift regime.⁶² This would require to take spectra of $\approx 8,000$ additional target objects distributed primarily over the Southern hemisphere.

⁶²The Milky Way covers roughly one quarter of the sky, such that the increase is only a factor of three not four.

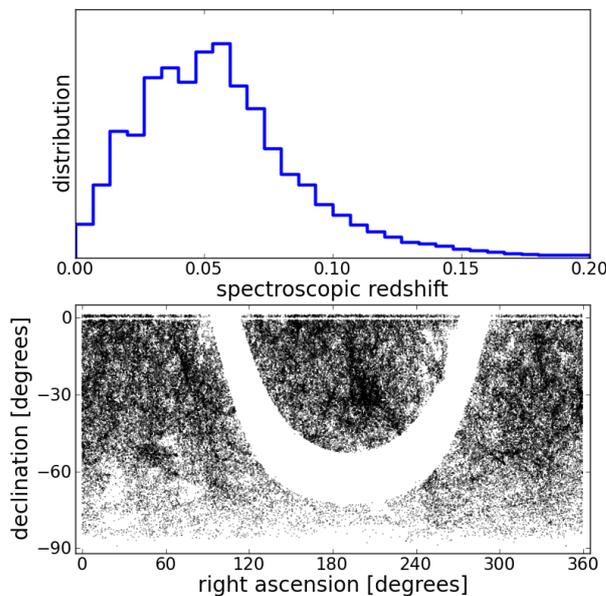


Figure 8.1: 6dF survey of spectroscopic redshifts on southern hemisphere. Top panel: Distribution of spectroscopic redshifts. Bottom panel: Distribution of target galaxies over the southern hemisphere.

As PanSTARRS is planned to be slightly deeper than SDSS, completeness will also hold up to redshifts slightly larger than $z \approx 0.02$, i.e., more distant Scd galaxies can be included. In total, the final number of Scd galaxies which require spectral redshifts to be estimated will not exceed 15,000 to 20,000. Increasing the data sample by approximately a factor of four therefore approximately halves the errors in the marginal autocorrelation estimates. Given the marginal autocorrelation estimates of angular-momentum-orientation vectors shown in Figs. 6.11 and 7.7, this would be a major breakthrough because then the currently weak indications might become detections of high statistical significance. Such an investigation would be highly relevant for our understanding of disc-galaxy formation in the cosmological context.

Unfortunately, spectral observations of 20,000 targets distributed over the southern hemisphere is an infeasible task. The southern hemisphere is 32,400 square degrees large, i.e., on average there is less than one target per square degree. Consequently, multi-object spectroscopy would not be possible and almost every target object would require its own telescope pointing. Given a typical exposure time of 20-30 minutes in order to obtain a decent spectrum, this would require 10,000 hours of observation time. Fortunately, many of these additional 20,000 Scd galaxies might be contained in the 6dF survey (Jones et al. 2004) which estimated spectroscopic redshifts of 150,000 galaxies in the southern hemisphere.⁶³ Figure 8.1 shows that the distribution of spectroscopic redshifts is very similar to the redshift distribution of SDSS shown in Fig. 5.1. However, the survey area of 6dF is twice as large as the survey area of SDSS though the 6dF catalogue contains a factor of ≈ 6 fewer objects than the SDSS spectroscopic sample. Consequently, the 6dF survey is much shallower than the SDSS survey, i.e., Scd galaxies will not be complete

⁶³<http://www.aao.gov.au/local/www/6df/>

CHAPTER 8. SUMMARY AND OUTLOOK

out to redshift $z \leq 0.02$. Nevertheless, joining photometric information from, e.g., PanSTARRS, and spectroscopic redshift estimates from the 6dF survey, we may be able to increase the database of Scd galaxies sufficiently in order to improve the statistical significance of autocorrelation estimates in the foreseeable future.

CHAPTER 8. SUMMARY AND OUTLOOK

Acknowledgements

First and foremost, I whole-heartedly thank my wife Ellen Andrae for her continuous support both privately and professionally. We had lots of discussions on my work, concerning pros and cons of different methods and the interpretation of the results. Furthermore, she was (almost) always very tolerant whenever a sudden “inspiration” overcame me and I sat down to work – regardless of late evenings or weekends. Of course, I also thank my newborn son Thomas for being such a cute, quiet, sleepy, and well-behaved baby. Both of them enabled me to largely concentrate on finishing my research and writing up my thesis.

This work would not have been possible without the support of my supervisor Knud Jahnke. He offered guidance and useful insights into the observer’s perspective whenever it was necessary. Otherwise, he allowed me to follow my own research interests and to independently navigate between dead ends, the falsification of previous results, and all sorts of superpositions of these two. Knud exhibited an admirable patience with a stubborn theoretician and methodologist like myself who was always ready at hand with extremist convictions about “suboptimal” scientific work. He also offered me the unique opportunity to participate in real astronomical observations at Calar Alto, Spain. In the beginning, I was not overwhelmingly excited about this, but once there the original fascination of astronomy returned and today I consider it as a very important and interesting experience. Furthermore, Knud bought a modern multi-core server for his group, whose considerable computational power I could use almost alone. This great technical facility certainly also contributed to the fast progress of my work.

I also thank my co-advisers Matthias Bartelmann and Coryn Bailer-Jones for their support in PhD committee meetings and whenever else I asked for their help. Matthias was the one who suggested that I should apply to the Klaus-Tschira Foundation for a PhD scholarship, which allowed me to concentrate on my research without having to care about finances and contracts. As former dean of the physics faculty, Matthias was also a valuable source of information concerning the bureaucratic issues of the PhD. Furthermore, after certain unpleasant (and unjustified) feedback, it was Matthias who restored my enthusiasm and zeal to continue with my scientific work. Likewise, Coryn supported me in my understanding of methodology as a crucial part of science and eventually offered me a great position in his Gaia group at MPIA that

perfectly matches my interests in statistical inference from data.

Many thanks also go to all with whom I had the pleasure to work together. Primarily, I would like to thank Peter Melchior who taught me a lot of how to construct scientific lines of arguments and especially the philosophy of mercilessly exposing any method to rigorous and profound testing. Furthermore, Peter also provided the LaTeX source for the fancy bottom lines at the end of each chapter. I am also very thankful for the numerous interesting discussions about data analysis with Tim Schulze-Hartung, especially concerning the falsification of a recently claimed “discovery” of an Earth-like planet in the habitable zone of Gliese 581. Tim told me a lot about the problems involved in the detection of exoplanets.

My current and former officemates also have their share in my thesis. In an office housing five students, work could have been very hard for various reasons. Fortunately, I never experienced any annoyances and it was always a great pleasure to enter the office and work there. Due to this excellent working atmosphere, I am much obliged to Marie-Helene Nicol, Jan Pitann, Dading Nugroho, Natalia Kudryavtseva, Tim Schulze-Hartung and Gabriele Maier.

I would also like to thank everyone in Knud’s Emmy-Noether group, namely Katherine Inskip, Dading Nugroho and Mauricio Cisternas. In our regular group meetings, I learned a lot of reducing observational data in particular spatially resolved spectra. They also were very patient listening to the sometimes somewhat theoretical excursions of my own work.

As member of the International Max-Planck Research School (IMPRS) I want to thank the IMPRS coordinator Christian Fendt for his commitment and ready help, e.g., concerning organisational and bureaucratic aspects. Furthermore, I would like to thank my fellow students from the fifth IMPRS generation: Ellen Andrae, Paul Boley, Federica Capranico, Gustavo Dopcke, Aram Giahi, Philipp Girichidis, Alessandra Grassi, Oleksiy Golubov, Meiert Grootes, Raoul Haschke, Mathias Jaeger, Fazeel Mahmood Khan, Ervin Kafexhiu, Natalia Kudryavtseva, Eva Lefa, Chia-Chun Lu, Johannes Ludwig, Iwona Mochol, Faviola Molina, Natalie Raettig, Daniel Seifried, Jochen Tackenberg, and Tessel van der Laan. Together, we endured long seminars, enjoyed post-seminar beers, explored the city of Amsterdam, made the Eurovision Song Contest sufferable with self-made cocktails, and baked German cookies before Christmas. Together, we also mourned for our fellow student Crystal Basseur who deceased in Spring 2011.

Last but not least, I also thank the Klaus-Tschira Foundation for their financial support and the internet community for helpful tips on LaTeX to design this thesis layout.

Bibliography

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, 182, 543
- Abraham, R. G., Tanvir, N. R., Santiago, B. X., et al. 1996, *MNRAS*, 279, L47
- Abraham, R. G., Valdes, F., Yee, H. K. C., & van den Bergh, S. 1994, *ApJ*, 432, 75
- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218
- Andrae, R., Bartelmann, M., Rix, H.-W., Gennaro, M., & Karim, A. in prep.
- Andrae, R., Melchior, P., & Bartelmann, M. 2010b, *A&A*, 522, A21+
- Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010c, arXiv e-prints 1012.3754
- Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, *MNRAS*, 393, 1324
- Barlow, R. 1993, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley VCH)
- Bartelmann, M. & Schneider, P. 2001, *Phys. Rep.*, 340, 291
- Becker, M. R., McKay, T. A., Koester, B., et al. 2007, *ApJ*, 669, 905
- Bellman, R. 1961, *Adaptive Control Processes: A Guided Tour* (Princeton University Press)
- Bernstein, G. M. 2010, *MNRAS*, 406, 2793
- Bernstein, G. M. & Jarvis, M. 2002, *AJ*, 123, 583
- Berry, R. H., Hobson, M. P., & Withington, S. 2004, *MNRAS*, 354, 199
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, *AJ*, 119, 2645

- Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
- Blake, C., Davis, T., Poole, G., et al. 2011, ArXiv e-prints 1105.2862
- Blazek, J., McQuinn, M., & Seljak, U. 2011, ArXiv e-prints
- Bridle, S., Balan, S. T., Bethge, M., et al. 2010, *MNRAS*, 405, 2044
- Bridle, S., Shawe-Taylor, J., Amara, A., et al. 2009, *Annals of Applied Statistics*, 3, 6
- Brunthaler, A., Reid, M. J., Falcke, H., Greenhill, L. J., & Henkel, C. 2005, *Science*, 307, 1440
- Cameron, E. 2010, ArXiv e-prints 1012.0566
- Caon, N., Capaccioli, M., & D’Onofrio, M. 1993, *MNRAS*, 265, 1013
- Chang, T., Refregier, A., & Helfand, D. J. 2004, *ApJ*, 617, 794
- Chemin, L., Carignan, C., & Foster, T. 2009, *ApJ*, 705, 1395
- Cioni, M.-R. L., Irwin, M., Ferguson, A. M. N., et al. 2008, *A&A*, 487, 131
- Ciotti, L. & Bertin, G. 1999, *A&A*, 352, 447
- Conselice, C. J. 2003, *The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories*
- Conselice, C. J., Bershadsky, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
- Corbelli, E. & Schneider, S. E. 1997, *ApJ*, 479, 244
- Crittenden, R. G., Natarajan, P., Pen, U., & Theuns, T. 2001, *ApJ*, 559, 552
- Csabai, I., Budavári, T., Connolly, A. J., et al. 2003, *AJ*, 125, 580
- de Vaucouleurs, G. 1948, *Annales d’Astrophysique*, 11, 247
- Dressler, A. 1980, *ApJ*, 236, 351
- Elmegreen, B. G., Elmegreen, D. M., Vollbach, D. R., Foster, E. R., & Ferguson, T. E. 2005, *ApJ*, 634, 101
- Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, *ApJ*, 665, 265
- Ferreras, I., Lisker, T., Pasquali, A., & Kaviraj, S. 2009, *MNRAS*, 395, 554
- Fluke, C. J., Barnes, D. G., Barsdell, B. R., & Hassan, A. H. 2011, *PASA*, 28, 15
- Fraix-Burnet, D., Davoust, E., & Charbonnel, C. 2009, *MNRAS*, 398, 1706
- Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, *ApJ*, 553, 47
- Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, *AJ*, 134, 579

- Gauci, A., Zarb Adami, K., & Abela, J. 2010, ArXiv e-prints 1005.0390
- Gelman, A. & Weakliem, D. 2009, *American Scientist*
- Glasser, C. 1962, *Journal of the American Statistical Association*, 57, 648
- Goldberg, D. M. & Bacon, D. J. 2005, *ApJ*, 619, 741
- Graham, A. W. & Driver, S. P. 2005, *Publications of the Astronomical Society of Australia*, 22, 118
- Hamilton, A. J. S. 1998, in *Astrophysics and Space Science Library*, Vol. 231, *The Evolving Universe*, ed. D. Hamilton, 185–+
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer-Verlag)
- Haynes, M. P. & Giovanelli, R. 1984, *AJ*, 89, 758
- Heavens, A. 2009, ArXiv e-prints 0906.0664
- Heidmann, J., Heidmann, N., & de Vaucouleurs, G. 1972, *MmRAS*, 75, 85
- Hirata, C. & Seljak, U. 2003, *MNRAS*, 343, 459
- Huertas-Company, M., Aguerri, J. A. L., Bernardi, M., Mei, S., & Sánchez Almeida, J. 2011, *A&A*, 525, A157+
- Iye, M. & Ozawa, T. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 182, *Galaxy Dynamics - A Rutgers Symposium*, ed. D. R. Merritt, M. Valluri, & J. A. Sellwood, 255–+
- Jiménez-Teja, Y. & Benítez, N. 2011, ArXiv e-prints 1104.0683
- Jones, D. H., Saunders, W., Colless, M., et al. 2004, *MNRAS*, 355, 747
- Kelly, B. C. & McKay, T. A. 2004, *AJ*, 127, 625
- Kelly, B. C. & McKay, T. A. 2005, *AJ*, 129, 1287
- Kennicutt, Jr., R. C. 1998, *ARA&A*, 36, 189
- Kroupa, P. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 285, *Modes of Star Formation and the Origin of Field Populations*, ed. E. K. Grebel & W. Brandner, 86–+
- Land, K., Slosar, A., Lintott, C., et al. 2008, *MNRAS*, 388, 1686
- Landsman, W. B., Roberts, M. S., Bohlin, R. C., et al. 1992, *ApJL*, 401, L83
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, *ApJS*, 192, 16
- Lee, J. 2011, *ApJ*, 732, 99
- Lee, J. & Erdogdu, P. 2007, *ApJ*, 671, 1248

- Lee, J. & Pen, U. 2008, *ApJ*, 681, 798
- Lin, Y., Shen, Y., Strauss, M. A., Richards, G. T., & Lunnan, R. 2010, *ApJ*, 723, 1119
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, 410, 166
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179
- Lisker, T. 2008, *ApJS*, 179, 319
- Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, *ApJ*, 672, 177
- Lotz, J. M., Madau, P., Giavalisco, M., Primack, J., & Ferguson, H. C. 2006, *ApJ*, 636, 592
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
- MacArthur, L. A., Courteau, S., & Holtzman, J. A. 2003, *ApJ*, 582, 689
- MacKay, D. 2008, *Information Theory, Inference, and Learning Algorithms*
- Macri, L. M., Stanek, K. Z., Sasselov, D. D., Krockenberger, M., & Kaluzny, J. 2001, *AJ*, 121, 870
- Massey, R. & Réfrégier, A. 2005, *MNRAS*, 363, 197
- Melchior, P., Andrae, R., Maturi, M., & Bartelmann, M. 2009, *A&A*, 493, 727
- Melchior, P., Böhnert, A., Lombardi, M., & Bartelmann, M. 2010, *A&A*, 510, A75+
- Melchior, P., Meneghetti, M., & Bartelmann, M. 2007, *A&A*, 463, 1215
- Miller, M. A., Amon, L. M., & Reinhardt, W. P. 2000, *Chemical Physics Letters*, 331, 278
- Moster, B. P., Macciò, A. V., Somerville, R. S., Johansson, P. H., & Naab, T. 2010, *MNRAS*, 403, 1009
- Neal, R. M. 2003, *The Annals of Statistics*, 31, 705
- Nelder, J. A. & Mead, R. 1965, *The Computer Journal*, 7, 308
- Ngan, W., van Waerbeke, L., Mahdavi, A., Heymans, C., & Hoekstra, H. 2009, *MNRAS*, 396, 1211
- Olsen, K. A. G. & Massey, P. 2007, *ApJL*, 656, L61
- Pasha, I. I. 1985, *Soviet Astronomy Letters*, 11, 1
- Pellerin, A. & Macri, L. M. 2011, *ApJS*, 193, 26
- Pen, U., Lee, J., & Seljak, U. 2000, *ApJL*, 543, L107
- Petrosian, V. 1976, *ApJL*, 209, L1

- Press, W., Teukolsky, S., Vetterling, W., & Flanner, B. 2002, *Numerical Recipes in C++* (Cambridge University Press)
- Putman, M. E., Peek, J. E. G., Muratov, A., et al. 2009, *ApJ*, 703, 1486
- Réfrégier, A. 2003, *MNRAS*, 338, 35
- Refregier, A. & Bacon, D. 2003, *MNRAS*, 338, 48
- Richards, J. W., Freeman, P. E., Lee, A. B., & Schafer, C. M. 2009, *ApJ*, 691, 32
- Rosa-González, D., Terlevich, E., & Terlevich, R. 2002, *MNRAS*, 332, 283
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2010, *ArXiv e-prints* 1012.3002
- Sargent, M. T., Carollo, C. M., Lilly, S. J., et al. 2007, *ApJS*, 172, 434
- Scarlata, C., Carollo, C. M., Lilly, S., et al. 2007, *ApJS*, 172, 406
- Schäfer, B. M. 2009, *International Journal of Modern Physics D*, 18, 173
- Schäfer, B. M. & Merkel, P. 2011, *ArXiv e-prints*
- Sérsic, J. L. 1968, *Atlas de galaxias australes* (Cordoba, Argentina: Observatorio Astronomico, 1968)
- Shen, Z.-Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P., & Zhao, J.-H. 2005, *Nature*, 438, 62
- Simmat, E., Tuffs, R. J., & Popescu, C. C. 2010, in *American Institute of Physics Conference Series*, Vol. 1240, *American Institute of Physics Conference Series*, ed. V. P. Debattista & C. C. Popescu, 87–88
- Slosar, A., Land, K., Bamford, S., et al. 2009, *MNRAS*, 392, 1225
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, 124, 1810
- Sugai, H. & Iye, M. 1995, *MNRAS*, 276, 327
- Toomre, A. & Toomre, J. 1972, *ApJ*, 178, 623
- Trujillo, I., Graham, A. W., & Caon, N. 2001, *MNRAS*, 326, 869
- van den Bosch, F. C., Abel, T., Croft, R. A. C., Hernquist, L., & White, S. D. M. 2002, *ApJ*, 576, 21
- van der Marel, R. P. & Cioni, M.-R. L. 2001, *AJ*, 122, 1807
- van der Wel, A., Bell, E. F., Holden, B. P., Skibba, R. A., & Rix, H. 2010, *ApJ*, 714, 1779
- van Dokkum, P. & Conroy, C. 2011, *ArXiv e-prints*
- van Dokkum, P. G. & Conroy, C. 2010, *Nature*, 468, 940

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

Walterbos, R. A. M. & Kennicutt, Jr., R. C. 1988, A&A, 198, 61

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579