Lens Models for Compact and Extended Sources

Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Physik
der Universität Hamburg

vorgelegt von
Olaf Wucknitz
aus Hamburg

Hamburg
2002

Gutachter der Dissertation: Prof. Dr. Sjur Refsdal
Dr. Neal Jackson

Gutachter der Disputation: Prof. Dr. Sjur Refsdal
Prof. Dr. Joachim Wambsganß

Datum der Disputation: 14. Mai 2002

Dekan des Fachbereichs Physik und
Vorsitzender des Promotionsausschusses: Prof. Dr. Friedrich-Wilhelm Büßer

Abstract

This work discusses the determination of cosmological parameters, especially the Hubble constant, from observations of gravitational lenses. The lens method has the advantage that it depends on the understanding of only very little astrophysics. This allows very robust results and makes estimates of the remaining uncertainties relatively simple. The most important contribution to possible errors is given by the mass models for the lensing galaxies. It is thus necessary to obtain good estimates of these uncertainties and to reduce them as much as possible. We present analytical calculations for a general family of power-law lens models with arbitrary angular shape plus external perturbations. The latter are parametrized as external shear. To include all constraints for optimal unresolved multiple image lens systems, we examine quadruple systems using the image positions and the three independent time-delays as constraints. It is well known that the radial mass distribution has important effects on the determination of the Hubble constant. Our calculations result in a generic and exact scaling relation for the dependence of the Hubble constant H_0 on the power-law exponent of the potential's radial part β. This scaling is the same for all lens systems in this family of models. Systematic errors in the assumed β will therefore lead to a systematic error in H_0. The effect of external shear is quantified by the new concept of a 'critical shear'. For an external shear exactly equal to this value, all time-delays vanish.

To improve the situation, the parameters of shear and radial mass distribution have to be measured accurately. Multiply imaged unresolved sources, which are commonly used for this purpose, can provide only a limited number of constraints. It is therefore important to study lens systems with extended sources, which can constrain the lens models much better. We use the lens system JVAS B0218+357 as an example and test case. This system has a measured time-delay and can thus be used to determine H_0. We show that 'classical' model fits, using only the two compact images in this system, are not sufficient to determine the position of the lensing galaxy and cannot be used to determine the Hubble constant. To exploit the extended structure of the Einstein ring which is part of this system, the LENSCLEAN algorithm can be used. The main part of this thesis is devoted to this method. We discuss a number of significant improvements of LENSCLEAN which were necessary to turn it into a useful tool for systems like B0218+357. The parameters of an isothermal elliptical mass model can now be constrained with sufficient accuracy to obtain a competitive result for the Hubble constant of $H_0 = (71 \pm 5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for an Einstein-de Sitter universe. The error bar is a $2\,\sigma$ confidence limit including uncertainties of time-delay and lens model. Only slightly different results are expected for non-isothermal models in the case of B0218+357.

We also present new VLBI observations of B0218+357 which for the first time show parts of the jet in the doubly imaged region. We argue that these data can be used to constrain the radial mass distribution with unprecedented accuracy. Together with the results from scheduled HST observations, B0218+357 will soon be the system with the best constrained lens model and the most robust result for the Hubble constant. It can then fulfill its expectations as a 'golden lens'.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Bestimmung kosmologischer Parameter, insbesondere der Hubble-Konstanten, mittels Beobachungen von Gravitationslinsen. Diese Methode hat den Vorteil, nur von wenigen astrophysikalischen Vorstellungen abhängig zu sein. Dadurch werden sehr zuverlässige Resultate ermöglicht, deren restliche Unsicherheiten relativ einfach abgeschätzt werden können. Der wichtigste Beitrag zu möglichen Fehlern ist auf die verwendeten Linsenmodelle zurückzuführen. Es ist deshalb nötig, diese Fehlerquellen genau zu untersuchen und sie so weit wie möglich zu reduzieren. Es werden Ergebnisse analytischer Untersuchungen einer sehr allgemeinen Familie von Linsen mit radialem Potenzgesetz und beliebiger Winkelabhängigkeit präsentiert. Störungen von außen werden durch eine externe Scherung beschrieben. Um alle Beobachtungseinschränkungen eines optimalen Systems mit kompakten mehrfach abgebildeten Quellen untersuchen zu können, betrachten wir Vierfachsysteme, bei denen die Bildpositionen und alle drei unabhängigen „time-delays“ als Einschränkungen für die Linsenmodelle wirken. Es ist lange bekannt, dass der radiale Dichteverlauf der Linsen einen wichtigen Effekt auf die ermittelten Werte der Hubble-Konstante H_0 haben kann. In unseren Rechnungen finden wir eine allgemeingültige Skalierung des gemessenen H_0 in Abhängigkeit vom Exponenten des Radialanteils des Potentials β. Diese Skalierung ist für alle derartigen Linsensysteme exakt gleich. Systematische Fehler des angenommenen β werden deshalb in ebensolchen systematischen Fehlern in H_0 resultieren. Um den Effekt der externen Scherung zu beschreiben, verwenden wir das neue Konzept der „kritischen Scherung“. Für eine externe Scherung, die diesem Wert exakt gleich ist, verschwinden alle time-delays.

Um die Situation zu verbessern, müssen externe Scherung und radialer Dichteverlauf exakt vermessen werden. Systeme mit mehrfach abgebildeten unaufgelösten Bildern können immer nur eine begrenzte Anzahl von Einschränkungen liefern. Es ist deshalb wichtig, auch ausgedehnte mehrfach abgebildete Quellen zu verwenden, die die Linsenmodelle deutlich besser einschränken können. Für unsere Untersuchungen verwenden wir das System JVAS B0218+357 als exemplarischen Testfall. Für dieses System ist der time-delay bekannt, so daß die Bestimmung von H_0 prinzipiell ermöglicht wird. Es zeigt sich, dass „klassische“ Linsenmodellierung, die nur Informationen der beiden kompakten Bilder verwendet, nicht ausreichend ist um die Position der Linsengalaxie zu bestimmen. Solche Modelle können deshalb nicht zur Bestimmung der Hubble-Konstanten verwendet werden.

Um zusätzlich die Strukturen des hier ebenfalls vorhandenen Einsteinrings ausnutzen zu können, verwenden wir den LENSCLEAN Algorithmus. Dies macht den Hauptteil dieser Arbeit aus. Es wird eine Anzahl von deutlichen Verbesserungen von LENSCLEAN besprochen, die nötig waren, um daraus ein nützliches Verfahren für Systeme wie B0218+357 zu machen. Hiermit können die Parameter eines elliptischen isothermen Massenmodells mit ausreichender Genauigkeit bestimmt werden um ein konkurrenzfähiges Ergebnis für die Hubble-Konstante von $H_0 = (71 \pm 5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ für ein Einstein-de Sitter-Universum zu erhalten. Die Fehlergrenzen geben einen $2\,\sigma$ Vertrauensbereich an. In ihnen sind die Unsicherheiten aus time-delay und Linsenmodell enthalten. In diesem speziellen Fall sind keine wesentlich unterschiedlichen Resultate zu erwarten, wenn nicht-isotherme Modelle verwendet werden.

Zudem präsentieren wir neue VLBI Beobachtungen von B0218+357, die zum ersten Mal doppelt abgebildete Bereiche des Jets dieser Quelle zeigen. Diese Daten können die radiale Masseverteilung mit bisher unerreichter Genauigkeit festlegen. Zusammen mit den Ergebnissen von bereits genehmigten HST Beobachtungen wird B0218+357 das System mit dem am genauesten bekannten Linsenmodell und dem besten Ergebnis für die Hubble-Konstante werden. B0218+357 ist damit auf dem besten Weg, seine Versprechen als „goldene Linse“ einzulösen.

Dedicated to my father

In the online version (PDF file) of this thesis, hyperlinks are provided for references to equations, sections, citations, etc. External WWW links are included in the bibliography (see note on page 238).

Contents

Chapter 1

Introduction

Gravitational lenses offer a unique possibility to determine cosmological distances and hence the Hubble constant in a relatively simple way (Refsdal, 1964) by using light travel time-delays between multiple images of one source as fundamental distance measure. The method avoids using any form of distance ladder and instead determines distances at high redshift directly in one step. Classical methods involve many possible sources of uncertainties and errors. In distance ladder methods, each step introduces its own problems, because the ‘standard candles’ or their equivalents such as the Tully-Fisher relation have to be understood in detail. Many areas of astrophysical research thus contribute to the results, but also to the possible errors. This includes stellar physics, galactic dynamics, physics of supernovae, structure formation, dynamical properties of the local universe, evolution of stars and galaxies, dust formation and many other topics. It is almost impossible to understand and estimate the possible errors (statistical and more important systematic) of all these contributions reliably. Results from different methods or groups which are highly incompatible with each other within the error estimates are a well known problem in the history of the Hubble constant[1]. Though much effort has been devoted to improving the classical methods, only during the last decade have they begun to produce consistent results. It is very important to continue this work to obtain results from as many independent methods as possible. This can not only provide a robust result for the Hubble constant, but can also serve as a test of the underlying concepts of all the methods. In the end, some of the methods might teach us more about the astrophysical problems involved than about cosmology.

Gravitational lensing is a very special method in the sense that it is almost independent of a deeper understanding of astrophysical processes. It relies on general relativity, including the cosmological fundament of a homogeneous Friedmann–Robertson-Walker universe, and needs mass models of the lenses to produce results. Apart from this, no complicated astrophysics is needed to apply the method. This has the advantage that possible statistical and systematic uncertainties can be controlled much better than with other methods. The possible errors from incomplete knowledge of the cosmological parameters can be estimated easily by repeating the calculations for a range of possible values. These uncertainties are currently at the 5 % level. More important is the influence of the mass models for the lenses. If these are based on other astrophysical research, e.g. studies of rotation curves and radial velocity distributions of nearby galaxies, unknown errors are again introduced, which are difficult to quantify. Even if typical mass distributions of nearby galaxies could be determined with satisfying accuracy, the extrapolation to high-redshift lenses, which are at different levels of evolution from today’s galaxies, is highly problematic. This would imply a model of structure formation, galaxy evolution and depend e.g. on the validity of the concept of cold dark matter (CDM).

---

[1] Mould et al. (2000) still differ from Sandage (1999) or Parodi et al. (2000) by about 30 %.

The importance of lens models imposes two requirements. First, the possible errors have to be analysed by studying very general mass distributions as lens models. Analytical and numerical methods can be used to quantify uncertainties implied by unknown model parameters. With the knowledge from this analysis, ways to improve the constraints on the relevant model parameters to reduce the errors can be developed. We follow the first line with analytical investigations of a very general family of power-law models. In that part of the work, we will concentrate on two effects. One is the radial mass distribution, which can lead to large uncertainties for all lens systems. An important result is the fact that systematic errors introduced by the assumption of an incorrect profile can lead to exactly the same errors for a large family of lens systems. These errors can therefore not be detected as scatter in the results from many lenses. External perturbing masses, adding to the effect of the main lens, are another important effect. We will introduce a new concept of a 'critical shear' which can be used to estimate the possible errors. Even small external perturbations can have very strong effects for certain lenses.

The effects of other cosmological parameters, besides H_0, on lensing properties can also be used to constrain these parameters from individual lenses or small number of systems. These cosmological tests are purely geometric and directly test the distance parameters. They do not involve detailed astrophysical models like structure formation and a priori unknown properties of the dark matter. This is again in contrast to most other methods, including the use of lens statistics. Two methods are known to determine cosmological parameters from individual lenses. One relies on measured time-delays. This method can only be used if the Hubble constant is known from other methods, or in combination of several lens systems which together constrain the Hubble constant *and* other cosmological parameters. Unfortunately, the effects are very small and a very high accuracy of the time-delays and lens models is needed to apply it successfully. An alternative method gained interest during the last years. It works without a known time delay but uses the image separation or the geometry of critical curves. This method relies on an alternative mass determination result, e.g. from velocity dispersion measurements. The fundamental problem with this method is that it depends very sensitively on the dynamical model of the lensing galaxy which is used to convert the velocity dispersion to a mass scale. It may turn out that this method is more valuable to study the dynamics of high redshift galaxies than to determine cosmological parameters.

Although the application of the lens effect to determine cosmological parameters will possibly play an important role in the future, we do not discuss it in detail in this thesis. Our work on lens models will, nevertheless, also help in this application. This leads us to the second task, the improvement of lens models to reduce uncertainties and errors.

The 'golden lens' B0218+357 serves as a test case for most of this work, although the analytical methods for quadruple systems do not apply to this double. B0218+357 is a radio lens system consisting of two bright images and an additional Einstein ring. The time-delay between the bright components has been measured and can be used to determine the Hubble constant. This system was used for our development of the LENSCLEAN algorithm which can be applied to constrain lens models from radio data of lenses with extended sources. This work was motivated by the only principal problem with B0218+357: It is the system with the smallest image separation of all known lenses. Because the lens galaxy is not seen at radio wavelengths, measurements of its position rely on optical observations, which are extremely difficult for such a small system. No accurate measurement for this fundamental parameter has been possible yet. We therefore use LENSCLEAN mainly to constrain the lens position to be able to estimate the Hubble constant from B0218+357's time delay. The method works successfully now and will in the near future be used to constrain parameters of detailed lens models for several radio lenses. One application will be a new VLBI data set of B0218+357 which provides strong constraints on the radial mass profile by showing rich substructure in the lensed jet of the source. This structure probes the lensing potential at a wide range of radial distances and thus provides far better constraints than data for multiply imaged compact sources.

Preview of this thesis

We start with theoretical considerations on model fitting and error statistics in chapter 2. An understanding of the theory of modelling is essential for the development of general lens modelling methods, but also for radio interferometry and LENSCLEAN. We concentrate on the statistical analysis of model fitting residuals and on confidence limits.

Chapter 3 comprises a brief introduction on gravitational lenses, including the necessary cosmological background. As promised, only little astrophysics is needed. The cosmology enters only in the calculation of angular size distances. We introduce the general concepts of lens equation, deflection angle and magnification. A simple derivation of the time-delay equation is also provided. Classical lens model fitting is discussed in chapter 4. We present the model families that are used throughout this work and the implementation of different types of observational constraints in the modelling process.

Many lens modelling procedures imply the inversion of the lens equation; it is necessary to find the position of all images for a given source position. This highly non-trivial problem can be solved analytically only in very special cases. Chapter 5 presents the new numerical algorithm LENTIL for the inversion, which was designed to work very reliable for any kind of lens model. Simpler alternatives are available for classical modelling. LENTIL was developed mainly for the use with LENSCLEAN, which relies on an extreme robustness of the lens equation inversion.

The analytical work on general power-law models is presented in chapter 6. We investigate these models for quadruply imaged systems with measurements of all three independent time delays to analyse possible model degeneracies for lenses which provide almost the optimum of constraints. It is shown that a very simple scaling relation exists for the dependence of H_0 on the power-law exponent. The new concept of the 'critical shear' is also introduced. For external perturbations equal to this value, all time delays vanish. This surprising fact is closely related to the formation of possible Einstein rings and to very high image multiplicities. A simple geometrical interpretation is presented, which can be used to obtain the critical shear directly from the image geometry. We also compare the analytical results with numerical models for a number of real lens systems.

Chapter 7 gives on overview of the lens B0218+357. We summarize the available observational data and possible constraints for lens models. We learn that, without the ring, the constraints are far from sufficient to determine the lens position accurately. As a result, no H_0 determination is possible without including further constraints. Published VLBI data resolve both images and show two subcomponents each. The positions of these can be used to constrain the radial mass distribution quite accurately.

Chapter 8 is devoted to the theory of radio interferometry, especially to the deconvolution of radio maps. Interferometers measure the Fourier transform of the celestial emission pattern. This measurement is not complete, so that the real brightness distribution cannot be derived uniquely from the data. The directly produced 'dirty maps' are convolutions with the 'dirty beam', which is the system's response for a point source ('point spread function' in optical astronomy). Several methods to solve the deconvolution problem are discussed theoretically and with numerical experiments. Particular attention is paid to 'CLEAN', which is by far the most important method in practical work. CLEAN works by successively subtracting shifted and scaled versions of the dirty beam from residual maps to produce deconvolved 'CLEAN maps'.

The modification of CLEAN for lensed sources, called LENSCLEAN, is discussed in chapter 9. This should be considered as the main part of this thesis. LENSCLEAN was already proposed by Kochanek & Narayan (1992) but is still not widely used. This is a result of the enormous technical difficulties and the high computational demands. The basic idea is simple. While components can be subtracted freely in unlensed CLEAN, LENSCLEAN only allows simultaneous subtraction of all images corresponding to a certain position in the source plane. The emission model built in this way is exactly compatible with the given lens model and the final residuals can be used to assess the ability of the lens model to fit the data and thus to find the best lens model. The original algorithm has many numerical problems. A number of modifications were required to convert the elegant concept of LENSCLEAN into a valuable tool for lens modelling in cases like B0218+357. We illustrate most of these improvements with artificial data sets, for which the true lens model is known. It is only the combination of many improvements which really helped to cope with the numerical problems of the original algorithm. The demands on computing power are very high. Many modern PCs running in parallel were used for the computations. LENSCLEAN is at the very limit of what can be done without special equipment.

We use LENSCLEAN mainly to determine the lens position relative to the lensed images in B0218+357 from a 15 GHz VLA data set. The final accuracy is much higher than the resolution of the data, although the surface brightness of the ring is not very high and no sharp features are present besides the two bright images. The position is then used to determine the Hubble constant with a competitive accuracy. This is the first time that LENSCLEAN is applied to a lens with measured time-delay to determine the Hubble constant. A second data set with higher resolution, taken at 5 GHz with the MERLIN array, could not be used successfully to improve the accuracy, although it was successfully used to produce the best maps of the radio ring yet. This is partly a result of possible scattering in the lensing galaxy, which is stronger at lower frequencies. Other problems are related to the calibration of this multi-frequency data set. The data are nevertheless valuable for LENSCLEAN if used in combination with the 15 GHz data. We use lens models obtained at 15 GHz and apply them to the data at 5 GHz to quantify the effects of scattering and a possible shift in position between the two, by fitting these as free parameters.

Most of the calculations were performed for isothermal lens models. Only preliminary results, obtained with a combination of LENSCLEAN and LENTIL, are available for different power-law models at the moment. It seems as if two effects of changes of the exponent cancel each other to a degree. The effect of the model itself is compensated by the shift of the lens centre induced by a modified exponent. We do therefore not expect a large error of the result for H_0 from the possibly incorrect radial profile used.

Finally, we present preliminary results of new high-resolution VLBI observations of the system B0218+357 in chapter 10. These data reveal structure in the jet unseen before. We use our best lens models to project the maps of both images of the jet back to the source plane to compare them directly. The lensed images look very different, because of the combined effects of different magnifications and finite resolution of the observations. The source plane images are compatible with each other, if the lens model is not exactly isothermal but has a slightly different radial exponent. This result is in agreement with the one from an older data set which does not show the jet but only the two central subcomponents. A future analysis of the new observations with LENSCLEAN will improve the constraints for the radial mass distribution considerably. Together with the lower-resolution data from the VLA and MERLIN which are very sensitive for the lens position, this will lead to very tight constraints of the possible lens models and to a very accurate and robust result for the Hubble constant.

Chapter 2

Model fitting and error statistics

The analysis of astronomical observations often involves inverse problems. This means, that we believe to understand how the observational data would look like if the true state of the system under investigation was known exactly. However, the task we have to perform is the inversion of this process: Deduce the state of the system (in our case the mass distribution of a gravitational lens or the brightness distribution of a radio source) from the observational data. A common approach to this kind of problems is model fitting. Given a parametrized model of the system, we determine the observational data expected if this model would be correct. Using this, we can minimize the deviations from the real observations to obtain the 'best' model.

2.1 Residuals

Suppose a model described by M parameters x_ν which shall be constrained by a number of N observational quantities y_j. Both can also be written as vectors $\mathfrak{x}$ and $\mathfrak{y}$. We assume to understand how ideal observations (without errors) should look like if the true model was known and write these as $\mathfrak{y}^{\mathrm{m}}$ ('m' for 'model'). In reality, observations always have errors, which we want to describe as additive noise $\mathfrak{n}$.

$$\mathfrak{y} = \mathfrak{y}^{\mathrm{m}} + \mathfrak{n} \tag{2.1}$$

This approach of a noise independent of the model is only useful for sufficiently small errors. Let the statistical properties of the noise be described by its expectation value $< \mathfrak{n} >= 0$ and its covariance matrix C:[1]

$$\mathsf{C} = \left\langle \mathfrak{n}\mathfrak{n}^{\dagger} \right\rangle \tag{2.2}$$

$$C_{jk} = \left\langle n_j n_k \right\rangle \tag{2.3}$$

For Gaussian noise, the distribution is uniquely defined by these quantities.

The idea of model fitting is now to define residuals in a way to discriminate between 'good' and 'bad' models. Throughout the later work, we will always use the following definition of the residuals R^2. We use a symmetric and positive definite matrix W to weigh the residual

---

[1] With $\mathsf{A}^{\dagger}$ and $\mathfrak{x}^{\dagger}$, we generally mean the Hermitian conjugate, which is a combination of transposition and complex conjugation. $(\mathsf{A}^{\dagger})_{jk} = A^{\star}_{kj}$. The matrices are real here, Hermitian conjugation is thus equivalent to transposition.

components and add them quadratically.

$$R^2 = (\mathbf{y}-\mathbf{y}^{\mathrm{m}})^\dagger \mathsf{W}\,(\mathbf{y}-\mathbf{y}^{\mathrm{m}}) \tag{2.4}$$

$$= \sum_{jk} W_{jk}\,(y_j - y_j^{\mathrm{m}})(y_k - y_k^{\mathrm{m}}) \tag{2.5}$$

A special case of this would be a diagonal matrix $\mathsf{W} = \mathrm{diag}(w_j)$, which leads to a simple sum of squares. For the (usually unknown) true model $\mathbf{x}_0$, the residuals would consist of noise only:

$$R^2 = \mathbf{n}^\dagger \mathsf{W}\mathbf{n} \tag{2.6}$$

For incorrect models, the difference to the true model would also contribute to the residuals. It is thus very likely, that the true model is located somewhere near the minimum of R^2, which we call $\mathbf{x}_{\mathrm{fit}}$. To quantify this statement, we want to discuss some aspects of the statistics in the following sections.

A special case of weighting is to use the 'natural weights' defined by $\mathsf{W} = \mathsf{C}^{-1}$. The residuals R^2 are then also called

$$\chi^2 = (\mathbf{y}-\mathbf{y}^{\mathrm{m}})^\dagger \mathsf{C}^{-1}\,(\mathbf{y}-\mathbf{y}^{\mathrm{m}}) \quad . \tag{2.7}$$

In the case of uncorrelated noise, where the covariance matrix is diagonal and $C_{jj} = \sigma_j^2$, a standard form in which χ^2 is often discussed can be recognized:

$$\chi^2 = \sum_j \frac{(y_j - y_j^{\mathrm{m}})^2}{\sigma_j^2} \tag{2.8}$$

The special properties of χ^2 will be discussed in detail in section 2.6.

2.2 Linear models

Linear model approaches are important in the context of this work for different reasons. First, they are often a good approximation even for nonlinear models for parameters near the best fit. Second, the CLEAN and LENSCLEAN algorithms, which will be described later, fit linear brightness models to the data. An understanding of general linear models will help in the analysis of these methods.

Using the same notation as before, the relation between model parameters $\mathbf{x}$ and observational quantities $\mathbf{y}$ can be written by using a matrix A:[2]

$$\mathbf{y}^{\mathrm{m}} = \mathsf{A}\mathbf{x} \tag{2.9}$$

With this approach, it is simple to minimize the residuals analytically by setting the variation to zero:

$$\delta R^2 = -2\,\delta\mathbf{x}^\dagger \mathsf{A}^\dagger \mathsf{W}\,(\mathbf{y}-\mathsf{A}\mathbf{x}) \tag{2.10}$$

$$= 0 \tag{2.11}$$

The solution is

$$\mathbf{x}_{\mathrm{fit}} = (\mathsf{A}^\dagger \mathsf{W}\mathsf{A})^{-1}\mathsf{A}^\dagger \mathsf{W}\mathbf{y} \quad , \tag{2.12}$$

$$R^2_{\mathrm{fit}} = \mathbf{y}^\dagger \mathsf{M}'\mathbf{y} \tag{2.13}$$

---

[2] An additional shift in $\mathbf{x}$ and/or $\mathbf{y}$ can be included in a simple way and is thus neglected in the discussion.

with

$$\mathsf{M} = \mathsf{W}\,\mathsf{A}\,(\mathsf{A}^{\dagger}\mathsf{W}\mathsf{A})^{-1}\,\mathsf{A}^{\dagger}\,\mathsf{W} \quad , \tag{2.14}$$

$$\mathsf{M}' = \mathsf{W} - \mathsf{M} \quad . \tag{2.15}$$

2.3 Regularization

The solution for $\boldsymbol{x}$ is unique for all $\boldsymbol{y}$ only if the matrix $\mathsf{A}^{\dagger}\mathsf{W}\mathsf{A}$ is invertible. A necessary condition for this is $M \leq N$. Obviously, we cannot determine more parameters than we have constraints. To define a uniquely solvable problem even in cases of a singular (or numerically close to singular) matrix, further constraints have to be imposed on the solution. One possibility is to search for the 'principal solution', defined by

$$\boldsymbol{x}^{\dagger}\boldsymbol{x} = \text{minimum} \quad . \tag{2.16}$$

This solution can be found in a very elegant and numerically stable way by using the method of 'singular value decomposition' (e.g. Press et al., 1992). Another method is the use of Lagrange multipliers.

Even in formally invertible cases, the solutions may sometimes be very unstable and change violently with small changes in the measurements. Clearly, we are not interested in these unrealistic variations but in the 'stable part' of $\boldsymbol{x}$. Several regularization methods are known to accomplish this task. The so-called 'linear regularization' is especially simple. The idea is to put further (typically smoothness) constraints on the solution and modify the fundamental equations accordingly. The general approach can be written as modified residuals R'^2, which then have to be minimized instead of the true residuals R^2.

$$R'^2 = R^2 + \lambda\,\boldsymbol{x}^{\dagger}\,\mathsf{S}\boldsymbol{x} \tag{2.17}$$

The positive definite symmetric matrix S describes the particular kind of regularization; the constant $\lambda \geq 0$ defines to what extent the regularization should be allowed to change the solutions. For infinitely small λ (we will call this 'weak regularization' later) and $\mathsf{S} = \mathsf{1}$, this leads to the exact principal solution. For finite λ, the models found in this way are not exact solutions of minimal R^2 anymore. We therefore do not only select one of the exact solutions but change the optimization problem itself. We refer to this methods as 'strong regularization' later. Regularization can be interpreted as a special kind of a Bayesian prior, which will be discussed in section 2.8.

For $\lambda > 0$, the positive definite matrix S shifts the vanishing eigenvalues of $\mathsf{A}^{\dagger}\mathsf{W}\mathsf{A}$ to finite values and thus allows to invert the equation.

$$\boldsymbol{x}_{\text{reg}} = (\mathsf{A}^{\dagger}\mathsf{W}\mathsf{A} + \lambda\mathsf{S})^{-1}\,\mathsf{A}^{\dagger}\mathsf{W}\boldsymbol{y} \tag{2.18}$$

The trade-off between small residuals R^2 and 'regularity' of the solution is represented by λ, which we can choose to our needs. If we want to interpret $\boldsymbol{x}_{\text{reg}}$ itself as a 'best' model and not as a smoothed version of some solution, λ should be chosen in a way to get reasonable residuals (see section 2.5).

2.4 The 'Maximum likelihood' approach

In 'maximum likelihood' model fitting, the 'best' model is defined to be the one with the highest likelihood. The likelihood is defined as the probability that, given the model, the observations are within a small (but constant) interval of the ones observed in reality. The likelihood is *not* the probability of the model being correct, although this is often confused. However, it is possible to relate the two using Bayes' theorem and an a priori probability distribution of models.

For Gaussian noise, the χ^2 derived from the true model $\mathbf{x}_0$ and a set of observations $\mathbf{y}$ is a measure for the probability density of these observations. This is the case because the difference consists of noise only and is weighted with the covariance matrix of the noise. The probability density of a certain measurement if given by

$$\mathrm{d}p \propto \mathrm{e}^{-\chi^2/2} \prod_j \mathrm{d}y_j \quad . \tag{2.19}$$

Minimal χ^2 therefore corresponds to maximal likelihood. We will see later, that the minimal χ^2 can be used to assess the goodness of the fit. The probability that the measured χ^2 is equal to or less than an observed value t is given by the cumulative χ^2-distribution. Changes of χ^2 depending on $\mathbf{x}$ are used to determine confidence limits of the model parameters.

We have to keep in mind, that minimizing R^2 with non-natural weighting is *not* equivalent to finding the model of maximal likelihood. We can nevertheless determine confidence limits using this alternative measure (see section 2.7).

2.5 Calculating the mean and variance of R^2

In this section we want to analyse some theoretical properties of model fitting residuals. The results will later be used to judge the goodness of fit and to determine confidence limits. For the analytical calculations, we assume a linear model following section 2.2. The results are also applicable for nonlinear models, if the errors are sufficiently small. For simplicity, we furthermore assume that the real model is $\mathbf{x}_0 = 0$. This can always be achieved by a shift in $\mathbf{x}$ and $\mathbf{y}$.

Let us now investigate the statistics for the residuals R^2_{fit} of the best fit model $\mathbf{x}_{\text{fit}}$, for the residuals R^2_0 of the true (unknown) model $\mathbf{x}_0$ and for their difference ΔR^2. Using the results from section 2.2 and the assumption that the true model is $\mathbf{x}_0 = 0$ and therefore also $\mathbf{y}^{\text{m}}_0 = 0$, we can write:

$$R^2_0 = \mathbf{n}^\dagger \mathsf{W} \mathbf{n} \tag{2.20}$$

$$R^2_{\text{fit}} = \mathbf{n}^\dagger \mathsf{M}' \mathbf{n} \tag{2.21}$$

$$\Delta R^2 = R^2_0 - R^2_{\text{fit}} \tag{2.22}$$

$$= \mathbf{n}^\dagger \mathsf{M} \mathbf{n} \tag{2.23}$$

$$\geq 0 \tag{2.24}$$

For M and M' we use the same definition as before in (2.14) and (2.15). For the case of arbitrary weights W, we only want to calculate the expectation value and the variance of these quantities. For natural weighting, we will present the complete probability distributions in 2.6.

The mean values can be calculated trivially. We only give the results for R_0, as those for R^2_{fit} and ΔR^2 can be obtained by substituting W with M' or M, respectively. This is also true for the variances.

$$\left\langle R^2_0 \right\rangle = \sum_{jk} W_{jk} C_{jk} \tag{2.25}$$

$$= \operatorname{Tr} \mathsf{W}\mathsf{C} \tag{2.26}$$

The trace of a matrix is defined as the sum of the diagonal elements,

$$\mathrm{Tr}\, M = \sum_j m_{jj} \quad . \tag{2.27}$$

For diagonal W and C, equation (2.26) reads

$$\langle R_0^2 \rangle = \sum_j w_j\, \sigma_j \quad . \tag{2.28}$$

To calculate the variances, we need the expectation value of $n_j\, n_k\, n_l\, n_m$. To derive this, it is best to write the covariance matrix as

$$\begin{aligned} \mathsf{C} &= \mathsf{S}'^{\dagger}\, \mathsf{D}\, \mathsf{S}' \quad , & (2.29) \\ &= \mathsf{S}^{\dagger}\, \mathsf{S} \quad , & (2.30) \end{aligned}$$

where D is diagonal and S' orthogonal. In the second step we have absorbed the eigenvalues D_{jj} into S' to form S. Using this, we can write the noise as

$$\mathfrak{n} = \mathsf{S}^{\dagger}\, \boldsymbol{\xi} \tag{2.31}$$

with an uncorrelated and normalized noise vector $\boldsymbol{\xi}$ satisfying

$$\langle \xi_j\, \xi_k \rangle = \delta_{jk} \quad . \tag{2.32}$$

The needed high order covariance is now

$$\langle n_j\, n_k\, n_l\, n_m \rangle = \sum_{abcd} S_{aj}\, S_{bk}\, S_{cl}\, S_{dm} \langle \xi_a\, \xi_b\, \xi_c\, \xi_d \rangle \quad . \tag{2.33}$$

The measure for $\boldsymbol{\xi}$ on the right hand side can be derived with (2.32) and symmetry arguments:

$$\langle \xi_a\, \xi_b\, \xi_c\, \xi_d \rangle = \delta_{ab}\, \delta_{cd} + \delta_{ac}\, \delta_{bd} + \delta_{ad}\, \delta_{bc} \tag{2.34}$$

The indices have to be equal in pairs, otherwise at least one factor would separate and vanish. For $a = b$, $c = d$ and $a \neq c$ the result is $< \xi_a^2 \xi_c^2 >= 1$. In the case $a = b = c = d$ it is 3, which is the correct value of $< \xi_a^4 >$. Using this, we arrive at

$$\langle n_j\, n_k\, n_l\, n_m \rangle = \sum_{ac} S_{aj}\, S_{ak}\, S_{cl}\, S_{cm} + \sum_{ab} S_{aj}\, S_{bk}\, S_{al}\, S_{bm} + \sum_{ab} S_{aj}\, S_{bk}\, S_{bl}\, S_{am} \quad , \tag{2.35}$$

and, with (2.30),

$$\langle n_j\, n_k\, n_l\, n_m \rangle = C_{jk} C_{lm} + C_{jl} C_{km} + C_{jm} C_{kl} \quad . \tag{2.36}$$

Now we can determine the variance of R_0^2 with (2.20).

$$\begin{aligned} \mathrm{var}(R_0^2) &= \langle R_0^4 \rangle - \langle R_0^2 \rangle^2 & (2.37) \\ &= \sum_{jklm} W_{jk} W_{lm} \Big(\langle n_j\, n_k\, n_l\, n_m \rangle - \langle n_j\, n_k \rangle \langle n_l\, n_m \rangle \Big) & (2.38) \\ &= \sum_{jklm} W_{jk} W_{lm} \Big(C_{jl} C_{km} + C_{jm} C_{kl} \Big) & (2.39) \end{aligned}$$

Using the symmetry of W and C, this simplifies to

$$\operatorname{var}(R_0^2) = 2\sum_{jklm} W_{jk} W_{lm} C_{jl} C_{km} \quad , \tag{2.40}$$

$$= 2\,\mathrm{Tr}\,(\mathsf{WC})^2 \quad , \tag{2.41}$$

or, for diagonal W and C,

$$\operatorname{var}(R_0^2) = 2\sum_j w_j^2\, \sigma_j^4 \quad . \tag{2.42}$$

For the case of natural weighting, where $R^2 = \chi^2$ and $\mathsf{WC} = \mathsf{1}$, the definitions of M and M' in (2.14) and (2.15) can be used to prove that $\mathsf{P}_1 = \mathsf{MC}$ and $\mathsf{P}_2 = \mathsf{M}'\mathsf{C}$ are projection operators with

$$\mathsf{P}_j^2 = \mathsf{P}_j \quad , \tag{2.43}$$

which project into a subspace. The trace of a projection operator is simply the dimension number of the subspace:

$$\mathrm{Tr}\,(\mathsf{WC})^2 = \qquad \mathrm{Tr}\,\mathsf{WC} = \qquad N \tag{2.44}$$

$$\mathrm{Tr}\,(\mathsf{MC})^2 = \qquad \mathrm{Tr}\,\mathsf{MC} = \qquad M \tag{2.45}$$

$$\mathrm{Tr}\,(\mathsf{M}'\mathsf{C})^2 = \qquad \mathrm{Tr}\,\mathsf{M}'\mathsf{C} = \qquad N-M \tag{2.46}$$

As a conclusion we find the most important statistical properties of the corresponding χ^2 distributions:

$$\left\langle \chi_0^2 \right\rangle = N \qquad \operatorname{var}(\chi_0^2) = 2N \tag{2.47}$$

$$\left\langle \chi_{\text{fit}}^2 \right\rangle = N-M \qquad \operatorname{var}(\chi_{\text{fit}}^2) = 2(N-M) \tag{2.48}$$

$$\left\langle \Delta\chi^2 \right\rangle = M \qquad \operatorname{var}(\Delta\chi^2) = 2M \tag{2.49}$$

These can directly be used to assess the goodness of a fit and to determine confidence limits. It can be shown, that the distributions not only have the same mean and variance as the χ^2-distributions with $\nu = N$, M and $N-M$ but really are identical to those distributions (see also section 2.6). The normalized (or reduced) χ^2 is defined by dividing it by the mean value:

$$\bar{\chi}^2 = \frac{\chi^2}{\nu} \tag{2.50}$$

$$\operatorname{var}(\bar{\chi}^2) = \frac{2}{\nu} \tag{2.51}$$

We can define an analogous quantity for R_0^2 by using (2.26) and (2.41):

$$\bar{R}_0^2 = \frac{R_0^2}{\mathrm{Tr}\,\mathsf{WC}} \tag{2.52}$$

$$= \frac{R_0^2}{\sum_j w_j\, \sigma_j^2} \quad \text{(for diagonal } \mathsf{W}, \mathsf{C}) \tag{2.53}$$

$$\operatorname{var}(\bar{R}_0^2) = \frac{2}{\nu_0'} \tag{2.54}$$

Here we used an effective number of degrees of freedom ν_0' defined by

$$\nu_0' = \frac{(\mathrm{Tr}\,\mathsf{WC})^2}{\mathrm{Tr}\,(\mathsf{WC})^2} \quad . \tag{2.55}$$

In the diagonal case it becomes

$$\nu_0' = \frac{\left(\sum w_j\,\sigma_j^2\right)^2}{\sum w_j^2\,\sigma_j^4} \quad . \tag{2.56}$$

It can be shown, that always $\nu_0' \leq \nu_0$. The two are equal only for natural weighting with $w_j = \sigma_j^{-2}$. The variance of $\overline{R_0^2}$ is therefore minimized with natural weighting. Analogous derivations are possible for χ^2_{fit} and $\Delta\chi^2$. With general (non-natural) weighting, the traces in equations (2.26) and (2.41) can usually not be calculated easily for M and M'. It is nevertheless possible, to find a good approximation for non-pathological cases. Let us suppose a model with only one parameter x, which acts on the N measurements as follows:

$$y_i = a_i\,x \tag{2.57}$$

$$\mathsf{A} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \tag{2.58}$$

For $a_i \equiv 1$, this corresponds to $y_i = x$, i.e. a number of N measurements of one and the same parameter. The best solution is then the weighted arithmetical mean. The general case can be interpreted as finding the scaling constant x between the given a_i and the corresponding measurements y_i. It is now easy to find the following result for diagonal W and C:

$$\left\langle \Delta R^2 \right\rangle = \mathrm{Tr}\,\mathsf{MC} \tag{2.59}$$

$$= \frac{\sum_j w_j^2\,a_j^2\,\sigma_j^2}{\sum_j w_j\,a_j^2} \tag{2.60}$$

$$= \left\langle w\,\sigma_j^2 \right\rangle' \tag{2.61}$$

The mean in (2.61) is weighted with $w_j\,a_j^2$. In cases of very many measurements, we can hope that the a_j are not significantly correlated with the $w_j\,\sigma_j^2$. We can then calculate the mean with weights w_j instead of $w_j\,a_j^2$. The following generalized results for M parameters are exact in a number of special cases.

$$M' = M \left\langle w\,\sigma_j^2 \right\rangle \tag{2.62}$$

$$= M\,\frac{\sum_j w_j^2\,\sigma_j^2}{\sum_j w_j} \tag{2.63}$$

$$N' = \sum_j w_j\,\sigma_j^2 \tag{2.64}$$

$$\left\langle R_0^2 \right\rangle = N' \tag{2.65}$$

$$\left\langle R_{\mathrm{fit}}^2 \right\rangle \approx N' - M' \tag{2.66}$$

$$\left\langle \Delta R^2 \right\rangle \approx M' \tag{2.67}$$

We will use these approximations to estimate confidence limits of LENSCLEAN results in chapter 9, especially section 9.7. Monte Carlo simulations for the final result in section 9.21.2 will confirm that the accuracy is remarkably high in this case.

2.6 The χ^2 distribution

Because the χ^2 distribution is so fundamental, we want to present its derivation here. If a random vector $\mathbf{n}$ consists of ν independent random quantities with probability densities w_j, the probability density of $\mathbf{n}$ is

$$w_n(\mathbf{n}) = \prod_{j=1}^{\nu} w_j(n_j) \quad . \tag{2.68}$$

Equation (2.19) is the special form of this for Gaussian measurement errors. The probability density of an arbitrary function f of $\mathbf{n}$ can be written as

$$w_f(t) = \int \mathrm{d}^\nu n \, w_n(\mathbf{n}) \, \delta\big(f(\mathbf{n}) - t\big) \quad . \tag{2.69}$$

In the case of uncorrelated errors, χ^2 is just a sum of ν normally distributed random numbers (2.8). In the general case, we can use (2.7) transformed to diagonal form with the same result. The same formalism was used in section 2.5. The function $f = \chi^2$ is therefore the sum of the components for n_j. We can now use (2.69) to calculate the (differential) probability distribution of χ^2:

$$w_{\chi^2}(t) = \frac{1}{(2\pi)^{\nu/2}} \int \mathrm{d}^\nu n \; \mathrm{e}^{-\sum\limits_{i=1}^{\nu} n_i^2/2} \, \delta\left(\sum_{i=1}^{\nu} n_i^2 - t\right) \tag{2.70}$$

This integral can be transformed to ν-dimensional spherical coordinates:

$$w_{\chi^2}(t) = \frac{1}{(2\pi)^{\nu/2}} S_\nu \int_0^\infty \mathrm{d}r \; r^{\nu-1} \, \mathrm{e}^{-r^2/2} \, \delta(r^2 - t) \tag{2.71}$$

Here S_ν is the ‘area’ of the surface of the ν-dimensional unit sphere (e.g. Forster, 1984):

$$S_\nu = \frac{2\pi^{\nu/2}}{\Gamma(\nu/2)} \tag{2.72}$$

We can now perform the integration using

$$\int \mathrm{d}x \; f(x) \, \delta\big(g(x)\big) = \sum_{g(x_0)=0} \frac{f(x_0)}{|g'(x_0)|} \quad , \tag{2.73}$$

where the sum is taken over all zeros x_0 of g within the integration interval, to get the result

$$w_{\chi^2}(t) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \, t^{\nu/2-1} \, \mathrm{e}^{-t/2} \quad . \tag{2.74}$$

The cumulative χ^2-probability of χ^2 being equal to or less than a certain value is then

$$P(\chi^2|\nu) = \int_0^{\chi^2} \mathrm{d}t \; w_{\chi^2}(t) \quad . \tag{2.75}$$

This distribution can be written as an incomplete Gamma function. The last equation follows the notation of Abramowitz & Stegun (1972), where tabulated values can be found as well. In the case of $\nu \gg 1$, the χ^2 distribution is approximated by a Gaussian distribution, which is completely characterized by its mean ν and variance 2ν.

2.7 Goodness of fit and confidence limits

Equally important as the knowledge of the 'best' solution to an inverse problem is to estimate the accuracy of this solution and the goodness of the fit itself. To assess the latter, we can calculate the probability $P(\chi^2|\nu)$ of χ_{fit} being equal to or less than the measured value. The number of degrees of freedom in this case is $\nu = N - M$ (see below). For many realizations, this probability should itself be equally distributed between 0 and 1. Values very close to 1 (very high χ^2) usually mean, that the model cannot fit the data, because it is an inappropriate description of nature, or that the errors σ_j have been underestimated or are not Gaussian. Very low values, on the other hand, are an indication for overestimated errors. As a first check, the result can be compared with the expectation value of ν, taking into account the standard deviation of $\sqrt{2\nu}$.

To estimate the confidence limits, we have to assume that nature can be described by the model approach and that χ^2_{fit} has a reasonable value. We then use the difference $\Delta\chi^2 = \chi^2_0 - \chi^2_{\text{fit}}$ and its statistical properties to determine the range of acceptable model parameters. If we knew the real model, the difference $\mathbf{y} - \mathbf{y}^{\text{m}}_0$ (remember it is N-dimensional) would consist of noise only and χ^2_0 would follow the χ^2-distribution with as many degrees of freedom as there are measurements ($\nu = N$). In the fit we minimize the residuals to obtain $\chi^2_{\text{fit}} \leq \chi^2_0$. It can be shown that this χ^2_{fit} follows a distribution with $\nu = N - M$, where M is again the number of parameters of the model. For the proof, we use the linear form of $\mathbf{y}^{\text{m}}(\mathbf{x})$ in the region of interest. We can then apply transformations to $\mathbf{x}$ and $\mathbf{y}$ to diagonalize C and let $\mathbf{x}$ influence only M elements of $\mathbf{y}^{\text{m}}$. By minimizing χ^2, these contributions to χ^2 become zero. We are then left with a sum of $\nu = N - M$ normally distributed quantities, leading to a distribution with ν degrees of freedom. Said in a simplified way, by fitting M parameters to our N measurements, we were left with $N - M$ measurement errors in χ^2_{fit}, while there were N in χ^2_0. The difference therefore consists of M errors and $\Delta\chi^2$ follows a distribution with $\nu = M$. For a calculation of mean and variance of these distributions for arbitrary weighting, see section 2.5.

Knowing the distribution of $\Delta\chi^2$, we can define confidence regions whose boundaries have constant χ^2. The probability that $\Delta\chi^2$ is equal to or less than a certain value is given by $P(\Delta\chi^2|M)$. As an example, the 68.3% confidence limit for $M = 1$ is defined by $\Delta\chi^2 = 1$. The confidence usually is a good approximation to the probability that the true model actually lies in this region. To calculate real probabilities of models given the measured data, we have to apply Bayesian statistics (see section 2.8). If we are interested in confidence limits for a subset of $m < M$ parameters, we have to fit the other parameters to obtain the boundary given by statistics for $\nu = m$ in the subspace of m parameters. In other words, we have to determine the χ^2 by varying the irrelevant $M - m$ parameters to minimize the residuals. Only the m parameters of interest contribute to the $\Delta\chi^2$ statistics.

For differently weighted residuals R^2, boundaries of constant R^2 can still be used as confidence limits. Unfortunately, the probabilities can *not* be determined using a generic function of just the number of parameters and measurements. They also depend on other properties of the problem in question. The mean and variance of R^2 can be calculated exactly or estimated using the results from section 2.5.

Using boundaries of constant naturally weighted χ^2 optimizes the results in a certain sense. The probability density is a monotonous function of χ^2, and the region around the maximum likelihood limited by constant χ^2 is thus the most compact one. The errors in the model parameters are minimal in this sense. The shapes of the confidence regions change for different weighting, which can sometimes be desirable.

2.8 Bayesian statistics

The concept of maximum likelihood model fitting is somewhat arbitrary. It has proven to be useful in many problems, but it offers no real to solution to the question that we are actually interested in: Which is the most *probable* model and how large is the region of allowed parameters to enclose a total probability of say 95 %? Up to now, the probabilities we dealt with were of the kind 'probability of a certain observation given a model', while we are really interested in the 'probability of a certain model given an observation'. These two can actually be related to each other using Bayes' theorem, which can be written like this:[3]

$$P(M_j|D) = \frac{P(M_j)\,P(D|M_j)}{\sum\limits_k P(D|M_k)\,P(M_k)} \tag{2.76}$$

Here $P(M_j|D)$ is the probability that the model M_j is the correct one, given that the observations lead to D. $P(M_j)$ is the so-called 'a priori probability' of the model M_j (also called the 'prior'), the probability of this model without knowing anything about the observations. $P(D|M_j)$ is the probability that the measurements lead to D given that a model M_j is correct. This was called 'likelihood' of the model before. The denominator can be thought of as a normalization constant for the probabilities to assure $\sum_j P(M_j|D) = 1$. It is important that the normalization does not depend on the model in question but only on the data which are fixed by the observations anyway.

Equation (2.76) is easy to prove. We start with the probability that M_j and D are true, which can be written in two ways as the product of probabilities of M_j and D:

$$P(M_j \cap D) = P(D)\,P(M_j|D) \tag{2.77}$$

$$= P(M_j)\,P(D|M_j) \tag{2.78}$$

As both expressions have to be equal, we immediately arrive at (2.76), if we take into account that

$$P(D) = \sum_j P(D|M_j)\,P(M_j) \quad . \tag{2.79}$$

For models and observations described by continuous data, we can use the same equation for probability densities. The sum in the denominator becomes an integral then.

For a flat prior, $P(M_j) = \text{const}$ in the region of interest, the probability of the model is directly proportional to the likelihood. Only in this case can the two be used exchangeably. If the measurement errors are small, any models with significantly different a priori probabilities will be so incompatible with the data, $P(D|M_j) \ll 1$, that they do not contribute to the result. We can then assume a flat prior without distorting the results too much. This is why the maximum likelihood approach works so well in most cases.

A much cleaner reasoning regarding the confidence limits is possible using Bayes statistics. A confidence region of say 95 % is simply a (not *the*) region in model parameter space, whose total integrated probability is 95 %. The shape of this region can be chosen arbitrarily to our needs, and the size is then determined by the error statistics.

---

[3] Think of a situation in everyday life: If you know that the reliability of the statements of a certain person is not 100 % and does not depend on the very subject of the statement, you would rather believe a claim like 'It is raining today.' than 'I just won 2 millions in the lottery.', because the a priority probability $P(M_j)$ of the former is much higher than that of the latter.

Finally, we want to show that the linear regularization method presented in section 2.3 can be interpreted as an application of Bayesian statistics. For natural weighting, we have $P(D|M) \propto \exp(-\chi^2/2)$. If we do not minimize χ^2 but $\chi^2 + \lambda \mathbf{x}^\dagger \mathsf{S} \mathbf{x}$, this can be seen as finding the most *probable* model with a prior $P(M)$ defined by the λ and S:

$$P(M|D) \propto P(M)\,P(D|M) \tag{2.80}$$

$$P(D|M) \propto \mathrm{e}^{-\chi^2/2} \tag{2.81}$$

$$P(M|D) \propto \mathrm{e}^{-(\chi^2+\lambda \mathbf{x}^\dagger \mathsf{S} \mathbf{x})/2} \tag{2.82}$$

$$P(M) \propto \mathrm{e}^{-\lambda \mathbf{x}^\dagger \mathsf{S} \mathbf{x}/2} \tag{2.83}$$

When thinking about this, more general regularization schemes come into mind. Priors of a different form can be used to determine the most probable models. Another possibility is to determine a *mean* model, which is defined by averaging over all possible models using the probability density given by Bayes' theorem. Such methods are very promising candidates for improved deconvolution algorithms in radio astronomy. They might be used as substitutes for CLEAN and LENSCLEAN in the future.

Chapter 3

Gravitational lenses

3.1 Introduction

In an attempt to save both paper and the reader's time[1], we do not repeat the introduction about gravitational lenses and the historical overview that can be found in other publications. Excellent review articles are available on the subject that can explain the whole area of current gravitational lens research much better (Blandford & Narayan, 1992; Refsdal & Surdej, 1994; Narayan & Bartelmann, 1999). The textbook on gravitational lenses by Schneider, Ehlers & Falco (1992) is also worth reading. This includes the exhaustive, though not up to date, list of references therein. A short historical summary can also be found in Wucknitz (1996).

To remind the reader of the most fundamental equations in lensing theory, but mainly to introduce the notation, we start with a very quick 'derivation' of the lens equation and the expression to calculate time-delays. Many more details can be found in Schneider et al. (1992) or Wucknitz (1996).

3.2 Angular size distances

Only little knowledge of cosmology is needed to understand the work presented in this thesis. Gravitational lenses have the advantage, that cosmology enters only in the definition of distances which define the lensing geometry. These distances depend on measured quantities (redshifts) and on the cosmological parameters. To some extent, the distances can be determined with the lensing effect. They can then be used to derive constraints for the cosmological parameters.

In a possibly curved and certainly expanding universe, Euclidian geometry can not be used to relate events which are separated by cosmological distances in space and/or time. At the present level of knowledge, homogeneous and isotropic geometrical models like the Robertson-Walker metric can be used with sufficient accuracy. In these coordinates, the Einstein field equations lead to the Friedman equations. They control the evolution of the scale factor in the metric and the evolution of the universe as a whole. Luckily, the complete theory is *not* needed for gravitational lens work. We will show that only very simple geometrical properties are relevant for out work. These can all be described by the angular size distances of source and lens.

Figure 3.1 shows the fundamental diagram for the definition of angular size distances. The

---

[1] Not to mention the author's time…

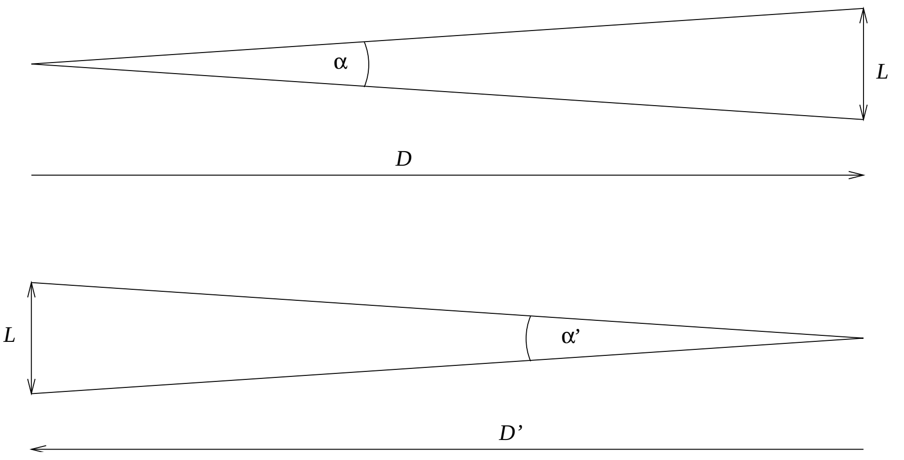


Figure 3.1: Definition of angular size distances. The distance parameter D is defined as L/α, where α (assumed small) is the angle at which an object of size L is seen from the starting point (usually the observer). Note that the distance generally depends on the direction, $D \neq D'$.

distance parameter D is defined in a way to expand the validity of

$$\alpha = \frac{L}{D} \tag{3.1}$$

to non-Euclidian geometry, where α (we assume $\alpha \ll 1$) is the angle under which an object of extension L is seen by a real or hypothetical observer at the starting point of the distance. In a general Riemannian geometry, the distance parameter can depend on the direction in which it is measured. Think of a universe which continues its expansion while the light is on the way from the object to the observer. For standard Friedmann–Robertson-Walker cosmologies, the evolution changes the scale factor R only. The normalized distances $D_{\mathrm{AB}}/R_{\mathrm{B}}$ are then independent of the direction[2]. The relation between the distances is therefore very simple:

$$\frac{D_{\mathrm{AB}}}{R_{\mathrm{B}}} = \frac{D_{\mathrm{BA}}}{R_{\mathrm{A}}} \tag{3.2}$$

$$\frac{D_{\mathrm{AB}}}{D_{\mathrm{BA}}} = \frac{R_{\mathrm{B}}}{R_{\mathrm{A}}} \tag{3.3}$$

$$= \frac{1+z_{\mathrm{A}}}{1+z_{\mathrm{B}}} \tag{3.4}$$

For a strictly homogeneous universe, the distances can be calculated directly from the global Robertson-Walker metric together with the Friedmann equations.

The presence of gravitational lenses proves, that inhomogeneities can at least in some cases be important for the light propagation in our universe. If a certain part α of the matter is distributed homogeneously, and the remaining $1-\alpha$ is taken into account as the lens effect of galaxies, clusters of galaxies etc., things become more difficult. The propagation of light bundles in a so-called evacuated cone (because $1-\alpha$ of the matter is removed) embedded in the surrounding global geometry of an evolving universe must now be calculated by (generally numerical) integration of a second-order differential equation. Details of this problem can be found

---

[2] D_{AB} is the distance parameter of an object B measured from A. The angle is measured at A, the length at B.

in the fundamental papers of Dyer & Roeder (1972, 1973) or elsewhere. The same recipe can also be used for more general cosmological models, including e.g. a cosmological constant Λ. See Kayser, Helbig & Schramm (1997) for the description of a code to calculate distances in such models and an overview of the theory.

The most general cosmological model widely discussed in the last years includes a generalization of the cosmological constant, called ‘quintessence’ or ‘dark energy’. The idea behind this concept is a scalar field, which leads to an equation of state according to

$$p_{\mathrm{X}} = w\rho_{\mathrm{X}} \quad , \tag{3.5}$$

which is a special simplified case of more physically motivated scalar field theories. The constant is restricted to $-1 \leq w < 0$. The lower limit is equivalent to a cosmological constant, the upper limit corresponds to dustlike pressureless matter. It is possible to include this new kind of ‘matter’ in the Friedmann equations and in the Dyer-Roeder equations for the distances. For details, we refer to the literature on the subject (e.g. Giovi & Amendola, 2001; Sereno et al., 2001). The resulting Dyer–Roeder-like differential equation for the normalized dimensionless distance parameter

$$r(z_1,z_2) = \frac{H_0}{c} D(z_1,z_2) \tag{3.6}$$

looks like this:

$$\begin{aligned}(1+z)^2\left(\Omega_{\mathrm{M}}(1+z)^3+\Omega_{\mathrm{X}}(1+z)^m+\Omega_k(1+z)^2\right)\frac{\mathrm{d}^2r}{\mathrm{d}z^2}&\\ +(1+z)\left(\frac{7}{2}\Omega_{\mathrm{M}}(1+z)^3+\frac{m+4}{2}\Omega_{\mathrm{X}}(1+z)^m+3\Omega_k(1+z)^2\right)\frac{\mathrm{d}r}{\mathrm{d}z}&\\ +\left(\frac{3}{2}\alpha_{\mathrm{M}}\Omega_{\mathrm{M}}(1+z)^3+\frac{m}{2}\alpha_{\mathrm{X}}\Omega_{\mathrm{X}}(1+z)^m\right)r=0&\end{aligned} \tag{3.7}$$

The parameters α_{M} and α_{X} describe the homogeneity of normal matter and quintessence (the latter is assumed to be distributed evenly, $\alpha_{\mathrm{X}} = 1$), and the index m is related to the equation of state by $m = 3(w+1)$. The lower limit of $m = 0$ corresponds to a cosmological constant, the upper limit $m = 3$ to normal matter. We also used the abbreviation $\Omega_k = 1 - \Omega_{\mathrm{M}} - \Omega_{\mathrm{X}}$ in the equation. Spatially flat universes have $\Omega_k = 0$. The initial conditions for the integration are

$$r(z_1,z)\big|_{z_1} = 0 \quad \text{and} \tag{3.8}$$

$$\left.\frac{\mathrm{d}r(z_1,z)}{\mathrm{d}z}\right|_{z_1} = \frac{1}{(1+z_1)\sqrt{\Omega_{\mathrm{M}}(1+z_1)^3+\Omega_{\mathrm{X}}(1+z_1)^m+\Omega_k(1+z_1)^2}} \quad . \tag{3.9}$$

A simple Runge-Kutta method was used to calculate distances for the cosmological models described in Table 3.1. The Einstein-de Sitter model (EdS) serves as a reference in many cases.

3.3 The cosmological lens equation

The theory used in this work is restricted to ‘thin lenses’, in which significant light deflection takes place only very close to the lens plane. This is not equivalent to a thin mass distribution. For a detailed analysis of the thin lens approximation, see Wucknitz (1996). It is valid in all cases of one-plane lenses, where all galaxies contributing to the deflection have approximately

model	Ω_M	Ω_X	m	Ω_k	α_M
flat Λ	0.3	0.7	0	0	1
flat Λ clumpy	0.3	0.7	0	0	0
no Λ	0.3	0		0.7	1
EdS	1.0	0		0	1
Quint 1	0.3	0.7	1	0	1
Quint 2	0.3	0.7	2	0	1
Quint 3	0.3	0.7	3	0	1

Table 3.1: Cosmological models used to calculate the distance parameters

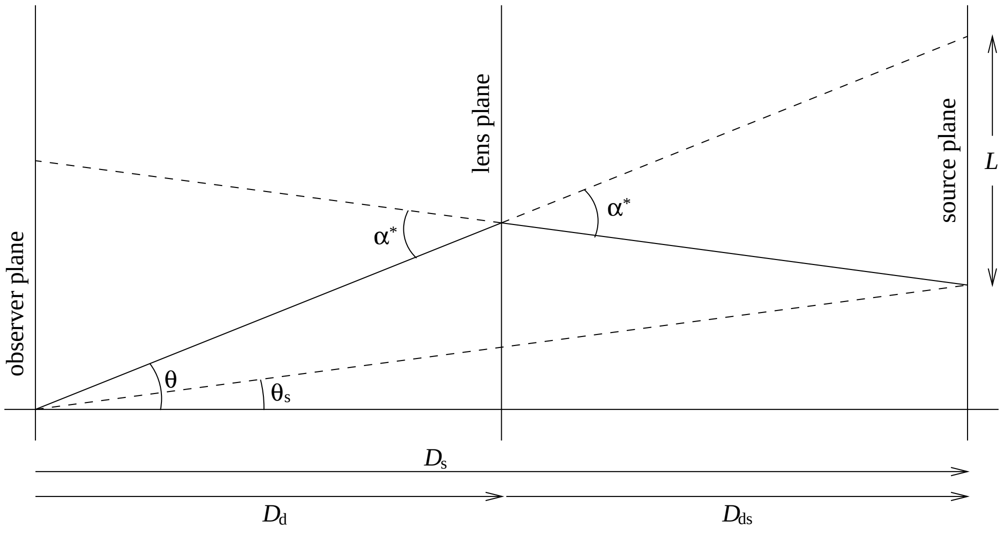


Figure 3.2: Sketch of the fundamental lensing geometry. The light ray is deflected by α^* in the lens plane. The observer sees the source at a position θ, which is displaced from the true position θ_s.

the same redshift. Multi plane lenses can be much more complicated, because the effects do not superpose linearly then. Small deflection angles are assumed throughout this work.

The lens equation can be derived from the geometrical diagram in Figure 3.2. The length L can be calculated from the true deflection angle α^* and from the apparent displacement $\theta - \theta_\mathrm{s}$,

$$\alpha^* D_\mathrm{ds} = (\theta - \theta_\mathrm{s}) D_\mathrm{s} \quad , \tag{3.10}$$

which directly leads to the lens equation

$$\theta_\mathrm{s} = \theta - \alpha \tag{3.11}$$

with the apparent deflection angle

$$\alpha = \frac{D_\mathrm{ds}}{D_\mathrm{s}} \alpha^* \quad . \tag{3.12}$$

All angles are two-dimensional vectors in reality.

3.4 Magnification and amplification

The Jacobian of the lens equation provides the magnification matrix M.

$$\frac{\partial \vartheta_s}{\partial \vartheta} = 1 - \frac{\partial \boldsymbol{\alpha}}{\partial \vartheta} \tag{3.13}$$

$$\mathsf{M} = \frac{\partial \vartheta}{\partial \vartheta_s} \tag{3.14}$$

$$= \left(1 - \frac{\partial \boldsymbol{\alpha}}{\partial \vartheta}\right)^{-1} \tag{3.15}$$

Sources are magnified linearly with M,

$$\Delta\vartheta = \mathsf{M}\Delta\vartheta_s \quad . \tag{3.16}$$

Because the lensing effect preserves surface brightness (e.g. Schneider et al., 1992), this magnification implies an apparent *amplification* of unresolved sources. This amplification is given by the area magnification, which is the determinant of the magnification matrix.

$$\mu = |\mathsf{M}| \tag{3.17}$$

The absolute value of μ is the flux amplification, the sign is the parity of the corresponding image (negative for a mirror image).

3.5 Deflection angle and potential

In the thin lens approximation, the effect of the lens can be described by a two-dimensional potential function Φ that describes a delay of light rays in the lens plane.

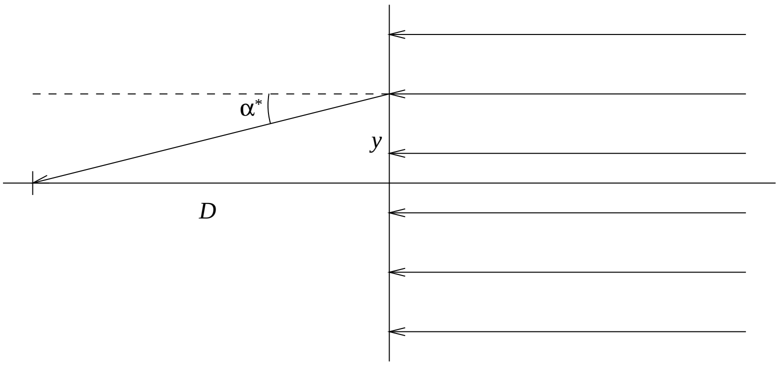


Figure 3.3: Deflection angle α^* in the lens plane

Figure 3.3 can be used to calculate the deflection angle from the potential. We assume incoming rays parallel to the optical axis for simplicity. The result would not change for other directions. The light travel time from the source to the observer, who is located at a cosmologically small distance D from the lens, is given by

$$T = \mathrm{const} + \frac{1}{c}\sqrt{D^2+y^2} + \Phi \tag{3.18}$$

$$\approx \mathrm{const} + \frac{D}{c} + \frac{1}{2c}\frac{y^2}{D} + \Phi \quad . \tag{3.19}$$

Following the theorem of Fermat (Kovner, 1990, for the cosmological case), this must be stationary:

$$0 = \frac{\mathrm{d}T}{\mathrm{d}y} \tag{3.20}$$

$$\approx \frac{1}{c}\frac{y}{D} + \frac{\mathrm{d}\Phi}{\mathrm{d}y} \tag{3.21}$$

$$\alpha^* \approx \frac{y}{D} \tag{3.22}$$

$$\approx -c\,\frac{\mathrm{d}\Phi}{\mathrm{d}y} \tag{3.23}$$

This result for the deflection angle is independent of D as expected. It can now be combined with Einstein's famous result of

$$\alpha^* = \frac{4GM}{c^2\,r} \tag{3.24}$$

for an impact parameter of r and a compact mass M (Einstein, 1915) to obtain the potential in this case:

$$\Phi(r) = -\frac{4GM}{c^3}\ln r \tag{3.25}$$

For a direct derivation of this result, the reader is referred to Wucknitz (1996). The potential for an arbitrary mass distribution is a linear superposition of all its parts. With $\mathbf{r} = (x,y)$ measuring the position in the lens plane, this reads

$$\Phi(\mathbf{r}) = -\frac{4G}{c^3}\int \mathrm{d}^2r'\ \Sigma(\mathbf{r}')\ln|\mathbf{r}-\mathbf{r}'| \quad , \tag{3.26}$$

with the projected surface mass density

$$\Sigma(x,y) = \int \mathrm{d}z\ \rho(x,y,z) \quad . \tag{3.27}$$

It is convenient to transform the integral and the derivative/gradient from dimensional vectors $\mathbf{r}$ to angular vectors $\boldsymbol{\theta}$ using $\mathbf{r} = D_\mathrm{d}\,\boldsymbol{\theta}$. If the deflection angle is at the same time transformed from the true value $\boldsymbol{\alpha}^*$ to the apparent one $\boldsymbol{\alpha}$, the equations can be written in a very simple form:

$$\boldsymbol{\theta}_\mathrm{s} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \tag{3.28}$$

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}\,\psi(\boldsymbol{\theta}) \tag{3.29}$$

The scaled version of the potential can be calculated with

$$\psi(\boldsymbol{\theta}) = -c\frac{D_\mathrm{ds}}{D_\mathrm{d}D_\mathrm{s}}\Phi(D_\mathrm{d}\boldsymbol{\theta}) \tag{3.30}$$

$$= \frac{1}{\pi}\int \mathrm{d}^2\theta'\ \sigma(\boldsymbol{\theta}')\ln|\boldsymbol{\theta}-\boldsymbol{\theta}'| \quad , \tag{3.31}$$

where the surface mass density was normalized with the so called 'critical density':[3]

$$\sigma = \frac{\Sigma}{\Sigma_\mathrm{c}} \tag{3.32}$$

$$\Sigma_\mathrm{c} = \frac{c^2}{4\pi G}\frac{D_\mathrm{s}}{D_\mathrm{d}D_\mathrm{ds}} \tag{3.33}$$

The name 'critical density' is due to the fact that a circular disc with a constant surface mass density equal to this value focuses all light rays coming from the source at the observer. Gauss' law can be used to prove that all mass inside the radius θ can be thrown into the centre for a radially symmetric mass distribution, without changing the deflection angle. This makes it simple to show that $\vartheta_\mathrm{s} \equiv 0$ for all ϑ, if $\sigma \equiv 1$. The Poisson equation for the potential in angular coordinates reads

$$\nabla_\vartheta{}^2 \psi = 2\sigma \quad . \tag{3.34}$$

3.6 Light travel times and time-delays

The physical light travel time t consists of the delay Φ in the lens plane, which is redshifted by $1+z_\mathrm{d}$, and a geometrical part depending on ϑ and ϑ_s. We work with a scaled version of t, called T, to eliminate the distance parameters from the equations.

$$t = \frac{D_\mathrm{eff}}{c} T \tag{3.35}$$

$$D_\mathrm{eff} = (1+z_\mathrm{d})\frac{D_\mathrm{d}D_\mathrm{s}}{D_\mathrm{ds}} \tag{3.36}$$

$$T = T_\mathrm{geom} + T_\mathrm{pot} \tag{3.37}$$

$$T_\mathrm{pot} = -\psi(\vartheta) \tag{3.38}$$

At this point, Fermat's theorem can again be used: t, or equivalently T, must be stationary if the lens equation is fulfilled. We thus set the gradient of T in (3.37) to 0 and use equation (3.38) together with the lens equation (3.28) and (3.29) to express the gradient of the geometrical part in terms of ϑ and ϑ_s:

$$\nabla_\vartheta T_\mathrm{geom} = \nabla_\vartheta \psi \tag{3.39}$$

$$= \vartheta - \vartheta_\mathrm{s} \tag{3.40}$$

This equation can now be integrated to obtain the absolute geometrical part. After adding the potential part, the total light travel time is given by

$$T = \mathrm{const} + \frac{1}{2}|\vartheta - \vartheta_\mathrm{s}|^2 - \psi(\vartheta) \quad . \tag{3.41}$$

Although Fermat's theorem applies only for the real image positions, this integration is allowed here. The expression in equation (3.41) is the only possible global solution which is quadratic in ϑ and ϑ_s. Higher order terms are neglected in the approximation of small angles without noticeable loss of accuracy. Equation (3.35) can be used to transform T to physical units.

---

[3] Often the designation κ is used instead of σ. We use κ only for additional and constant normalized surface mass densities.

3.7 Applications

The best known cosmological application of gravitational lenses is the determination of H_0 with the time-delay. This method was proposed by Refsdal (1964). If the distances D_i in (3.35) are substituted by their normalized counterparts d_i, with

$$D_i = \frac{c}{H_0} d_i \quad , \tag{3.42}$$

where the d_i can be calculated from the redshifts and cosmological parameters without H_0 alone, the equation can be written as

$$H_0 t = d_{\mathrm{eff}} T \quad , \tag{3.43}$$

using the effective distance parameter D_{eff} from equation (3.36) in its normalized form. Although it is not possible to measure the light travel times themselves, *differences* of light travel times for double or multiple images of one source (so-called 'time-delays') *can* be measured, if the source is variable. Equation (3.43) can also be used to determine the Hubble constant H_0 from a measured time-delay in physical units Δt and a lens model, which provides the quantity ΔT. In a simple geometrical interpretation of the equations, the time-delay is used to determine the scale of the lens geometry. All angles in the configuration can either be measured or determined with lens models, so that the measurement of a single length, in this case the difference of two light paths, determines all distances at once. Apart from cosmological corrections included in the definition of d_{eff}, the Hubble constant is proportional to redshift divided by distance and therefore $H_0 \propto t^{-1}$.

If both H_0 and the time-delay Δt are known, the same equation can be used to constrain the combination of distances d_{eff}, which depends on the cosmological parameters. This method to determine cosmological parameters was first discussed by Refsdal (1966). Another possibility to estimate cosmological parameters directly from the lens effect is to compare the lens models with other methods of mass determination. A singular isothermal sphere (see section 4.1.1) with a one-dimensional (e.g. radial) velocity dispersion of σ_v has a constant apparent deflection angle of

$$\alpha_0 = 4\pi \frac{D_{\mathrm{s}}}{D_{\mathrm{ds}}} \frac{\sigma_v^2}{c^2} \quad . \tag{3.44}$$

This deflection angle can be determined from the image configuration. If, on the other hand, the velocity dispersion is known, the distance ratio $D_{\mathrm{s}}/D_{\mathrm{ds}}$ can be determined and used to constrain cosmological parameters *without knowing a time-delay*. This method was proposed by the author of this thesis (Wucknitz, 1996, 1997) and others. One problem with this idea is the high accuracy which is needed for the *mass models* and measurements of σ_v to discriminate different cosmological models and the fact that the isothermal sphere is only one possible *dynamical model* for the lensing galaxy. If the real dynamical state differs from this simple assumption, (3.44) is not necessarily valid anymore. This problem is discussed by e.g. Saha (2000).

With many lenses, modelled with high accuracy, it may nevertheless be possible to apply this method in the near future. The dynamical state of the lenses may depend on their redshifts but not on the redhift of the lensed sources. An ensemble of lenses with comparable z_{d} but with a wide range of z_{s} could therefore in principle be used to break the degeneracy. A variant works with several sources at different redshifts, which are lensed by one and the same lens. The combination of different redshifts replaces the dynamical mass determination and can thus avoid this source of uncertainty. The only known lenses for this method are massive clusters of galaxies, which act as lenses for a high number of background sources, typically galaxies. The difficulty lies in the measurements of the source redshifts and in determining accurate models of the lens.

For all (cosmological and other astrophysical) applications of gravitational lensing, reliable mass models of lenses are needed. This thesis is concerned with explaining why accurate mass modelling is not trivial, and how the situation can be improved generally or for at least some favourable lens systems.

Chapter 4

Lens modelling

4.1 Models

The conventional approach in lens modelling is to use simple parametric mass models and fit the parameters to observational data. The choice of a family of models is often determined by the available number of constraints. This method has the advantage that unique solutions are possible in many cases, but the disadvantage that the parameter space of realistic mass distributions is not fully probed and that possible errors can be dramatically underestimated in some cases.

An alternative approach of non-parametric mass models has been proposed by Saha & Williams (1997) and Williams & Saha (2000). They describe the mass distribution by a large number of discrete pixels with a priori independent and arbitrary densities. Some physically motivated constraints are then used to reduce the freedom to a more realistic degree. These calculations have shown (not unexpectedly), that many different mass models can describe the observations of typical lens systems and that the uncertainties in any parameters derived from these can by huge. Especially the large range of values allowed for the Hubble constant is sobering.

Because it is known that some of the possible non-parametric models are still very unrealistic mass distributions for real galaxies, some intermediate approach should be used to obtain useful and accurate results without underestimating the errors. At least three approaches are possible. One could start from the non-parametric models and include more constraints for the allowed mass distributions or use more general parametric models which cover all realistic mass distributions. Alternatively, one could use analytical considerations to understand possible model degeneracies better and find ways to break them with more and better observations. Parts of this thesis belong to the last two categories.

Many parametrized mass models have been used during the last decades to fit known lenses. The most popular class of models is based on the isothermal sphere. Variants of this are pseudo-isothermal ellipsoidal mass distributions and others. These models are generally seen as very good approximations for the mass distributions in real galaxies. We therefore start our own investigations from this base and generalize only so far as to include the most important possible deviations. This includes more general asymmetries than simple elliptical mass distributions and modified radial mass distributions, which can have a dramatic effect, especially on the Hubble constant. The latter is done by using power-law potentials, which are good approximations for even more general models. These days it is still difficult to determine the radial mass slope of power-law models directly from lensing observations. Possible deviations from power-law models are a higher order effect, which only becomes important when the constraints are improved beyond the present level.

The general asymmetries are investigated by analytical calculations for semi-parametric models (section 6). The radial density distribution is again taken as a power-law, whilst the azimuthal dependence can have arbitrary (and thus non-parametric) forms.

The building blocks for our models will be described briefly in the following sections. More details can be found in the literature.

4.1.1 The singular isothermal sphere

The singular isothermal sphere (SIS) has a potential, deflection angle and surface mass density like follows ($r = |z|$):

$$\psi(z) = \alpha_0 r \tag{4.1}$$

$$\alpha(z) = \alpha_0 \frac{z}{r} \tag{4.2}$$

$$\kappa(z) = \frac{\alpha_0}{2r} \tag{4.3}$$

It is thus characterized by a constant absolute deflection angle α_0, which is related to the measurable velocity dispersion by equation (3.44). The popularity of this model is not only motivated by the accuracy with which it describes real galaxies, but also by the very convenient properties it has for the lensing effect.

4.1.2 External shear

Perturbations of the spherical lens models can be described in several ways. On is the perturbation caused by additional external masses. If the effect is sufficiently small and varies only slowly in the regions of interest (if the masses are sufficiently far away), the effect can be described by a quadratic approximation for the potential, which is equivalent to a linear approximation for the deflection angle.

$$\psi(z) = \frac{1}{2} z^\dagger \Gamma z \tag{4.4}$$

The shear matrix Γ includes both the convergence κ and the shear $\boldsymbol{\gamma}$ itself.

$$\Gamma = \begin{pmatrix} \kappa - \gamma_x & -\gamma_y \\ -\gamma_y & \kappa + \gamma_x \end{pmatrix} \tag{4.5}$$

$$\boldsymbol{\gamma} = \gamma \left(\cos 2\theta_\gamma, \sin 2\theta_\gamma \right) \tag{4.6}$$

The constant term is irrelevant for lensing, the linear term is equivalent to a constant shift of source positions and can be neglected as well. The direction of shear θ_γ is defined to point in the direction of the perturbing mass or equivalently in the opposite direction. Another definition, rotated by $\pi/2$, is also commonly used in the literature.

4.1.3 Spherical power-law models

The power-law model is used as a generalization of the isothermal sphere more for analytical than for physical reasons. Other approaches are possible, but the power-law models are much easier to analyse. To be as independent of astrophysical knowledge or prejudice as possible, models should be chosen to be very general and cover all physically sensible possibilities. We therefore favour power-law models before other generalizations. We do not discuss central mass concentrations like central black holes, cusps or finite core radii in this thesis. Otherwise we would have used modifications like softened power-laws, double power-laws or added point masses. A power-law exponent β is used for the potential:

$$\psi(\boldsymbol{z}) = a_0 r^{\beta} \tag{4.7}$$

$$\boldsymbol{\alpha}(\boldsymbol{z}) = \beta\, a_0\, r^{\beta-1}\frac{\boldsymbol{z}}{r} \tag{4.8}$$

$$\kappa(\boldsymbol{z}) = \frac{\beta^2 a_0}{2} r^{\beta-2} \tag{4.9}$$

The limiting cases of $\beta \to 0$ and $\beta = 2$ correspond to a point mass and a constant surface mass density, respectively. Isothermal models are another special case between the two with $\beta = 1$. We use angular coordinates, which are dimensionless quantities. If units like arcsec or mas are used, r should be normalized by some unit length and a_0 corrected accordingly.

4.1.4 Elliptical potentials

A relatively simple approximation to generalize spherical models is the elliptical potential. It starts from the spherical potential $\psi(r)$, which is then transformed to an elliptical radial coordinate

$$\psi = \psi(r_{\text{ell}}) \quad , \tag{4.10}$$

$$r_{\text{ell}} = \sqrt{\frac{x^2}{(1+\varepsilon)^2} + \frac{y^2}{(1-\varepsilon)^2}} \tag{4.11}$$

in the principal axes system. These models were analysed and compared with true elliptical mass distributions (see next section) by Kassiola & Kovner (1993). For small ellipticities ε, they are a good approximation, but they lead to unphysical density functions for large ellipticities. We compared the two families of models in many practical calculations and found very good agreement in most cases. Elliptical potentials have the advantage that their deflection angles are analytic and simple to calculate. This will be especially important for the LENSCLEAN calculations (chapter 9). The deflection angle can be calculated with

$$\boldsymbol{\alpha} = \frac{\psi'(r_{\text{ell}})}{r_{\text{ell}}} \left(\frac{x}{(1+\varepsilon)^2}, \frac{y}{(1-\varepsilon)^2} \right) \quad , \tag{4.12}$$

where ψ' denotes the derivative with respect to the argument. A deeper discussion of these models can be found in Kassiola & Kovner (1993). We discuss the correspondence to elliptical mass distributions for small ellipticities in section 6.14.

4.1.5 Elliptical mass distributions

These are considerably more complicated to handle than elliptical potentials, but they are of course the more realistic mass distributions. We follow the complex formalism of Kassiola & Kovner (1993) for these models. Complex quantities are written as vectors $\boldsymbol{z}$, real and imaginary parts represent the x and y components, respectively. Details of this formalism are discussed in

Wucknitz (1996). Analytical deflection angles are only known for a small number of values of β. For pseudo-isothermal models (isothermal with a finite core radius r_c), the deflection angle reads like this:[1]

$$R = \sqrt{r_\mathrm{c}^2 + r_\mathrm{ell}^2} \tag{4.13}$$

$$\boldsymbol{\alpha} = \frac{(1-\varepsilon^2)\,\alpha_0}{2\mathrm{i}\sqrt{\varepsilon}} \ln \frac{\frac{1-\varepsilon}{1+\varepsilon}x - \mathrm{i}\frac{1+\varepsilon}{1-\varepsilon}y + 2\mathrm{i}\sqrt{\varepsilon}R}{\boldsymbol{z}^\star + 2\mathrm{i}r_\mathrm{c}\sqrt{\varepsilon}} \tag{4.14}$$

$$\psi = \boldsymbol{z}\cdot\boldsymbol{\alpha} - \frac{(1-\varepsilon^2)\,\alpha_0 r_\mathrm{c}}{2} \ln \frac{\left((1-\varepsilon^2)\,R + (1+\varepsilon)^2\,r_\mathrm{c}\right)^2 + 4\varepsilon x^2}{(1+\varepsilon)^2(1+\varepsilon^2)} \tag{4.15}$$

$$\kappa = \frac{\alpha_0}{2R} \tag{4.16}$$

The expression for the potential is a significantly simplified version of the result in Kassiola & Kovner (1993). A treatment with real vectors can be found in Kormann, Schneider & Bartelmann (1994).

4.2 The main algorithm

Only some simple lens model families allow a direct analytical fitting of model parameters to observations. Usually numerical methods have to be used. For these, it is necessary to define a function, which assesses the goodness of fit. Residuals in the form of a χ^2 are preferred, because they lead to a maximum likelihood result and confidence limits can be calculated easily with the methods discussed in section 2.7. Once the residual function is defined and implemented in the modelling code, different minimization algorithms can be used to find the best lens model. We generally used the Powell minimization technique, but in some very unstable cases, when the calculation of the residual function very often fails, the more primitive downhill simplex method was more reliable. Press et al. (1992) give a good overview of some of the standard methods. Details depend on the constraints used in the system in question.

4.3 Constraints

4.3.1 Lens position

If the lens position is measured directly, it can either be used as fixed or it is formally included as free parameter and the difference to the measurements is included in the χ^2 function. The latter approach is generally the preferred, but the former can be used if the measurement accuracy is very high or to find a first model which can then be refined with the other method.

4.3.2 Image positions

These are the most fundamental lensing constraints. They can be included into the modelling with several different methods. The conceptually simplest but numerically most complicated and unstable method is to include the source position as parameter and calculate the positions of all images for each tested lens model. This involves the inversion of the lens equation, which is a highly non-trivial problem. In some cases it can be performed analytically (see section 4.5),

---

[1] Complex conjugation is denoted by $\boldsymbol{z}^\star$.

but generally a numerical search is necessary. The measured image positions can then be related to their corresponding model image positions, and the χ^2 contribution can be calculated. The numerical burden is high compared to other methods, but the residuals do exactly describe the difference between model and measurements.

In classical lens modelling, the lens inversion is made simpler by the fact that approximate image positions are already provided by the measurements. It is in some cases nevertheless necessary to find any *additional* images to compare their data with measured flux limits for possible additional images.

The new numerical algorithm LENTIL will be discussed in chapter 5. This method was developed to solve the lens equation for very many source positions for each lens model, which is needed for LENSCLEAN. It can also be used for classical model fitting but is not the most effective method for this purpose.

A zero-order approximation to the full algorithm is the source plane minimization proposed by Kayser (1990). The image positions are directly taken from the measurements and the lens model is fitted by minimizing the dispersion of the projected sources positions from these images. There is no simple way to combine these residuals with those from other constraints, and the projection can change the error statistics considerably. The results are therefore *not* equivalent to the full algorithm. We only used this very simple approach to find a first model and in cases where the image positions were fitted exactly anyway. If the dispersion is zero, it makes no difference if it is measured in the lens plane or in the source plane.

Kochanek (1991) proposed to project the differences back to the lens plane with the magnification matrix, but he used the isotropic amplification to approximate the matrix. In what follows, we generalize this method to work with the exact magnification matrix. The lens equations for model image positions z_i^{m} and a model source position $z_{\mathrm{s}}^{\mathrm{m}}$ read

$$z_{\mathrm{s}}^{\mathrm{m}} = z_i^{\mathrm{m}} - \alpha(z_i^{\mathrm{m}}) \quad . \tag{4.17}$$

The projected source positions z_{s}^i for the measured images z_i are

$$z_{\mathrm{s}}^i = z_i - \alpha_i \quad , \tag{4.18}$$

$$\alpha_i = \alpha(z_i) \quad . \tag{4.19}$$

The difference between measurement and model can be linearly approximated as

$$z_{\mathrm{s}}^i - z_{\mathrm{s}}^{\mathrm{m}} \approx \mathsf{M}_i^{-1}(z_i - z_i^{\mathrm{m}}) \quad , \tag{4.20}$$

with the magnification matrix

$$\mathsf{M}_i = \mathsf{M}(z_i) \quad . \tag{4.21}$$

With this equation, we can transform the residual contribution from the lens plane to the source plane:

$$\chi^2_{\mathrm{pos}} = \sum_i (z_i - z_i^{\mathrm{m}})^\dagger \mathsf{C}_i^{-1} (z_i - z_i^{\mathrm{m}}) \tag{4.22}$$

$$= \sum_i (z_{\mathrm{s}}^i - z_{\mathrm{s}}^{\mathrm{m}})^\dagger \mathsf{M}_i \mathsf{C}_i^{-1} \mathsf{M}_i (z_{\mathrm{s}}^i - z_{\mathrm{s}}^{\mathrm{m}}) \tag{4.23}$$

C_i is the error covariance matrix of the ith image position. The source position minimizing the residuals can now be calculated easily. The resulting $z_{\mathrm{s}}^{\mathrm{m}}$ is a weighted mean of the individual projections z_s^i:

$$z_s^m = \left(\sum_i \mathsf{M}_i \mathsf{C}_i^{-1} \mathsf{M}_i\right)^{-1} \sum_i \mathsf{M}_i \mathsf{C}_i^{-1} \mathsf{M}_i z_s^i \tag{4.24}$$

The χ^2 contribution for the numerical outer minimization loop can then be calculated from (4.23), the model image positions from the inverse of (4.20). A similar formalism was independently developed and published by Trotter, Winn & Hewitt (2000).

With an extended formalism, it is also possible to simultaneously fit the position of the lens centre analytically. We do not present the slightly more complicated equations here, because this modification works not very reliably and destabilizes the algorithm in some cases.

4.3.3 Flux ratios

Measured flux ratios f_A/f_B can be compared with model values with a contribution of

$$\chi^2_{\text{flux}} = \frac{(f_A/f_B - \mu_A/\mu_B)^2}{\sigma^2} \quad . \tag{4.25}$$

In the case of double images, the inclusion of the negative sign (doubles always have different parities) helps in rejecting certain configurations with higher multiplicities. If no information about image parities is available, the absolute values of f_A/f_B and μ_A/μ_B have to be used.

4.3.4 Relative positions of subcomponents

Some lens systems show substructure of the images on a very small scale. The subcomponents can be treated as different sources, but a linear approach is more appropriate, because these subcomponents can directly constrain only the (relative) magnification matrices. If Δz denote the relative displacement of subcomponents, the equations for the mapping of the model values can be written as

$$\Delta z_i^m = \mathsf{M}_i \Delta z_s^m \tag{4.26}$$

with $i = \text{A,B}$. The contribution to the residuals is

$$\chi^2_{\text{subpos}} = \sum_{i=\text{A,B}} \left(\Delta z_i - \mathsf{M}_i \Delta z_s^m\right)^\dagger \mathsf{C}_i^{-1} \left(\Delta z_i - \mathsf{M}_i \Delta z_s^m\right) \quad , \tag{4.27}$$

with the error covariance matrices C_i. The analytical solution for Δz_s is a weighted mean of the transformed Δz_i:

$$\Delta z_s^m = \left(\sum_{i=\text{A,B}} \mathsf{M}_i \mathsf{C}_i^{-1} \mathsf{M}_i\right)^{-1} \sum_{i=\text{A,B}} \mathsf{M}_i \mathsf{C}_i^{-1} \Delta z_i \tag{4.28}$$

$$\chi^2_{\text{subpos}} = \sum_{i=\text{A,B}} \Delta z_i^\dagger \mathsf{C}_i^{-1} \Delta z_i - \left(\Delta z_s^m\right)^\dagger \left(\sum_{i=\text{A,B}} \mathsf{M}_i \mathsf{C}_i^{-1} \mathsf{M}_i\right) \Delta z_s^m \tag{4.29}$$

4.3.5 Subcomponent shapes

If subcomponent shapes are described as ellipses, the formalism explained in appendix A.2 can be used to exploit these data as constraints. This Cartesian formalism proved to be more stable and reliable than the normally used polar description.

4.3.6 Time-delays

If several time-delays are known, their ratios can be used as additional constraints in a straightforward way. It is then better to use the Hubble constant as additional parameter and use the time-delays directly (not the ratios) as constraints. See chapter 6 for details and further analytical considerations.

4.4 Error estimates

To estimate the accuracy of one ore more (m) model parameters, their values can be fixed at a range of test values and the residuals can be minimized varying only the remaining parameters. The difference of the residuals to the minimal value follows a χ^2 distribution with m degrees of freedom. The methods discussed in section 2.7 can thus be used to calculate confidence limits from the residuals.

More sophisticated methods have to be used for the accuracy of parameters that are not directly part of the lens model. An important example is the Hubble constant, which is determined from a combination of other parameters and cannot be fixed easily. For parameters like this, Lagrangian methods can be used. A multiplier λ is used to calculate confidence limits for a parameter p by adding another contribution to the residuals:

$$\chi^2_\lambda = \lambda\,(p - p_0) \tag{4.30}$$

The usual Lagrange method would now minimize the total χ^2 by varying the model parameters *and* λ. The result would be the best fit with the secondary condition $p = p_0$. This can be repeated for a range of values of p_0 to estimate the confidence limits from the χ^2 distribution.

A better approach is to use constant λ and $p_0 = 0$. Varying λ has the same effect as varying the parameter p_0 then and can thus be used to scan the possible values without controlling p explicitly.

4.5 Analytical lens equation inversion

Different methods to avoid the full lens equation inversion have been discussed in section 4.3.2. In some cases (especially for the LENSCLEAN algorithm), this is however not possible. In these cases, certain simple lens models can be used, for which the inversion can be done analytically.

The typical problem in LENSCLEAN is to find all 'secondary' (additional) images of a source position, for which *one* image position is already known. This information can be used to simplify the search for the other images. Explicit solutions have been calculated for spherical singular isothermal models with external shear and for singular isothermal elliptical potentials. Results for the latter will be discussed in chapter 9, so we only discuss the lens inversion for this case here. We start with the lens equation for an elliptically modified isothermal model (sections 4.1.1 and 4.1.4). The lens equation for x and y in the principal axes system reads like follows:

$$x_\mathrm{s} = x\left(1 - \frac{\alpha_0}{(1+\varepsilon)^2}\,\frac{1}{r_\mathrm{ell}}\right) \tag{4.31}$$

$$y_\mathrm{s} = y\left(1 - \frac{\alpha_0}{(1-\varepsilon)^2}\,\frac{1}{r_\mathrm{ell}}\right) \tag{4.32}$$

The equations can be combined by solving the first for x, the second for y and calculate r_ell^2 from this. The result is a quartic equation for r_ell. If one solution (x_1, y_1) with corresponding $r_{\mathrm{ell}1}$ is already known, it can be used to eliminate the source position. After that, the quartic polynomial can be divided by $r_{\mathrm{ell}} - r_{\mathrm{ell}1}$ to remove the known solution from the polynomial. The resulting cubic equations may also be useful for other applications and shall be presented here:

$$\begin{aligned}
&-(1-\varepsilon^2)^4\, r_{\mathrm{ell}}^3 \\
&+(1-\varepsilon^2)^2 \left(4\alpha_0(1+\varepsilon^2) - r_{\mathrm{ell}1}(1-\varepsilon^2)^2\right) r_{\mathrm{ell}}^2 \\
&+\left(2\alpha_0\, r_{\mathrm{ell}1}\,(1-\varepsilon^2)^2\,(1+2\varepsilon m_1+\varepsilon^2) - \alpha_0^2\left(4\varepsilon m_1\,(1+\varepsilon^2)+5-2\varepsilon^2+5\varepsilon^4\right)\right) r_{\mathrm{ell}} \\
&+2\alpha_0^3\,(1+2\varepsilon m_1+\varepsilon^2) - \alpha_0^2\, r_{\mathrm{ell}1}\left(4\varepsilon m_1\,(1+\varepsilon^2)+1+6\varepsilon^2+\varepsilon^4\right) = 0
\end{aligned} \tag{4.33}$$

$$m_1 = \frac{1}{r_{\mathrm{ell}1}^2}\left(\frac{x_1^2}{(1+\varepsilon)^2} - \frac{y_1^2}{(1-\varepsilon)^2}\right) \tag{4.34}$$

This equation can be solved analytically. The result for r_{ell} can then be used to calculate x and y with equations (4.31) and (4.32). This approach is more robust than most numerical methods, but even the analytical solution sometimes fails very close to the axes.

4.6 Examples

We used the numerical methods described in the last sections as well as analytical calculations to fit mass models for several lens systems. Most of this work is not in direct relation to the rest of this thesis and shall only be summarized here to point the reader to the corresponding publications.

4.6.1 HE 1104–1805

The lens system HE 1104–1805 raised some questions to the theoreticians, when the position of the lensing galaxy could first be measured. It turned out that the centre of the lens is close to the brighter of the two images and not to the fainter one, which simple models would demand. First interpretations favoured a model where the brighter image consists of two unresolved very closely separated images, probably very near to a radial critical curve. We found out, however, that even much simpler models are able to reproduce the strange properties (Remy et al., 1998). A singular isothermal sphere model is sufficiently modified by an external shear to adjust the amplification ratio to the one observed.

The system remained very interesting. In Wisotzki et al. (1998), we used the dispersion minimization method (Pelt et al., 1994, 1996) to determine a first time-delay from spectrophotometric monitoring data. The best result was 0.73 years, but the accuracy was quite low. We also used ellipsoidal mass distribution models *plus* external shear to estimate the (then still unknown) redshift of the lens from the time-delay and a canonical value for H_0. If the lens galaxy is associated with one of the absorption line systems detected in the QSO light, the one at $z = 0.73$ is the most probable. Later Gil-Merino, Wisotzki & Wambsganß (2002) published a more detailed study of the time-delay. Their methods are, however, still under discussion (Refsdal, priv. comm.).

4.6.2 RX J0911.4+0551

The system RX J0911.4+0551 was discussed in Burud et al. (1998). Our first modelling attempts showed that this very interesting quad system can be explained with elliptical models and external shear. The direction of the shear is compatible with the direction of a nearby rich cluster of galaxies. Some more details of this systems will be discussed in section 6.13.4. Particularly interesting is the fact that measurements of the mass of the perturbing cluster (Kneib, Cohen & Hjorth, 2000) seem to be very low compared with the needs of the lens models.

4.6.3 HE 2149–2745

For a publication on the system HE 2149–2745 (Lopez, Wucknitz & Wisotzki, 1998), we contributed first lens models that used the proposed measurement of the galaxy position. It seemed to be located almost exactly between the two images. We developed analytical methods to study singular isothermal ellipsoidal mass distributions with external shear in cases where the images are located exactly on the major axis of the lens. Although it turned out later that the measured galaxy position was highly inaccurate (Kochanek et al., 2002), these analytical calculations are still of some interest. To the knowledge of the author, this was the first publication that discussed possible configurations with *up to eight* images for these relatively simple lens models.

4.6.4 1517+656

The BL Lac object 1517+656 is a special case. This source is not lensed but the host galaxy of the AGN seems to *act* as a lens itself. Scarpa et al. (1999) reported the discovery of a fragmented Einstein ring around the object with a radius of $2\overset{\prime\prime}{.}4$. No accurate astrometry or redshift of the ring are available, but from the ringlike geometry and the assumption that it is indeed caused by a lensed background object, it was possible to estimate the mass and velocity dispersion of the lens as a function of the source redshift. The enclosed mass is of the order $2\cdot 10^{10}M_\odot$ or more, the velocity dispersion is at least $350\,\mathrm{km\,s^{-1}}$. Both values have to be even much higher, if the source redshift is 2 or less. This is, to the knowledge of the author, the only direct measurement of a BL Lac host galaxy mass and thus of high astrophysical importance. Unfortunately, the small size of the system and the low surface brightness make a deeper observational investigation very difficult.

Chapter 5

A new technique to solve the lens equation (LENTIL)

5.1 Introduction

A fundamental building brick of many lens modelling algorithms is a method to invert the lens equation (3.28), i.e. to find all images z_i for a given source position z_s. Only for some relatively simple lens models can the inversion be performed analytically (see e.g. section 4.5). Generally, numerical methods have to be used. Things are relatively simple in classical lens modelling (chapter 4), where approximate solutions are given by the observed image positions. Refining these to obtain exact values is a numerical standard problem. If the algorithm fails in some cases, not all observed images are found at their correct positions, which will generally increase the residuals for these models, and the overall minimization method will move to slightly modified lens models that are otherwise equivalent. If this happens very often, the minimization can become very inefficient, but failed fits are always easily identified as such.

The problem becomes much more serious when the LENSCLEAN method is used (see chapter 9). In the course of this algorithm, the lens equation has to be inverted for each source position in the source emission model and for each tested lens model. Typical model fits thus rely on a very high number (typically 10^7) of lens equation inversions. In contrast to classical lens models, incorrect results often increase the freedom of the emission models and lead to slightly reduced residuals. If the reduction is stronger than the true variations at this stage of the minimization, the main fitting algorithm gets stuck at one of the lens models for which only one lens equation inversion failed.

For our model fits, we generally scanned a subspace of the possible parameter space and fitted the remaining parameters. If failed fits occur only in a very small fraction of the calculations (say a few per cent), they can again be identified and rejected from the further analysis. If the fraction becomes too large, reliable fits are made impossible.

To summarize: An extremely reliable (failure less often than one in 10^8 runs) algorithm is needed. Even analytical solutions like the one discussed in section 4.5 can reach this level of robustness only with some extra effort. The solution of this problem is made even more difficult by the demands on the typical execution times. If the total CPU time needed for LENSCLEAN should increase by less than a factor of two by the use of the numerical lens equation inversion technique, the algorithm has a budget of only a few milliseconds (typically 3 msec) for each lens model and source position.

Most numerical methods that are in use today have problems in reliably finding *all* images, which is absolutely essential for LensClean. A discussion of some methods and the problem in general can be found in Schneider et al. (1992).

Our solution to the problem was inspired by a tiling algorithm used in a publicly available lens modelling software package (Keeton, 2001a,b). That algorithm is based on an adaptive tiling of the lens plane (finer close to the critical curves). The tiles are then projected back to the source plane, and the lens plane positions of those tiles, that enclose the source position in question, are used as starting values for a standard numerical refinement algorithm. This method works very efficiently and is appropriate for the needs of classical lens model fitting. Unfortunately, the robustness is by far not sufficient for the use with LensClean (Keeton, priv. comm.). We therefore had to develop a new algorithm that is designed to be as reliable as possible without becoming to slow for complicated lens models.

An advantage of the LensClean application is that very many inversions have to be done for each lens model. Some overhead in the preparation of the model (once for each model but for all source positions) is thus acceptable.

5.2 One-dimensional root finding

The lens equation can be written as a two-dimensional function, whose roots are the solutions. In one dimension, conceptually very simple methods are readily available. They can even be combined with more efficient modifications of the Newton method. The idea of a robust one-dimensional algorithm is sketched in Figure 5.1.

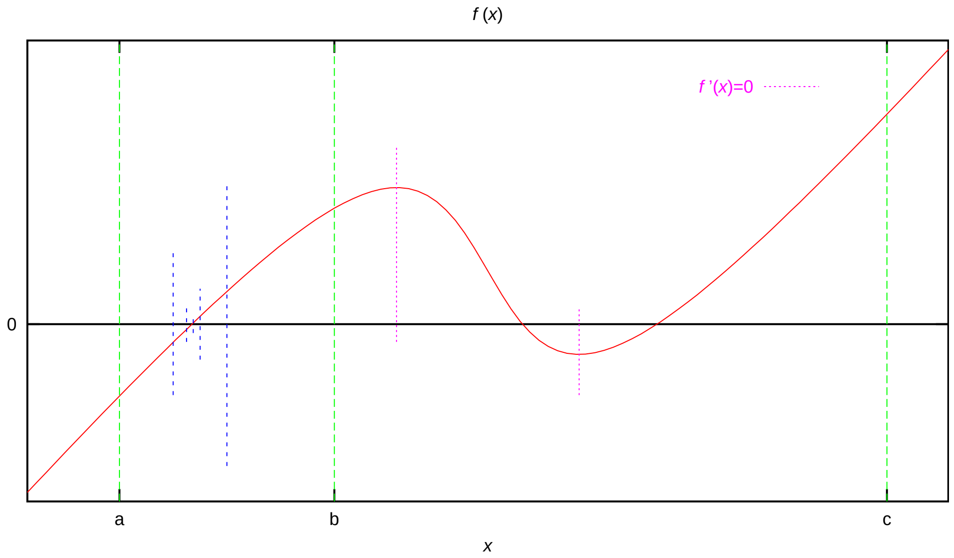


Figure 5.1: One-dimensional root finding. Shown are the bracketing of the zero between a and b and the stationary points with $f'(x) = 0$.

Let the function $f(x)$ be continuous, and take the starting values a and b as given. Because the signs of $f(a)$ and $f(b)$ are different, it is known that a and b bracket (at least) one zero of the function. One can then successively pick values of x midway between the limits and replace one of the old limiting values by the new value, depending on the sign of f at this position. Each step halves the size of the interval, and at the limit the zero is bracketed with infinite accuracy.

This strategy is not very efficient but cannot fail, if a and b bracket a solution at the beginning.

If the initial interval encloses more than one zero (e.g. a and c), the algorithm always finds one of the solution but never both. A recipe to find *all* zeros could be the following: Start with determining all stationary points ($f' = 0$). Then use as initial intervals the ones given by these points. At the left and right side, use a sufficiently high $|x|$ to be sure that no zeros are lost. For each of these intervals perform the subdivision until the required accuracy is reached.

In the lensing problem, the selection of appropriate lower and upper limits for x is not problematic. More serious is the problem of finding the zeros of $f'(x)$. In lensing, it is usually known how many zeros can at most occur. It is then sufficient to start with points in each domain defined by the sign of f' (positive or negative slope of f). The domain borders, whose total number is known, can then also be found by bracketing the zeros of f'. This strategy could not be used for the function f directly, because the number of domains depends on the source position. One can then never be sure that all domain borders are included in the initial starting points and could easily miss close double images. This is not true for the first derivative, because it only depends on the lens model but not on the source position.

The initial search for the domain borders has to be performed only once for each lens model. The same borders can be used as starting intervals for all values of z_s.

5.3 The basic idea of LENTIL

The idea for our new algorithm is to use as many of the ideas from the one-dimensional case as possible. The intervals translate to tiles, which cover the lens and source plane, hence the name Lens Tiling algorithm or 'LENTIL'.[1]

Several possibilities have been tried as generalization of the subdivision of intervals. In the final version, triangular tiles are used for several reasons. They are the simplest possible tilings, which makes bookkeeping less cumbersome, and they stay convex under any kind of mapping. A convex triangle in the lens plane is always mapped onto a convex triangle in the source plane, which is fundamental for the algorithm. Just as the intervals cover the total range in the one-dimensional case, is the whole region of interest in the lens plane covered by the tiling. Each point is covered by exactly one sub-tile at all stages of the algorithm.

See Figure 5.2 for how the subdivision is performed. One side (or edge) of the old tiling is subdivided (in this case in the middle), and a new vertex is introduced at this position. This produces four new tiles from the two old ones neighbouring the side. In the source plane, the new vertex is not necessarily located on one of the old sides. It can even happen that it falls outside of the two old tiles. In this case tiles with incorrect parity (different from the mapping parity) are produced. An important difference of our algorithm from the one presented in Keeton (2001a) is that the subdivision of a certain tile also affects the neighbouring tiles. No gaps or doubly covered regions can therefore form in the process of successive subtiling.

5.4 The initial tiling

The first step in the one-dimensional case was to find the domain borders. The same has to be done in the two-dimensional case. The domains are now defined by the sign of the (inverse) amplification $\mu^{-1} = |\partial z_\mathrm{s}/\partial z|$, which is also called the 'parity' of the images. The borders are

---

[1] The german words for 'lens' and 'lentil' are the same („Linse").

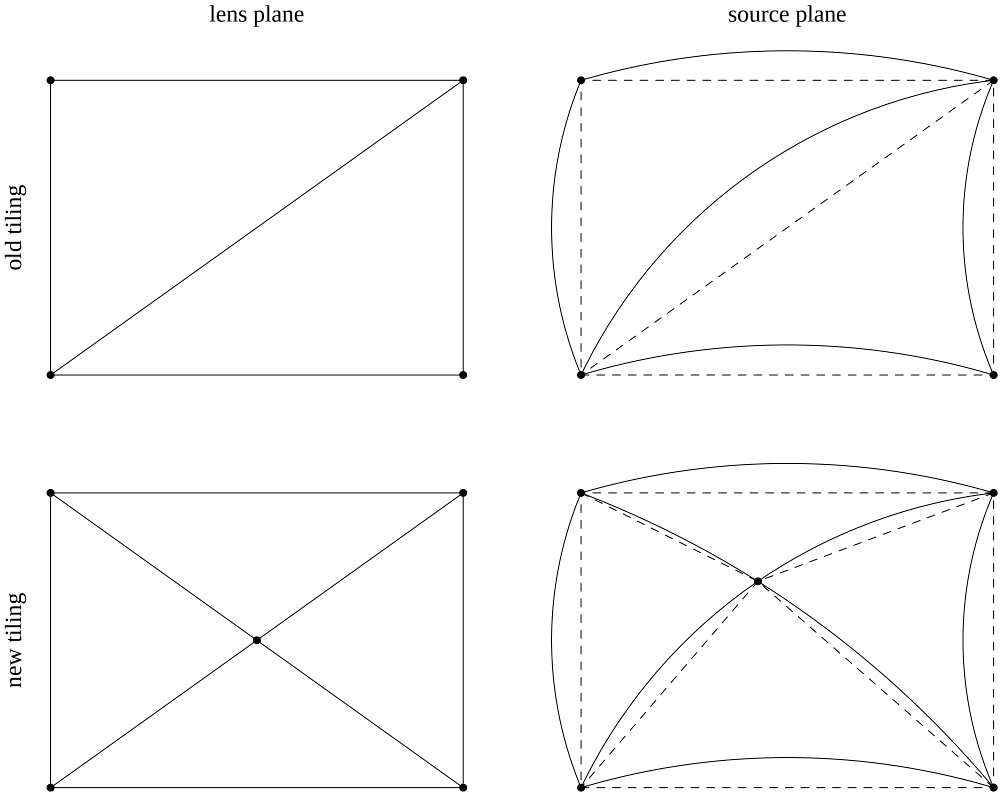


Figure 5.2: Subtiling by subdivision of one side. The left plots show the lens plane, where the subtiling is performed. The right plots show the projection into the source plane. Solid lines show projections of tile sides, while straight dashed lines represent the source plane tiling.

thus lines of infinite amplification or ‘critical curves’. They can be parametrized as a sufficiently dense collection of points.

To start, each domain has to be populated with at least one vertex. This can only be done knowing the properties of the lens model. A point mass e.g. has a small region of negative parity around the singularity and positive parity outside. An isothermal sphere with finite core radius has positive parity in the centre, then a region of negative parity and again positive parity outside. With this knowledge, a radial search can easily find representative points for all domains. Problems are only expected with multiple component lens models, where the other components can disturb the search for one component. In the test models used so far, the final algorithm always worked successfully.

This first stage of the initial tiling for an elliptical isothermal model with finite core radius is shown in Figure 5.3. We also included an algorithm to encapsulate the centres (there can be many for multi-component models) by successively splitting the edges of tiles that connect the centres with outer areas. This helps to make the algorithm more efficient for singular lenses.

In the next step, those tile edges that connect vertices of different parities are split exactly at the critical curve (Figure 5.4). This is done with a one-dimensional bracketing algorithm along the direction of the edge. After this step, all edges connect vertices with equal parities. The only exception are those which connect zero parity, located on the critical curves, with arbitrary

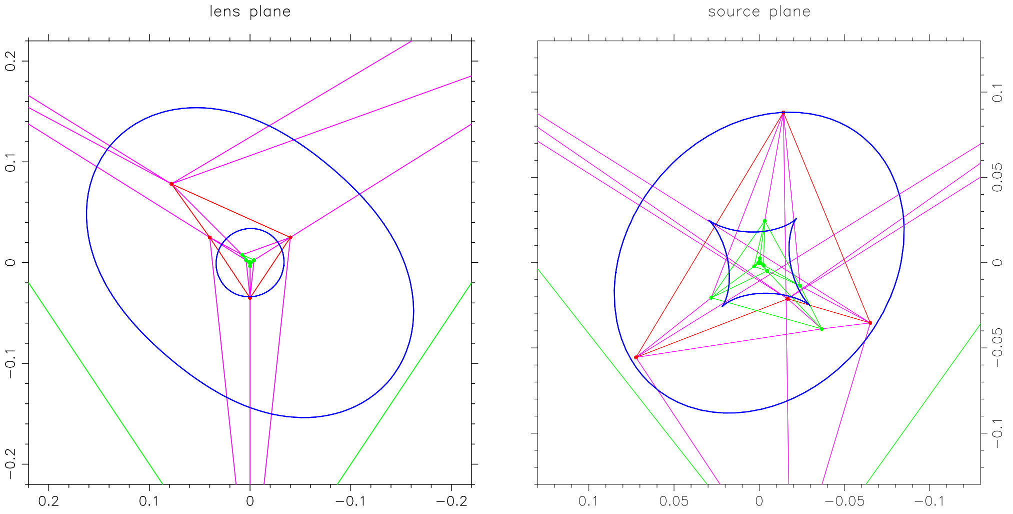


Figure 5.3: First stage of the initial tiling. All domains are occupied by at least one vertex. The plots also include the critical curves (left) and caustics (right). In the colour version, green marks positive and red marks negative parity. Critical curves are blue.

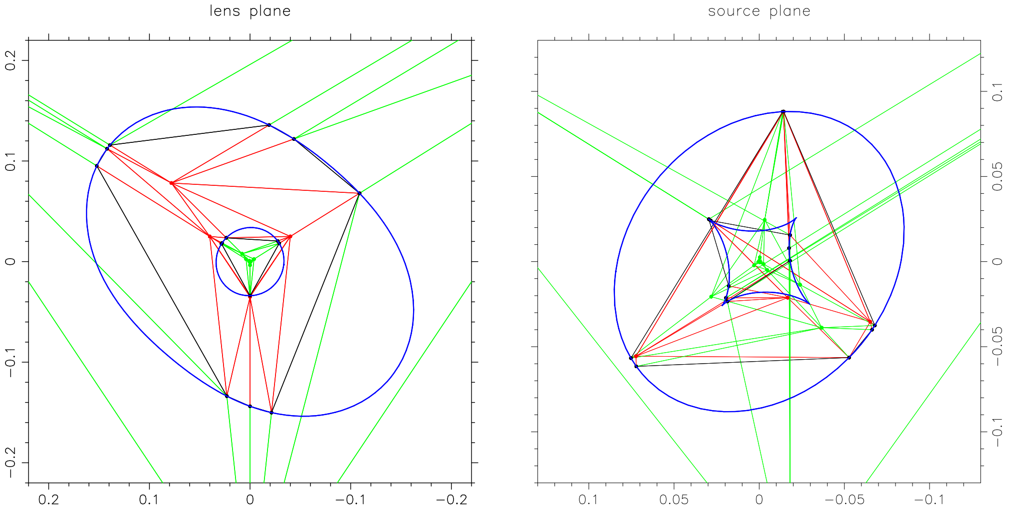


Figure 5.4: Second stage of the initial tiling. Edges with different parities of the vertices are split at the critical curve. The plots also include the critical curves (left) and caustics (right). In the colour version, green marks positive and red marks negative parity. Critical curves are blue.

parity.

Finally, the sampling of the critical curve is refined by successively splitting zero-zero parity edges until a limiting length is reached (Figure 5.5). The size of this length is chosen according to the required accuracy of finding all images close to the critical curve. Figure 5.6 illustrates situations in which close double images can be lost near the critical curve. The sampling has to be fine enough, that these images are either excluded anyway due to amplification limits

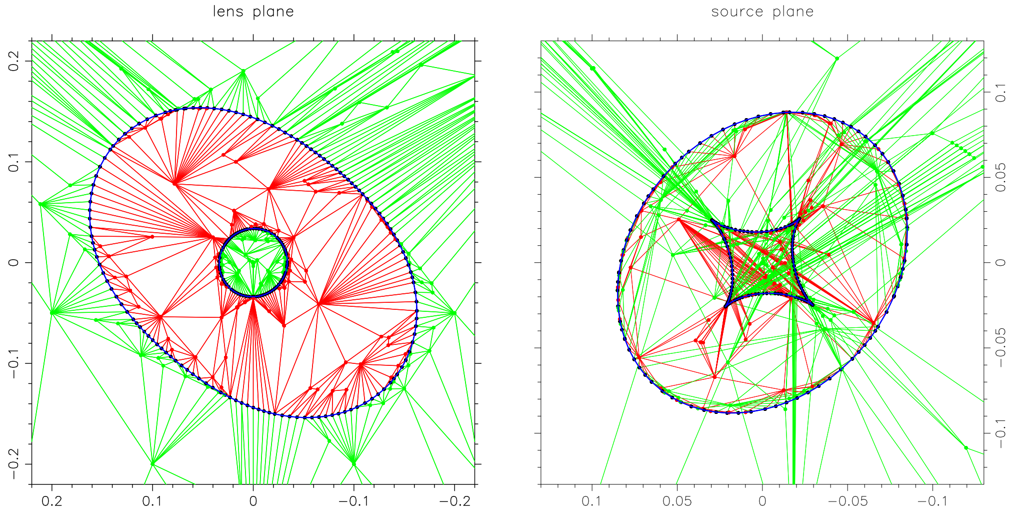


Figure 5.5: Third (last) stage of the initial tiling. Edges connecting two vertices of parity 0 (on the critical curve) are split until a limiting length is reached. The plots also include the critical curves (left) and caustics (right). In the colour version, green marks positive and red marks negative parity. Critical curves are blue.

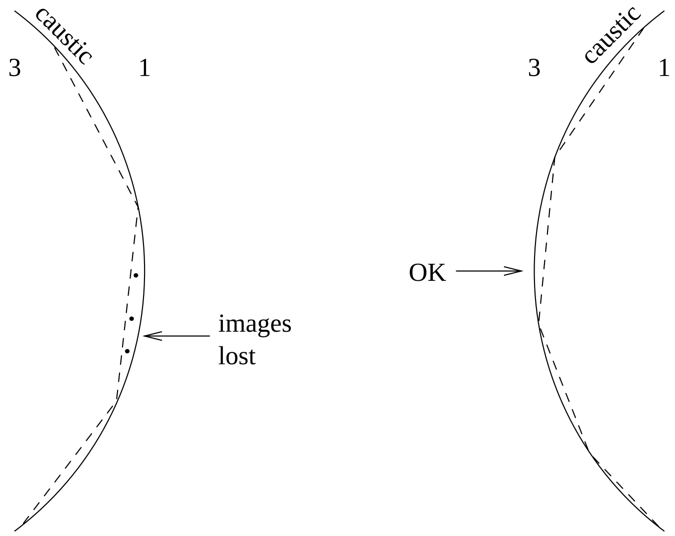


Figure 5.6: If the sampling of the critical curve is too coarse, images can be lost close to the curve. The left figure shows regions of higher multiplicity that are enclosed by the curved caustic (solid) but not by the tiling (dashed). Images will be lost for source components between the tiling's edges and the critical curve. The situation at the right is different. At the first stages, phantom images appear between edges and critical curve. They vanish in the further subdivision.

(very close images are very bright) or that small changes of the model parameters, which would make these images disappear, are below the required accuracy. Typical lengths of 10 mas or less worked successful in LENSCLEANing B0218+357.

5.5 Finding the images

Once the initial tiling is built, it can be used to search all images for a given source position. First, all source plane tiles are tested if they enclose the source position (with correct or incorrect parity). It is essential, how this test is performed to avoid loosing or doubling images close to the edges. The test consists of three tests whether the position z_s is located to the left or right of the straight lines connecting the vertices (the extensions of the tile's edges). If all three tests give the same result, the tile encloses z_s. Whether the position is enclosed clockwise or anti-clockwise usually corresponds to the parity of the mapping. Only sometimes is the handedness of the projected tile not consistent with the mapping parity. We call these cases tiles with incorrect parity.

If a position is located almost exactly on an edge, it makes no difference for which of the neighbouring tiles it is counted, but the test *must* be consistent to avoid errors. The tests are therefore always performed in the same direction, which is defined by the numbering of the vertices. The direction of the edge relative to the tile is taken into account afterwards. With this method, the test 'z_s is left of the line A-B' always leads to a consistent (opposite) result with the test 'z_s is left of the line B-A'. Problems can still occur close to the vertices, where no simple binary decisions can be made. These cases are identified by source positions lying exactly *on* at least two edges of a tile.

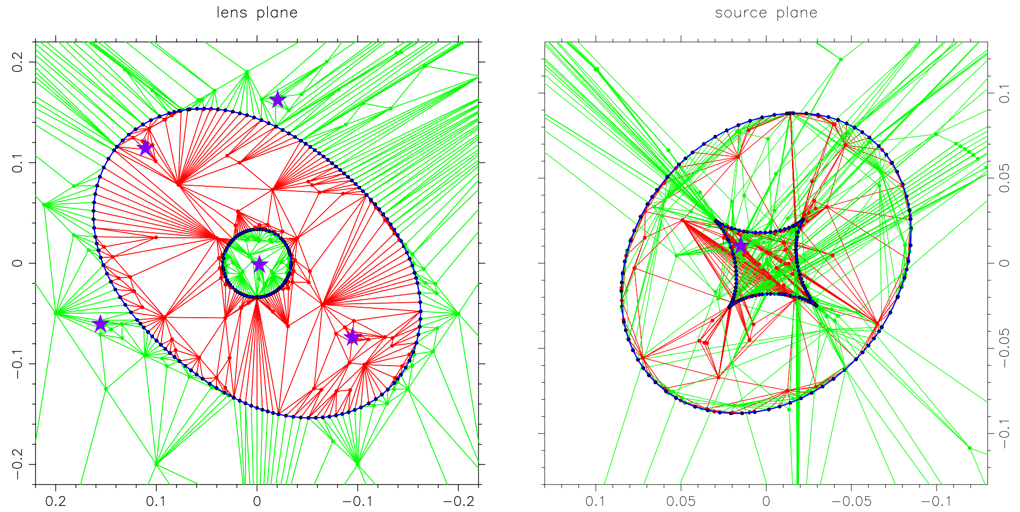


Figure 5.7: Image search with the LenTil algorithm. In addition to the initial tiling, this plot includes the subtilings performed in the search of images for one source position, shown as star symbols. The plots also include the critical curves (left) and caustics (right). In the colour version, green marks positive and red marks negative parity. Critical curves are blue.

The tiles enclosing the source position are selected for the following subdivisions. Some of the selected tiles may still contain phantom images and/or enclose the position with an incorrect parity. These problems vanish in the following course of further subdivisions at the centres of the longest edges of the selected tiles. The resulting image positions are not used as starting values for conventional refinement algorithms, which are not sufficiently reliable, but the subdivision is continued until the required accuracy is reached.

At the end, final conservative tests are performed to assure correct image positions. In the (very rare) case of numerical problems, e.g. incorrect parity of one of the final tiles, the source position is shifted by a very small amount much below the required accuracy and further attempts are made. If secondary images for a known first image are searched for, it is the position of the first image that is shifted, to make the final identification of the primary image from the list of images easier.

Figure 5.7 shows the same initial tiling as before but includes the subdivision for one source position together with the five image positions that were found with this method.

5.6 Technical details

Only the most important details shall be discussed here. There are many other small modifications in the algorithm, which either help in making it more efficient or in avoiding failure under special circumstances. The development will probably continue when the new algorithm has been used in more test cases. Problems are expected for special lens models. With non-exotic models, the method has already survived very strict tests with LENSCLEAN.

5.6.1 Avoiding degeneracies

The successive subdivision can only converge, if the tiles are not degenerated, i.e. they enclose a non-vanishing area. Figure 5.8 shows two kinds of degenerated tilings, that can occur with LENTIL. If the tiles once start to become degenerated, the problem can become even worse in the next subdivision steps, until the tiles become exactly one-dimensional lines. This effect must be avoided, if failure of the algorithm in many cases can not be accepted. Several tests are included in our algorithm for this purpose. If at the next subdivision a *neighbouring* tile would become too degenerate, this neighbouring tile is split at its longest edge, before the tile in question is subdivided itself. This is done recursively: If the neighbour of the neighbour is even more degenerated, it is split first etc. With this approach, true problems with degeneracies occur only very infrequently and are not considered a problem anymore.

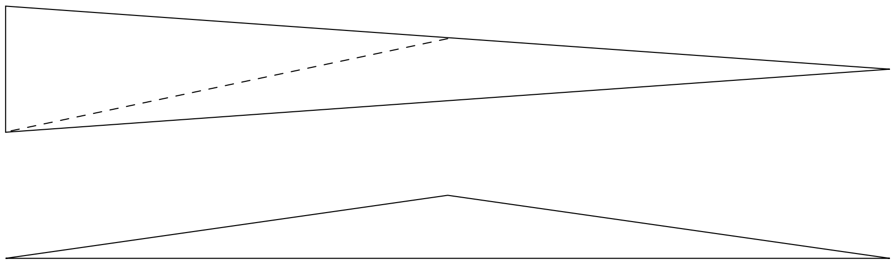


Figure 5.8: Two kinds of almost degenerated tiles. The first kind seems to be less serious at first sight, but it leads to degeneracies of the second kind (bottom) after further subdivision (dashed).

5.6.2 Avoiding parity changes

In the initial tiling, all parity changing edges were split to avoid edges connecting vertices of different parity. During the subdivision, new vertices can again cause such problems. We therefore shift new vertices, which would belong to parity changing edges, onto the critical curve. If this is not possible, because the new tiles would become degenerated, further subdivisions are made until the problem disappears.

5.6.3 Treatment of singularities

Singularities appear regularly in ordinary lens models and have to be treated accordingly in the algorithm. The problem can again be illustrated best in the one-dimensional case (Figure 5.9).

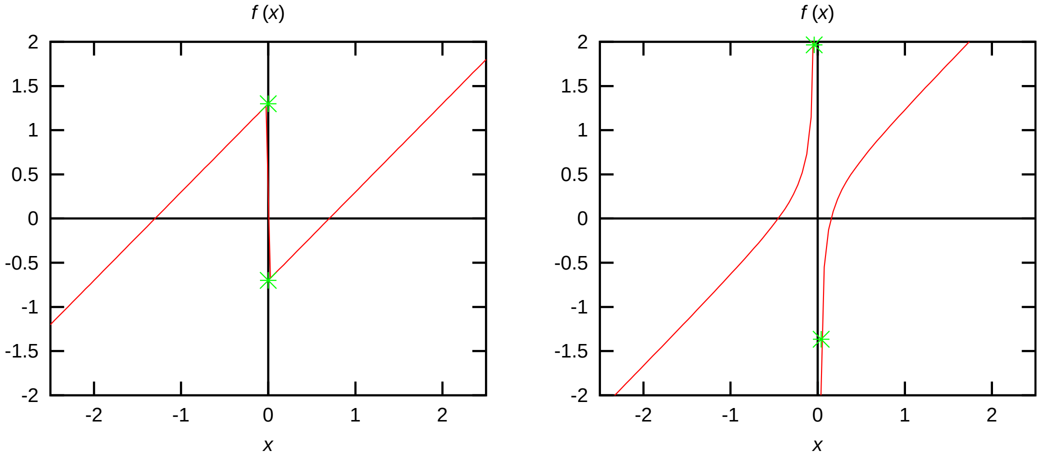


Figure 5.9: One-dimensional root finding with singularities. Shown is a SIS-like (left) and a pointmass-like singularity (right). The marked points have to be included as limits of intervals and the central interval has to be excluded from the root search.

Two kinds of singularities are common in gravitational lenses. The first is the SIS-like singularity, where deflection angles have a constant absolute value at the centre and do not diverge. In this case, a small interval around the singularity in x must be excluded from the root search. The limits, which mark the 'cut' or 'pseudo-caustic' in the source plane[2], are used as limits for the intervals close to the centre. The pointmass-like singularities have diverging deflection angles. The same recipe can be used in this case, but the x_s values of the limits diverge. They have to be chosen large enough to cover the region of interest in the source plane.

In the two-dimensional case, LENTIL cuts small circles around the singularities and marks all vertices inside of these as singular. These circles are treated similar to critical curves. Edges crossing the circles are split at the circles and edges connecting vertices on the circles are subdivided until the sampling is sufficiently fine. In the case of SIS-like singularities, the mapping into the source plane marks the 'cuts'. The fineness limits are here also applied in the source plane to assure a fine sampling of the cut. This is not necessary for pointmass-like singularities, because the corresponding lines in the source plane are far away from the centre and outside the region of interest. The sampling has to be dense enough only to surround the singularity completely. All singular tiles are deleted from the list at the end to save CPU time in the later subdivision stages. Images inside the singular circles would be deamplified so much, that they can be treated as non-existent without loss of accuracy.

---

[2] Caustics are lines of infinite amplification. The number of images changes by two when crossing the caustic. A 'cut' (Kovner, 1987) is a limiting case for SIS-like lens model singularities. When crossing a cut, the number of images changes by one. This image has zero amplification close to the cut.

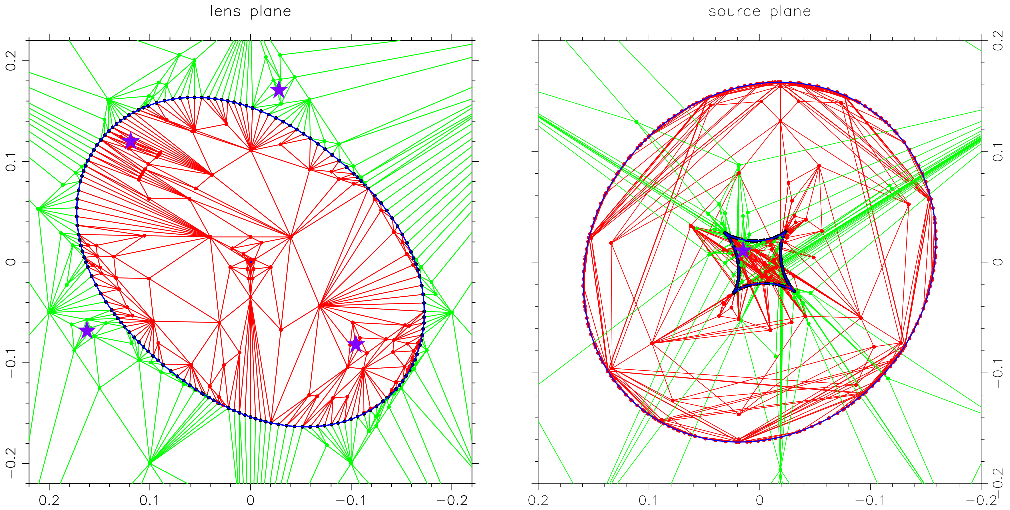


Figure 5.10: Image search with the LenTil algorithm for a SIS model. In addition to the initial tiling, this plot includes the subtilings in the search of images for one source position, shown as star symbols. The plots also include the critical curves (left) and the cut (right). In the colour version, green marks positive and red marks negative parity. The critical curve and cut are blue. The small circle enclosing the singularity can not be seen on this scale. It is shown enlarged in Fig. 5.11.

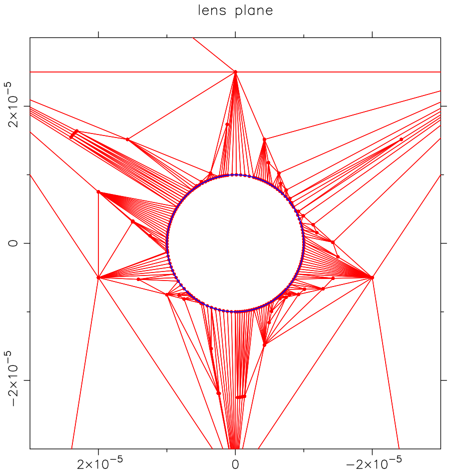


Figure 5.11: Image search with the LenTil algorithm for a SIS model, enlarged version of the region around the singularity. Larger sections of the lens plane as well as the source plane are shown in Fig. 5.10.

Figures 5.10 and 5.11 show the tiling for a model similar to the one used before but with a singular centre. The projected version of the small circle enclosing the singularity marks the cut in the source plane. Images inside of the small circle are lost.

Chapter 6

Degeneracies in general power-law lens models

This chapter is based on a publication about lens model degeneracies (Wucknitz, 2002a). Only half-hearted attempts have been made to change the style from a paper to a thesis' chapter. Some more explicit solutions of the equations are included here, as well as some thoughts about correspondences of elliptical mass distribution and elliptical potentials. These will be needed for the interpretation of lens modelling results.

6.1 Introduction

In this chapter, we consider a family of mass models with separable radial and angular dependence of the potential. External perturbations are included as an external shear. In this way, the different parts can be studied independently. For the radial dependence, we choose a simple power-law and are especially interested in the influence the radial slope has on the results. Several authors (Wambsganß & Paczyński, 1994; Witt, Mao & Schechter, 1995; Wucknitz & Refsdal, 2001) have found approximate scaling relations of $H_0 \propto 2-\beta$ for spherical models with external shear. The same dependence was found before by Chang & Refsdal (1976) for doubles in spherical models without shear (see also Refsdal & Surdej, 1994). The work of Williams & Saha (2000) showed that the radial slope also has important effects in non-parametric models. As a result of the very general nature of these models, no exact scaling law could be found for them.

Modelling constraints provided by observational data are never sufficient to fix the (otherwise free) mass distribution uniquely (Saha & Williams, 1997; Williams & Saha, 2000). Simple parametric models, based on the knowledge about typical galaxies in the local universe or desired mathematical properties, are normally used to overcome this difficulty. This approach, while leading to consistent results, has the disadvantage of hiding the underlying uncertainties and making it difficult to quantify them.

We follow an intermediate approach by using an arbitrary function only for the angular part. The models include elliptical mass distributions and potentials as well as other shapes. It has been observed in several lens systems that the external shear required to fit the data with simple elliptical models is much higher than expected. This might indicate a kind of asymmetry of the galaxy that cannot be accounted for by elliptical models. The general angular part of the potential we use can describe this 'internal shear' more accurately than simple parametric approaches.

We extend and generalize the work of Witt, Mao & Keeton (2000) by including the three independent time-delays in quadruple systems as constraints for the models. We do not use magnification ratios for several reasons. One is the observational problem of reliably measuring correct flux ratios. Fluxes are influenced by microlensing and extinction. These effects can be very strong in the optical and in some cases they are still significant for radio wavelengths. The other reasons are related to our formalism and will be discussed later.

With this approach, we can derive explicit solutions for the Hubble constant as a function of the observables, the power-law exponent β, and the external shear γ. We use the results to find a simple but rigorous scaling law describing the dependence of $H_0\Delta t$ on β in lenses with quadruple images. This scaling law of $H_0\Delta t \propto (2-\beta)/\beta$ is also valid for more special parametric power-law models within the allowed range of β and is therefore inherent to these models and not an artifact of the general angular dependence we discuss.

The qualitative effect of this scaling law is easily understood when comparing it with the $2-\beta$ scaling discussed before and with the mass-sheet degeneracy, which leads to $H_0\Delta t \propto 1-\kappa$. In all cases, a shallower density profile (larger β or κ) leads to smaller values of $H_0\Delta t$. The flat part of the density distribution (κ in the mass-sheet degeneracy) amplifies the deflection angles but leaves the time-delays unaffected. To fit the given geometry, the total deflection angles have to be constant, therefore the time-delays ($H_0\,\Delta t$) decrease.

We also discuss the effect of the external shear on the time-delays and the measured Hubble constant. If the shear is changed, the internal asymmetries of the mass distribution have to be adjusted to compensate for the changes. The total effect of these two contributions will be described by the new concept of a 'critical shear' γ_{c}. The measured Hubble constant is a linear function of the external shear and becomes zero for $\gamma = \gamma_{\mathrm{c}}$. We will present a simple interpretation of the critical shear in terms of the image geometry.

The main goal of this work is to investigate the uncertainties in measurements of H_0 or cosmological parameters from time-delays in gravitational lenses. The results will also help in minimizing possible errors, either by selecting lenses with the least uncertainties or by using constraints of the model parameters which are lensing-independent.

Finally, we apply the formalism to several known lens systems, with and without measured time-delays, to compare our analytical results with those from numerical model fitting using parametric models. For 2237+0305, we compare with own numerical models and find a very good agreement even though the time-delays are not used as constraints in this case.

Our results can be used directly to determine H_0 from time-delays without explicit modelling but should not be used as a substitute for it. They are rather meant to provide an explanation for the degeneracies and scaling relations that have been found for several families of lens models. Nevertheless, we show that the direct application to real systems is possible.

Additionally, we discuss possible Einstein rings in our family of models. Interestingly, infinitely small sources can be mapped as elliptical rings for arbitrary values of the external shear, if the other parameters are chosen in a special way. The ellipticity of these rings is directly determined by the shear.

6.2 The lens model

We use a power-law approach (compare with the spherical power-law models in section 4.1.3) with a general separable angular part for the main lensing galaxy.

$$\psi^{(\mathrm{g})}(\boldsymbol{z}) = r^{\beta}\,F(\theta) \tag{6.1}$$

This family of models includes both elliptical mass distributions and elliptical potentials with arbitrary radial mass index β. In the following, we assume all image positions as known. For the chosen model, this implies also the knowledge of the position of the galaxy centre itself, because coordinates are relative to this centre. Later we will see, however, that some of the equations are translation invariant, leaving the results unchanged when shifting the galaxy. The normalized surface mass density of the model is

$$\begin{aligned} \sigma(\boldsymbol{z}) &= \frac{1}{2}\boldsymbol{\nabla}^2\psi^{(\mathrm{g})} && (6.2)\\ &= \frac{r^{\beta-2}}{2}\Big(\beta^2 F(\theta)+F''(\theta)\Big) \quad . && (6.3)\end{aligned}$$

Here and in the rest of this chapter, primes indicate derivatives with respect to θ. A very simple relation holds for the radial derivative of the potential which we will need in the time-delay equations later:

$$\boldsymbol{z}\cdot\boldsymbol{\nabla}\psi^{(\mathrm{g})} = \beta\,\psi^{(\mathrm{g})} \qquad (6.4)$$

To account for nearby field galaxies or the contribution of a cluster, we include an external shear plus an additional constant mass density or convergence κ into our models. This κ can not be determined from observations as a consequence of the so-called mass-sheet degeneracy, first discussed by Falco, Gorenstein & Shapiro (1985) and Gorenstein, Shapiro & Falco (1988). In the following, we therefore always use a fixed κ. We parametrize the shear as in section 4.1.2. Note that this part of the potential is a special case of the power-law with $\beta = 2$. We denote the shear contribution to the potential by $\psi^{(\gamma)}$.

6.3 Time-delays

The light travel time t for a certain image at $\boldsymbol{z}$ can be calculated for arbitrary lens models by using equation (3.43). To eliminate d_{eff} from the equation, we now use a scaled version of the Hubble constant,

$$h = \frac{H_0}{d_{\mathrm{eff}}} \quad , \qquad (6.5)$$

which includes the cosmological factors and will later be determined by the lensing effect. This h is directly proportional to the Hubble constant itself when the cosmology is fixed. For two images i and j, we get a time-delay of Δt_{ij} with

$$\begin{aligned} h\,\Delta t_{ij} &= h\,(t_i - t_j) && (6.6)\\ &= \frac{1}{2}\Big(|\boldsymbol{\alpha}_i|^2 - |\boldsymbol{\alpha}_j|^2\Big) - (\psi_i - \psi_j) \quad . && (6.7)\end{aligned}$$

By using the lens equation (3.28), we can transform this expression into a linear functional of the deflecting potential and its derivatives. Mixed terms like $\boldsymbol{z}_i\cdot\boldsymbol{\alpha}_j$ for $i \neq j$ can be eliminated, so that the resulting time-delay can again be written as the difference of the light travel times t_i themselves.

$$h\,t_i = -\frac{1}{2}r_i^2 + \boldsymbol{z}_i\cdot\boldsymbol{\nabla}\psi_i - \psi_i - C \qquad (6.8)$$

Although it is not immediately noticeable, this equation still is invariant under translations of coordinates. To prove this, one has to apply the lens equation.

For the special potential we discuss here, the relation (6.4), which also holds for $\psi^{(\gamma)}$ with $\beta = 2$, can be used to eliminate the derivatives of the potential and obtain a simple expression only depending on both parts of the potential at the image position:

$$h t_i = -\frac{1}{2} r_i^2 - (1-\beta)\psi_i^{(\mathrm{g})} + \psi_i^{(\gamma)} - C \tag{6.9}$$

We retain using the notion of light travel times for each image instead of time-delays *between* the images to keep the equations simple. The t_i are defined except for an overall additive constant that can be absorbed into the constant C.

The last equation was already presented in Witt et al. (2000) for the special cases of $\beta = 1$ including external shear and for the general shearless power-law model.

6.4 Counting constraints and parameters

Before solving the equations, we want to discuss how many parameters of the model can at most be determined from observations of lens systems with n images, especially $n = 4$. See Table 6.1 and 6.2 for a list of parameters and constraining equations. As the time-delays do not change when adding the same constant to all t_i, we have to include the constant C as parameter and thus have n constraints with one parameter for this. We might as well have used only the (uniquely defined) $n-1$ time-delays without C. Both possibilities result in $n-1$ more constraints than parameters just for the time-delays.

parameters		number
h	scaled Hubble constant	1
γ	external shear	2
β	power-law exponent	1
F_i	angular part $F(\theta_i)$	n
F_i'	$\mathrm{d}F/\mathrm{d}\theta$ at θ_i	n
z_s	source position	2
C	constant in light travel times	1
total	without fluxes	$2n+7$

Table 6.1: Parameters for the chosen family of models for a system with n images. The F_i' are needed in the lens equations.

constraints		number
z_i	image positions	$2n$
t_i	light travel times	n
total	without fluxes	$3n$

Table 6.2: Constraints from observations of image positions and time-delays for a system with n images.

Even for systems with 4 images of one source, it will be impossible to determine all parameters. We therefore decide to fix β in the following calculations so that the results can be used to study the dependence of h on the radial mass slope. We will see that, with fixed β, all equations stay linear.

Another critical parameter is the external shear, which seems to be higher than expected in many detailed lens models. Fixing γ for the moment, we will be able to investigate the influence of any uncertainties in the external shear. In the special case of $\beta = 1$, it will even be possible to determine $\boldsymbol{\gamma}$ from the constraints, because a number of other parameters do not contribute then.

We do not include fluxes or magnifications, because they would provide only $n-1$ constraints (the independent flux ratios) and at the same time add n more parameters (the second derivatives of the angular part of the potential at the image positions F_i''). The effect that models get less constrained, when more observations are included in the analysis, is of course unknown in normal parametric models, where the number of parameters is fixed. Our models have an infinite number of parameters, of which only a finite subset is needed to compare with observations. The number of relevant parameters changes when we include more constraints.

Another reason (besides the difficulties in determining fluxes already discussed) is that, in contrast to deflection angles and time-delays, magnifications and magnification ratios are non-linear functionals of the lensing potential, complicating the equations considerably.

6.5 Lens equations

To utilize the information given by the image positions, we have to insert the power-law potential with shear into the lens equation (3.28). The derivative of the galaxy part of the potential can be obtained most simply by rotating its polar form to Cartesian coordinates using the transformation matrix

$$\begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta/r \\ \sin\theta & \cos\theta/r \end{pmatrix} \begin{pmatrix} \partial_r \\ \partial_\theta \end{pmatrix} \quad . \tag{6.10}$$

Written in a form to take into account the role of $\boldsymbol{\gamma}$, F_i and F_i' as unknowns of the equations, this leads to the following equation.

$$\boldsymbol{z}_\mathrm{s} = (1-\kappa)\,\boldsymbol{z}_i - r_i^{\beta-2} \begin{pmatrix} x_i & -y_i \\ y_i & x_i \end{pmatrix} \begin{pmatrix} \beta F_i \\ F_i' \end{pmatrix} + \begin{pmatrix} x_i & y_i \\ -y_i & x_i \end{pmatrix} \begin{pmatrix} \gamma_x \\ \gamma_y \end{pmatrix} \tag{6.11}$$

It is easily seen that this set of $2n$ equations can be used to determine F_i and F_i', assuming $\boldsymbol{\gamma}$ as known:

$$\beta F_i = r_i^{-\beta}\Big((1-\kappa)\,r_i^2 - x_i x_\mathrm{s} - y_i y_\mathrm{s} + \gamma_x\,(x_i^2 - y_i^2) + 2\,\gamma_y\,x_i y_i\Big) \tag{6.12}$$

$$F_i' = r_i^{-\beta}\Big(y_i x_\mathrm{s} - x_i y_\mathrm{s} - 2\,\gamma_x\,x_i y_i + \gamma_y\,(x_i^2 - y_i^2)\Big) \tag{6.13}$$

6.6 The general set of linear equations

We now use (6.12) to express the galaxy potential in the time-delay equations (6.9). We decide to use the light travel times themselves rather than the H_0-independent time-delay ratios as constraints. In this way we can keep the equations linear and the analysis much simpler. As a result, the following set of equations is obtained for $i = 1\ldots n$:

$$(1-\kappa)\,r_i^2 \;+\; (x_i^2 - y_i^2)\,\gamma_x \;+\; 2x_i y_i\,\gamma_y \;+\; \frac{2\beta}{2-\beta}\,(t_i h + C) \;=\; 2\frac{1-\beta}{2-\beta}\,\boldsymbol{z}_i \cdot \boldsymbol{z}_\mathrm{s} \tag{6.14}$$

The most interesting fact, besides the linearity, is the simple way the mass index β appears in the equations. By combining information from the time-delay and lens equations the way we did, it was possible to remove the terms with β-dependent exponents. Now β only contributes in the scaling factors of the unknown parameters. Having solved the system for one value of β, we immediately find the general solution by scaling h, z_s and C with the appropriate factors.

6.7 The isothermal model

In the case $\beta = 1$, the equations (6.14) obviously degenerate with respect to the source position z_s. It is then impossible to constrain the latter, but the same information can now be used to determine the external shear. In this case, the inclusion of the lens equations to determine the parameters F_i of the galaxy potential was not really necessary, because $\psi^{(\mathrm{g})}$ does not contribute to the time-delay equations (6.9). We can now directly invert the latter or (6.14) to obtain solutions for h_iso and γ_iso. Using Cramer's rule to solve (6.14) for $\beta = 1$, the result reads like follows:

$$A := \frac{1}{2}(1-\kappa)\begin{vmatrix} t_1 & x_1^2 - y_1^2 & x_1 y_1 & 1 \\ t_2 & x_2^2 - y_2^2 & x_2 y_2 & 1 \\ t_3 & x_3^2 - y_3^2 & x_3 y_3 & 1 \\ t_4 & x_4^2 - y_4^2 & x_4 y_4 & 1 \end{vmatrix}^{-1} \tag{6.15}$$

$$h_\mathrm{iso} = A \begin{vmatrix} x_1^2 & y_1^2 & x_1 y_1 & 1 \\ x_2^2 & y_2^2 & x_2 y_2 & 1 \\ x_3^2 & y_3^2 & x_3 y_3 & 1 \\ x_4^2 & y_4^2 & x_4 y_4 & 1 \end{vmatrix} \tag{6.16}$$

$$\gamma_\mathrm{iso} = A \left(-\begin{vmatrix} t_1 & r_1^2 & x_1 y_1 & 1 \\ t_2 & r_2^2 & x_2 y_2 & 1 \\ t_3 & r_3^2 & x_3 y_3 & 1 \\ t_4 & r_4^2 & x_4 y_4 & 1 \end{vmatrix}, +\begin{vmatrix} t_1 & r_1^2 & (x_1^2 - y_1^2)/2 & 1 \\ t_2 & r_2^2 & (x_2^2 - y_2^2)/2 & 1 \\ t_3 & r_3^2 & (x_3^2 - y_3^2)/2 & 1 \\ t_4 & r_4^2 & (x_4^2 - y_4^2)/2 & 1 \end{vmatrix} \right) \tag{6.17}$$

The transition from exactly isothermal to nearly isothermal models deserves some attention. Even for models with β differing only slightly from unity, γ is a free parameter, while it is fixed by the observational data for $\beta = 1$. An incorrect external shear for almost isothermal models is usually compensated for by diverging source positions and F_i and F_i', leading to unrealistic models. It therefore seems appropriate to use the correct isothermal shear even for models that do not exactly obey $\beta = 1$. However, one has to take into account possible measurement uncertainties that introduce errors into γ_iso. Especially the time-delays, which all contribute to the solution, can introduce significant uncertainties.

6.8 Solution for the general model

In the general case with $\beta \neq 1$, (6.14) can be solved directly to determine h for a given shear γ. Even without writing the explicit solution, we see that

$$h \propto \frac{2-\beta}{\beta}(1 + g_x \gamma_x + g_y \gamma_y) \tag{6.18}$$

for some constants g_x and g_y. The Hubble constant scales with $(2-\beta)/\beta$ and is a linear (but not proportional) function of the shear. Notably, the isothermal model plays no special role in this equation, despite the fact that, strictly speaking, $\boldsymbol{\gamma}$ can not be chosen freely in this case. Since the isothermal shear is usually only weakly constrained due to the limited measurement accuracy, we may, however, use different values of $\boldsymbol{\gamma}$ even for $\beta = 1$. This is always true when the constraints we used here are not all available. Considering this, one might well apply equation (6.18) regardless of the value of β.

To obtain explicit expressions for the coefficients $\mathbf{g} = (g_x, g_y)$, we can again use Cramer's rule:

$$h = B\,\frac{2-\beta}{\beta}\,(1 + g_x\gamma_x + g_y\gamma_y) \tag{6.19}$$

$$B = -\frac{1-\kappa}{2}\begin{vmatrix} t_1 & x_1 & y_1 & 1 \\ t_2 & x_2 & y_2 & 1 \\ t_3 & x_3 & y_3 & 1 \\ t_4 & x_4 & y_4 & 1 \end{vmatrix}^{-1} g_0 \tag{6.20}$$

$$g_0 = \begin{vmatrix} r_1^2 & x_1 & y_1 & 1 \\ r_2^2 & x_2 & y_2 & 1 \\ r_3^2 & x_3 & y_3 & 1 \\ r_4^2 & x_4 & y_4 & 1 \end{vmatrix} \tag{6.21}$$

$$\mathbf{g} = \frac{1}{g_0}\left(\begin{vmatrix} x_1^2-y_1^2 & x_1 & y_1 & 1 \\ x_2^2-y_2^2 & x_2 & y_2 & 1 \\ x_3^2-y_3^2 & x_3 & y_3 & 1 \\ x_4^2-y_4^2 & x_4 & y_4 & 1 \end{vmatrix}, \begin{vmatrix} 2x_1y_1 & x_1 & y_1 & 1 \\ 2x_2y_2 & x_2 & y_2 & 1 \\ 2x_3y_3 & x_3 & y_3 & 1 \\ 2x_4y_4 & x_4 & y_4 & 1 \end{vmatrix}\right) \tag{6.22}$$

6.9 The ‘Critical shear’

Since the Hubble constant is linear in $\boldsymbol{\gamma}$, there has to be a one-dimensional family of values of the shear with vanishing h. The shear with the smallest absolute value from this family will now be called the ‘critical shear’ $\boldsymbol{\gamma}_{\mathrm{c}}$ (see Figure 6.1). If we denote the shearless value with h_0, we can write the effect of external shear as

$$\frac{h}{h_0} = 1 - \frac{\boldsymbol{\gamma}\cdot\boldsymbol{\gamma}_{\mathrm{c}}}{\gamma_{\mathrm{c}}^2} \quad . \tag{6.23}$$

The critical shear can be calculated using equations (6.19) to (6.22).

$$\boldsymbol{\gamma}_{\mathrm{c}} = -(1-\kappa)\frac{\mathbf{g}}{g_x^2 + g_y^2} \tag{6.24}$$

If $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_{\mathrm{c}}$ point in the same direction, this scaling factor can be written as $1-\gamma/\gamma_{\mathrm{c}}$ which is analogous to the scaling of $1-\sigma$ for an additional mass sheet Σ with critical density Σ_{c} and $\sigma = \Sigma/\Sigma_{\mathrm{c}}$. The critical shear does not dependent on β or the time-delays t_i and can be calculated from the image positions alone. If an upper bound of the external shear called $\gamma_{\max}$ can be assumed, this translates to a range of

$$h_{\binom{\max}{\min}} = \left(1 \pm \frac{\gamma_{\max}}{\gamma_{\mathrm{c}}}\right) h_0 \quad . \tag{6.25}$$

The critical shear is thus a measure for the amount of external shear that can change h significantly. The larger it is, the smaller the influence of uncertainties in the shear on the determined Hubble constant. Shear in the direction of $\boldsymbol{\gamma}_{\mathrm{c}}$ contributes maximally, in a perpendicular direction the influence vanishes. With the 'direction of shear', we mean in this context the orientation $2\theta_\gamma$ of the vector $\boldsymbol{\gamma}$. This is *not* the direction of θ_γ towards the perturbing mass, cf. eq. (4.6). External field galaxies located perpendicular to the direction of maximal effect change h by the same amount but with opposite sign. There are, however, four directions where external masses do not contribute. See section 6.13 for actual numbers of the critical shear in real observed lens systems.

Geometrical interpretation

For constant light travel times or $h = 0$, equation (6.14) describes an ellipse (compare appendix A.1) whose axes a and b are related to the external shear by

$$\frac{\gamma}{1-\kappa} = \frac{a^2-b^2}{a^2+b^2} \quad . \tag{6.26}$$

For $\gamma > 1-\kappa$, this becomes a hyperbola. The position angle of the *minor* axis is the same as that of the perturbing mass responsible for the shear (θ_γ). This means that each ellipse/hyperbola passing through all images corresponds to a value of $\boldsymbol{\gamma}$ with $h/h_0 = 0$. According to equation (6.23), for these values

$$\boldsymbol{\gamma}\cdot\boldsymbol{\gamma}_{\mathrm{c}} = \gamma_{\mathrm{c}}^2 \tag{6.27}$$

holds. The values of $\boldsymbol{\gamma}$ for all these conic sections span the complete subspace for which $h/h_0 = 0$ (see Fig. 6.1). The critical shear is the smallest of these values and can therefore be calculated with equation (6.26) for the 'roundest' ellipse passing through all images. In Figure 6.2, we show several such ellipses for the lens systems that are discussed in sections 6.13.2 to 6.13.5.

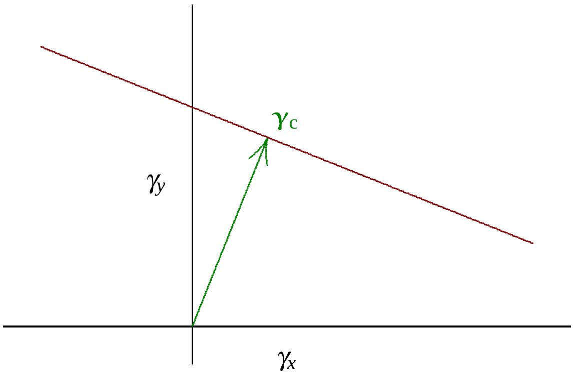


Figure 6.1: Example critical shear $\boldsymbol{\gamma}_{\mathrm{c}}$ (arrow) and values of $\boldsymbol{\gamma}$ for which $h/h_0 = 0$ (straight line). Note that the direction of $\boldsymbol{\gamma}$ is $2\theta_\gamma$ and *not* the direction to the perturbing mass θ_γ.

We conclude that for certain values of $\boldsymbol{\gamma}$ (see Figure 6.1), e.g. for $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{\mathrm{c}}$, all observed time-delays (or alternatively H_0) vanish. Each of the fitting ellipses can act as an Einstein ring for the correct value of $\boldsymbol{\gamma}$ if the function F is chosen accordingly. The light travel time is the same for all parts of such rings as Fermat's theorem requires. In our consideration, the potential is

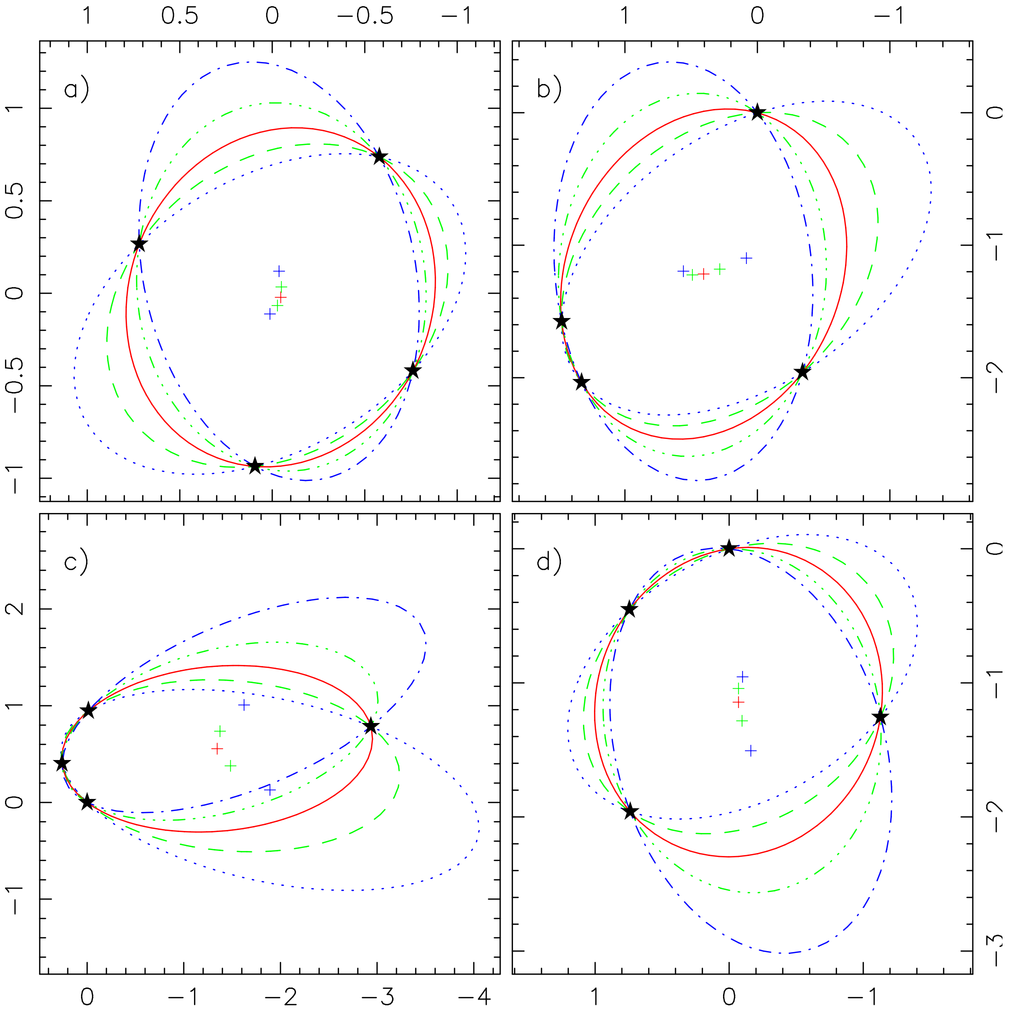


Figure 6.2: Ellipses passing through all images (stars) of four lens systems. The ‘roundest’ one, which corresponds to the critical shear, is shown as solid line. The centres are shown as crosses. For each ellipse, a shear for which $h/h_0 = 0$ can be calculated by using equation (6.26). (a) Q 2237+0305, (b) PG 1115+080, (c) RX J0911.4+0551, (d) B1608+656

only constrained at the image positions and arbitrary in other directions. The Einstein ring will therefore usually break up and form a number of discrete images with still vanishing time-delays (see section 6.15).

6.10 Shifting the lensing galaxy

Surprisingly, most of the quantities we determined do not depend on the position of the lens centre. A shift of the centre is equivalent to adding a constant displacement to the vectors z_i and z_s. If we look at the general set of equations (6.14), we see that such a shift adds terms linear in z_i and a constant to the equations. The constant term can be absorbed in C and (for $\beta \neq 1$) the linear terms in z_s. As C and z_s are of no interest, the equations do not change when applying this shift. The same result could be derived from the explicit solutions and using the general properties and rules for determinants.

This result means that h and the critical shear are translation invariant and can be determined even if the lens position is not known. This is only true for the family of models we analyse here. Simple parametric models usually only fit the data for a specific position of the lens centre. Because of the degeneracy for $\beta = 1$, the results for isothermal models in section 6.7 are *not* translation invariant. This is not in contradiction with the general behaviour, because the external shear is not a free parameter for isothermal models.

6.11 Spherical models for nearly Einstein ring systems

For spherical models, the equations are overdetermined. There are nevertheless systems, which can be fit accurately with this kind of model. Without showing all the calculations here, we only want to present the result

$$\frac{h_{\mathrm{sph}}}{h_0} = \frac{r_0^{2-\beta}}{g_0} \begin{vmatrix} r_1^\beta & x_1 & y_1 & 1 \\ r_2^\beta & x_2 & y_2 & 1 \\ r_3^\beta & x_3 & y_3 & 1 \\ r_4^\beta & x_4 & y_4 & 1 \end{vmatrix} \quad . \tag{6.28}$$

It is easy to show that h_{sph}/h_0 becomes zero for $\beta \to 0$ to cancel the vanishing denominator in (6.18) and ensure a finite h_{sph}. It can also be shown that $\boldsymbol{\gamma}$ is parallel to $\boldsymbol{\gamma}_{\mathrm{c}}$ for arbitrary β. Taken together, this means that for point mass models, we always have $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{\mathrm{c}}$. The geometrical interpretation can be used to determine the direction of external shear for spherical models without calculations. It is given by the orientation of the minor axis of the roundest ellipse passing through all images. For the systems discussed below, even the absolute value of γ can be determined from this ellipse for point mass lenses.

For systems where the n images are all located very close to the Einstein ring at r_0, we can recover another scaling relation. In this case, the power-law can be interpreted as a local approximation to *any* radial mass profile, like softened power-law spheres or other models. We assume, that one fitting reference model is known. It is then possible to find a family of other models which are also consistent with the observations.

This was done numerically by Wambsganß & Paczyński (1994) and led to a scaling of $h_{\mathrm{sph}} \propto 2-\beta$. Wucknitz & Refsdal (2001) presented a simple interpretation of this fact in terms of the well known mass-sheet degeneracy (Gorenstein et al., 1988). If we multiply the lens equation $\boldsymbol{z}_{\mathrm{s}} = \boldsymbol{z} - \boldsymbol{\alpha}$ with $1-\kappa$, we get another lens equation with the source position and lensing potential (or deflection angle) scaled with the same factor, plus an additional constant convergence κ. This means that lens models given by the deflection angles $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}} = (1-\kappa)\,\boldsymbol{\alpha} + \kappa\boldsymbol{z}$ are equivalent, but source positions, time-delays, etc. are scaled by $1-\kappa$ in the latter model.

As reference model, we choose an exponent $\beta_0 = 1$. This model is now transformed as just described and then approximated locally near r_0 by a modified power law with exponent β. Figure 6.3 illustrates this idea. For the new exponent, we find

$$2-\beta = 1-\kappa \quad . \tag{6.29}$$

As time-delays, source position and external shear all scale with $1-\kappa$, we find the general scaling laws

$$\boldsymbol{z}_{\mathrm{s}} \propto 2-\beta, \quad h_{\mathrm{sph}} \propto 2-\beta, \quad \boldsymbol{\gamma} \propto 2-\beta \tag{6.30}$$

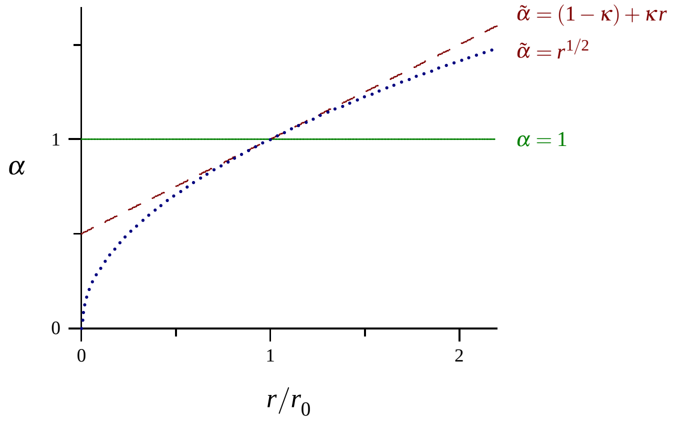


Figure 6.3: The deflection angles for the three equivalent lens models. The isothermal ($\beta = 1$) reference model is shown solid, the transformed model dashed and the final equivalent power-law model dotted.

for an arbitrary mass index β. The time-delay ratios do not change.

With $\gamma = \gamma_c$ for point mass systems, the shear for arbitrary β becomes $\gamma = \gamma_c\,(2-\beta)/2$, leading to a ratio of the Hubble constants for the spherical and shearless case of

$$\frac{h_{\mathrm{sph}}}{h_0} = \frac{\beta}{2} \quad . \tag{6.31}$$

The two models differ significantly for a realistic range of β. Systems with images very close to an Einstein ring should therefore not be used to determine H_0 because of the high uncertainties introduced by the external shear.

6.12 Influence of the radial mass index β

One of the interesting properties of equation (6.18) is the very simple dependence of the results on β. The determined value of the Hubble constant h simply scales with the factor $(2-\beta)/\beta$. The most alarming fact is that this factor does not depend on the geometry of the lens, the time-delay ratios or the amount of external shear. When using the models described here to determine H_0 from lens time-delays, the error due to the assumption of an incorrect β will be *exactly the same* for all lens systems as long as the real β is more or less equal for all lensing galaxies.

Witt et al. (2000) numerically found a scaling of $(2-\beta)/\beta$ in the case of a power-law model without external shear for orthogonal image pairs ($\theta_i - \theta_j \approx 90°$), while these computations lead to $2-\beta$ for opposed images. In contrast to their work, we have used all time-delays as constraints so that they cannot scale differently. The common scale factor of all time-delays shows as a scaling of h in our calculations.

The reader might feel as uneasy about the seemingly diverging time-delays or h in the limit $\beta \to 0$ as the author did. This limit is equivalent to point mass models and one should not observe diverging time-delays for this kind of lens. The point causing trouble here is the fixed external shear in our considerations. To fit the data with a point mass model, the shear has to be equal to the critical shear γ_c. Taking this into account, the result gets multiplied by a vanishing $h/h_0 \propto \beta$ which cancels the $1/\beta$ factor. If we now change the shear by a small amount, the potential and thus h will immediately diverge in the limit of small β. This has no direct physical implications, because the mass models will become extremely unphysical to compensate for the shear effect. Small relative differences in the F_i, which will be introduced by an incorrect shear, will lead to enormous asymmetries in the mass distribution. This is related to the fact that realistic compact mass distributions can provide only almost spherically symmetric potentials. Any multipole moments would radially decrease more rapidly than the monopole term and must be very strong to have any effect. We will discuss this in more detail for a special case in section 6.13.1.

For spherical systems with images near the Einstein ring we confirmed the approximate scaling of $h_{\mathrm{sph}} \propto 2-\beta$. This seems to be incompatible with the general scaling law of $h_0 \propto (2-\beta)/\beta$ at first sight. With the factor of $h_{\mathrm{sph}}/h_0 = \beta/2$ from equation (6.31), by which the shear changes the result in spherical models, these two results are, however, in perfect agreement.

6.13 Application to special cases

To illustrate the results, we have presented here, and to test their relevance for real lenses, we want to apply the formalism to systems with a special Einstein cross like symmetry and to some of the known real systems that are either useful to actually determine H_0 (1115+080, 0911+0551, 1608+656) or are interesting because they are very well studied systems like 2237+0305. The calculations will show that both scaling relations, $h_{\mathrm{sph}} \propto 2-\beta$ and $h_0 \propto (2-\beta)/\beta$, are relevant for the determination of H_0 from time-delays. The detailed numerical models for 2237+0305 will furthermore show that the scaling also applies if the time-delays themselves are not used to constrain the models. All time-delays scale almost exactly as predicted by our analytical work. We will also see that parametric models may fit only for a limited range of β. The scaling relations are then valid only within this range.

Besides the effects of the exponent β, the possible strong effects of any external shear will also be confirmed by the numerical models.

6.13.1 Symmetric Einstein cross like systems

A rather special example of systems shall be discussed explicitly in this section. We consider a lens with vanishing time-delays of opposing image $\Delta t_{12} = \Delta t_{34} = 0$ and the following image positions:

$$x_{1,2} = \pm(1-\varepsilon)\,r_0 \qquad y_{1,2} = 0 \tag{6.32}$$

$$x_{3,4} = 0 \qquad y_{3,4} = \pm(1+\varepsilon)\,r_0 \tag{6.33}$$

From these data, we immediately conclude $z_{\mathrm{s}} = 0$. The y-component of the shear does not contribute at all and cannot be constrained. The equation determining the Hubble constant now reads

$$h\,\Delta t_{13} = (1-\kappa)\,r_0^2\,\frac{2-\beta}{\beta}\,2\,\varepsilon\left(1-\frac{\gamma_x}{\gamma_{\mathrm{c}}}\right) \quad , \tag{6.34}$$

with the critical shear

$$\gamma_{\mathrm{c}} = (1-\kappa)\,\frac{2\,\varepsilon}{1+\varepsilon^2} \quad . \tag{6.35}$$

The special symmetry makes it possible to choose any value of external shear even for isothermal models. Furthermore, it is possible to exactly reproduce the data with spherically symmetric models plus external shear. In this case, the shear is uniquely defined:

$$\gamma = (1-\kappa)\,\frac{(1+\varepsilon)^{2-\beta}-(1-\varepsilon)^{2-\beta}}{(1+\varepsilon)^{2-\beta}+(1-\varepsilon)^{2-\beta}} \tag{6.36}$$

$$= (1-\kappa)\,(2-\beta)\,\varepsilon + \mathcal{O}(\varepsilon^2) \tag{6.37}$$

The time-delay equation now becomes somewhat more complicated than in the non-spherical case:

$$h_{\mathrm{sph}}\,\Delta t_{13} = (1-\kappa)\,r_0^2\,\frac{2-\beta}{\beta}\,\frac{(1+\varepsilon)^{\beta}-(1-\varepsilon)^{\beta}}{(1+\varepsilon)^{\beta-2}+(1-\varepsilon)^{\beta-2}} \tag{6.38}$$

$$= (1-\kappa)\,r_0^2\,(2-\beta)\,\varepsilon + \mathcal{O}(\varepsilon^2) \quad . \tag{6.39}$$

Comparing this with (6.34), we recover the factor of $\beta/2$ between the spherical and shearless case. Both models fit the data exactly and, in this special case, the models are even compatible with highly popular elliptical mass distributions. That means that without any independent information about the external shear (or equivalently the ellipticity of the galaxy itself), we have a factor of two uncertainty even when only considering these two simplest models for $\beta = 1$. The real situation may be even much worse, when we consider models with internal *and* external shear. In this case, any small unknown contribution of external shear of the order of γ_c (which for very symmetric systems becomes arbitrarily small) will change the result significantly.

Witt et al. (1995) discussed exactly the same type of systems with spherical models plus shear. With γ fixed, they also derived a scaling law of $(2-\beta)/\beta$ (see their equation 8[1]). When the shear is constrained by the lens equations, the scaling changes to the $2-\beta$ form.

We finally want to discuss the consequences of diverging time-delays in the fixed shear case for $\beta \to 0$ due to equations (6.18) and (6.34). For simplicity, we assume $\gamma = 0$, $\kappa = 0$ and $\varepsilon \ll 1$, but the argument is generally also true for other values. We write the potential as a multipole expansion[2]

$$\beta\,F(\theta) = r_0^{2-\beta}\left(1-\sum_{k=1}^{\infty} a_k \cos 2k\theta\right) \quad . \tag{6.40}$$

To be compatible with equation (6.12), the coefficients have to meet the condition

$$\sum_{k=1}^{\infty} a_k = (2-\beta)\,\varepsilon \quad . \tag{6.41}$$

We notice that densities (6.3) can become negative near the axes. To minimize the angular density contrast, we have to keep only the monopole and quadrupole terms and set all higher coefficients to 0. The potential is then equivalent to a density of

$$\sigma = \frac{\beta}{2}\left(\frac{r}{r_0}\right)^{\beta-2}\left(1+\varepsilon\,\frac{(2-\beta)^2\,(2+\beta)}{\beta^2}\cos 2\theta\right) \quad , \tag{6.42}$$

which is everywhere positive only for sufficiently high values of β. When using realistic mass models, we can therefore expect a lower bound for β to achieve acceptable fits. This applies not

---

[1] The exponent of the first term in equation (8) in Witt et al. (1995) is incorrect; it should be the same as that in the second term. (S. Mao, private communication)

[2] We use an expansion for the principal axes of the lens system. The $\sin 2k\theta$ terms might be included as well but they would not change the density on these axes, which is what we are interested in.

only to this special symmetric lens system but is true in general. Numerical models presented in the next section will confirm this result for $\gamma = 0$ (see Figure 6.5).

Another aspect of the correspondence of potential and density will be discussed in section 6.14.

6.13.2 The Einstein cross Q 2237+0305

This lens (Figure 6.4) is not usually taken into consideration when thinking about determination of H_0, because the time-delays are expected to be very small and can therefore not be determined easily. Here we show that even if all three time-delays were known exactly, constraints for the Hubble constant would still be very weak.

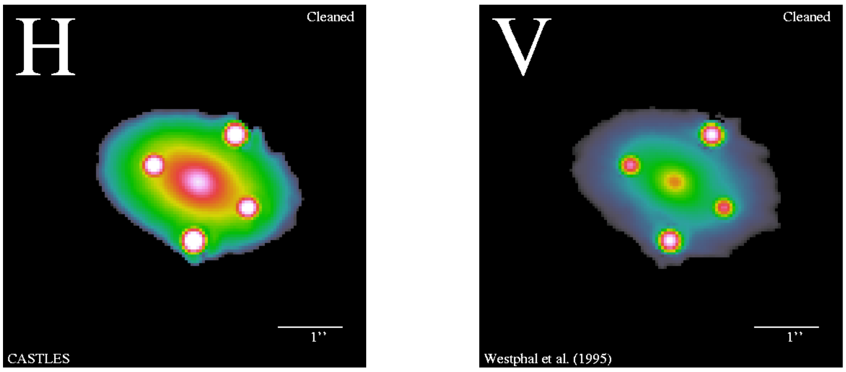


Figure 6.4: IR and optical HST images of Q 2237+0305 (Kochanek et al., 2002)

Table 6.3 shows the distance parameters calculated with the redshifts from Huchra et al. (1985). The small lens redshift leads to a very small $d_{\rm eff}$ and small time-delays. The high symmetry of this system adds to this expectation.

world model	$d_{\rm d}$	$d_{\rm s}$	$d_{\rm ds}$	$d_{\rm eff}$	$d_{\rm ds}/d_{\rm s}$
flat Λ	0.0376	0.4077	0.3933	340	0.9645
flat Λ clumpy	0.0376	0.4729	0.4525	343	0.9568
no Λ	0.0371	0.3688	0.3504	340	0.9501
EdS	0.0368	0.2901	0.2759	338	0.9510
Quint 1	0.0373	0.3763	0.3619	339	0.9617
Quint 2	0.0371	0.3358	0.3215	338	0.9574
Quint 3	0.0368	0.2901	0.2759	338	0.9510

Table 6.3: Distance parameters for Q 2237+0305 calculated with $z_{\rm d} = 0.0394$ and $z_{\rm s} = 1.695$. See sections 3.2 and 3.6 for explanations. Units of $d_{\rm eff}$ are $\rm km\,s^{-1}\,Mpc^{-1}\,days\,arcsec^{-2}$.

The degeneracy caused by the unknown mass index β was already discussed by Wambsganß & Paczyński (1994) for spherical models plus external shear. The authors found the $2-\beta$ scaling by using numerical models. We now want to investigate how strong the assumption of a spherical main galaxy influences these results.

No time-delays are available for 2237+0305 and they may never be determined. We can nevertheless calculate the critical shear defined before and compare it with the typical value one gets for spherical models. Positions including error bars used for this were taken from Crane et al. (1991) to make results comparable with Wambsganß & Paczyński (1994).

The critical shear as calculated from these positions[3] is $\boldsymbol{\gamma}_{\mathrm{c}} = (0.092 \pm 0.044, 0.096 \pm 0.045)$ or $|\boldsymbol{\gamma}_{\mathrm{c}}| = 0.13$, where the errors are 1σ bounds from Monte Carlo simulations. Numerical modelling results in a shear of 0.0696 almost exactly parallel to $\boldsymbol{\gamma}_{\mathrm{c}}$ for isothermal spherically symmetric potentials. We therefore expect the time-delays (or h if we take Δt as known) of the spherical model with shear to be a factor 0.475 smaller than in the shearless case. For the moderately small shear of ≈ 0.07, this is a huge effect. This factor is in good agreement with the expected value of $1/2$ for idealized systems.

To compare results in the general case, we performed numerical model fitting with an elliptical potential plus external shear.

$$F(\theta) \propto \left(\left(\frac{\cos\theta}{1+\varepsilon} \right)^2 + \left(\frac{\sin\theta}{1-\varepsilon} \right)^2 \right)^{\beta/2} \tag{6.43}$$

Elliptical potentials are known to be unphysical for large ellipticities. Although ε is small in our case, we may expect unrealistic solutions for small values of β, because the limit of acceptable ellipticities vanishes for $\beta \to 0$ (cf. last section). In fact the fitted ε even increases with decreasing β.

We decided not to use invented time-delays (calculated for a reference model) to fit the models. In this way, we can check the validity of our results even for cases where multiple time-delays are not used as constraints. Plots of the residuals, ellipticity, shear and time-delay between component A and B are shown in Figure 6.5.

The non-vanishing residuals at $\beta \to 2$ might be worrying at first, because a spherical model without external shear can fit any image configuration for $\beta = 2$. The result would be a sheet of constant density equal to the critical density. There would be no isolated images but an area of constant (and very high) surface brightness. One might thus naively think that residuals should be very small near $\beta = 2$. This is not the case. It is true that the deviations of the projected images will become arbitrary small in the source plane. On the other hand, however, the magnifications diverge in the limit, causing the deviations in the lens plane to stay finite.

Three families of models have to be discussed. First, we fixed the ellipticities at 0 to compare with the results from Wambsganß & Paczyński (1994). The residuals are almost constant for spherical models. The shear and the time-delays scale very accurately with $2 - \beta$ as in the idealized considerations.

More interesting in our context is the behaviour in the shearless case where we expect to find a $(2-\beta)/\beta$ scaling of the time-delays. The χ^2 gets unacceptably large for $\beta \ll 1$. In the other cases ($\beta \gtrsim 0.6$), the agreement with the theoretical predictions also shown in Figure 6.5 is very good. For isothermal models, the ratio of time-delays calculated for the two models (spherical/shearless) is 10–20 per cent larger than predicted by the critical shear. This is still a very good agreement considering that the time-delays were not used as constraints for the numerical models.

Real lensing galaxies usually are elliptical and also embedded in an external shear field. We therefore also let both ε and γ vary freely in order to minimize χ^2. Counting the formal number of constraints and parameters, we expect a minimum of $\chi^2 = 0$. The fact that χ^2 does not vanish is a confirmation of the degeneracy involving shear and ellipticity already discussed

---

[3] All coordinates in this thesis: x to east (positive right ascension) and y to north (positive declination).

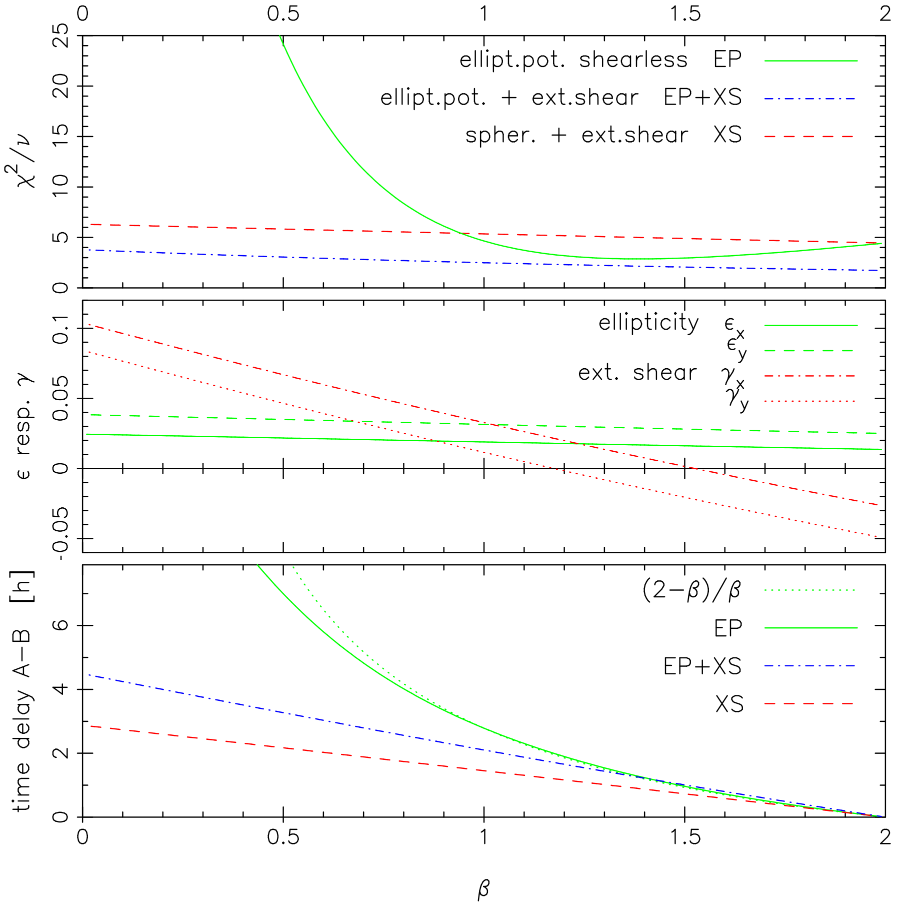


Figure 6.5: Results of numerical model fitting for Q 2237+0305. The top panel shows the reduced χ^2 for the best models. Here β was counted as fixed, resulting in $\nu = 3/3/1$ degrees of freedom for the spherical / shearless elliptical potential / general elliptical potential model, denoted as XS / EP / EP+XS. In the middle, both components of the shear and ellipticity are plotted for the EP+XS model. The last panel presents the time-delay A−B in hours for $H_0 = 75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and Einstein-de Sitter cosmology. We also included a curve scaling with $(2-\beta)/\beta$ to compare with the EP case.

by Witt & Mao (1997). The effective number of parameters is therefore smaller then the formal one. Because we notice the ellipticity changing only slightly with β, a scaling like that in the spherical case is expected. This can indeed be seen in the plot, where the time-delay scales with $2-\beta$. Contrary to the spherical case, γ does not scale proportional to $2-\beta$ but is additionally shifted by a constant offset. This is due to the fact that part of the shear has been transformed to ellipticity.

We conclude that both scaling laws for h or the time-delays can be relevant, depending on the family of models used. In the case of 2237+0305, however, the influence of external shear is stronger than the effect of β for any realistic values of the latter.

6.13.3 PG 1115+080

Models and their degeneracies for this quadruple system (see Figure 6.6) have been studied extensively (e.g. Courbin et al., 1997; Impey et al., 1998; Keeton & Kochanek, 1997; Saha & Williams, 1997; Schechter et al., 1997; Zhao & Pronk, 2001), leading to a variety of more or less realistic mass distributions and a range of values for the Hubble constant between about 40 and $80\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

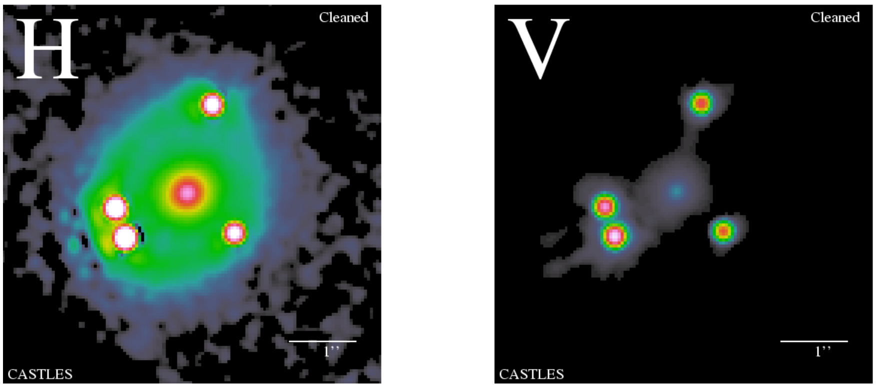


Figure 6.6: IR and optical HST images of PG 1115+080 (Kochanek et al., 2002)

All these authors agree on the importance of taking the effect of a nearby galaxy group into account. Our formalism includes this group as an external shear of the order $\gamma \approx 0.1$. This shear is not well constrained, however. Keeton, Kochanek & Seljak (1997) showed that the residuals do not change much for ranges of $\gamma \approx 0.06$–0.2. With our general family of models, the effect of unknown external shear can be quantified by using the critical shear and equation (6.23). To calculate γ_c, we only need image positions relative to an arbitrary reference centre. The uncertainty in the galaxy position, which is usually much higher than in the image positions, does not affect the result. Using the HST observations from Impey et al. (1998) with their claimed accuracy of $0\farcs002$ as a basis for Monte Carlo simulations, we obtain a critical shear of $\boldsymbol{\gamma}_\mathrm{c} = (0.142 \pm 0.002, 0.167 \pm 0.003)$. Although the ground-based positions from Courbin et al. (1997) are not compatible with the HST results within the formal error bars, the critical shear from these data is about the same, $\boldsymbol{\gamma}_\mathrm{c} = (0.143 \pm 0.003, 0.160 \pm 0.005)$. This means that any shear of the order 0.1 can change the results for H_0 significantly. Since the uncertainties in γ are of this order of magnitude, large effects on h result even for fixed β.

The formalism we present in this chapter was not developed to directly determine H_0 from observations but to study the model degeneracies and scaling laws. One might nevertheless try to use the prescription from section 6.7 to obtain an estimate for the Hubble constant and the external shear for an isothermal model. The errors of the observational data of course have to be taken into account.

Time-delays derived from the same light curves have been published by Schechter et al. (1997) and Barkana (1997). The A component was not resolved in these observations, and time-delays were only determined relative to the sum of A_1 and A_2. This is justified by the small time-delay between the two, which is expected to be the order of hours. For our Monte Carlo simulations, we assumed a time-delay between the A images of (0 ± 1) days.

Results for H_0 in the isothermal case differ depending on which set of time-delays and po-

world model	d_d	d_s	d_ds	d_eff	$d_\mathrm{ds}/d_\mathrm{s}$
flat Λ	0.2194	0.4075	0.3019	3 257	0.7409
flat Λ clumpy	0.2210	0.4744	0.3319	3 473	0.6997
no Λ	0.2034	0.3690	0.2502	3 299	0.6781
EdS	0.1928	0.2894	0.1966	3 121	0.6794
Quint 1	0.2107	0.3761	0.2747	3 171	0.7304
Quint 2	0.2017	0.3354	0.2384	3 122	0.7106
Quint 3	0.1928	0.2894	0.1966	3 121	0.6794

Table 6.4: Distance parameters for PG 1115+080 calculated with $z_\mathrm{d} = 0.310$ and $z_\mathrm{s} = 1.722$. See sections 3.2 and 3.6 for explanations. Units of d_eff are $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ days $\mathrm{arcsec^{-2}}$.

sitions is used. With the redshifts $z_\mathrm{s} = 1.722$ and $z_\mathrm{d} = 0.310$ (Tonry, 1998), we calculated the distance parameters given in Table 6.4. For an Einstein-de Sitter universe, we obtain values of H_0 between 47 and $58\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with errors between 12 and 30 per cent (1σ). The external shear is only weakly constrained, but seems to be of the order 0.1. We have to stress that this result includes all possible isothermal models with arbitrary angular dependence and is thus much more general than elliptical mass distributions. As long as different determinations of the positions and time-delays are not consistent with each other within their error bars, any results for H_0 have to be interpreted with care, of course.

6.13.4 RX J0911.4+0551

This quad, initially discovered as a triple (Bade et al., 1997), has unique geometrical properties and is a strong candidate for time-delay determination, although no result has been published yet[4]. Rapid variability has been detected in the X-ray regime, providing the possibility of a determination of all three time-delays with unprecedented accuracy (Chartas et al., 2001). Figure 6.7 shows images of this interesting system. Our first models (elliptical potential plus shear, $\chi^2 \approx 1$) were presented in Burud et al. (1998). The external shear in the best-fitting model is $\gamma = 0.32$ and points almost exactly in the direction of a nearby cluster of galaxies. The redshift and velocity dispersion of this cluster was measured by Kneib et al. (2000) to $z = 0.769$ and $\sigma_v = (836^{+180}_{-200})\,\mathrm{km\,s^{-1}}$. From this, they derive an absolute shear of $\tilde{\gamma} = 0.11^{+0.05}_{-0.04}$ using a SIS model for the cluster.

To compare with lens models, we have to use the reduced shear $\gamma = \tilde{\gamma}/(1-\kappa)$, because the convergence κ caused by the cluster was not taken into account explicitly. The mass-sheet degeneracy simply scales all parameters in (6.14) with $1/(1-\kappa)$. For SIS models, $\tilde{\gamma} = \kappa$ holds and we obtain $\gamma = 0.12^{+0.05}_{-0.04}$. Contrary to the claim in Kneib et al. (2000), this measurement is not in good agreement with the model from Burud et al. (1998). The reduced shear differs by about 0.2, the direction of the two being almost identical. A possible explanation for this discrepancy is the presence of a second galaxy close to the main lens that might change the potential considerably. It is also possible that the internal asymmetry of the main galaxy itself can not be described as an elliptical mass distribution.

To estimate the uncertainty in time-delays or the Hubble constant derived from them, we use the critical shear again, which is $\boldsymbol{\gamma}_\mathrm{c} = (-0.553 \pm 0.013, 0.101 \pm 0.005)$ or $|\boldsymbol{\gamma}_\mathrm{c}| = 0.56$. Even with this very large critical shear, the uncertainty in the real γ has significant effects, because it

---

[4] After finishing this work, a time-delay of $\Delta t = (150 \pm 6)$ days was presented in the PhD thesis of Burud (2002).

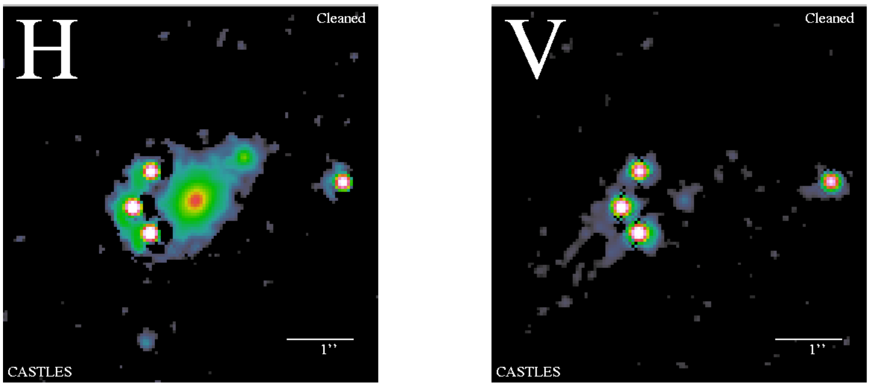


Figure 6.7: IR and optical HST images of RX J0911.4+0551 (Kochanek et al., 2002)

is large as well.

world model	$d_{\rm d}$	$d_{\rm s}$	$d_{\rm ds}$	$d_{\rm eff}$	$d_{\rm ds}/d_{\rm s}$
flat Λ	0.3567	0.3783	0.2123	9 439	0.5611
flat Λ clumpy	0.3712	0.5069	0.2401	11 640	0.4736
no Λ	0.3179	0.3601	0.1730	9 828	0.4804
EdS	0.2805	0.2563	0.1257	8 494	0.4905
Quint 1	0.3330	0.3501	0.1950	8 877	0.5571
Quint 2	0.3072	0.3083	0.1653	8 510	0.5361
Quint 3	0.2805	0.2563	0.1257	8 494	0.4905

Table 6.5: Distance parameters for RX J0911.4+0551 calculated with $z_{\rm d} = 0.769$ and $z_{\rm s} = 2.8$. See sections 3.2 and 3.6 for explanations. Units of $d_{\rm eff}$ are km s^{-1} Mpc^{-1} days arcsec^{-2}.

Even though no time-delay was known when writing this chapter, we present the distance parameters in Table 6.5. The redshift of the source is taken from Bade et al. (1997), the redshift of the lens from a reference in Kneib et al. (2000). With these data, the expected time-delay between the widely separated images was about 210 days. Small time-delays of the order 1 day are expected for the close triple. The short time-delays depend on the lens model more sensitively than the longer one.

The recently measured time-delay from Burud (2002) significantly deviates from the expectation. It will, together with H_0 measurements from other lenses, help in determining the real mass of the nearby cluster. The discrepancy between the lens modelling results and the dynamical measurement (Kneib et al., 2000) can then probably be resolved. Detailed new modelling efforts are necessary to acomplish this task.

6.13.5 B1608+656

This system (Figure 6.8) is the first and up to now only quad for which all three independent time-delays have been measured (Fassnacht et al., 1999). This offers the unique possibility to apply our method to a system providing the complete set of constraints. HST images show a main lensing galaxy but also a weaker second galaxy between the four images (Jackson et al., 1998). We nevertheless apply the method to B1608+656, if only to see if the effect of a secondary lens can be detected in this way, e.g. by pretending there is a very large external shear. Data for the positions of the images and the main lensing galaxy were taken from Koopmans & Fassnacht (1999). The formal accuracy of the image positions is extremely high, of the order 2–12 µarcsec. For the Monte Carlo simulations, we used 1 mas scatter in each coordinate to account for possible shifts by local density fluctuations, caused for example by globular clusters (Mao & Schneider, 1998).

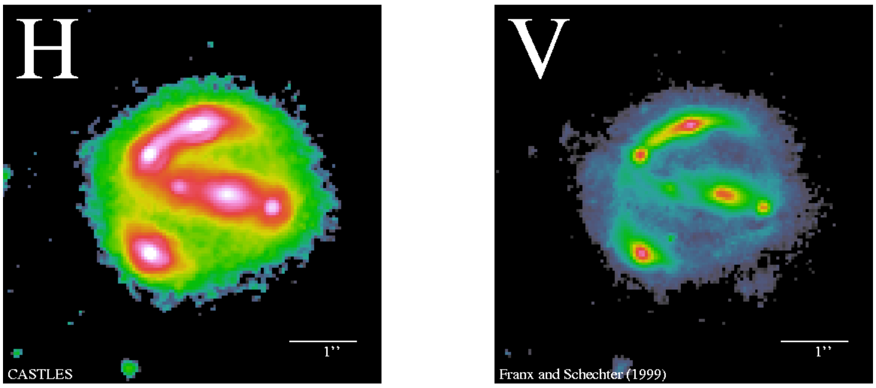


Figure 6.8: IR and optical HST images of B1608+656 (Kochanek et al., 2002). Four images (outer ring) and two lensing galaxies (central part) are seen in the H band image. The weaker galaxy is almost lost in the V band.

We can use the general solution in equations (6.19)–(6.21) to determine the Hubble constant for shearless models. With the redshifts of $z_d = 0.6304$ and $z_s = 1.394$ and standard Einstein-de Sitter cosmology, we obtain a value of $H_0 = (37 \pm 5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for $\beta = 1$. Values for other cosmological models can be calculated with the distance parameters in Table 6.6.

world model	d_d	d_s	d_ds	d_eff	$d_\mathrm{ds}/d_\mathrm{s}$
flat Λ	0.3293	0.4060	0.1817	10 067	0.4477
flat Λ clumpy	0.3386	0.4532	0.1869	11 236	0.4124
no Λ	0.2953	0.3633	0.1430	10 272	0.3935
EdS	0.2660	0.2955	0.1143	9 408	0.3869
Quint 1	0.3092	0.3750	0.1644	9 652	0.4385
Quint 2	0.2879	0.3371	0.1411	9 415	0.4185
Quint 3	0.2660	0.2955	0.1143	9 408	0.3869

Table 6.6: Distance parameters for B1608+656 calculated with $z_d = 0.6304$ and $z_s = 1.394$. See sections 3.2 and 3.6 for explanations. Units of d_eff are $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ days $\mathrm{arcsec^{-2}}$.

The critical shear is $\gamma_c = (0.072 \pm 0.001, 0.069 \pm 0.001)$. For isothermal models, equation (6.17) predicts a shear of about $\gamma = (-0.32 \pm 0.02, -0.11 \pm 0.01)$, slightly depending on which HST image is used to determine the galaxy position. The result for the Hubble constant is $H_0 = (130 \pm 15)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from equation (6.16).

The enormous differences in both H_0 predictions is a consequence of the large external shear

$\gamma \gg \gamma_{\mathrm{c}}$. No external shear at all is needed to fit the data when including the influence of the second galaxy in the field. The models in Koopmans & Fassnacht (1999) even predict velocity dispersions for both galaxies which are of the same order of magnitude. This is surprising, since the secondary galaxy is much weaker in all bands in the optical images (see Figure 6.8).

6.14 Correspondence of elliptical potentials and elliptical mass distributions

Elliptical potential lens models are used as an approximation to elliptical mass distributions. To interpret any results for these models, the correspondence has to be investigated in some detail. Here we only want to calculate the ellipticity of the mass distribution that is approximated by a certain elliptical potential.

For the elliptical potential, we use the approach from section 4.1.4, which in the formalism of this chapter was presented in equation (6.43). We rewrite this expression to find an approximation for small ellipticities of the potential ε_{P}:

$$
\begin{aligned}
\psi_{\mathrm{EP}} &= {r_{\mathrm{ell}}}^{\beta} && (6.44)\\
r_{\mathrm{ell}} &= r\,\frac{\sqrt{1+\varepsilon_{\mathrm{P}}^2}}{1-\varepsilon_{\mathrm{P}}^2}\sqrt{1-\frac{x^2-y^2}{r^2}\,\frac{2\varepsilon_{\mathrm{P}}}{1+\varepsilon_{\mathrm{P}}^2}} && (6.45)\\
&= r\,(1-\varepsilon_{\mathrm{P}}\cos 2\phi)+\mathcal{O}(\varepsilon_{\mathrm{P}}^2) && (6.46)\\
\psi_{\mathrm{EP}} &= r^{\beta}\,(1-\beta\varepsilon_{\mathrm{P}}\cos 2\phi)+\mathcal{O}(\varepsilon_{\mathrm{P}}^2) && (6.47)
\end{aligned}
$$

The density responsible for this potential can now easily be calculated with equation (6.3):

$$
\sigma_{\mathrm{EP}} = \frac{\beta^2}{2} r^{\beta-2}\left(1+\frac{4-\beta^2}{\beta}\varepsilon_{\mathrm{P}}\cos 2\phi\right)+\mathcal{O}(\varepsilon_{\mathrm{P}}^2) \qquad (6.48)
$$

This must now be compared with a true elliptical mass distribution normalized to the same scale

$$
\begin{aligned}
\sigma_{\mathrm{EMD}} &= \frac{\beta^2}{2}{r_{\mathrm{ell}}}^{\beta-2} && (6.49)\\
&= \frac{\beta^2}{2} r^{\beta-2}\Big(1+(2-\beta)\varepsilon_{\mathrm{M}}\cos 2\phi\Big)+\mathcal{O}(\varepsilon_{\mathrm{M}}^2) \quad . && (6.50)
\end{aligned}
$$

The ellipticity of the mass distribution is denoted by ε_{M} in this equation. We learn that the two models are equivalent for small ellipticities, with the relation

$$
\varepsilon_{\mathrm{M}} = \frac{2+\beta}{\beta}\,\varepsilon_{\mathrm{P}} \quad . \qquad (6.51)
$$

For isothermal models ($\beta = 1$), this reads $\varepsilon_{\mathrm{M}} = 3\varepsilon_{\mathrm{P}}$. The ellipticity of the mass distribution is three times as high as that of the potential. The axial ratio is given by $b/a = (1-\varepsilon_{\mathrm{M}})/(1+\varepsilon_{\mathrm{M}})$, see also appendix A.1.

6.15 Einstein rings and high image multiplicities

An interesting property of the general power-law models we used in the main part of this paper is the possibility to produce Einstein rings from point sources for arbitrary values of the external shear. For an Einstein ring parametrized by $r = r(\theta)$, all points on this ring must have the same light travel time to meet Fermat's theorem. For this section, we set κ and γ_y to 0 for simplicity. For an arbitrary direction of the shear, we only have to replace θ by $\theta - \theta_\gamma$. We also assume that $\gamma < 1$.

With constant light travel times, e.g. $t \equiv 0$, the general equations (6.14) describe an ellipse. This ellipse is centred on the lens in the special cases $z_\mathrm{s} = 0$ or $\beta = 1$.

$$r = r_0\,(1 + \gamma \cos 2\theta)^{-1/2} \tag{6.52}$$

The minor and major axes are $r_0/\sqrt{1 \pm \gamma}$, compatible with equation (6.26). The potential is in this case an elliptical one

$$F(\theta) = \frac{1}{\beta}\, r_0^{2-\beta}\,(1 + \gamma \cos 2\theta)^{\beta/2} \quad . \tag{6.53}$$

With respect to the tangential caustic, the effects of ellipticity and shear cancel in these models and the caustic degenerates to a point. This is qualitatively different from elliptical mass distributions where the caustic is deformed and overlaps itself, producing areas of higher multiplicities (see below). With arbitrary z_s and β, the centre of the ellipse is shifted to a position z_ER with[5]

$$\begin{pmatrix} 1+\gamma_x & \gamma_y \\ \gamma_y & 1-\gamma_x \end{pmatrix} z_\mathrm{ER} = \frac{1-\beta}{2-\beta} z_\mathrm{s} \quad . \tag{6.54}$$

For $\beta \neq 1$, this shift can take any value if z_s is varied. To obtain a globally unique function $F(\theta)$, the centre of the lens has to be located inside of the ellipse.

Even for lens systems with four images, it is always possible to find an ellipse passing through all of them, which can act as an Einstein ring for the corresponding value of γ given by the ellipticity. Of course this does not mean that we always see an Einstein ring for this special value of external shear. Since $F(\theta)$ is not constrained for angles between the images, the ring will usually break up and form only a number of discrete images.

Small deviations from the Einstein ring case can lead to an arbitrary number of images near the former elliptical ring. Special cases of these systems (singular isothermal ellipsoidal mass distributions with shear) with up to eight images were mentioned by us in Lopez et al. (1998), by Witt & Mao (2000) and discussed in detail by Keeton, Mao & Witt (2000). Evans & Witt (2001) present results for shearless models with arbitrary $F(\theta)$. From (6.12) and (6.13) we obtain the following condition for an image of a source at $z_\mathrm{s} = 0$:

$$\frac{F'}{F} = -\frac{\beta\,\gamma \sin 2\theta}{1 + \gamma \cos 2\theta} \tag{6.55}$$

A global solution for this differential equation is given by (6.53), which leads to the elliptical ring we discussed before. For a number of discrete images, a more general solution is possible:

$$F(\theta) = f(\theta)\,(1 + \gamma \cos 2\theta)^{\beta/2} \tag{6.56}$$

At the positions of the images θ_i, equation (6.55) has to be met, leading to the simple condition

$$f'(\theta_i) = 0 \quad . \tag{6.57}$$

As $f(\theta)$ is an arbitrary function, we can easily construct systems with any number of images. The radial coordinates of the images can then be determined to be

$$r(\theta_i) = \big(\beta\, f(\theta_i)\big)^{1/(2-\beta)} (1 + \gamma \cos 2\theta_i)^{-1/2} \quad . \tag{6.58}$$

---

[5] This equation is valid for arbitrary directions of γ.

Chapter 7

The radio lens JVAS B0218+357

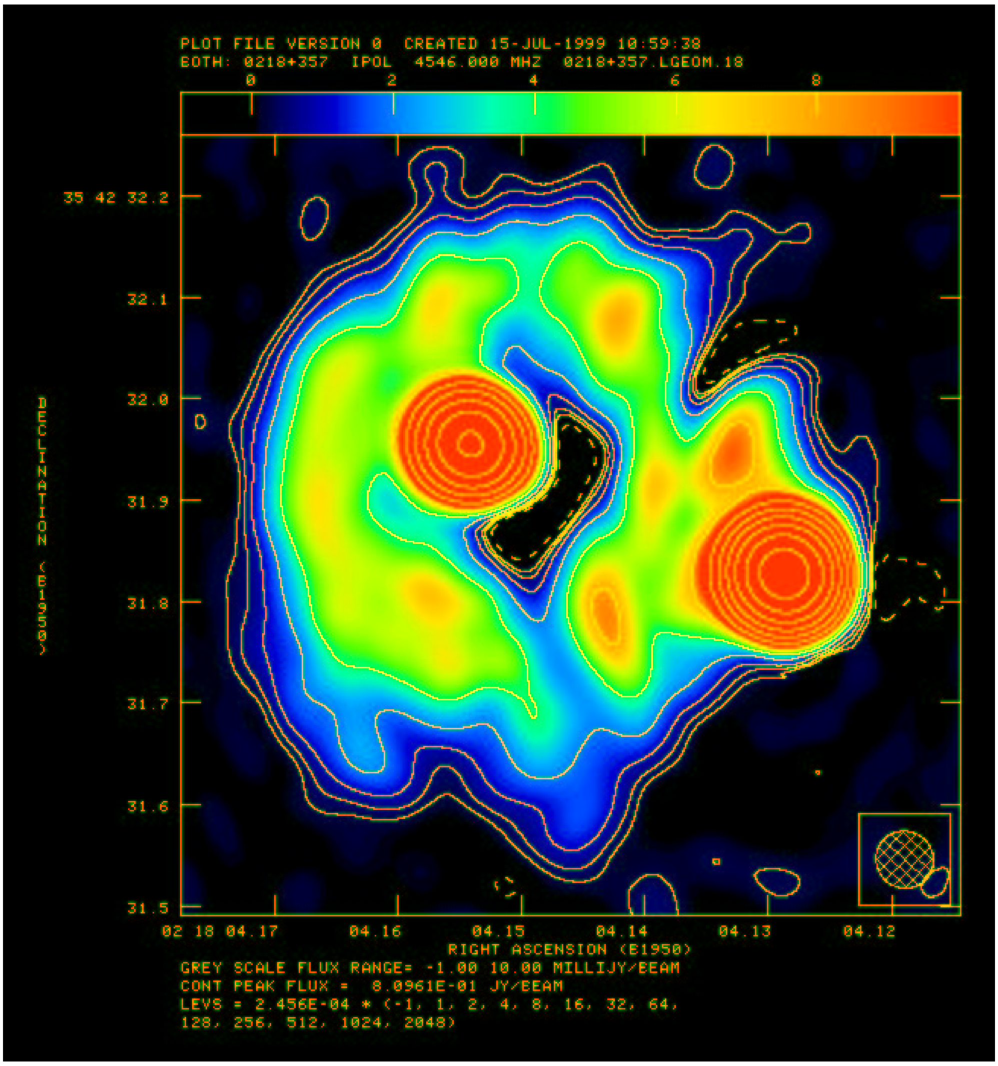


Figure 7.1: 5 GHz MFS map of B0218+357 from combined MERLIN and VLA data (from Biggs et al., 2001). See section 7.10 for details of these data.

7.1 Introduction

The lens JVAS B0218+357 plays a central role in this work, because it was the main motivation to start the work on the LENSCLEAN algorithm. The special properties of this system make it uniquely valuable for cosmological applications, especially for determining the Hubble constant H_0. We will show in the following sections that the lens models cannot be constrained strongly enough with 'classical' lens modelling, when only the compact images (see Fig. 7.1) are used. Before this work was started, no good result for H_0 could be obtained, despite the fact that B0218+357 is such a well studied system with a wealth of data at different frequencies and resolutions available. In chapter 9, we will use the LENSCLEAN algorithm to constrain all model parameters sufficiently tight to obtain a robust and competitive result for the Hubble constant. LENSCLEAN is able to take advantage of the rich structure of the radio ring (see Figure 7.1) to fit mass models of the lens. This information has not been used before for this purpose. B0218+357 is now the first system for which LENSCLEAN has been applied to determine the Hubble constant.

7.2 Discovery

B0218+357 was discovered in the 'Jodrell/VLA Astrometric Survey' (JVAS), which was carried out at 8.4 GHz to select phase calibrator sources for the MERLIN array and to search for gravitational lenses (Patnaik et al., 1992, 1993). With two images separated by 335 mas, it is still the lens system with the smallest separation known. The lensed image consists of two bright flat-spectrum compact components (A and B) and a steep-spectrum radio ring with the same diameter as the image separation (see Figures 7.1 and 7.2). Variability of the two images was detected soon, so that the chance to determine the Hubble constant from this system became apparent.

The compact images are strongly polarized (up to more than 10 %) with a very high differential Faraday rotation. This is a first strong hint that the lensing galaxy probably is a spiral with a very active interstellar medium along at least one of the light paths.

VLBA observations at 15 GHz (Patnaik, Porcas & Browne, 1995) for the first time resolved the substructure of the two compact images. Both consist of two subcomponents each, with a separation of about 1.4 mas (see Figure 7.3 on page 74 and the very similar maps in Figure 7.2). The positions and to some extent also the shapes of these subcomponents can potentially provide constraints for the lens models.

7.3 Data for classical lens model fitting

7.3.1 Image positions

The most fundamental data are the positions of the two bright images. Because there are no other bright sources in the field and absolute astrometry is not accurate enough to compare different measurements, only relative positions are meaningful. Unless otherwise noted, we measure all positions relative to the A component. All coordinates are in the J2000 equinox, x is measured positive eastwards (positive right ascension), y northwards (positive declination).

Table 7.1 (page 72) shows a compilation of known relative positions of the A and B components. For the high-resolution VLBI data, the numbers are given for each of the two subcomponents. With the exception of the 5 GHz EVN data, these results are in very good agreement

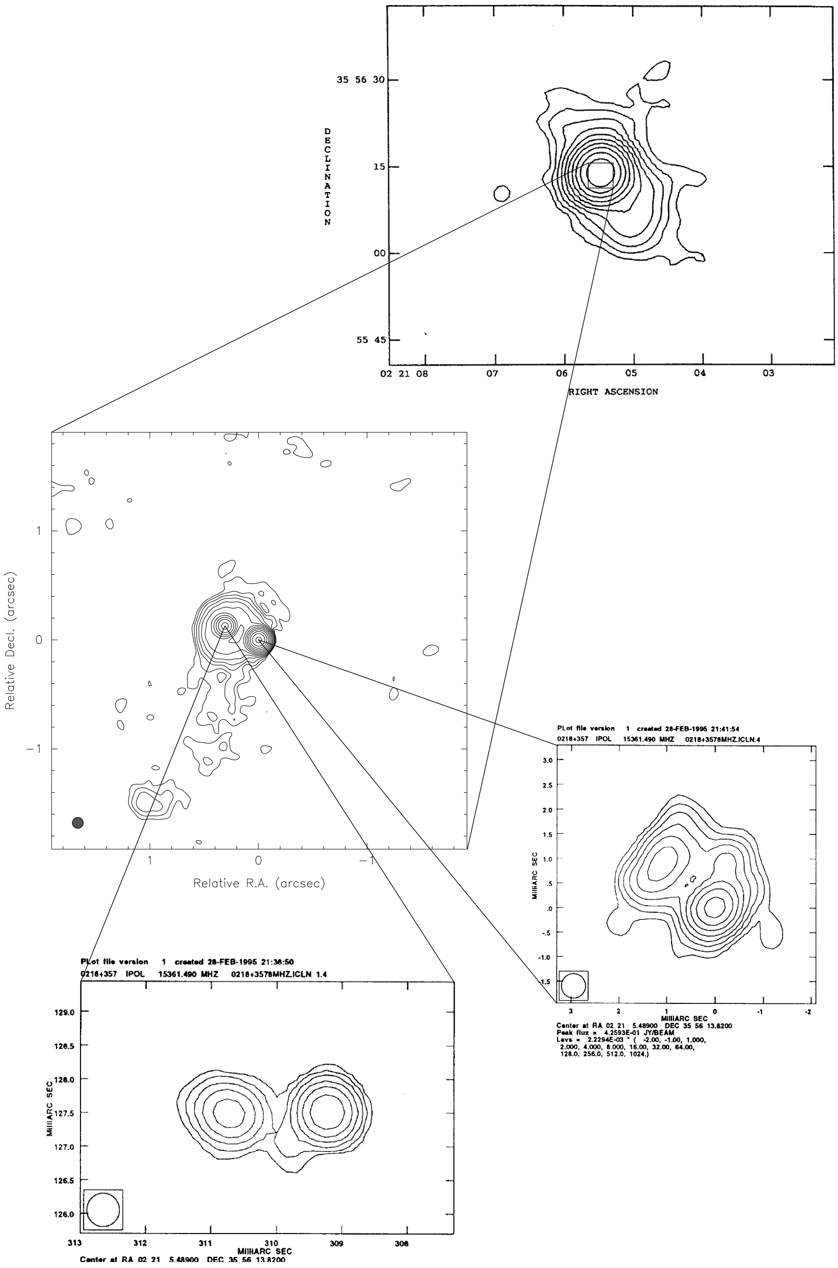


Figure 7.2: B0218+357 at different scales from tens of arcseconds down to milliarcseconds. Top: VLA 1.5 GHz (Patnaik et al., 1993), middle: VLA 15 GHz (Biggs et al., 1999), bottom and bottom right: VLBA 15 GHz (Porcas & Patnaik, 1996a). Composite taken from Biggs (1999).

with each other. An apparent shift of the source with changing frequency has been discussed for some time (see section 7.8). In this scenario, the source should move in the direction along the jet (from subcomponents 1 to 2) with decreasing frequency. For the change in the relative B−A position, an effect opposite to the observed one should then be observed.

The VLBA positions for component 1 are used for our lens models. Formal statistical errors

data set			x [mas]	y [mas]	
5 GHz	EVN	[a]	308.5	130.3	
8.4 GHz	VLBI	[b]	309.00	127.30	subcomponents B1−A1
			309.32	126.37	subcomponents B2−A2
15 GHz	VLBA	[c]	309.2	127.4	subcomponents B1−A1
			309.6	126.6	subcomponents B2−A2
15 GHz	VLA	[d]	310.56	127.11	
optical	HST	[e]	307	126	

---

[a] from Patnaik et al. (1993), given to 0.6 mas accuracy
[b] from Kemball, Patnaik & Porcas (2001), accuracy 0.09 mas in each coordinate
[c] from Patnaik et al. (1995), given to 0.6 mas accuracy
[d] own fit (uniform weighting), same data as for LENSCLEAN
[e] from Lehár et al. (2000)

Table 7.1: Relative position of the B component with respect to A.

of positions from radio data are not very meaningful in cases like this, because calibration errors and confusion with other parts of the source are often more important for the errors than the thermal noise of the observations. To calculate χ^2 residuals for the lens models, we use an accuracy of $\sigma = 0.1$ mas for each coordinate. This value is not critical for the lens models and the results do not change significantly even if very different values are used.

The VLA position was determined with model fitting of two δ-components in DIFMAP. The data were reweighted to uniform before, to reduce the influence of the larger scale structures of the ring on the result. This is not critical for the position, but is of some importance for the flux ratios.

7.3.2 Flux density ratio A/B

Determination of flux densities are difficult for several reasons. Luckily, we are only interested in the ratio A/B, which involves less problems, and not in the absolute values. Nevertheless, the flux density ratio in B0218+357 is still a matter of debate. The most serious problem is its apparent dependence on the frequency. This is disturbing, given the fact that lensing is an achromatic process. Two possible explanations for this behaviour will be discussed later: other physical effects in the lensing galaxy and frequency dependence of the structure and position of the source (see section 7.8).

Another problem is the different light travel time for the two components. The observations detect the source flux at different epochs which, in combination with the strong variability, can lead to incorrect results from single epoch observations. In the process of determining the time-delay, the real flux ratio is obtained for free (see section 7.5), however.

A selection of different measurements is shown in Table 7.2 and Figure 7.13 (page 86). For higher frequencies, the ratio is ca. 3.7 and goes down to 2.6 at 1.7 GHz. For our classical lens models, we use the value 3.75, unless otherwise noted. The data do not constrain the models uniquely anyway, therefore the exact value does not make an important difference.

data set			epoch	ratio
1.7 GHz	VLBI	[a]	1992 Jun 19	2.62
5 GHz	MERLIN	[b]	1991 Aug 26	2.976
5 GHz	MERLIN	[b]	1992 Jan 13	3.23
5 GHz	MERLIN	[b]	1992 Mar 27	3.35
5 GHz	EVN	[b]	1990 Nov 19	3.185
8.4 GHz	VLA	[b]	1991 Aug 1	3.247
8.4 GHz	VLA	[c]	1996/1997	3.57 ±0.01
8.4 GHz	VLA	[d]	1996/1997	3.2 ±0.35
8.4 GHz	VLBI	[e]	1995 May 9	3.18 ±0.17
		[f]		3.72 ±0.20
15 GHz	VLA	[b]	1991 Aug 1	3.690
15 GHz	VLA	[g]	1992 Nov 18	3.79
15 GHz	VLA	[h]	1996/1997	3.73 ±0.01
15 GHz	VLA	[d]	1996/1997	4.3 ±0.65
15 GHz	VLBA	[i]	1994 Oct 3	3.623±0.065
22 GHz	VLA	[b]	1991 Aug 1	3.636
5550 Å	HST	[j]		0.14

[a] from Patnaik & Porcas (1999)
[b] from Patnaik et al. (1993)
[c] from Biggs et al. (1999), simultaneously fit with time-delay, used for their models
[d] from Cohen et al. (2000), varying part simultaneously fit with time-delay
[e] subcomponents A1/B1 (Kemball et al., 2001)
[f] subcomponents A2/B2 (Kemball et al., 2001)
[g] own fit (uniform weighting), same data as for LensClean
[h] from Biggs et al. (1999), simultaneously fit with time-delay
[i] from Patnaik et al. (1995)
[j] see Table 7.5

Table 7.2: Flux density ratio A/B for different frequencies and arrays. Figure 7.13 (section 7.8) shows a plot of the data.

7.3.3 Relative positions of subcomponents

Figure 7.3 presents a high-resolution VLBA map of the A and B components and shows that they consist of two subcomponents each. The relative displacements of the subcomponents are given in Table 7.3. VLBA observations at the same and at higher frequencies (15, 22, 43 GHz) were presented by Porcas & Patnaik (1996b). Unfortunately no results for the positions of the subcomponents were included. A more recent publication (Kemball et al., 2001) presented results from 8.4 GHz VLBI observations, which are also included in Table 7.3.

7.3.4 Shapes of subcomponents

The subcomponents are marginally resolved in 15 GHz observations, so that there is some hope that their shapes and sizes can also be used as lens model constraints. Patnaik et al. (1995) fit elliptical Gaussians to the data with results given in Table 7.4 (page 75). At 8.4 GHz, the resolution is not sufficient to determine shape parameters for both subcomponents. Kemball et al. (2001) therefore only present numbers for the A1 and B1 component.

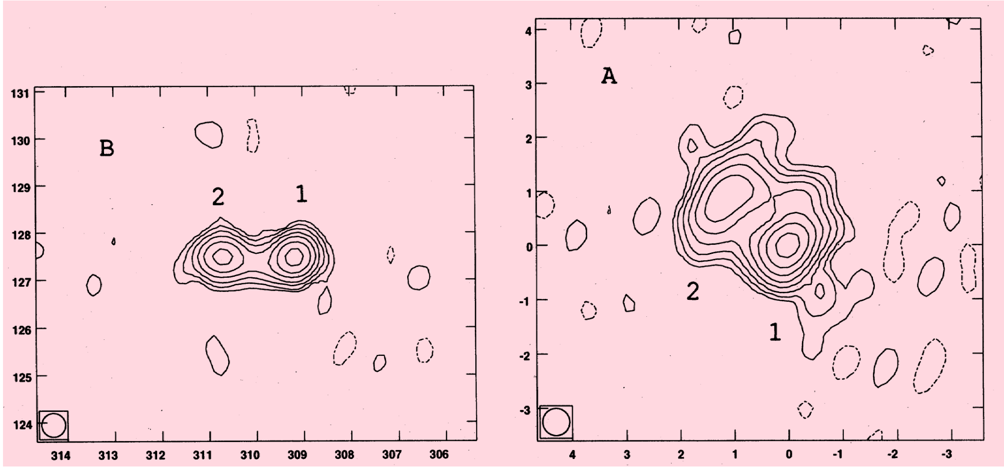


Figure 7.3: 15 GHz VLBA maps of the B and A components of B0218+357 (from Patnaik et al., 1995). Tick marks are separated by 1 mas each.

	Δx [mas]	Δy [mas]	Δx [mas]	Δy [mas]
A2-A1	1.072	0.868	1.18 ± 0.05	0.87 ± 0.05
B2-B1	1.470	0.000	1.50 ± 0.05	-0.06 ± 0.02
frequency	15 GHz		8.4 GHz	

Table 7.3: Relative positions of the subcomponents in A and B at 15 GHz (Patnaik et al., 1995) and 8.4 GHz (Kemball et al., 2001)

7.3.5 Position of the lens

No radio emission has been detected from the lensing galaxy itself or from a central image so far. The only hope to estimate the galaxy position directly (without modelling the mass distribution) is therefore in the optical. It is, however, extremely difficult to determine the position with the required accuracy of a few milliarcseconds. The task is made even more difficult by the two compact images, which are of a brightness comparable to the lensing galaxy (see Table 7.5). Especially the B component, which lies at the very close distance of about 50 mas from the expected centre of the galaxy and which is the brighter one in the optical, is a problem. With ground based telescopes, the situation is quite hopeless. But even the capabilities of the HST have not been sufficient to obtain a useful result yet.

Lehár et al. (2000) presented results from NICMOS observations, which are not considered very reliable by the authors themselves. We cite the positions only for completeness in Table 7.6. The difference between the results from two observations is 46 mas, much more than can be tolerated for H_0 determination. The images themselves are shown in Figures 7.4 (page 76) and 7.5 (page 77). Especially the residuals after subtraction of the compact images may be more illuminating than any formal statistical errors. Although the infrared-sensitive NICMOS detector may be better suited to determine the lens position, WFPC2 is able to produce more pleasing pictures. We present the H band image in Figure 7.6 on page 77.

The conclusion is that no useful position of the lensing galaxy is available and that this parameter must be determined from other lensing constraints. This is the main problem with

	major axis [mas]	minor axis [mas]	p.a. [deg]
A1 [a]	0.58 ± 0.05	0.28 ± 0.05	-37 ± 5
A2 [a]	1.02 ± 0.05	0.54 ± 0.05	-47 ± 5
B1 [a]	0.36 ± 0.05	0.16 ± 0.05	-65 ± 15
B2 [a]	0.61 ± 0.05	0.23 ± 0.05	83 ± 10
A1 [b]	1.90 ± 0.30	1.12 ± 0.30	157 ± 6
A2 [b]	≥ 2.0	≥ 2.0	
B1 [b]	0.72 ± 0.30	0.35 ± 0.04	73 ± 12
B2 [b]	≥ 0.73	≥ 0.73	

---

[a] 15 GHz (Patnaik et al., 1995)
[b] 8.4 GHz (Kemball et al., 2001)

Table 7.4: Shapes of the subcomponents

band	A	B	galaxy
H=F160W	17.52 ± 0.03	16.94 ± 0.03	17.50 ± 0.04
I =F814W	21.83 ± 0.11	19.39 ± 0.03	20.06 ± 0.18
V=F555W	23.28 ± 0.21	21.11 ± 0.06	21.95 ± 0.29

Table 7.5: Optical magnitudes measured with the HST (Kochanek et al., 2002; Lehár et al., 2000)

B0218+357, because almost all information from the radio observations is needed to determine the lens position and very little is left for other parameters. Because the radio data are potentially sufficient to constrain very general mass models, if only the position of the mass centre was known, every effort should be made to obtain better optical images of the system. Observations with the new ACS camera on board the HST are planned for this purpose (PI: Neal Jackson) and will by carried out in cycle 11. Extensive simulations were performed to estimate the possible accuracy, with the result that the position can be determined with uncertainties of about 5–10 mas. When these images are available, more detailed models can be used and lead to the best constrained mass model of all time-delay lenses and to a very robust and accurate determination of the Hubble constant.

7.4 Redshifts and the lensing galaxy

Redshifts of the lens and source are needed to calculate the distance parameters, which provide the scaling factor to determine the Hubble constant from time-delays and lens model parameters (see section 3.6). If velocity dispersion measurements are available as independent constraints of the mass scale of the lens, the redshift of the lens alone is sufficient (eg. Wucknitz, 1996).

Browne et al. (1993) published a measurement of the lens redshift from narrow emission and absorption lines in optical spectra of B0218+357. Their result of $z_\mathrm{d} = 0.6847$ motivated the search for H I absorption in radio spectra, which was found by Carilli, Rupen & Yanny (1993) at $z = 0.68466 \pm 0.00004$. Wiklind & Combes (1995) and other groups later found strong absorption by different molecules at the same redshift.

component	x [mas]	y [mas]	
A	0	0	
B	307 ± 3	126 ± 3	
galaxy	181 ± 30	69 ± 30	mean
	178 ± 25	46 ± 25	NIC2
	184	92	NIC1

Table 7.6: Positions for the lensing galaxy from HST/NICMOS observations Lehár et al. (2000). The images used for these results are shown in Figures 7.5 (NIC2) and 7.4 (NIC1, Jackson et al., 2000).

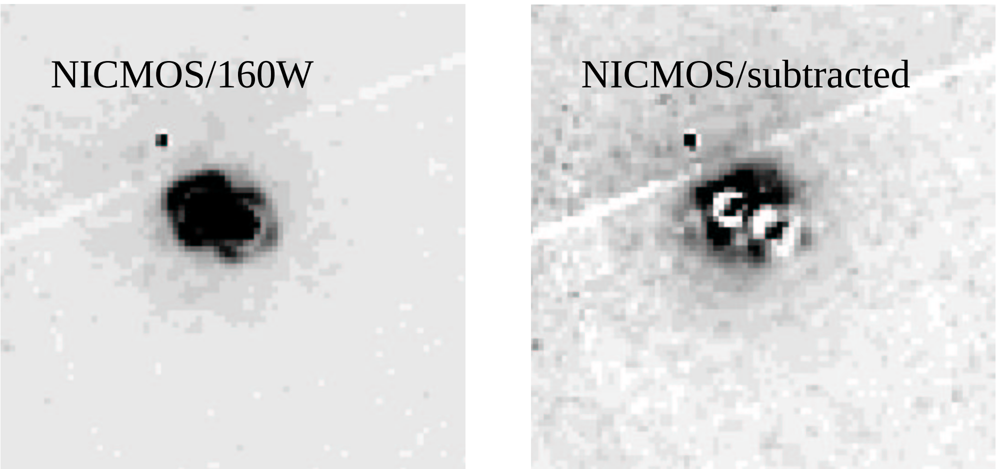


Figure 7.4: NIC1 images from Jackson et al. (2000) used for the results in Table 7.6

We conclude that the lensing galaxy is a late-type gas-rich disk galaxy, presumably a spiral. The inclination seems to be low, so that we do not expect large ellipticities of the mass distribution.

Ground based optical observations (Grundahl & Hjorth, 1995) showed the first sign that in optical bands B is more prominent than A. The hypothesis, which is consistent with the optical data and the molecular absorption lines, is that the A component is covered by a molecular cloud in the lensing galaxy, which causes the absorption lines and an optical extinction of several magnitudes. This cloud may also be responsible for the scattering at low radio frequencies.

The redshift of the source was determined by Lawrence (1996) as $z_s = 0.96$. With these data, we can eventually calculate the normalized angular size distances and the cosmological scaling factor d_{eff} (see sections 3.2 and 3.6). Results are shown in Table 7.7 on page 78.

7.5 The time-delay

The first estimate of a time-delay between the A and B components was presented by Corbett, Browne & Wilkinson (1996) at about the same time as the final assessment of the time-delay in B0957+561 (Kundic et al., 1997). These authors used percentage polarization at 15 GHz

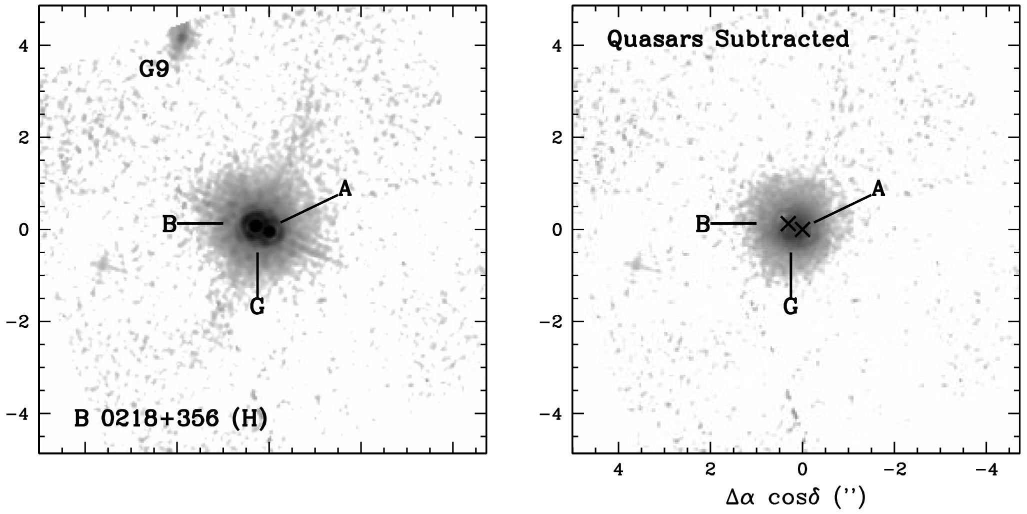


Figure 7.5: NIC2 images from Lehár et al. (2000) used for the results in Table 7.6

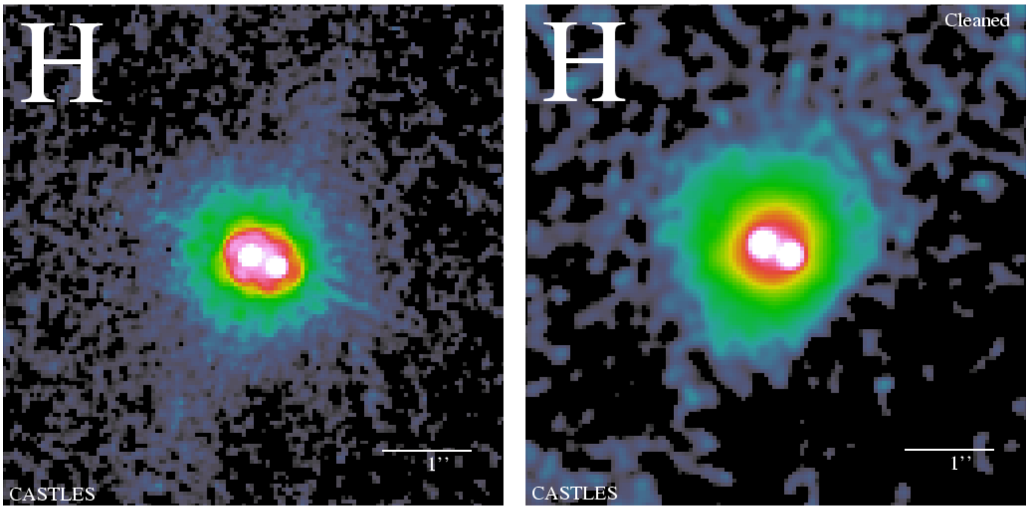


Figure 7.6: WFPC2 image (Kochanek et al., 2002)

observed with the VLA. This measure has the advantage that the a priori unknown amplification ratio has no influence, which reduces the number of parameters to fit. This allows a time-delay determination even from a linear trend in the curves, without any other features. This simple picture is, however, not exactly true in B0218+357, because the A component seems to be depolarized by the inhomogeneous interstellar medium in the lensing galaxy. Different parts of the image suffer different Faraday rotation, which in the integrated measurements reduces the total polarization.

The result of these first efforts was $\Delta t = (12 \pm 3)$ days, with A leading, where the error is

world model	d_d	d_s	d_ds	d_eff	$d_\mathrm{ds}/d_\mathrm{s}$
flat Λ	0.3411	0.3819	0.0887	20 773	0.2322
flat Λ clumpy	0.3523	0.4052	0.0890	22 674	0.2198
no Λ	0.3050	0.3394	0.0693	21 111	0.2043
EdS	0.2725	0.2915	0.0573	19 609	0.1965
Quint 1	0.3195	0.3547	0.0800	20 028	0.2256
Quint 2	0.2964	0.3239	0.0692	19 618	0.2136
Quint 3	0.2725	0.2915	0.0573	19 609	0.1965

Table 7.7: Distance parameters for B0218+357 calculated with $z_\mathrm{d} = 0.6847$ and $z_\mathrm{s} = 0.96$. See sections 3.2 and 3.6 for explanations. Units of d_eff are $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ days $\mathrm{arcsec^{-2}}$.

the $1\,\sigma$ confidence interval. This accuracy is not sufficient for an accurate determination of the Hubble constant, of course.

A great improvement was brought by Biggs et al. (1999), who performed an extensive analysis of different kind of ‘light curves’. They used VLA monitoring data from between October 1996 and January 1997 at 15 and 8.4 GHz. Data were analysed for total flux density, percentage polarization and polarization angle. All these measures at 15 GHz and total flux density only at 8.4 GHz were used to determine the time-delay. In the separate results, all but the one from percentage polarization agree very well with each other. For the latter, however, the fit is not very good at all, so that the formal error bars might be underestimated. The result of the combined analysis is

$$\Delta t = (10.5 \pm 0.4)\,\text{days} \quad , \tag{7.1}$$

where the error is the $2\,\sigma$ (95 %) confidence interval. At the $1\,\sigma$ level, the expected error is 0.2 d. The time-delay is between 10.0 and 11.1 d with 99 % confidence. At the current level of uncertainties of the lens models, this accuracy of the time-delay (2 % at $1\,\sigma$) is sufficient for the determination of H_0. When better data are available for the modelling, more effort might be needed to improve the time-delay, too.

Figure 7.7 shows the combined light curves, in which the time-delay was applied to the A data and results were corrected for the flux density ratio, depolarization fraction in A and for differential Faraday rotation. These plots might be even more convincing than any formal error statistics. Agreement between the shifted A and the B curve is very good in all cases. For percentage polarization, the noise is so large that the constraints on the time-delay are only weak.

The flux density ratios, which are another result of the analysis, are 3.73 for 15 GHz and 3.57 for 8.4 GHz with $1\,\sigma$ errors of 0.01 for both. These can be used as constraints for the lens models, because they are not disturbed by source variability.

Biggs et al. (1999) also presented a preliminary lens model (singular isothermal ellipsoidal mass distribution) constrained by the result for the 8.4 GHz flux ratio and the 15 GHz VLBI substructure (positions and shapes of the subcomponents) with the mass centre as a free parameter. Their result is $x_0 = 252^{+15}_{-\ 9}$ mas and $y_0 = 115^{+4}_{-6}$ mas ($1\,\sigma$ limits) with a corresponding Hubble constant of $H_0 = (69^{+13}_{-19})\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

With exactly the same constraints, we cannot confirm these results in our own calculations (see section 7.6). Their best model is compatible with the data, but the uncertainties seem to be underestimated significantly.

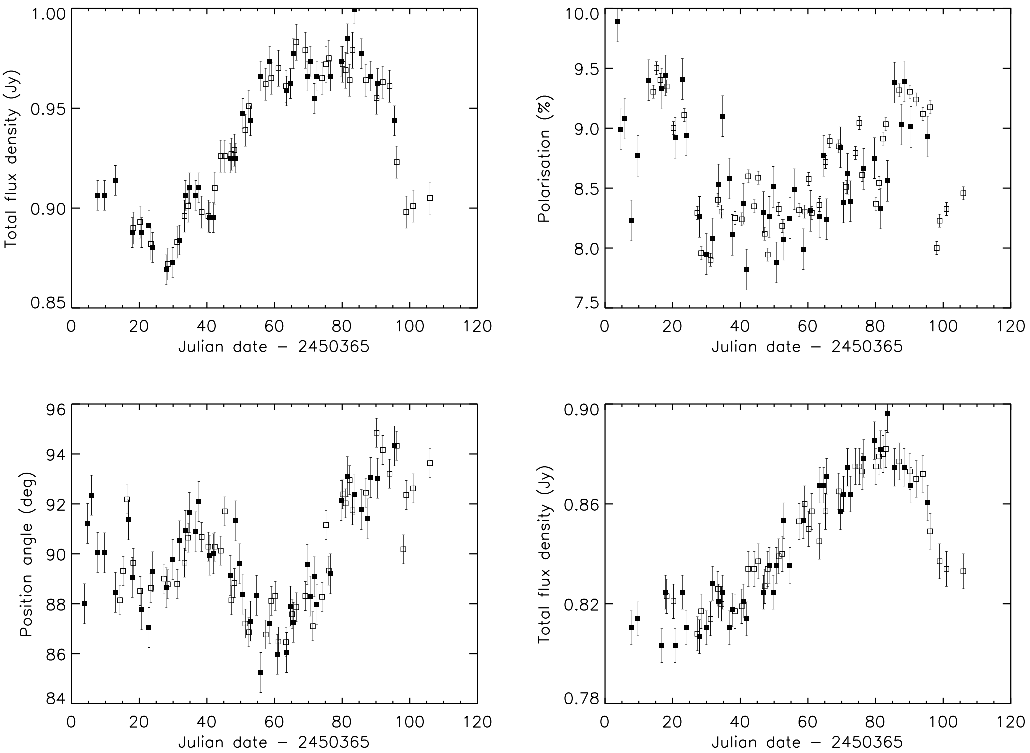


Figure 7.7: Combined 'light curves', produced by delaying the A component by 10.5 d. (from Biggs et al., 1999). (a) Total flux density, 15 GHz, (b) Percentage polarization, 15 GHz, (c) Polarization position angle, 15 GHz, (d) Total flux density, 8.4 GHz. Component A and B measurements are represented by open and filled squares respectively.

Cohen et al. (2000) presented the analysis of an independent data set, taken at the same epoch with the same instrument at the same frequencies. Their result of $\Delta t = (10.1^{+1.5}_{-1.6})$ days (95 % confidence) is compatible with the result from Biggs et al. (1999), but the error bars are much larger. This is partly due to the more general model of source variability and partly due to the different method in determining the time-delay. They assumed that the measured fluxes really consist of two parts: a constant part (probably due to parts of the ring) and the variable part (probably the variable core of the source). The flux ratios for the two are independent free parameters, because the amplification of the core might be different from parts of the ring which contaminate the measurements. It is also possible that the varying part is more or less obstructed by the molecular cloud in front of A than the non-varying part.

The data reduction of Cohen et al. (2000) is more sophisticated, because they try to correct for bias effects from varying *uv* coverage by calculating correction factors with an ideal data set convolved with the beam of each observation.

For the ratio of the variable part, the authors obtain $4.3^{+0.5}_{-0.8}$ at 15 GHz and $3.2^{+0.3}_{-0.4}$ at 8.4 GHz. Within the error bars, the two are compatible with each other. The difference (especially of the error bars) to the earlier result is easy to understand; Biggs et al. (1999) use the total flux to determine the ratio, while Cohen et al. (2000) only use the varying part, which is a relatively small (ca. 10 %) part of the total flux. Parts of the flux can be shifted from the varying to the constant parts and vice versa without changing the light curves significantly. In other words: A scaled light curve looks very similar to one with an added offset. Hence, there is a degeneracy between the ratio of both parts. It is not clear (and doubtable), whether this fact is also responsible for the larger uncertainty in the time-delay of Cohen et al. (2000).

7.6 'Classical' lens models

B0218+357 has the advantage that the expected external shear is very small. Lehár et al. (2000) estimate the convergence and shear caused by field galaxies and the cosmic shear caused by the large scale structure. The values are the smallest for all ten systems discussed in that publication. The contribution from field galaxies is expected to be $\kappa = 0.023$ and $\gamma = 0.014$, a typical value for the cosmic shear for these redshifts is $\gamma = 0.02$. All these contributions are very small compared to the accuracy that is typically achievable for the lens models. We therefore neglect external shear and convergence and work with a single galaxy lens model. The low level of external shear is consistent with the low ellipticity of the ring. It was shown in section 6.15 that the ellipticity of Einstein rings caused by small sources is a direct measure for the external shear. The ring in B0218+357 is not infinitely narrow and the accuracy of the general argument is therefore limited to the ring width.

In classical modelling, the number of constraints is too small to determine all parameters of detailed models uniquely. By far the largest uncertainty is caused by the unknown lens position. Other details of the model will only become relevant when the galaxy position can be determined accurately. To investigate the possibility of determining this parameter, we concentrate mainly on isothermal lens models.

Because of the high accuracy of the image positions, the linear source plane approximation (see section 4.3.2) can be used without noticeable changes of the results. This method is much more stable than full lens plane model fitting and was used to produce all plots in this context. One might even have used an exact source plane solution for the images. This is possible, because changes of the image positions within their error bars do not change the other χ^2 contributions significantly. Even in full numerical fits, the measured image positions are reproduced almost exactly.

7.6.1 Isothermal models

There is strong observational and theoretical evidence that isothermal mass distributions are a good approximation for real galaxies. Isothermal elliptical lens models are thus a very realistic approach. The number of parameters (not counting the source position here) is 5; 2 for the lens position, 2 for the ellipticity and position angle and one for the total mass scale. Without the VLBI data, the number of constraints is only 3; 2 for the relative position of the images[1] and one for the flux density or amplification ratio. If the effective numbers are the same as the real ones (no further degeneracies), two model parameters can be chosen freely and only the remaining 3 are constrained by the observations.

The most natural choice of the independent parameters to investigate the lens models for B0218+357 is the lens position. Mass scale and ellipticity can then be determined for each possible lens position. A much less realistic but very simple approach would be to use only spherically symmetric lens models.

---

[1] One might as well count *both* image positions as constraints. In that case, the source position has to be counted as parameter, too. The difference of the numbers of constraints and parameters does not change in this interpretation.

For models exactly fitting the data, explicit numerical modelling is not even necessary to determine the Hubble constant. This fact was first made clear explicitly by Witt et al. (2000). Chapter 6 presented a deeper analysis of the properties of a superset of the elliptical isothermal models (see also Wucknitz, 2002a). Equation (6.9) provides a relation between the product of time-delay and Hubble constant, called ΔT, and the image positions. This relation becomes particularly simple for isothermal, possibly elliptical, models without external shear. For absolute image positions z_i and a galaxy position of z_0, the model's time-delay has the following form:

$$\begin{aligned} \Delta T &= T_{\mathrm{B}} - T_{\mathrm{A}} && (7.2) \\ &= \frac{|z_{\mathrm{A}} - z_0|^2 - |z_{\mathrm{B}} - z_0|^2}{2} && (7.3) \\ &= (z_{\mathrm{B}} - z_{\mathrm{A}}) \cdot \left(z_0 - \frac{z_{\mathrm{A}} + z_{\mathrm{B}}}{2} \right) && (7.4) \end{aligned}$$

In the case of B0218+357, the image positions z_i are known with very high accuracy. The Hubble constant is then a linear function of the unknown galaxy position z_0. If the lens centre is located exactly midway between A and B, ΔT becomes 0.

An upper limit of ΔT can be calculated without constraints for z_0, because the lens centre has to be located somewhere between the images in a standard lensing scenario[2].

$$\begin{aligned} |\Delta T| &\le \frac{1}{2} |z_{\mathrm{A}} - z_{\mathrm{B}}|^2 && (7.5) \\ &= 0.0558\,\mathrm{arcsec}^2 && (7.6) \end{aligned}$$

For an Einstein-de Sitter universe, this corresponds to a maximal possible Hubble constant of $|H_0| \le 104\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Figure 7.8 shows the Hubble constant as function of the lens position together with some estimates of the latter.

To have at least a chance to constrain the position of the lens centre, more observations have to be included into the modelling. The best data available are the positions of the subcomponents from VLBI observations (see Table 7.3). The subcomponents are separated so closely, that a linear mapping between the images can be used for the calculations. The relative displacements do constrain the relative magnification matrix. To include them in the modelling, the methods from section 4.3.4 are used.

The observational constraints consist of two 2-dimensional vectors, but two more parameters (the displacement in the source plane) have to be included in the models on the other hand. This effectively results in two additional constraints for the mass model. If no degeneracies are present, this should be sufficient to constrain the two parameters of the lens centre z_0 at least to some degree. We have to keep in mind, however, that the flux ratio, which is also a function of the relative magnification matrix, is already included. In addition, isothermal models have a constant magnification of unity in certain directions (radial directions in spherical models). If the lens is moved along these lines, no changes of the magnification matrix are expected.

Figure 7.9 shows the results with the subcomponent positions as further constraints. The standard value of 3.75 was used for the flux ratio in both cases. The most striking result is, that the aforementioned degeneracy does indeed prevent sufficient constraints for the lens position. In the radial direction, almost along the straight line connecting A and B, the residuals are more or less constant. Perpendicular to this direction, the position is constrained quite tightly. Unfortunately, it is the radial direction which is important to determine H_0. The results for

---

[2] This rule is obeyed in all known lens systems, although there are theoretically possible exceptions.

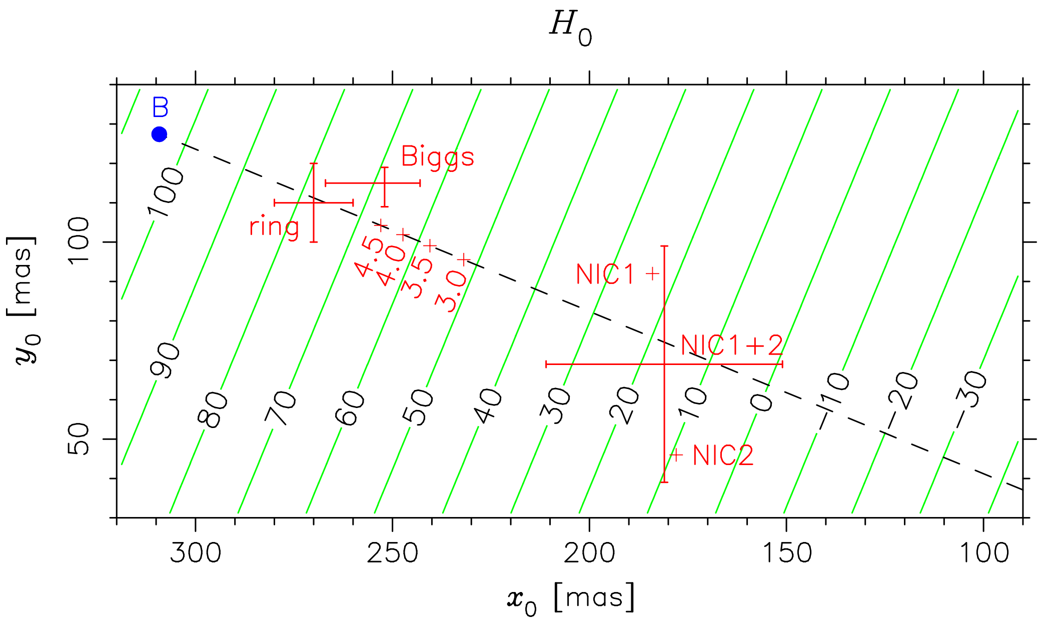


Figure 7.8: H_0 in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ as a function of the galaxy position. Shown are also different estimates for z_0. NIC1 and NIC2 are optical positions from Lehár et al. (2000). At the upper left are the results from Biggs et al. (1999) as well as the centre of the ring (Patnaik et al., 1993). The labelled positions on the dashed line connecting components A and B are spherical models for flux density ratios of 3.0 to 4.5.

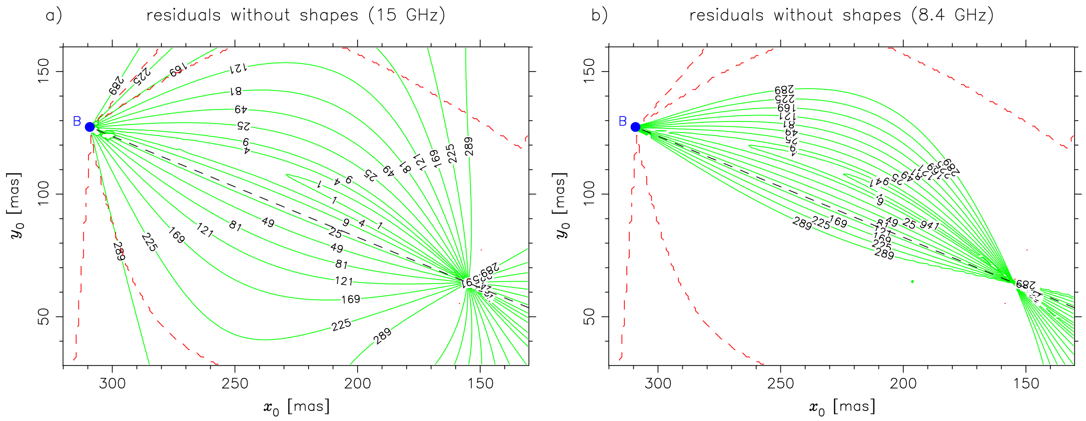


Figure 7.9: Residuals of classical lens model fits for B0218+357 without using the shapes of the subcomponents as constraints. Both plots use the VLBI subcomponent positions from Table 7.3. (a) is for the 15 GHz data (Patnaik et al., 1995), (b) for the 8.4 GHz data (Kemball et al., 2001). Contour lines are plotted for $\chi^2 = 1^2, 2^2, 3^2, 5^2, 7^2, 9^2, \ldots, 17^2$. The dashed curve separates allowed regions with two images in the central parts of the plot from regions where models consistent with the data of the two observed images would additionally show two more images. The same dashed curve was used for (a) and (b), because the differences of the real curves are very small.

the two subcomponent data sets are very similar. Close to the minimum, the χ^2 residuals are less than 1. The uncertainties for the 8.4 GHz data (Kemball et al., 2001) are smaller than for the 15 GHz data (Patnaik et al., 1995). This leads to faster increase of χ^2 and to better constraints with the former data. The results can be compared with those in Fig. 5 of Lehár et al. (2000). Their conclusion, that the classical constraints are not sufficient to determine the Hubble constant, is confirmed by our calculations. This is even true if the subcomponents shapes are included in the modelling.

Our plot also shows the line separating regions where double images are possible from regions where the lens would produce quadruple images. The latter have very high residuals anyway. Otherwise they should have been excluded explicitly. A formal negative amplification ratio (-3.75) was used in the numerical calculations, because double images systems always have different parities. This automatically excludes most of the quadruple lenses, when doubles are possible.

Biggs et al. (1999) present a lens model with much smaller error bars (see Figure 7.8). They used the shapes of the subcomponents as additional constraints. We repeated the calculations with the inclusion of these data as well. To allow robust and reliable model fits, the Cartesian linear formalism explained in appendix A.2 was used. To calculate the ellipse matrices for the subcomponents, a Monte Carlo approach was used. A large number (10^6) of noise realizations following the data in Table 7.4 was used to calculate the Cartesian ellipticity parameters $R_\pm$, R_x and R_y. Mean values and the covariance matrix were then calculated for these results. Simple linear error propagation would lead to very similar results. With this approach, the best source component shapes and the contribution to the residuals can be calculated with standard linear least-squares methods.

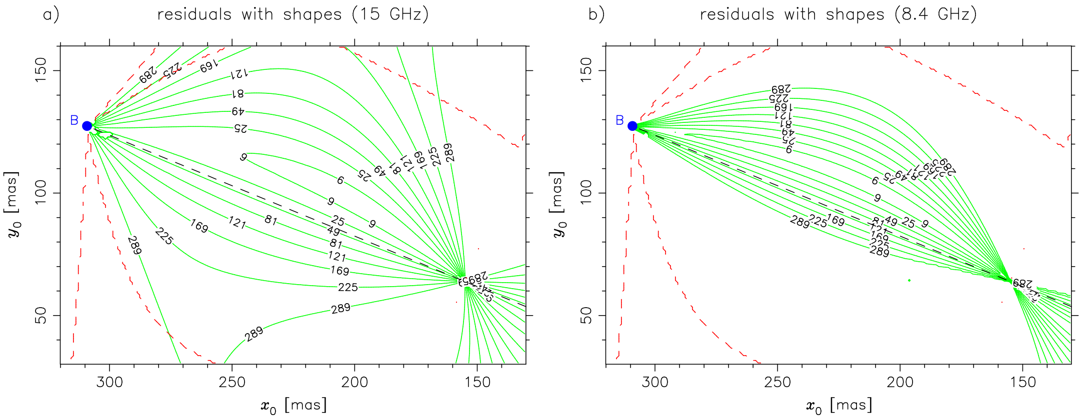


Figure 7.10: Residuals of classical lens model fits for B0218+357 including shapes of the subcomponents as constraints. Subcomponent positions are from Table 7.3, shape parameters from Table 7.4. (a) is for the 15 GHz data (Patnaik et al., 1995), (b) for the 8.4 GHz data (Kemball et al., 2001). Contour lines are plotted for $\chi^2 = 3^2, 5^2, 7^2, 9^2, \ldots, 17^2$. As before, the dashed line separates the double image region in the central part from quadruple image models.

Results are shown in Figure 7.10. For numerical reasons, the lens models were fitted without the shape constraints, and their contribution to χ^2 was added later. Tests on a coarser grid proved that this does not change the results significantly, but the algorithm works more reliable then.

We learn that the subcomponent shapes do not help in constraining the galaxy position. We can therefore not confirm the result from Biggs et al. (1999). Their best fitting lens model may well be the correct one, but the possible errors are greatly underestimated.

The fact that even the residual minimum is quite high, is not regarded as a problem. Estimating uncertainties of subcomponent shapes from VLBI data is a very difficult task and the results can be affected by confusing components which are not included in the model fitting.

It is easy to understand, why the relative magnification matrix cannot constrain the lens position well. If the lens is located exactly on the straight line connecting A and B, the non-diagonal elements of the magnification matrices have to vanish because of the symmetry. There is no radial magnification, because isothermal models have constant deflection angles. The tangential magnification, on the other hand, is already included in the models with the flux ratio. Any changes of the non-diagonal elements with z_0 are only of second order in the displacement. Because the real lens position has to be very close to this 'line of no information', it is not possible to constrain the lens position from small-scale substructure of the images. For larger structures, the second order effects become significant, which improves the situation.

As a conclusion, we have to admit that the Hubble constant cannot be determined from B0218+357's time-delay with classical lens modelling alone. In a central part of this thesis, we will show that the ring can be used to constrain the lens position and H_0 with the LENSCLEAN method (chapter 9).

7.6.2 Non-isothermal models

The disadvantage of the radial direction of the subcomponents is, that the relative magnification is not affected by the lens position for isothermal models. This can be turned into an advantage for non-isothermal models. If the radial mass distribution is described by a power-law model (see section 4.1.3), the radial magnification is influenced by the power-law exponent β *and* the lens position. The fact that the data can be fitted quite well with isothermal models already gives evidence that the true mass distribution is not far from isothermal. To illustrate the influence of β on the residuals, we fitted model parameters for a range of fixed β and x_0 values. The results are shown in Figure 7.11. As before, the subcomponent shapes do not help in constraining β and were thus not used. Power-law models with $\beta > 1$ can produce a central image, which is not observed in B0218+357. The plot also shows a flux limit of the central image of $f_{\mathrm{central}}/f_{\mathrm{A}} \leq 4.3 \cdot 10^{-4}$. This is a $5\,\sigma$ limit from 5 GHz MERLIN data (Norbury et al., 2000 and Norbury, priv. comm.). For models with moderate χ^2 residuals, this limit plays no role. If the radial mass distribution is slightly different near the centre than further outwards, the limit does not apply, anyway. Cuspy models are able to suppress the central image very effectively even for high values of the outer β.

For realistic values of $x_0 \approx 0.26$, the constraints on β are quite strong. If the lens centre can be determined by other means, e.g. optical observations (HST/ACS observations will be done soon, see section 7.3.5) or with the LENSCLEAN algorithm (see chapter 9), the limits for β can become even better. Figure 7.12 shows results for fixed lens positions from Tables 9.2 and 9.3 for demonstrative purposes. Please note that the best LENSCLEAN results will also shift with varying β (see section 9.26). These plots are therefore *not* meant as final results but only to show how strong the constraints can be. The real β will, however, not be far from the minima in the plots, because the minima are very insensitive to small shifts of z_0. We also included statistical $1\,\sigma$ limits for the 8.4 GHz data with the UNI lens model.

In the future, much better constraints can be calculated from new VLBI data (see chapter 10). These data show not only two subcomponents in each image but several ones with very different relative displacements. A preliminary analysis favours a value of $\beta \approx 1.04$, which would be compatible with the discussion above. With these new data, it may even be possible to constrain more general models than simple power laws.

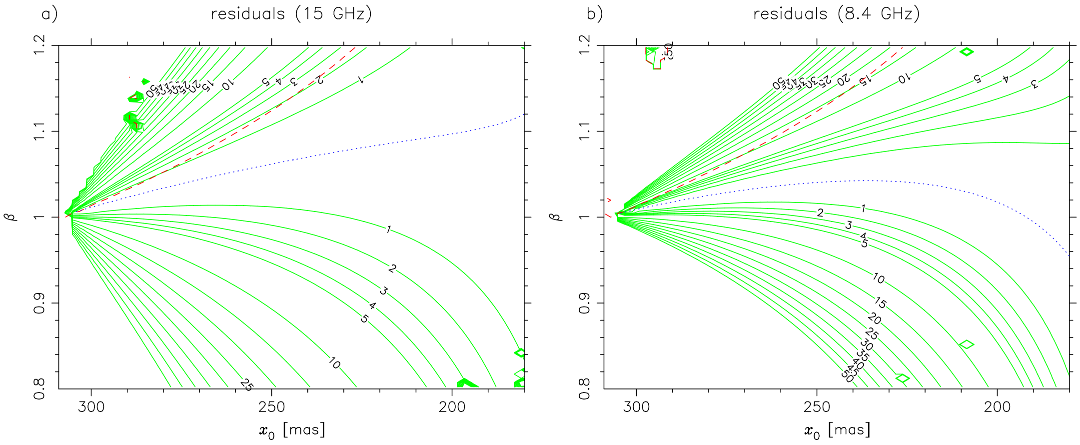


Figure 7.11: Residuals of classical lens model fits for B0218+357. The power-law exponent β and x_0 were fixed for each fit. Subcomponent positions are from Table 7.3. (a) is for the 15 GHz data (Patnaik et al., 1995), (b) for the 8.4 GHz data (Kemball et al., 2001). For $\beta > 1$, the models produce a third central image. The dashed line marks the flux limit for this image. The dotted line shows the best β for each x_0.

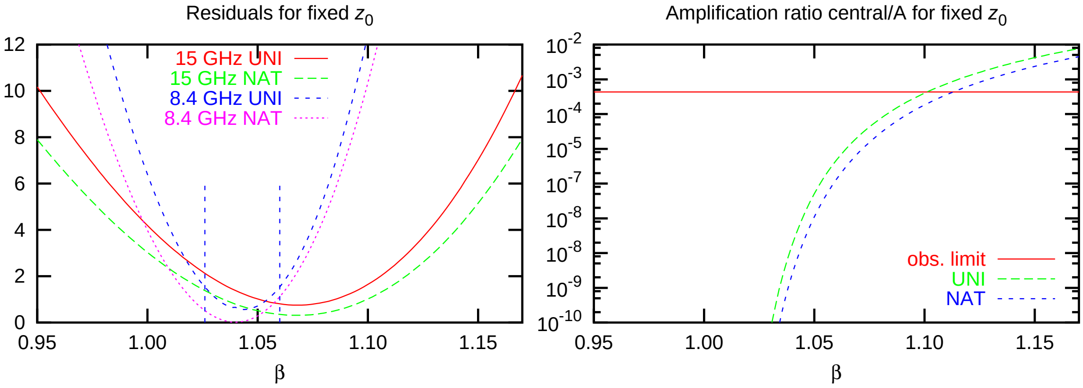


Figure 7.12: Results of classical lens model fits for B0218+357. The position of the lens centre was fixed at the values given in Table 9.2 (UNI model) and 9.3 (NAT model). (a) residuals, 1σ limits for 8.4 GHz UNI, (b) amplification ratio $f_{\mathrm{central}}/f_{\mathrm{A}}$ with the observational upper limit. The different VLBI data are indistinguishable in (b).

7.7 Exploiting the extended emission

Several methods have been developed to use extended emission as constraints for lens models. The most natural and simple approach seems to be the ‘ring cycle’ (Kochanek et al., 1989), which is based on work by Kayser & Schramm (1988). The idea is to use the fact that lensing preserves surface brightness. If certain regions of the source are multiply imaged and well resolved, the observed surface brightness should be the same in all images.

The ring cycle algorithm samples the lens and source plane on a regular grid and calculates as figure of merit the scatter of observed multiple images corresponding to the same source position. Some more constraints are put on the brightness distribution and the lens model parameters to stabilize the minimization. Kochanek et al. (1989) apply the algorithm to the Einstein ring

MG 1131+0456 and are able to constrain the lens models very well.

Unfortunately, the method relies on the fact that measured surface brightnesses represent the true ones very well. This is not the case, if unresolved or marginally resolved components are present in the source. The observed surface brightness maps are then a convolution of the true brightness distribution with a CLEAN beam (or PSF in optical observations). Using such maps with the ring cycle would lead to serious bias effects and errors.

We tried to apply the algorithm to the MERLIN/VLA 5 GHz maps without much success. The two bright images are so dominant, that their smearing by the beam destroys the signal from the ring completely. This result was confirmed with artificial data, which included the resolution effects. The ring cycle was not able to recover the lens model used to build the data.

7.8 Frequency dependent flux ratios

One astrophysically interesting property of B0218+357 is the fact that the flux density ratio of the compact images A and B shows a strong and systematic dependence on the frequency at which it is observed (see Figure 7.13 and Table 7.2 on page 73). Two possible explanations have been proposed for this effect, which both have relevant consequences in the context of lens model fitting with LENSCLEAN.

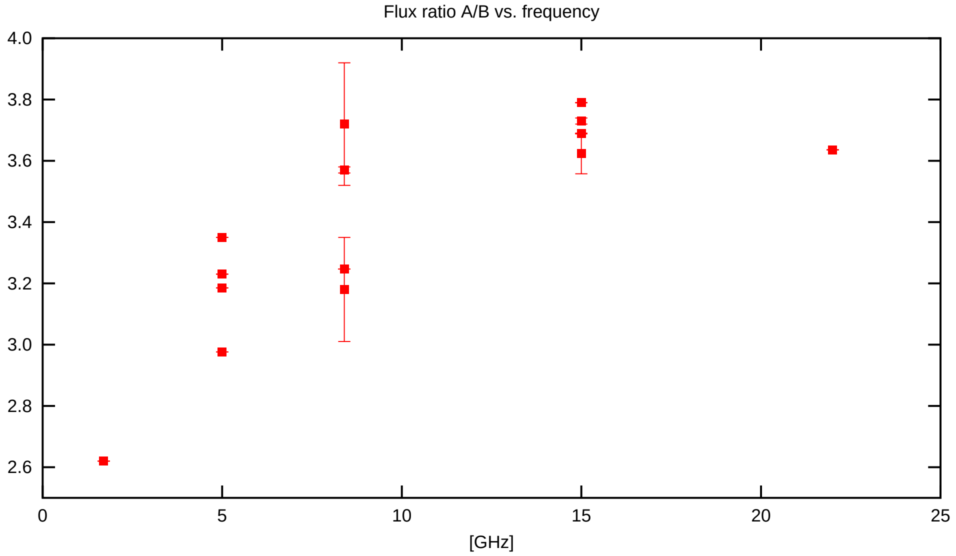


Figure 7.13: Flux density ratio A/B for different frequencies (data from Table 7.2, page 73)

One possibility is, that the interstellar medium, which is more dense and optically active in front of the A component, produces scattering in this light path and effectively removes some of the flux density and thus causes extinction. The effect on the refractive index of a plasma is proportional to λ^2, which will generally produce a stronger effect for lower frequencies. This could explain the smaller flux ratio. We know that the A component is broadened (probably by scattering) at lower frequencies on scales of milliarcseconds resolved by VLBI, because the peak in the maps is higher in B than in A. For resolved sources, the intensity in the map resembles the surface brightness and woulds be equal in the two components. For unresolved sources, on the other hand, the map's peak is a measure of the total flux and should therefore be higher in A, whose total flux density is higher by a factor of ~ 3. The very high Faraday rotation is also good evidence for a very active interstellar medium in the lensing galaxy.

In the optical, the flux ratio is 1:7 (V band, see Tab. 7.2 and 7.5). The extinction relative to the 15 GHz ratio is thus almost a factor of 30 or $3\overset{\mathrm{m}}{.}6$. Even in the H band, it is still a factor of 6 or 2 mag.

For LENSCLEAN, the possible extinction has to be taken into account, because the fitted lens models will otherwise not reproduce the real amplification ratio but the one distorted by the extinction.

The other possible explanation is a frequency-dependent structure and size of the source, which in combination with the strong amplification gradients near the A and B image can produce significant changes in the observed flux ratio. In this scenario, the ratio would not only depend on the frequency but also on the resolution of the observations. The simplest source structure would consist of two components (or groups of subcomponents). The one close to the core would naturally have a flatter spectral index, while the other one would be part of the jet and have a steeper spectrum. The source would then shift 'down the jet' when the frequency decreases and be located in regions with a smaller amplification ratio.

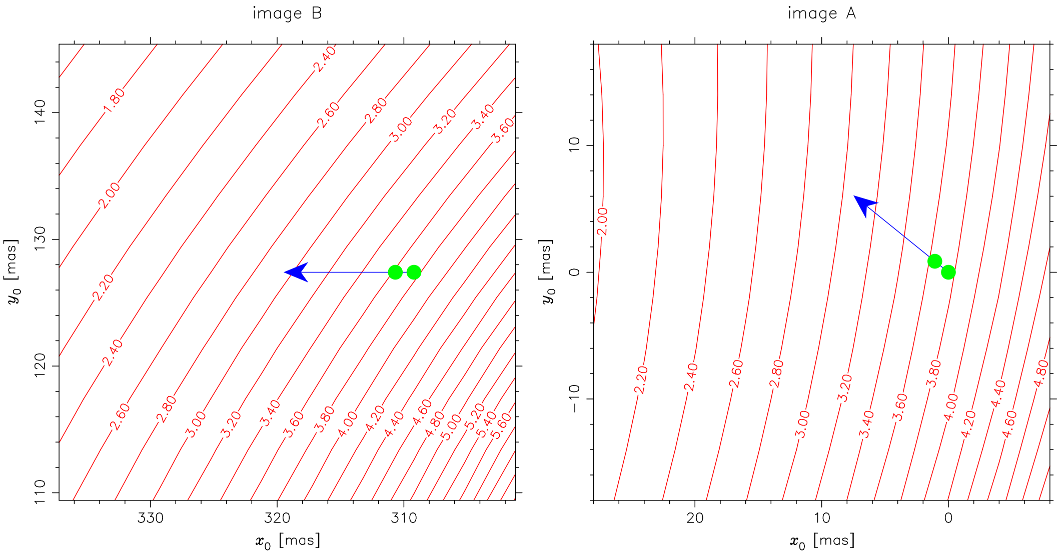


Figure 7.14: Relative amplification (A/B) in a best fitting lens model for B0218+357 with $x_0 = 0\overset{\prime\prime}{.}26$. We only show regions surrounding the two components. The filled circles represent the A1/2 and B1/2 components, the arrows extend in the same direction and symbolizes the jet (compare with Fig. 7.3).

Figure 7.14 shows the relative amplifications for a best fitting lens model, which was derived for a fixed x coordinate of the lens position $x_0 = 0\overset{\prime\prime}{.}26$. This value is close to the best results from LENSCLEAN, see section 9.21.2. We used the data given in section 7.3 as constraints for this model. The resulting lens centre is

$$z_0 = \left(0\overset{\prime\prime}{.}26, 0\overset{\prime\prime}{.}1175\right) \quad , \tag{7.7}$$

the residuals are $\chi^2 = 1.8$. We see, that even though the relative amplification gradient is quite high, a very large source shift of the order 10 mas is needed to explain the decrease of the flux

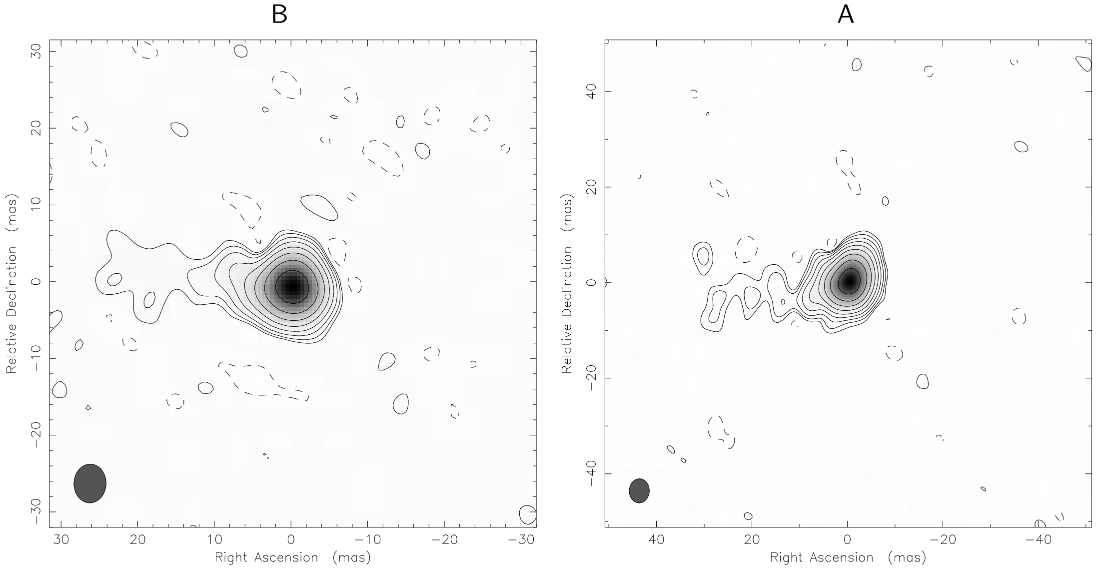


Figure 7.15: 5 GHz VLBI maps of the B and A component from the first Caltech–Jodrell Bank VLBI survey CJ1 (Wilkinson, priv. comm.)

ratio from 3.75 to ca. 2. If this shift is caused by the effect explained above, the source components which become dominant at lower frequencies should also be seen at higher frequencies with deep observations. Comparison of VLBI observations at 5, 8.4 and 15 GHz (shown in Fig. 7.15, 10.2, 7.3), however, show the same structure of two relatively compact subcomponents with distances of the order 1.5 mas (unresolved at 5 GHz). There is no reason to believe that these components are not the same at all frequencies. Otherwise the 5 GHz components should at least be detectable at higher frequencies and vice versa. The resolution of all the maps is better than the required shift of ca. 10 mas, so that a simple shift of centroids of the same components would not be sufficient to explain the observed change of the flux ratio. A small shift of the source was estimated by Porcas & Patnaik (1996a). It seems as if the the centroids at 5 GHz are located about midway between the subcomponents 1 and 2, while they shift to the position of 2 at 1.7 GHz. This shift of the order 1 mas would clearly not be sufficient to explain the observed flux ratio changes.

Another explanation for a possible source shift is refraction by the interstellar medium in the lensing galaxy. This should show in the relative position B−A, while no significant shift is detected in the data. If the shift is real at all, it has to be about the same in A and B. It therefore must be caused very close to the observer, in the interstellar medium of our own galaxy. Very strong gradients of the refractive index are necessary for this scenario, probably far beyond realistic values.

7.9 The 15 GHz VLA data

This data set was used for the LENSCLEAN work which will be presented in chapter 9. It is part of a larger program of VLA observations (Program ID AB 631, PI: A. Patnaik). Observations were done in the widest (A) configuration of the VLA. Total on-source integration time was slightly less than 6 hours in full polarization. The initial 10 sec integrations were further binned to 1 min in our calculations to reduce the amount of data and the computation times.

7.10 The 5 GHz MERLIN/VLA data

Biggs et al. (2001) presented a very high quality map made from a combined MERLIN+VLA observation at 5 GHz (see Figure 7.1). The combination of the two arrays with very different baselines and the multi frequency synthesis technique (MFS) resulted in a very good *uv* coverage. VLA observations were done at two frequencies (4.835 and 4.885 GHz) in August 1995. Three frequencies (4.546, 4.866 and 5.186 GHz, later referred to as 1,2,3) were used with MERLIN in June 1995. Since the *uv* values are the product of telescope spacing and frequency, the MFS technique can improve the *uv* coverage considerably.

Unfortunately, this data set has some disadvantages for LENSCLEAN. While the resolution is better than in the 15 GHz VLA data, and the flux in the ring is higher due to the lower frequency, it was not possible to obtain reliable constraints for the lens model with LENSCLEAN. One problem is the possibility of significant scattering at 5 GHz, which might change the structure of the ring and, more important, change the flux ratio of the two compact components by introducing an effective extinction in the A component. This possible extinction must be taken into account in LENSCLEAN, which adds at least one more free parameter to the problem. Another very serious problem is the frequency-dependence of the emission, which is different for the ring and the compact images. Biggs et al. (2001) approximately corrected for this effect by first mapping the three frequencies independently. The CLEAN components responsible for the compact images were than subtracted from the data so that only the ring remains. The three data sets were rescaled in amplitude to obtain a consistent total flux density at all frequencies. Finally, the compact components from the central frequency were added back to the complete data set. This process may introduce some distortions to the lowest and highest frequencies, because it combines data for the ring from these frequencies with data for the compact components from the central frequency. No such corrections were necessary for the VLA data set, because the frequencies are not as different as in the MERLIN data. The combined data were then successively mapped and self-calibrated to obtain the final data set and the map in Figure 7.1.

We used the resulting *uv* data set as basis for our computations. The spectral index correction seems to work very well in making maps, but introduces errors in LENSCLEAN which are very difficult to analyse.

Since we are mainly interested in the high resolution of the data, we did generally not use the VLA part. It would provide no further information about the lens model but might introduce further calibration problems in combination with the MERLIN part.

7.10.1 Correcting the weights

The weights in the *uv* data set were taken from the scatter in 32 sec intervals over which the visibilities were integrated. This leads to some visibilities with unrealistically high weights, which showed in striping in the dirty beam. To correct the weights, we first used DIFMAP to build emission models for the three frequencies and LL and RR polarizations separately. We then completely rejected the old weights and used the difference between the model and the measured visibilities as basis for a statistical analysis. We fitted these errors with station-based weights constant over time. This means that the weight of each visibility on a baseline consisting of the telescopes i and j was set to the product of the corresponding station weights, $w_{ij} = w_i w_j$. A χ^2 minimization was used to find the optimal station weights. The true RMS scatter could be reproduced by this approach with an accuracy of 5–9 %. The resulting weights showed the expected dependence on the quality and size of the telescopes, which was used as another plausibility check.

Another iteration of this procedure, using an emission model built with the corrected data set, did not introduce significant changes to the weights and was therefore not used for the computations.

When producing maps with DIFMAP and the corrected data of the three frequencies separately (without self-calibration), typical values of χ^2 were 1.03 for Stokes I and 0.985 for LL and RR. The value for I is slightly larger than for LL and RR, which is probably due to different calibration errors in LL and RR. The combination of the two cannot be fitted with normal emission models as well as the two separately. Using the complete MFS data set, residuals increased to 1.030, 1.591 and 1.379 for Stokes LL, RR and I. The combination should therefore be treated with particular caution in LENSCLEAN.

7.10.2 Differences of the sub-data sets

We will see later (section 9.22.2), that the results of the LENSCLEAN algorithm differ significantly for the three frequencies and Stokes parameters. To get a feeling for the differences in the data, we used exactly the same automated DIFMAP-procedure to make maps from the sub-data sets. More attention was paid to treat all sub-data sets equally than to produce the best maps possible.

The result for all three frequencies (plus all combined) and Stokes LL, RR and I is shown in Figure 7.16. Especially the differences between the frequencies are obvious. Simultaneous self-calibration for all frequencies might help in obtaining a consistent complete data set, but this does not help in constraining the lens models, which is the main goal of our work.

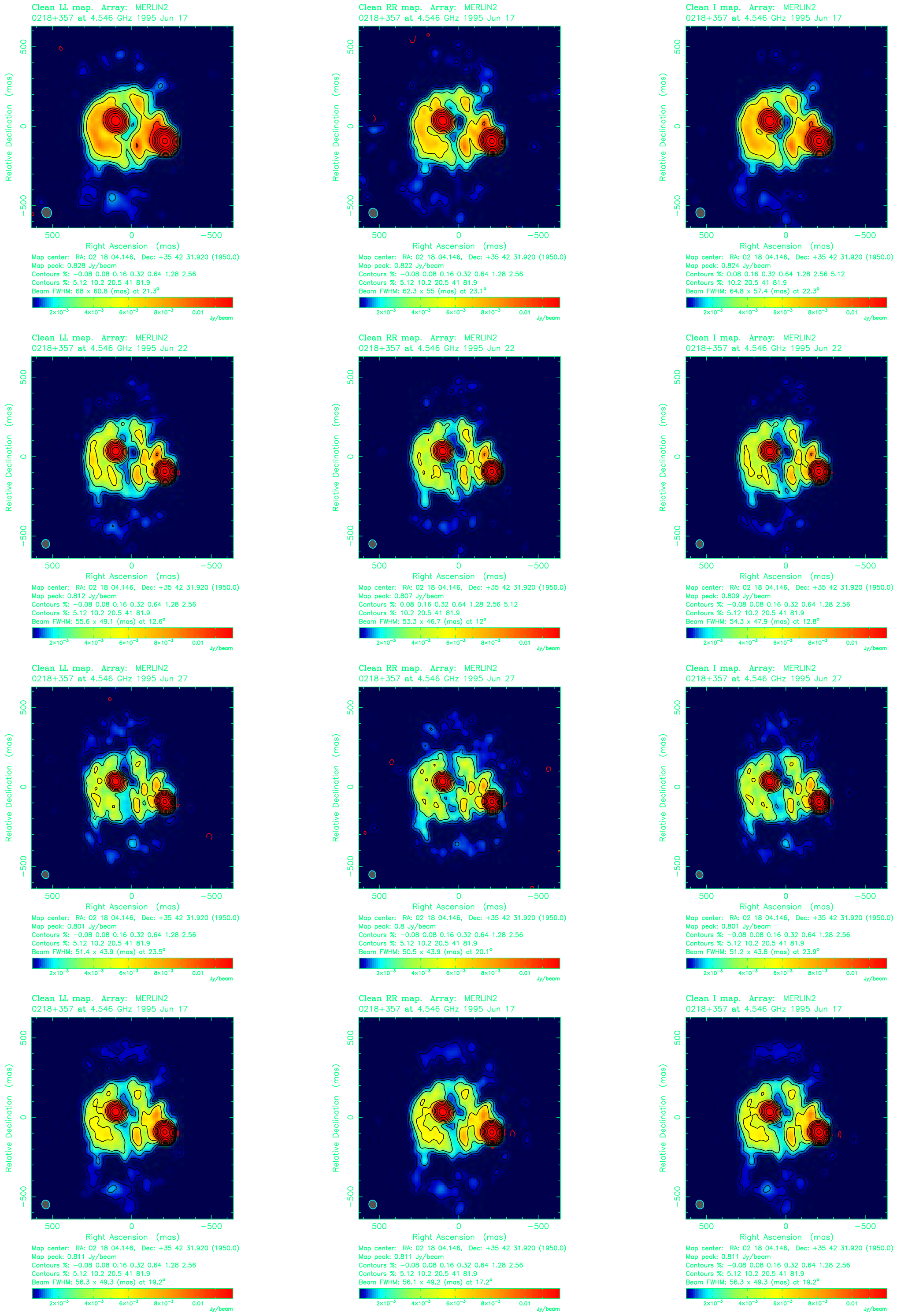


Figure 7.16: Naturally weighted maps from the 5 GHz MERLIN data set, produced with DIFMAP. From left to right, we show Stokes LL, RR, I. The first three rows are calculated for frequencies 1, 2, 3, the bottom row for the combined MFS data set.

Chapter 8

Radio interferometry

To help in understanding the LENSCLEAN algorithm which will be described in chapter 9, we want to give a brief introduction to the theory of radio interferometry. The intention is not to explain all technical details that are important in the reduction of radio data, nor to provide instructions for calibration and mapping. We merely want to make the following chapters accessible for scientists working on gravitational lenses but without experience in radio astronomy. We therefore skip everything that is, in our opinion, not required to know in the context of this work. To actually prepare a dataset for the use with LENSCLEAN or to produce a map using the standard techniques, a lot more has to be known about the technical details and possible problems of interferometric data.

For a deeper discussions of many aspects of radio interferometry, we refer the reader to the NRAO synthesis imaging summerschool proceedings (Perley, Schwab & Bridle, 1986, 1989; Taylor, Carilli & Perley, 1999) and other textbooks on the subject (e.g. Thompson, Moran & Swenson, 1986).

8.1 Why radio interferometry?

The main reason for using interferometers for radio observations of gravitational lenses is resolution. A single, diffraction limited, telescope of a diameter D observing at a wavelength λ has an angular resolution of about

$$\theta \approx \frac{\lambda}{D} \quad , \tag{8.1}$$

where the exact value depends on how it is defined. For the largest fully steerable telescopes ($D = 100\,\mathrm{m}$) operating at a typical wavelength of $\lambda = 6\,\mathrm{cm}$, the resolution would be of the order $2'$. This is about the effective resolution of the human eye and a factor of a few hundred worse than achievable with optical telescopes and needed for gravitational lens study, where typical image separations are of the order $1''$.

Because it is technically difficult and prohibitively expensive to build single radio telescopes of the required size, the idea of combining several telescopes of moderate size at different locations was born. One might think of the elements of a radio interferometer as sections of a giant parabolic reflector. Since the telescopes are usually not located on the surface of a paraboloid, electronic delay lines have to be used to compensate for the different travel lengths of the radiation. With a single telescope, the radiation (be it optical or radio) interferes at the focus to form an image of the celestial radiation pattern. With interferometers, the signals are combined electronically and in the computer to produce a similar result.

8.2 Visibilities measured by an interferometer

Let us for simplicity assume monochromatic radiation with a wavelength of λ coming from an infinitely distant point source. Radiation of a broad bandwidth from extended sources can later be described as a superposition of these plane monochromatic waves. If the position of the source on the celestial sphere is denoted by the unit vector $\hat{\mathbf{k}} = (l, m, n)$, the electrical (or likewise the magnetic) field at a position $\mathbf{r} = (x, y, z)$ at the time t can be written as

$$E(\mathbf{r}, t) = E_0 \, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}+\omega t)} \quad , \tag{8.2}$$

where the length of the wave vector is

$$k = \frac{2\pi}{\lambda} \quad . \tag{8.3}$$

In reality, the electrical field is a vector, but the polarization components[1] can be described independently as scalar fields in this context. We think of E as a complex quantity, the real part of which is the physical electrical field itself.

In the following, we always assume that the relevant emission is confined to a small area on the sky near $\hat{\mathbf{k}} = (0, 0, 1)$, therefore $|l| \ll 1$, $|m| \ll 1$ and $n \approx 1$. This assumption is well justified in lensing studies. If we now have two telescopes, measuring the field at different positions $\mathbf{r}_1$ and $\mathbf{r}_2$, we can combine the signals from the two in the following form, where $E^\star$ denotes the complex conjugate of the field E.

$$\left\langle E(\mathbf{r}_1, t) E^\star(\mathbf{r}_2, t) \right\rangle = |E_0|^2 \, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \tag{8.4}$$

Mathematically speaking, this expectation value (or mean value over a limited time interval, typical 1 sec) is the correlation of the two fields. For strictly monochromatic radiation, the averaging would not be necessary, because the product is constant then. The device that measures this quantity is called a correlator. For connected element interferometers, like the VLA or MERLIN, this is done in real-time during the observations, while VLBI correlators read the recorded signals of the different stations from magnetic tape and compute the correlations weeks after the observations. We now define the so-called *uv*-coordinates of the baseline as

$$u = \frac{x_1 - x_2}{\lambda} \quad , \qquad v = \frac{y_1 - y_2}{\lambda} \quad , \qquad w = \frac{z_1 - z_2}{\lambda} \quad , \tag{8.5}$$

and can then define the measured ‘visibility’ as

$$\tilde{I}(u, v) = \left\langle E(\mathbf{r}_1) E^\star(\mathbf{r}_2) \right\rangle \mathrm{e}^{-2\pi \mathrm{i} w n} \quad . \tag{8.6}$$

Writing the squared electrical field $|E_0|^2$ as intensity I_0 in appropriate units and using $n \approx 1$, we can combine (8.4) and (8.6) to obtain

$$\tilde{I}(u, v) = I_0 \, \mathrm{e}^{2\pi \mathrm{i}(ul+vm)} \quad . \tag{8.7}$$

The correlator provides us with measurements of the visibilities determined using (8.6), which will now be used to infer the properties of the source using (8.7).

For fixed telescope positions, the product $E_1 E_2^\star$ is independent of time t. In reality, the telescopes move as a result of the rotation of the earth. This has two important effects. First,

---

[1] We only need left-handed and right-handed circular polarization in this work, denoted by Stokes LL and Stokes RR, respectively. The total intensity, Stokes I, can be calculated by $I = (LL + RR)/2$.

the correlator has to compensate for the effect of changing w coordinate during the integration interval. Second (and more relevant in our context), we do not only measure a visibility for fixed uv-coordinates but for a number of integrations on a track of slowly changing u and v. The movement of the telescopes also limits the allowed averaging intervals to typical values of a few seconds or less.

To return to the analogue of segments of a large paraboloid, we can think of the $\exp(-2\pi\mathrm{i}\,w)$ factor in (8.6) as a compensating delay which is applied by the correlator to correct for the deviations of the telescope position from a true paraboloid[2].

8.3 Finite bandwidth

Let us now think of two superposed signals E^{A} and E^{B} with different circular frequencies ω_{A} and ω_{B}. The correlation of the two is then, analogously to (8.4), given by the following equation:

$$\begin{aligned}
&\left\langle \left(E^{\mathrm{A}}(\mathbf{r}_1,t)+E^{\mathrm{B}}(\mathbf{r}_1,t)\right)\left(E^{\mathrm{A}\star}(\mathbf{r}_2,t)+E^{\mathrm{B}\star}(\mathbf{r}_2,t)\right)\right\rangle \\
&\quad = \left(|E_0^{\mathrm{A}}|^2+|E_0^{\mathrm{B}}|^2\right)\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} + 2\,\mathrm{Re}\left(E_0^{\mathrm{A}}E_0^{\mathrm{B}\star}\left\langle \mathrm{e}^{\mathrm{i}\,(\omega_{\mathrm{A}}-\omega_{\mathrm{B}})\,t}\right\rangle\right)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \qquad (8.8)\\
&\quad \approx \left(I_0^{\mathrm{A}}+I_0^{\mathrm{B}}\right)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \qquad (8.9)
\end{aligned}$$

We see that the electrical field adds in quadrature and the intensity linearly, as long as the integration interval is much larger than $1/$bandwidth to make the second term in (8.8) negligibly small.

For finite bandwidth signals, the phase factor $\exp(-2\pi\mathrm{i}\,w)$ in equation (8.6) is very different for the contributing frequencies. In real correlators, it must therefore be applied as a delay of the signals to retain coherence.

8.4 Visibilities as Fourier transform

The results from the previous section show, how the visibilities are related to the properties of a point source. An area of extended (incoherent) emission can be thought of as a superposition of infinitely many point sources. SInce the visibilities depend linearly on the sources flux density, we can simply integrate the contributions from these point sources to obtain the visibility for a brightness distribution $I(l,m)$:

$$\tilde{I}(u,v) = \iint \mathrm{d}l\,\mathrm{d}m\,I(l,m)\,\mathrm{e}^{2\pi\mathrm{i}\,(ul+vm)} \qquad (8.10)$$

We notice that the visibilities are a simple Fourier transform of the real brightness distribution. The inverse transform reads

$$I(l,m) = \iint \mathrm{d}u\,\mathrm{d}v\,\tilde{I}(u,v)\,\mathrm{e}^{-2\pi\mathrm{i}\,(ul+vm)} \quad . \qquad (8.11)$$

For some fundamentals of Fourier transforms, see appendix B. Unfortunately, it is not possible to use the inverse transform directly to determine the brightness distribution from the visibilities.

---

[2] This is not exactly true, since the correction factor becomes unity if the telescopes are located on a plane and *not* on a paraboloid. The paraboloid is only used to combine the radiation in the focus with constant signal travel time. In radio interferometry, the signals are detected directly at the telescopes, so the ideal positions would be on a plane.

The visibilities are known only for a discrete subset of the *uv* plane and the integral in (8.11) can therefore not be computed.

The coverage of the *uv* plane (also known as 'aperture coverage') depends on the number of telescopes used, their positions on the earth, the position of the source on the sky, the time and duration of the observations, and on the frequencies used. For N telescopes, we measure $N(N-1)/2$ (the number of pairs of telescopes) independent visibilities per integration interval and frequency band. A typical coverage of the *uv*-plane for the VLA is shown in Figure 8.1.

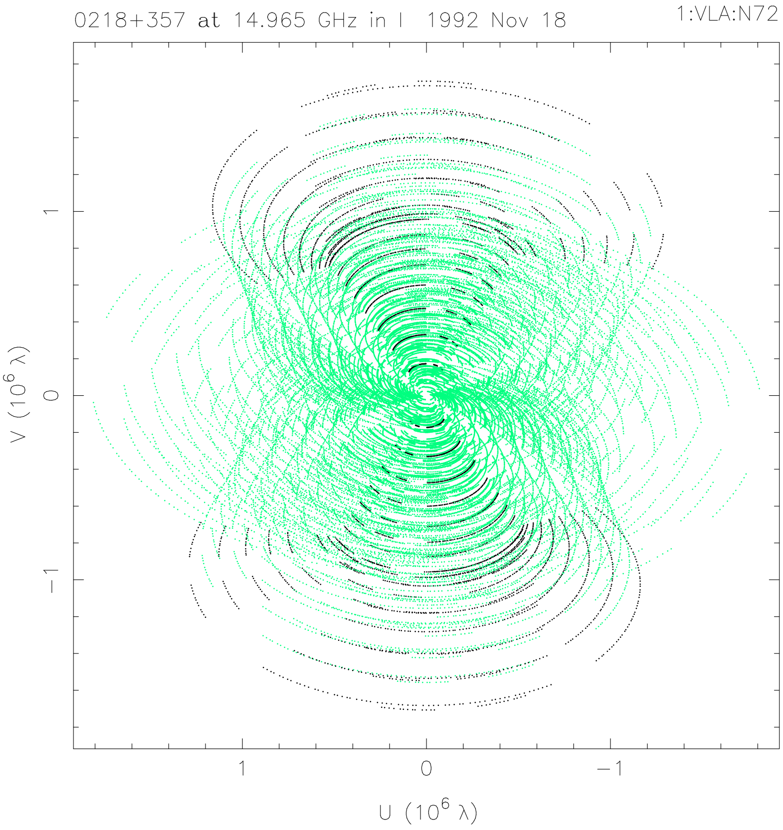


Figure 8.1: UV coverage of the 15 GHz VLA observations of B0218+357, which will be use with LensClean later. The tracks are due to the rotation of the earth during the observations, which continuously changes the antenna positions relative to the source. The baselines involving the telescope N72 are highlighted in the colour version.

8.5 Noise statistics

The equations presented in the last sections are valid only in absence of calibration errors and noise. We want to fit models to the data using a maximum likelihood approach later. To allow this, we have to understand the noise and its statistical properties. The so-called thermal noise in radio data is a combination of contaminating emission (from the ground, the telescopes themselves, and the atmosphere) and noise originating from the receivers and amplifiers. All these contributions have in common that they act on the signals from the telescopes directly and not on the correlated baseline signals. We do not take into account errors caused by the correlator itself, which are usually negligible compared to thermal noise.

Let us assume a monochromatic thermal noise contribution to E_j described by the complex quantity n_j with a vanishing mean and a variance of σ_j^2. Since the noise is incoherent and uncorrelated with the signal, we can study the noise independently from the signal here (cf. section 8.3). The noise contribution to the visibility $\tilde{I}_{jk}$ of a baseline j,k is then

$$n_{jk} = \mathrm{e}^{-2\pi \mathrm{i}\,(w_j - w_k)}\, n_j n_k^\star \quad . \tag{8.12}$$

The phase factor can be absorbed in the station based noise without changing its statistical properties. We therefore do not include it in the following equations.

We see immediately that the noise in the visibilities of different baselines is not independent. For N telescopes with $N(N-1)/2$ visibilities, we have only N independent noise signals. Nevertheless, it can be shown that the correlations of the noise of different baselines vanish, even for monochromatic noise. This is not true for higher order correlations, e.g. the correlation of the magnitude (absolute value) of the noise.

Real noise consists of a wide spectrum of independent monochromatic contributions. Any higher order correlations remain only for noise contributions of the same frequency and are thus negligible for typical bandwidths $\Delta\nu$ of a few MHz and integration intervals ΔT of the order a second. The correlations are suppressed by factors of $1/(\Delta T\,\Delta\nu) \approx 10^{-7}$.

We can therefore treat the noise of different baselines as independent and, using the formalism from chapter 2, write the residuals R^2 for a model $\tilde{I}^{\mathrm{m}}$ as

$$R^2 = \sum_j w_j\, |\tilde{I}_j - \tilde{I}_j^{\mathrm{m}}|^2 \quad , \tag{8.13}$$

where the sum is taken over all baselines and w_j are the (diagonal) weights of the visibilities. We do not need the w component of the uv coordinates anymore and can therefore use the symbol for the weights. For ‘natural weighting’ we use $w_j = \sigma_j^{-2}$ to obtain a normal χ^2.

To calculate the statistics correctly, σ^2 has to be the variance of the real or imaginary part which is half the total variance $\sigma^2 = \left\langle |n|^2 \right\rangle/2$. The number of independent contributions to the sum is twice the number of visibilities. Another subtlety is related to the symmetry of visibilities. A measurement with a baseline of telescopes j and k can either be included as baseline $j-k$ or as $k-j$. The two visibilities are complex conjugates of each other, see equation (B.13). The most elegant approach to take this into account is to use *both* possible measurement but divide the corresponding weights by 2. This does not change the residuals but assures that the dirty beams and dirty maps, defined in the next section, will always be real quantities. A lot of trouble with otherwise necessary selection of real parts can be saved with this approach, which is therefore used in all following calculations. The symmetric completion doubles the number of visibilities which is then equal to the number of real constraints. This modification of the weights cancels with the one discussed before, so that $w_j = \sigma^{-2} = 1/\left\langle |n|^2 \right\rangle$ can be used for natural weighting. This formalism can also be applied for visibilities with zero spacing. It then automatically corrects for non-vanishing imaginary parts which would not be compatible with any brightness distribution.

8.6 Dirty beam and dirty map

We remember from section 8.4, that the real brightness distribution is the inverse Fourier transform of the visibilities. The only problem with this is the incomplete uv coverage. A crude approximation to the real brightness distribution can be calculated by using only the known visibilities. In this way, the integral of the transformation becomes a discrete sum over all the discrete visibilities. This approximation resembles the true brightness distribution only very crudely. It is thus called the 'dirty map':

$$I_{\mathrm{D}}(\boldsymbol{x}) = \frac{1}{\sum_j w_j} \sum_j w_j \tilde{I}_j \, \mathrm{e}^{-2\pi \mathrm{i} \mathbf{u}_j \cdot \boldsymbol{x}} \tag{8.14}$$

Here we used the notation $\tilde{I}_j = \tilde{I}(\mathbf{u}_j)$, $\boldsymbol{x} = (l, m)$ and $\mathbf{u}_j = (u_j, v_j)$. The weights w_j assigned to the visibilities are the same as in the definition of the residuals R^2 in equation (8.13). We will see later that this leads to a very useful formalism.

Another important concept is that of the 'dirty beam' which is defined to be the dirty map of an ideal point source of unit flux density located at $\boldsymbol{x} = 0$. The visibilities for this would be $\tilde{I}_j \equiv 1$, leading to a dirty beam of

$$B(\boldsymbol{x}) = \frac{1}{\sum_j w_j} \sum_j w_j \, \mathrm{e}^{-2\pi \mathrm{i} \mathbf{u}_j \cdot \boldsymbol{x}} \quad . \tag{8.15}$$

The normalization with the sum of weights is conventionally applied to obtain a unity peak of the dirty beam. In this way the dirty map directly shows an approximation of point source flux densities in real units.

Without the symmetric completion of visibilities (previous section), we would have to take the real part in the definitions of dirty map and beam. An example of dirty beams and maps together with CLEAN maps using 15 GHz VLA data of B0218+357 is shown in Figure 8.2.

8.6.1 Noise in the dirty map

If the visibilities consist of uncorrelated noise only, with vanishing mean and variances of σ^2, the mean of the dirty map will also vanish. Its variance will be

$$\mathrm{var}(I_{\mathrm{D}}) = \frac{\sum_j w_j^2 \left\langle \left| \tilde{I}_j \, \mathrm{e}^{-2\pi \mathrm{i} \mathbf{u}_j \cdot \boldsymbol{x}} \right|^2 \right\rangle}{\left(\sum_j w_j \right)^2} \tag{8.16}$$

$$= \frac{\sum_j w_j^2 \sigma_j^2}{\left(\sum_j w_j \right)^2} \quad . \tag{8.17}$$

The square root of this is the RMS noise. It can be shown easily that the noise is minimal for natural weighting with $w_j \propto \sigma_j^{-2}$. The variance then becomes

$$\mathrm{var}(I_{\mathrm{D}}) = \frac{1}{\sum_j \sigma_j^{-2}} \quad . \tag{8.18}$$

Since the number of visibilities scales proportionally with the observation time but approximately quadratically with the number of telescopes, it is more effective to use more telescopes for a shorter time than less telescopes for a longer time, if the cost of the correlator is not taken into account.

To interpret the maps, the correlation of the noise must also be analysed. This will be done with a simple matrix formalism in section 8.8.1.

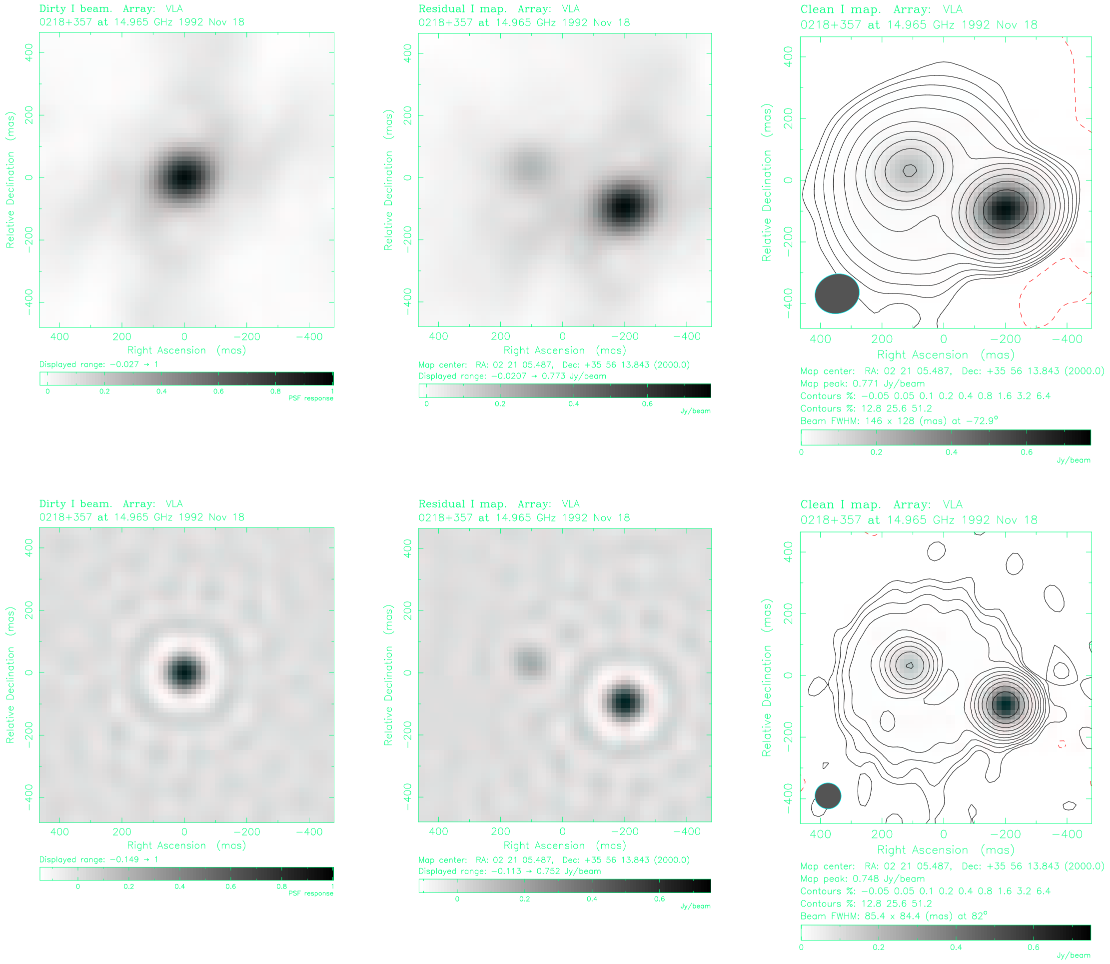


Figure 8.2: From left to right: Dirty beam, dirty map and a very crude CLEAN map of B0218+357 (VLA 15 GHz). The upper row is with natural, the lower one with uniform weighting (see section 8.7).

8.6.2 Shape and size of the dirty beam

The resolution of the observations is determined by the central part of the dirty beam (see section 8.11.2). The centre of B always is stationary ($\boldsymbol{\nabla} B = 0$), the second derivatives (the Hessian matrix of B) can be calculated easily using equation (8.15).

$$\frac{\partial^2 B}{\partial \mathbf{x}^2} = -4\pi^2 \frac{1}{\sum_j w_j} \sum_j w_j \begin{pmatrix} u_j^2 & u_j v_j \\ u_j v_j & v_j^2 \end{pmatrix} \mathrm{e}^{-2\pi \mathrm{i}\, \mathbf{u}_j \cdot \mathbf{x}} \tag{8.19}$$

In the centre, it reads

$$\left. \frac{\partial^2 B}{\partial \mathbf{x}^2} \right|_{\mathbf{u}=0} = -4\pi^2 \begin{pmatrix} \langle u_j^2 \rangle & \langle u_j v_j \rangle \\ \langle u_j v_j \rangle & \langle v_j^2 \rangle \end{pmatrix} \quad , \tag{8.20}$$

where the mean values are calculated with the weights w_j. We see that the central part of the beam is related to the distribution of visibilities in a very simple way. As expected, the beam becomes more compact and centrally concentrated, when more high-frequency visibilities (high $|\mathbf{u}|$) are present or when the weights of these visibilities are increased.

We will later need a 'CLEAN beam' with the same properties of the centre as the dirty beam. Conventionally, a Gaussian with a central peak of unity is used:

$$G(\mathbf{x}) = \mathrm{e}^{-\mathbf{x}^\dagger \mathsf{G} \mathbf{x}/2} \tag{8.21}$$

The shape of this beam is described by the symmetric and positive definite matrix G. To have the same second derivatives at the centre as the dirty beam, the coefficients have to be

$$G_{xx} = 4\pi^2 \left\langle u_j^2 \right\rangle \quad , \tag{8.22}$$

$$G_{yy} = 4\pi^2 \left\langle v_j^2 \right\rangle \quad , \tag{8.23}$$

$$G_{xy} = 4\pi^2 \left\langle u_j\, v_j \right\rangle \quad . \tag{8.24}$$

The major and minor axes can be calculated using the eigenvalues λ of G,

$$\lambda_{\mathrm{max/min}} = \frac{G_{xx} + G_{yy} \pm \sqrt{(G_{xx} - G_{yy})^2 + 4\, G_{xy}^2}}{2} \quad , \tag{8.25}$$

the area of the ellipse is proportional to $|\mathsf{G}|^{-1/2}$. The FWHM[3] are then

$$\mathrm{FWHM} = \sqrt{\frac{8 \ln 2}{\lambda}} \quad . \tag{8.26}$$

For radially symmetric beams ($< u_j^2 > = < v_j^2 > = < |\mathbf{u}|^2 > /2$ and $< u_j\, v_j > = 0$), this becomes

$$\mathrm{FWHM} = \frac{1}{\pi} \sqrt{\frac{4 \ln 2}{\left\langle |\mathbf{u}|^2 \right\rangle}} \quad . \tag{8.27}$$

The DIFMAP software uses an approximation of the factor and thus underestimates beam sizes by about 7 %.

The position angle θ of the major axis, measured in a mathematical sense from positive x to positive y, can be calculated using

$$\sin 2\theta \propto -2\, G_{xy} \quad , \tag{8.28}$$

$$\cos 2\theta \propto -(G_{xx} - G_{yy}) \quad . \tag{8.29}$$

The astronomical position angle (north through east) is p.a. $= \pi/2 - \theta$.

8.7 Weighting schemes

The most natural weights from a statistical point of view are $w_j = \sigma_j^{-2}$. With this choice, the noise is minimized (see section 8.6.1). Another important weighting scheme is the so-called 'uniform weighting', in which the weights are normalized with the local density of visibilities in the *uv* plane. In this way, a constant weight density is achieved in the *uv* regions where measurements are available. This usually increases the weights in the outer parts relative to the central parts and thus increases the resolution. The price paid for this is an increased noise. There are different possibilities to include the real statistical weights σ_j^{-2} in the calculation of the uniform weights. Uniform weighting also has some nice analytical properties because the dirty

---

[3] 'Full Width Half Maximum', the width at a level of half the peak.

beam becomes a projection operator. The dirty map and beam are invariant under convolution with the dirty beam then.

An intermediate approach called 'robust weighting' is often used in practical work. This is very similar to uniform weighting, but the weights are not divided by the density of visibilities itself but by the density plus some constant. In this way, one avoids to assign very high weights to very noisy visibilities, only because the local density is low.

8.8 The deconvolution problem

8.8.1 Formalism

In Fourier space, the dirty map is the product of the dirty beam (the weights) and the real brightness distribution (the visibilities). This becomes a convolution in image space:

$$I_{\mathrm{D}}(\boldsymbol{x}) = (I * B)(\boldsymbol{x}) \tag{8.30}$$

In general, the solution of the deconvolution problem is not unique. It is always possible to add a brightness distribution with Fourier transform vanishing at the positions of measured visibilities without changing the dirty map ('invisible solutions'). Uniqueness can be achieved by regularization or by confining the flux to a limited region of the map.

It is enlightening to write the deconvolution problem in terms of the general linear model fitting of chapter 2, especially section 2.2. For convenience, we write the model brightness distribution $I(\boldsymbol{x})$ as a vector $\mathbf{I}$,

$$I(\boldsymbol{x}) = \sum_{\nu} I_{\nu} \delta(\boldsymbol{x} - \boldsymbol{x}_{\nu}) \quad , \tag{8.31}$$

where we allow emission at the M positions $\boldsymbol{x}_{\nu}$.[4] Other shapes than δ functions may be used for the components, leading to interesting alternative deconvolution methods. A continuous model is included in this approach as a limiting case. The model visibilities for a given $\mathbf{I}$ can be calculated with (8.10), which in the new formalism reads

$$\tilde{I}_j^{\mathrm{m}} = \sum_{\nu} A_{j\nu} I_{\nu} \tag{8.32}$$

with the Fourier transform matrix

$$A_{j\nu} = \mathrm{e}^{2\pi \mathrm{i}\, \mathbf{u}_j \cdot \boldsymbol{x}_{\nu}} \quad . \tag{8.33}$$

The number of real measured quantities N is the same as that of the symmetrically completed complex visibilities. We can now use the formalism of chapter 2 with a better understanding of what the terms in the equations actually mean.

$$\mathsf{B} = \frac{1}{\mathrm{Tr}\,\mathsf{W}} \mathsf{A}^{\dagger} \mathsf{W} \mathsf{A} \tag{8.34}$$

$$B_{\mu\nu} = B(\boldsymbol{x}_{\mu} - \boldsymbol{x}_{\nu}) \tag{8.35}$$

$$\mathbf{I}_{\mathrm{D}} = \frac{1}{\mathrm{Tr}\,\mathsf{W}} \mathsf{A}^{\dagger} \mathsf{W} \tilde{\mathbf{I}} \tag{8.36}$$

$$(I_{\mathrm{D}})_{\nu} = I_{\mathrm{D}}(\boldsymbol{x}_{\nu}) \tag{8.37}$$

---

[4] We use greek indices on the model side and latin indices on the measurement side.

The normalization factor is written as trace of the matrix W now. For the diagonal errors in radio interferometry,

$$\sum_j w_j = \operatorname{Tr} \mathsf{W} \tag{8.38}$$

holds. In one dimension with a regular grid, the dirty beam matrix B has the structure of a Toeplitz matrix $B_{\mu\nu} = b_{\mu-\nu}$, which can help in inverting the matrix more efficiently. Even in two dimensions, a regular grid helps in keeping the number of independent components of B to a minimum.

description general	description here	chapter 2	here	dimensions
model vector	brightness distrib.	$\mathfrak{x}$	$\mathbf{I}$	M
obs. vector	visibilities	$\mathfrak{y}$	$\tilde{\mathbf{I}}$	N
model matrix	Fourier transform	A	A	$N \times M$
weight matrix	weights	W	W	$N \times N$
	dirty beam	$\propto \mathsf{A}^\dagger\mathsf{W}\mathsf{A}$	B	$M \times M$
	dirty map	$\propto \mathsf{A}^\dagger\mathsf{W}\mathfrak{y}$	$\mathbf{I}_\mathrm{D}$	M

Table 8.1: Comparison of the general notation (chapter 2) and the one used here. N is the number of symmetrically completed visibilities and thus equal to the number of real measurements.

An overview of the important quantities in terms of the general considerations and of the radio deconvolution problem can be found in Table 8.1. It is an interesting fact, that the varying part of the residuals can be written completely in terms of the image plane:

$$\begin{aligned} R^2 &= (\tilde{\mathbf{I}} - \mathsf{A}\mathbf{I})^\dagger \mathsf{W} (\tilde{\mathbf{I}} - \mathsf{A}\mathbf{I}) && (8.39)\\ &= \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} + \mathbf{I}^\dagger \mathsf{A}^\dagger \mathsf{W}\mathsf{A}\mathbf{I} - 2\mathbf{I}^\dagger \mathsf{A}^\dagger \mathsf{W} \tilde{\mathbf{I}} && (8.40)\\ &= \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} + \operatorname{Tr} \mathsf{W} \left(\mathbf{I}^\dagger \mathsf{B} \mathbf{I} - 2\mathbf{I}^\dagger \mathbf{I}_\mathrm{D}\right) && (8.41)\end{aligned}$$

The first term in (8.41) only depends on the measurements and not on the model. It can therefore be neglected in the fitting process. The simple form of the residuals written in terms of quantities in image space suggests that suitable fitting algorithms might work using the image plane alone, without referring to the visibilities directly. An example of this is CLEAN (see section 8.8.3). To have a minimum of R^2, the equation

$$\mathsf{B}\mathbf{I} = \mathbf{I}_\mathrm{D} \tag{8.42}$$

has to be fulfilled. In the case of a regular grid, this is equivalent to the convolution equation (8.30). If $|\mathsf{B}| \neq 0$, the beam matrix can be inverted and the equation has a unique solution. Obviously, the rank of B cannot be larger than M or N. A unique solution is therefore never possible if $M > N$ (more model components than visibilities). To define a unique solution, regularization as described in section 2.3 can be used. Linear regularization with a matrix S is equivalent to adding $\lambda\mathsf{S}$ to the dirty beam B. The factor λ is now scaled by $1/\operatorname{Tr}\mathsf{W}$ with respect to (2.17).

Several possible methods to solve the deconvolution problem will be discussed in the following sections. The most important algorithm for practical work is still the classical CLEAN method. More direct algebraic methods like NNLS have gained interest recently, because the computing power available to astronomers now allows the application of these methods to realistic problems. In this thesis, we are mainly interested to find *one* solution of the problem (see chapter 9) and determine the residuals. Since the problem of finding *the best* solution is so important in radio interferometry in general, we want to present some thoughts about the problem here. This is also important in the reconstruction of the true source brightness distribution, which we want to perform later (see section 9.24).

Finally we want to use the matrix formalism to calculate the noise correlation properties of the dirty map. If the dirty map consists of noise only, the autocorrelation function of I_D can be calculated directly from the properties of $\mathbf{n}$.

$$\left\langle \mathbf{I}_\mathrm{D} {\mathbf{I}_\mathrm{D}}^\dagger \right\rangle = \frac{1}{(\mathrm{Tr}\,\mathsf{W})^2} \mathsf{A}^\dagger \mathsf{W} \left\langle \mathbf{n}\mathbf{n}^\dagger \right\rangle \mathsf{W}\mathsf{A} \tag{8.43}$$

In the case of natural weighting,

$$\left\langle \mathbf{n}\mathbf{n}^\dagger \right\rangle = \mathsf{W}^{-1} \tag{8.44}$$

holds, and the autocorrelation function can be expressed by the dirty beam and the variance of I_D taken from equation 8.18:

$$\left\langle \mathbf{I}_\mathrm{D} {\mathbf{I}_\mathrm{D}}^\dagger \right\rangle = \frac{1}{(\mathrm{Tr}\,\mathsf{W})^2} \mathsf{A}^\dagger \mathsf{W}\mathsf{A} \tag{8.45}$$

$$= \mathrm{var}(I_\mathrm{D})\,\mathsf{B} \tag{8.46}$$

For arbitrary weighting, the autocorrelation function is a dirty beam calculated with modified weights of $w_j' = w_j^2\,\sigma_j^2$. Note that the autocorrelation function of the dirty map is not the same as that of the final deconvolved map. The latter is much more difficult to analyse, especially for non-linear algorithms like CLEAN.

8.8.2 (Non-)Equivalence of image space and *uv* space residuals

For the interpretation of fits, it is important to understand the difference of image space and *uv* space residuals. If the model is already subtracted from the measured visibilities $\tilde{I}$ and the dirty map I_D, the two kinds of residuals are defined as

$$R_{uv}^2 = \sum_j w_j |\tilde{I}_j|^2 \quad , \tag{8.47}$$

$$R_\mathrm{image}^2 = \sum_\nu |I_{\mathrm{D}\nu}|^2 \quad . \tag{8.48}$$

For a first comparison, we consider the gridded visibilities and the corresponding periodic image (see appendix B.2). The gridding intervals are Δx and $\Delta u = 1/(n\Delta x)$ and the grid points are at $x_\nu = \nu\,\Delta x$ and $u_j = j\,\Delta u$. We use a one-dimensional grid for simplicity.

$$I_{\mathrm{D}\nu} = \frac{1}{\sum_j w_j} \sum_j w_j\,\tilde{I}_j\,\mathrm{e}^{-2\pi\mathrm{i}\,\nu j/n} \tag{8.49}$$

This expression can be used to find a relation between the two kinds of residuals:

$$R_\mathrm{image}^2 = \frac{1}{\left(\sum_j w_j\right)^2} \sum_{jj'} w_j w_{j'}\,\tilde{I}_j \tilde{I}_{j'}^\star \sum_\nu \mathrm{e}^{2\pi\mathrm{i}\,\nu(j-j')/n} \tag{8.50}$$

$$= \frac{n}{\left(\sum_j w_j\right)^2} \sum_j w_j^2\,|\tilde{I}_j|^2 \tag{8.51}$$

In the last step, the relation

$$\sum_k \mathrm{e}^{2\pi\mathrm{i}\,\nu(j-j')/n} = n\,\delta_{jj'} \tag{8.52}$$

was used. All sums are taken over a range from 0 to $n-1$. We learn that the residuals are exactly proportional only if $w_j = w_j^2$ for all j. Because the w_j represent the gridded weights, this is equivalent to uniform weighting. One has to keep in mind, however, that the *uv* space residuals in this case are not exactly the usual χ^2, because they are summed over the gridded visibilities and not the measured visibilities themselves. For an optimal fit, they therefore become zero. This would not be the case for the measured ungridded visibilities, because the gridded emission can usually not represent all ungridded visibilities. This is the reason why χ^2 does not approach zero in normal CLEAN but becomes close to unity if normalized.

For general weighting, the above formalism is not very meaningful, because the gridded weights are related to the ungridded weights in a complicated way. In the following, we want to repeat the same calculations without referring to a grid. We start with the dirty map from (8.14) and apply it again to the difference between data and model. We use a one-dimensional formalism which can be generalized trivially for two dimensions. We assume that the dirty map is sampled infinitely fine in an interval $L/2 \le x \le L/2$:

$$R^2_{\text{image}} = \frac{1}{\left(\sum_j w_j\right)^2} \sum_{jj'} w_j w_{j'}\,\tilde{I}_j\,\tilde{I}^\star_{j'} \int\limits_{-L/2}^{L/2} \mathrm{d}x\,\mathrm{e}^{2\pi\mathrm{i}\,(u_j-u_{j'})x} \tag{8.53}$$

$$= \frac{L}{\left(\sum_j w_j\right)^2} \sum_{jj'} w_j w_{j'}\,\tilde{I}_j\,\tilde{I}^\star_{j'}\,\operatorname{sinc} L(u_j-u_{j'}) \tag{8.54}$$

The sinc function is defined in equation (B.25) in appendix B. In the limit of very large L (infinitely large field), only equal u_j contribute to the sum:

$$R^2_{\text{image}} = \frac{L}{\left(\sum_j w_j\right)^2} \sum_j w_j^2\,|\tilde{I}_j|^2 \tag{8.55}$$

Compared with the *uv* space residuals in (8.47), we see again that the weights are modified. Image space residuals weighted with w_j are equivalent to *uv* space residuals weighted with w_j^2. This may be the reason, why weighting with $w_j = \sigma_j^{-1}$ instead of σ_j^{-2} often leads to more pleasing results in image space. The image space residuals are then equivalent to naturally, and thus in a way optimally, weighted *uv* space residuals.

For finite L, the convolution of the visibilities with the sinc function corresponds to an optimal gridding convolution with total alias suppression. The image space residuals are equivalent to the *uv* space residuals *after convolution* and with modified weights. For small fields, even optimal image space solutions cannot represent the visibilities exactly. The *uv* space residuals are therefore always non-vanishing due to the small-scale scatter of the measured visibilities. The same effect can be seen in a reduced χ^2 of approximately unity even if the image space residuals are almost zero. The result of equivalence of both kinds of residuals in the case of uniform weighting can also be recovered from this very general formalism.

8.8.3 The CLEAN algorithm

The CLEAN algorithm was invented as a heuristic method to deconvolve dirty maps (Högbom, 1974). The simple idea is very obvious when one is looking at dirty maps and beams, e.g. Figure 8.2 on page 99. The dirty maps of B0218+357 are dominated by two scaled copies of the dirty beam, shifted to the positions of the bright images of this source. If the system consisted only of these two images, nothing but noise should be left when the beams are subtracted. Otherwise, the remaining structure should become more distinct. If more compact components are visible in the residual map, then further dirty beams can be subtracted at their positions. This leads directly to the CLEAN method, which works as follows.

- Start with a residual map equal to the dirty map, with an empty list of CLEAN components, and iterate the following loop until the residuals become sufficiently small.
  - Find the peak in the residual map.
  - Shift the dirty beam to the peak position and scale it to match the flux density of the peak.
  - Subtract the scaled and shifted beam from the dirty map.
  - Add the position and flux density to the list of CLEAN components.
- Build the CLEAN map from the list of CLEAN components by adding the so-called CLEAN beam at the given positions with the given flux densities. This is equivalent to a convolution with the CLEAN beam.
- Add the final residual map to the CLEAN map, if the residuals are still not negligible.

The convolution with the CLEAN beam is performed to reduce the high frequencies in the map, which cannot be reconstructed from the observations reliably. The canonical CLEAN beam is a Gaussian whose central part is fitted to the dirty beam to resemble its resolution (see section 8.6.2). Ideally, the resulting CLEAN map is then a convolution of the true brightness distribution with the CLEAN beam.

In practical work, the dirty beams to subtract are not scaled to 100 % of the peak flux density but to a fraction γ, which is called the loop gain and is of the order 0.1 or smaller. This helps to avoid ringlike residuals around bright compact images and generally leads to more realistic results.

If the algorithm converges, which it always does for non-pathological beams, the resulting clean map before the final convolution is obviously one solution of the deconvolution problem. It is optimized in a way to produce a solution which is a collection of a moderate number of point sources. No rigorous mathematical formulation of this is known to date.

8.8.4 Mathematical foundation of CLEAN

If we want to modify classical CLEAN as a tool for lens modelling, we first have to understand its mathematical foundation, which was laid by Schwarz (1978). For an algorithm of successively subtracted dirty beams, we have to give reasons to choose the position and flux density of the peak in the residual map for the next component to subtract. The idea is to minimize the residuals in each iteration.

Let the residual dirty map before the next subtraction be given by $I_{\mathrm{D}}(\boldsymbol{x})$, the residual visibilities by $\tilde{I}$ and use the same notation as in section 8.6. The residuals after subtracting a component at $\boldsymbol{x}_0$ with flux density S are given by

$$R^2 = \sum_j w_j \left| \tilde{I}_j - S\,\mathrm{e}^{2\pi\mathrm{i}\,\mathbf{u}_j\cdot\boldsymbol{x}_0} \right|^2 \quad . \tag{8.56}$$

To minimize these residuals, we first rewrite them using the definition of the dirty map (8.14):

$$R^2 = R_0^2 - \sum w_j \, \Delta R^2 \tag{8.57}$$

$$R_0^2 = \sum_j w_j \, |\tilde{I}_j|^2 \tag{8.58}$$

$$\Delta R^2 = 2S I_{\mathrm{D}}(\mathbf{x}_0) - S^2 \tag{8.59}$$

The term R_0^2 does not depend on $\mathbf{x}_0$ or S and can therefore be neglected here. For a given position $\mathbf{x}_0$, the residuals are minimized for maximal ΔR^2, which leads to

$$S = I_{\mathrm{D}}(\mathbf{x}_0) \tag{8.60}$$

and

$$\Delta R^2 = I_{\mathrm{D}}^2(\mathbf{x}_0) \quad . \tag{8.61}$$

We learn that the optimal position is the highest absolute peak in the map and that the optimal flux density is the residual flux at this peak. This is exactly the rule used by CLEAN. If we take into account the loop gain γ in selecting the optimal position $\mathbf{x}_0$, the result would still be the same. One might also think of adding or subtracting (depending on the sign of the residuals) a constant but very small flux density. The result for $\mathbf{x}_0$ would still be the same. The best maps are usually produced by choosing a very small loop gain and going to the limit of a continuous process.

The considerations in this section used continuous functions for the maps instead of applying the formalism developed in section 8.8.1. This is not necessary and was only done to make the modifications that are later needed for LENSCLEAN (chapter 9) more straightforward. In the algebraic framework, CLEAN can be interpreted as follows. In the first iterations, only the one component of **I** with maximal gradient of the residuals is changed in the direction towards minimal residuals. At some point, another component reduces the residuals more effectively and is therefore changed next. For very small loop gain, the first component will then again be more effective, so that the two are alternately changed. This is effectively equivalent to changing the two in the direction of strongest descent of the residuals. At some point, a third component comes into play, and so on. All in all, CLEAN always minimizes in the direction of the constrained gradient for n components until the next component becomes more effective. It then continues with $n+1$ components and so on. The formal calculations in the algebraic formalism are presented in the context of LENSCLEAN in section 9.23.1.

The flux may be confined to only a limited region in the lm plane or to the points of a grid (or both). The strategy does not change then. Confining the flux to a limited number of positions can be used to obtain a unique solution of the deconvolution problem and usually gives better results when a priori knowledge of where emission is expected exists.

Apart from LENSCLEAN, an analysis of the CLEAN algorithm is very important to be able to interpret its results. To the knowledge of the author, no general mathematical definition of the optimization problem solved by CLEAN has been published yet[5]. It is clear that the CLEAN result before the final convolution is *one* solution of the convolution equation (8.30). *Which* of these solution is produced by CLEAN and especially the error properties are not fully understood but an analysis is greatly desired. See section 8.9 for numerical examples of CLEAN and other methods.

---

[5] See Marsh & Richardson (1987) for one step in this direction.

8.8.5 The principal solution

The definition of the principal solution was already presented in 2.3. It is the one solution that minimizes $\sum I_\mu^2$ and can be calculated by using equation (2.17) with $\mathsf{S} = \mathsf{1}$ and small λ. The convolution equation (8.42) is then modified to become

$$(\mathsf{B} + \lambda\,\mathsf{S})\,\mathbf{I} = \mathbf{I}_\mathrm{D} \quad . \tag{8.62}$$

For positive definite S, it is invertible uniquely. The result for $\mathbf{I}$ is a linear function of $\mathbf{I}_\mathrm{D}$ or the measured visibilities, making the analysis of this solution simpler than for CLEAN.

The SSC method (smoothness stabilized CLEAN), proposed by Cornwell (1983), is closely connected to this approach. A CLEAN algorithm is used to solve (8.62) for finite λ. The only difference to standard CLEAN is the modified dirty beam B. The addition of a scaled unit matrix is equivalent to adding a δ-peak to the dirty beam $B(\boldsymbol{x})$. However, the idea behind SSC is *not* to find the exact principal solution, but to switch from the normal dirty beam to the modified one with certain values of λ to stabilize the results. Due to Cornwell (1983), real convergence should not be aimed for.

8.8.6 Strong linear regularization

The principal solution can, in some cases, be very sensitive to noise in the visibilities. To reduce the effects of noise, we may use the same equation (8.62) but now with finite λ. This parameter should be chosen in a way to obtain *uv* residuals in the expected range. In this way, an optimal solution in the sense of minimizing noise and systematic errors to the same level is found. For large λ, significant bias effects will occur. The resulting $\mathbf{I}$ will have systematically too low fluxes.

One possible way to reduce this bias is to modify the method and find the solution with minimal difference to an a priori expected solution $\mathbf{I}_\mathrm{p}$:

$$R'^2 = R^2 + \lambda\,(\mathbf{I} - \mathbf{I}_\mathrm{p})^\dagger(\mathbf{I} - \mathbf{I}_\mathrm{p})\ \mathrm{Tr}\,\mathsf{W} \tag{8.63}$$

The approach used before is equivalent to a prior of $\mathbf{I}_\mathrm{p} = 0$. Minimizing the modified R'^2 leads to the equation

$$(\mathsf{B} + \lambda\mathsf{1})\mathbf{I} = \mathbf{I}_\mathrm{D} + \lambda\,\mathbf{I}_\mathrm{p} \tag{8.64}$$

or equivalently

$$(\mathsf{B} + \lambda\mathsf{1})(\mathbf{I} - \mathbf{I}_\mathrm{p}) = \mathbf{I}_\mathrm{D} - \mathsf{B}\mathbf{I}_\mathrm{p} \quad . \tag{8.65}$$

This last equation gives a simple recipe to find the wanted solution. First subtract the prior convolved with the dirty beam from the dirty map, then find the strongly regularized solution for this modified dirty beam, finally add the prior back to the solution. The result depends on $\mathbf{I}_\mathrm{p}$ only for finite values of λ. The modification does not make a difference for weak regularization.

8.8.7 The NNLS solution

The non-negative-least-squares method (NNLS) was for the first time used in radio interferometry by Briggs (1995). The idea is to regularize the problem by using a the physically motivated constraint of non-negative fluxes. The NNLS solution is a non-negative **I** with minimal residuals R^2 as in equation (8.41).

$$R^2 \propto \text{const.} + \mathbf{I}^\dagger \mathsf{B} \mathbf{I} - 2\mathbf{I}^\dagger \mathbf{I}_\mathrm{D} \tag{8.66}$$

This is *not* the same as minimizing the image space residuals

$$R'^2 = |\mathsf{B}\mathbf{I} - \mathbf{I}_\mathrm{D}|^2 \tag{8.67}$$

$$= \text{const.} + \mathbf{I}^\dagger \mathsf{B}^2 \mathbf{I} - 2\mathbf{I}^\dagger \mathsf{B} \mathbf{I}_\mathrm{D} \quad , \tag{8.68}$$

which was used by Briggs (1995). They are equivalent only for uniform weighting, where $\mathsf{B}^2 = \mathsf{B}$, and in cases where an exact fit can be found. See section 8.8.2 for a discussion of the difference of residuals in image and *uv* space.

Further regularization as for the unconstrained direct inversion is possible by adding a $\lambda\mathsf{S}$-term to B. This is less necessary than in the unconstrained inversion, because the non-negativity constraint already reduces the freedom of the models.

8.9 Numerical experiments in one dimension

To compare some deconvolution methods, numerical experiments were performed. The data used for these experiments are shown in Figure 8.3. A one-dimensional setting was used for two reasons. First, the computations are much faster in one dimension. No sophisticated optimization was needed for the experiments. Direct algebraic inversion of the convolution equation is possible on normal PCs here. Second, the results can be interpreted more easily in one dimension, because the data can be shown as curve plots.

As true brightness distribution we used a simple rectangular function which is a quite hard test for all of the methods. It tests the behaviour at sharp edges, constant surface brightness regions and empty regions around the sources. Other brightness distributions lead to different results.

We used a very fine sampling of the image plane with $M = 100$ for all the methods to come close to the continuous limit. A *uv* coverage with $N = 10$ as shown in Figure 8.3(d) was used. To simulate a realistic *uv*-coverage, a gap near zero spacings was included, and the visibility density is higher near the centre than for larger spacings. To show the effect of thermal noise, we also included deconvolution results for noisy data which are also shown in Figure 8.3(c)+(d). The expected *uv*-residuals from the noise properties are $\chi^2 = 20 \pm 6.3$. Calculated from the actual noise realization, they are $\chi^2 = 15.2$. The expected RMS noise in the dirty map is 0.63, the real noise 0.58.

The results of the deconvolutions are shown in Figures 8.4 to 8.11 (pages 113–120). Here is the extended caption for these plots:

(a) Reconstructed map (as is), noiseless and noisy data compared with true brightness distribution.
(b) Same as (a) but convolved with CLEAN beam.
(c) Difference of reconstructed and true map. Noiseless and noisy data as well as difference between the two to show the effect of noise.
(d) Same as (c) but convolved with CLEAN beam.
(e) Fourier transform of reconstructed map compared with true Fourier transform. For an exact fit (converged CLEAN or no strong regularization), these should be equal.

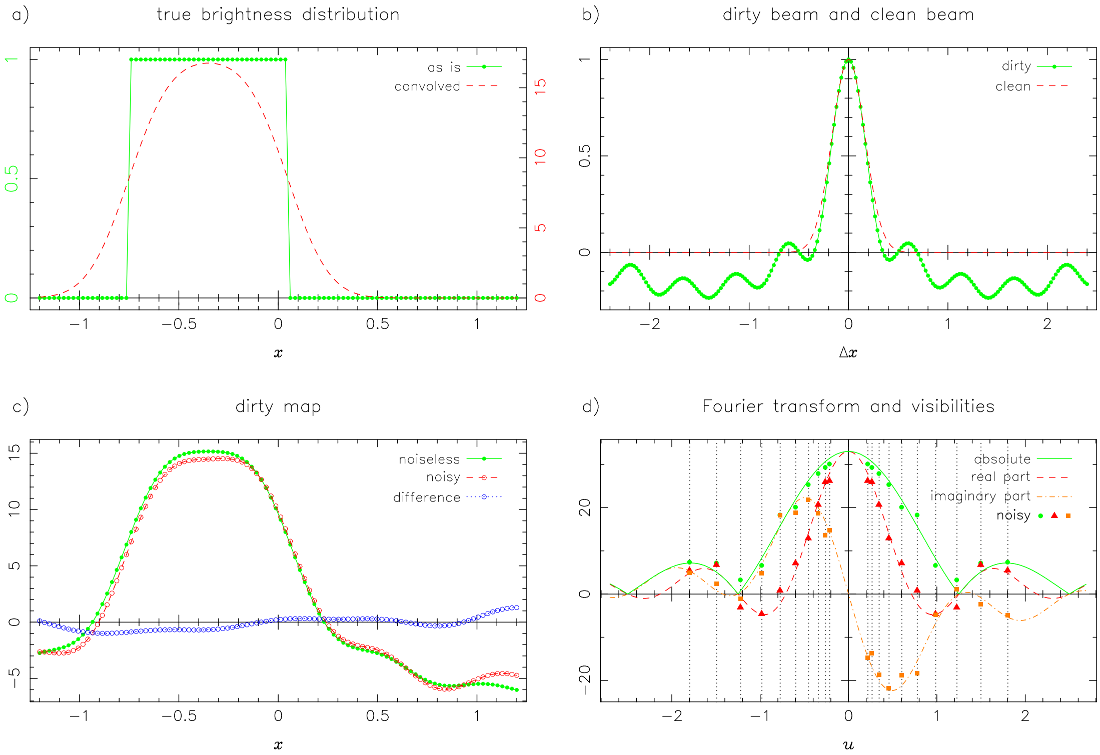


Figure 8.3: Data for one-dimensional experiments. (a) True brightness distribution ('map'), as is and convolved with CLEAN beam. (b) Dirty beam and centrally fitted CLEAN beam. (c) Dirty map from noiseless and noisy data and difference of the two. (d) Fourier transform of the brightness distribution (lines) and the measured visibilities with noise (symbols). The vertical dotted lines show the *u*-positions of the measured visibilities.

(f) Residuals in image space $I_{\mathrm{D}} - B * I^{\mathrm{m}}$. These vanish for an exact fit.

(g) Fourier transform of difference between reconstructed and true map (noiseless case). This should vanish at the measured *uv*-spacings for an exact fit.

(h) Same as (g) but for noisy data. Points give the noise of the visibilities. For an exact fit, they are located on the lines.

The vertical dotted lines in (e), (g) and (h) show the *u*-positions of the measured visibilities. Some important numerical properties of the results are shown in Table 8.2.

As the main goal of our work is to constrain lens models and not to produce the very best map of the source, we do not present results for other parameter combinations here. However, a much deeper analysis of the properties of different methods should be performed in the future to be able to extract all information from interferometry data.

8.9.1 CLEAN with small gain

Most observers agree that small loop gains γ generally produce 'better' maps. We therefore used a gain of $\gamma = 0.01$, which leads to results very similar to the limiting case of infinitely small γ. For the experiment in Figure 8.4 (page 113), we used 10^7 iterations to come sufficiently close to convergence. An interesting result of our numerical experiments with different parameters was the strong dependence of the required numbers of iterations on the number of measured

method	χ^2	residuals (RMS)	RMS diff. (as is)	RMS diff. (convolved)	Figure
line 1	noiseless				
line 2	noisy				
line 3	noise only				
CLEAN	4.520e−05	3.446e−06	9.243e−01	1.208e−01	8.4
10^7 its.	5.239e+00	4.500e−04	2.245e+00	7.315e−01	
	9.764e+00		2.232e+00	6.694e−01	
CLEAN	1.238e−02	5.691e−03	9.234e−01	1.309e−01	8.5
5000 its.	6.757e+00	4.003e−02	8.254e−01	5.281e−01	
	8.076e+00		6.710e−01	5.070e−01	
princ. sol.	1.938e−15	5.605e−13	1.256e−01	1.166e−02	8.6
$\lambda = 10^{-12}$	2.199e+00	2.158e−08	2.158e+04	6.395e+03	
	1.280e+01		2.158e+04	6.395e+03	
lin. reg.	1.287e+01	7.844e−01	1.588e−01	1.130e+00	8.7
$\lambda = 1.6$	1.979e+01	7.636e−01	1.720e−01	1.454e+00	
	2.449e+01		5.620e−02	4.729e−01	
lin. reg.	1.372e+01	6.044e−01	1.874e−01	1.464e+00	8.8
$\lambda = 0.85$	1.882e+01	6.109e−01	2.201e−01	1.368e+00	
$I_\mathrm{p}(x) \equiv 1$	1.870e+01		7.034e−02	4.950e−01	
NNLS	2.840e−14	1.302e−08	1.043e+00	3.355e−04	8.9
	7.682e+00	1.699e−01	1.402e+00	5.209e−01	
	7.151e+00		1.922e+00	5.208e−01	
NNLS	8.596e−15	1.269e−08	2.247e−02	1.074e−05	8.10
lin. reg.	7.682e+00	1.699e−01	1.402e+00	5.209e−01	
$\lambda = 10^{-12}$	7.151e+00		1.401e+00	5.209e−01	
NNLS	8.550e+00	7.084e−01	1.318e−01	7.073e−01	8.11
lin. reg.	1.874e+01	7.666e−01	1.372e−01	9.839e−01	
$\lambda = 1.4$	1.711e+01		3.981e−02	3.651e−01	
	g/h	f	c	d	subfigure

Table 8.2: Numerical results of the one-dimensional experiments, which are shown in Figures 8.4 to 8.11. Residuals in *uv* space χ^2 (noiseless/noisy/noisy result − true), image plane residuals $I_\mathrm{D} - B * I^\mathrm{m}$ (noiseless/noisy/noisy − noiseless), RMS of image plane difference (reconstructed − true) as is and convolved with CLEAN beam (noiseless/noisy/noisy − noiseless). The last row points to the panel number of the plots.

visibilities. For equal brightness distributions, a number of $N = 5$ leads to $\sim 10^5$ iterations to reach a residual level of 10^{-10}, but for $N = 10$ (as used in the shown results), even 10^8 iterations cannot push the residuals below $1.5 \cdot 10^{-6}$ (noiseless) and $2 \cdot 10^{-4}$ (noisy). These numbers are much larger than the number of CLEAN iterations used in any realistic mapping process. Interestingly, the results look better long before convergence is reached in some cases. This can bee seen in comparison of Figures 8.4 (10^7 iterations) and 8.5 on page 114 (5000 iterations). In image space, the latter represents the true brightness distribution much better then the first (subfigure c+d). Especially for noisy data, the results get worse with too many iterations[6]. Even though the *uv* and image plane residuals are smaller with more iterations, the difference between the reconstructed and the true brightness distribution in image space gets worse if too many iterations are performed. The RMS difference in the noisy case is 2.25 for 10^7 iterations, but only 0.83 for 5000. Convolved with the CLEAN beam, these numbers are 0.73 and 0.58.

With the fine sampling used here, the direct results of CLEAN shown in Figures 8.4(a) and 8.5(a) do not represent the true brightness distribution very well, since they consist of a number of well separated positive and negative peaks. These peaks (especially in the noisy case and far from the source) get stronger when reaching convergence. In Figure 8.4(a) they are almost as strong as on the source itself. With less iterations, Figure 8.5(a), these noisy peaks are much weaker. After convolution with the CLEAN beam, the peaks vanish and the result is consistent with the convolved true brightness distribution, Figures 8.4(b) and 8.5(b). We have to keep in mind that CLEAN used in this way does not provide a fit to the true brightness distribution itself but to the smoothed map.

The non-linearity of CLEAN has the effect that the noise can not be calculated independent of the signal itself. Figures 8.4(c)+(d) and 8.5(c)+(d) show that the noise properties are not constant along the x axis. The noise seems to be higher in regions with significant flux. The interpolation of the visibilities performed by CLEAN, see Figure 8.4(e),(g),(h) and 8.5(e),(g),(h) is quite good. The extrapolation to higher u is on the other hand poor (see the outer parts of subfigures e). Therefore the convolution with the CLEAN beam is absolutely essential to obtain useful results.

The *uv* residuals get close to zero in the noiseless case and reach $\chi^2 = 5.2/6.8$ for noisy data (see Table 8.2). Even the latter is less than the expected residuals. We can interpret CLEAN with a finite number of iterations (and including the CLEAN beam convolution) as a method to find a strongly regularized solution of the deconvolution problem. This kind of regularization is not easy to analyse mathematically. The number of iterations plays the role of a regularization parameter. The less iterations, the stronger the regularization. The approach of adjusting the number of iterations to obtain residuals in the expected range has not been used so far but might be useful to understand the kind of regularization performed by CLEAN better.

8.9.2 Principal solution and strong linear regularization

For this experiment we used equation (8.62) to calculate the weakly regularized solution. The results obtained with $\lambda = 10^{-12}$ are shown in Figure 8.6 (page 115). To find the solution we used the method of LU decomposition (Press et al., 1992). For this method, the whole B matrix and its decomposition matrices have to be kept in memory. This is not a problem for our examples with $M = 100$. For realistic two-dimensional problems, this approach would be very ineffective. Other methods like conjugate gradients are more appropriate then.

We see from Figure 8.6, that weak regularization works very well for the noiseless data. The *uv* residuals now reach their theoretical limit of $\chi^2 = 2.2$. Even measured in terms of the difference between the reconstructed and true brightness distribution, this solution is superior to CLEAN. The RMS difference is 0.13 (0.012 in the convolved case).

In *uv* space, we notice the far better interpolation and extrapolation than with CLEAN. In image space, the solution shows weak oscillations superposed on the correct brightness distribution. The sharp edges cannot be reproduced exactly. After convolution with the CLEAN beam,

---

[6] This is *not* the well known effect of gridding errors leading to diverging CLEAN results. For our experiments, we performed the Fourier transform directly, without any gridding. Nevertheless, numerical errors might play a role for very many iterations.

these oscillation become very small and the solution is almost perfect.

For noisy data, the solution shows ridiculously strong fluctuations in image space, even though the fit is quite good in terms of *uv* residuals ($\chi^2 = 2.2$). The noise is amplified extremely and shifted to frequencies which are not measured. We can use strong regularization to reduce these effects and still keep χ^2 sensibly small. For Figure 8.7 (page 116), we used $\lambda = 1.6$ to bring χ^2 (19.8 for noisy data) close to the expected value. Unfortunately, the result now shows a strong bias in the direction of low fluxes/visibilities. This is noticeable in image space, where it leads to stronger oscillations around the correct flux, and in Fourier space, where the reconstructed values are systematically too low, especially for small spacings, where the gap in the measurements makes an accurate interpolation of the visibilities difficult. Although systematic effects and the noise should contribute in equal parts to the errors (the residuals are in their theoretical range), the deviations in image space seem to be dominated by the strong regularization (see subfigure d).

To show the bias effect, we include in Figure 8.8 the result of another strongly regularized solution which was calculated as described in section 8.8.6. As prior, a constant $I_\mathrm{p}(x) \equiv 1$ was used. This corresponds to the plateau of the rectangular function of the real data. A regularization parameter of $\lambda = 0.85$ was used to produce residuals comparable to the solution for an empty prior (compare with Figure 8.7). The bias to small fluxes on the plateau is now suppressed. Now we have a bias to positive fluxes in the empty regions.

8.9.3 NNLS solution

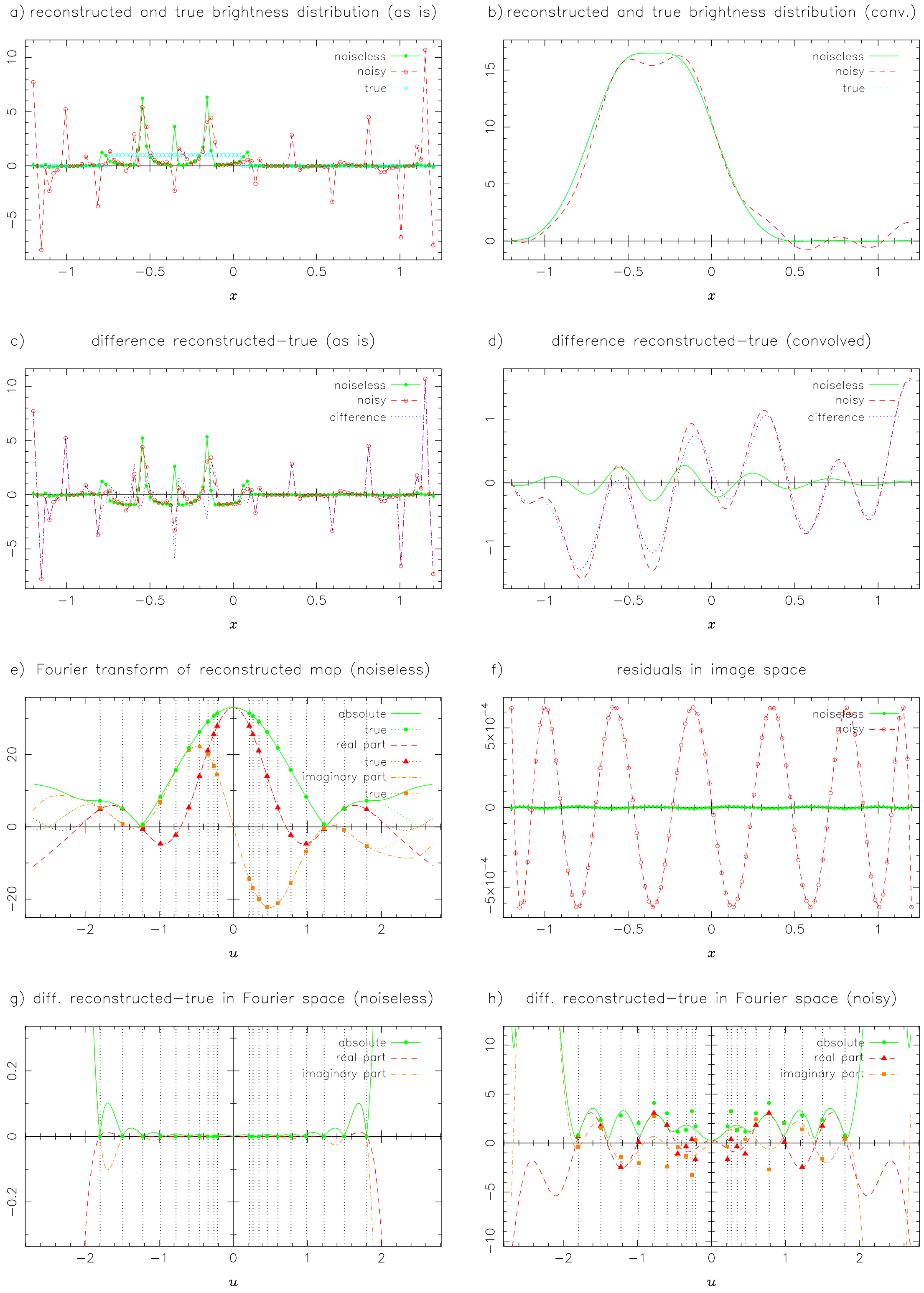
We used the algorithm of Lawson & Hanson (1974) as given in Briggs (1995). The results are shown in Figure 8.9 (page 118). The direct solution shows spikes similar to CLEAN, but the convolved version is (in the noiseless case) now much better than the CLEAN result. With noise, the convolved result is comparable to prematurely stopped CLEAN. The image space residuals are 0 without noise and ≤ 0 with noise. This is due to the fact that the gradient of the residuals has to be zero for components where the non-negativity constraint is not active (regions with positive flux) and negative for regions with zero flux.

Figure 8.10 (page 119) shows a weak regularization with $\mathsf{S} = 1$ and $\lambda = 10^{-12}$. Even without the CLEAN beam convolution, the solution for the noiseless case represents the true brightness distribution very well. To achieve the same even for the noisy data, we used $\lambda = 1.3$ for Figure 8.10. As in the unconstrained strongly regularized case, we see a bias to small fluxes. The oscillations in the empty regions are now suppressed by the non-negativity constraint. At the plateau, the constraint is not active, and the oscillations are still noticeable.

8.9.4 Summary

Without the CLEAN beam convolution, CLEAN is not competitive with the other methods. Without noise, NNLS with weak regularization is by far the best. With noise, strong regularization has to be used. Strongly regularized NNLS is now the best method, but the direct linear deconvolution is almost as good. With NNLS, the noise is shifted from the empty regions to the regions with true flux, while with unconstrained deconvolution, the noise is independent from the true signal (linearity of the method).

After convolution with the CLEAN beam, NNLS with weak regularization is best, but pure NNLS is almost as good and the principal solution is also very good. Including noise, the weakly and unregularized NNLS have the smallest RMS error, but the prematurely stopped CLEAN is almost as good.



a) reconstructed and true brightness distribution (as is)

b) reconstructed and true brightness distribution (conv.)

c) difference reconstructed−true (as is)

d) difference reconstructed−true (convolved)

e) Fourier transform of reconstructed map (noiseless)

f) residuals in image space

g) diff. reconstructed−true in Fourier space (noiseless)

h) diff. reconstructed−true in Fourier space (noisy)

Figure 8.4: One-dimensional experiment using CLEAN. 10^7 iterations. For details, see section 8.9, especially page 108 and section 8.9.1.

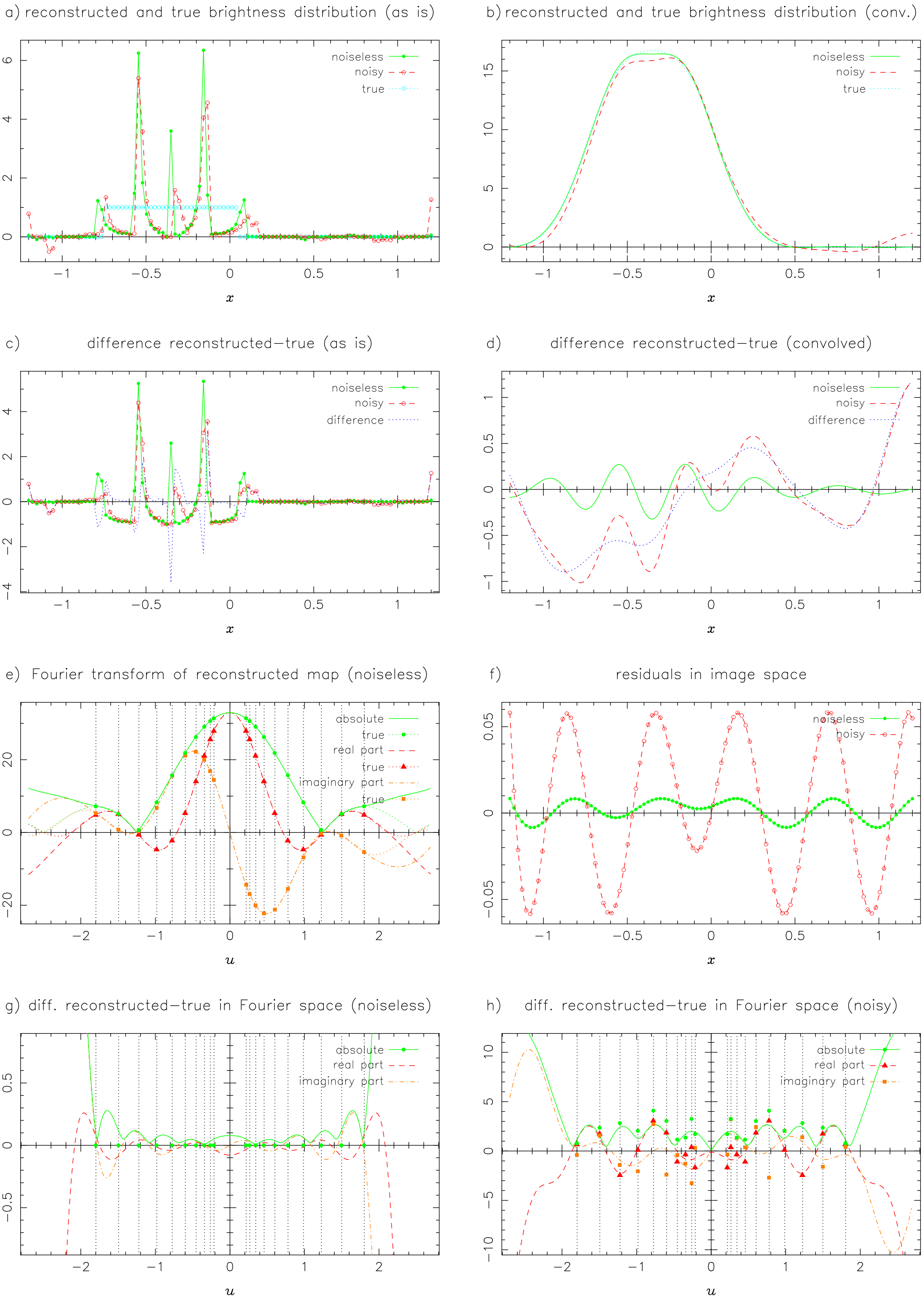


Figure 8.5: One-dimensional experiment using CLEAN. 5000 iterations. For details, see section 8.9, especially page 108 and section 8.9.1.

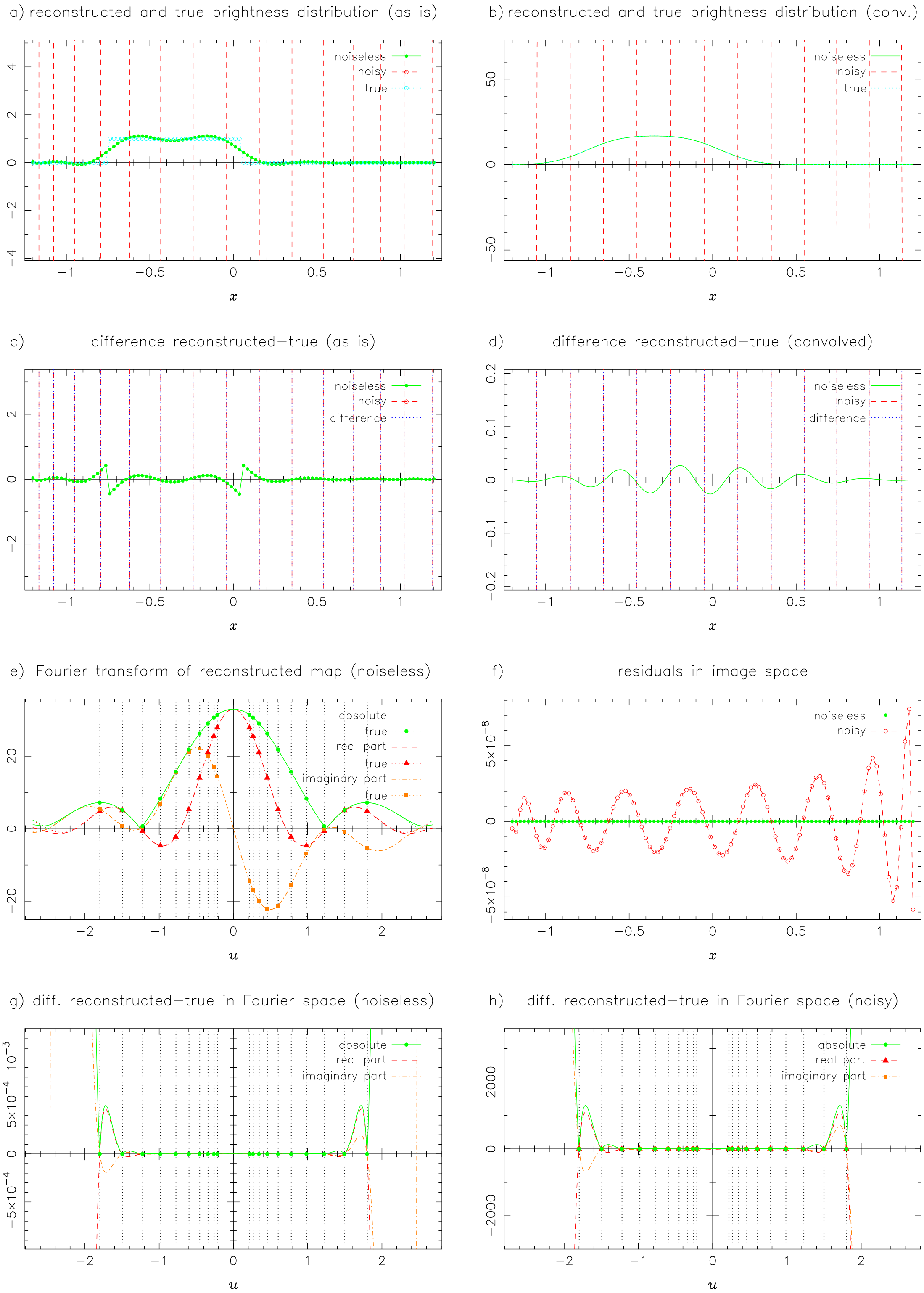
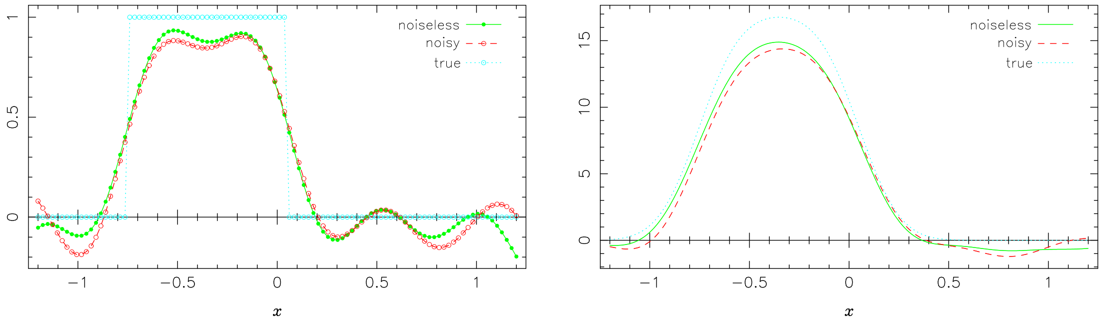
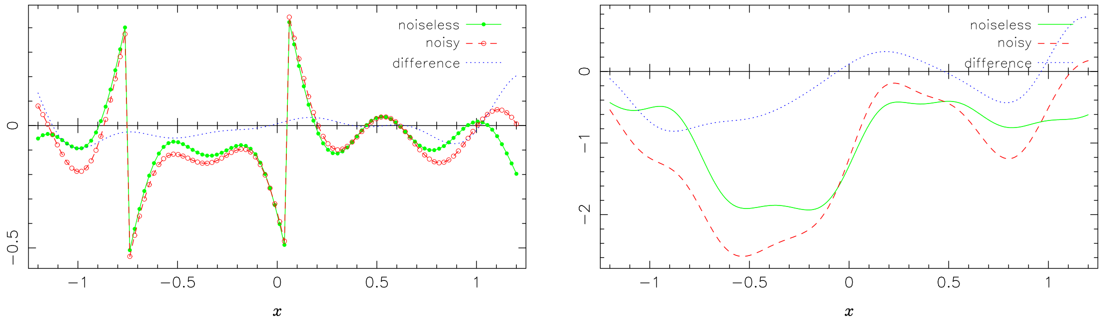


Figure 8.6: One-dimensional experiment using the principal solution (weakly regularized with $\lambda = 10^{-12}$). For details, see section 8.9, especially page 108 and section 8.9.2.



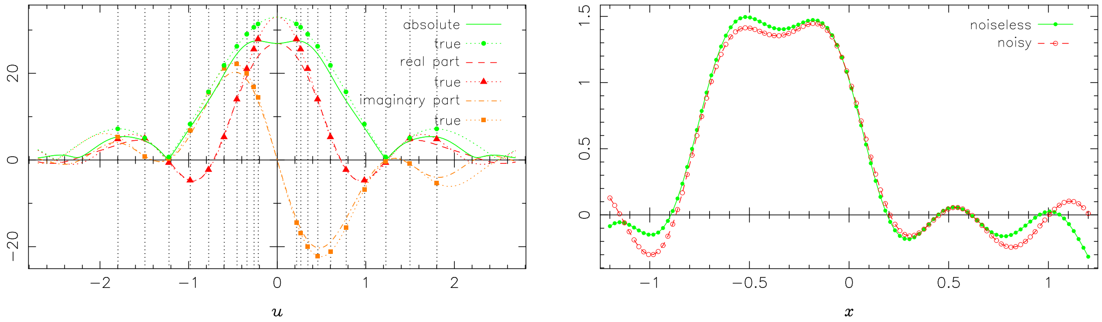
a) reconstructed and true brightness distribution (as is)

b) reconstructed and true brightness distribution (conv.)



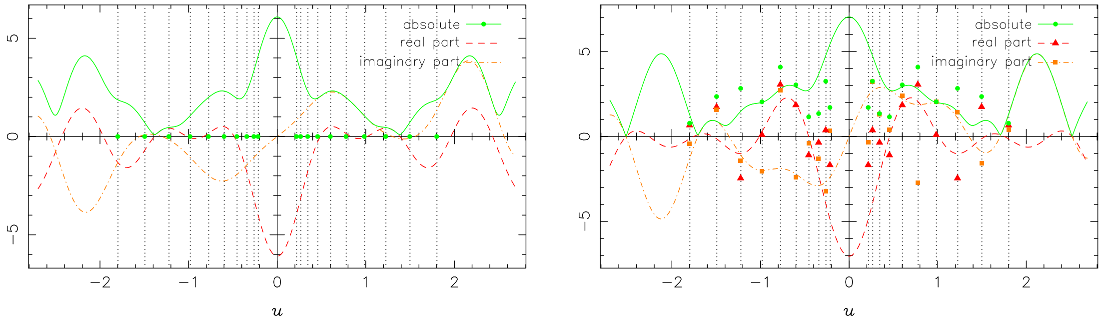
c) difference reconstructed−true (as is)

d) difference reconstructed−true (convolved)



e) Fourier transform of reconstructed map (noiseless)

f) residuals in image space



g) diff. reconstructed−true in Fourier space (noiseless)

h) diff. reconstructed−true in Fourier space (noisy)

Figure 8.7: One-dimensional experiment using direct algebraic inversion strongly regularized with $\lambda = 1.6$. For details, see section 8.9, especially page 108 and section 8.9.2.

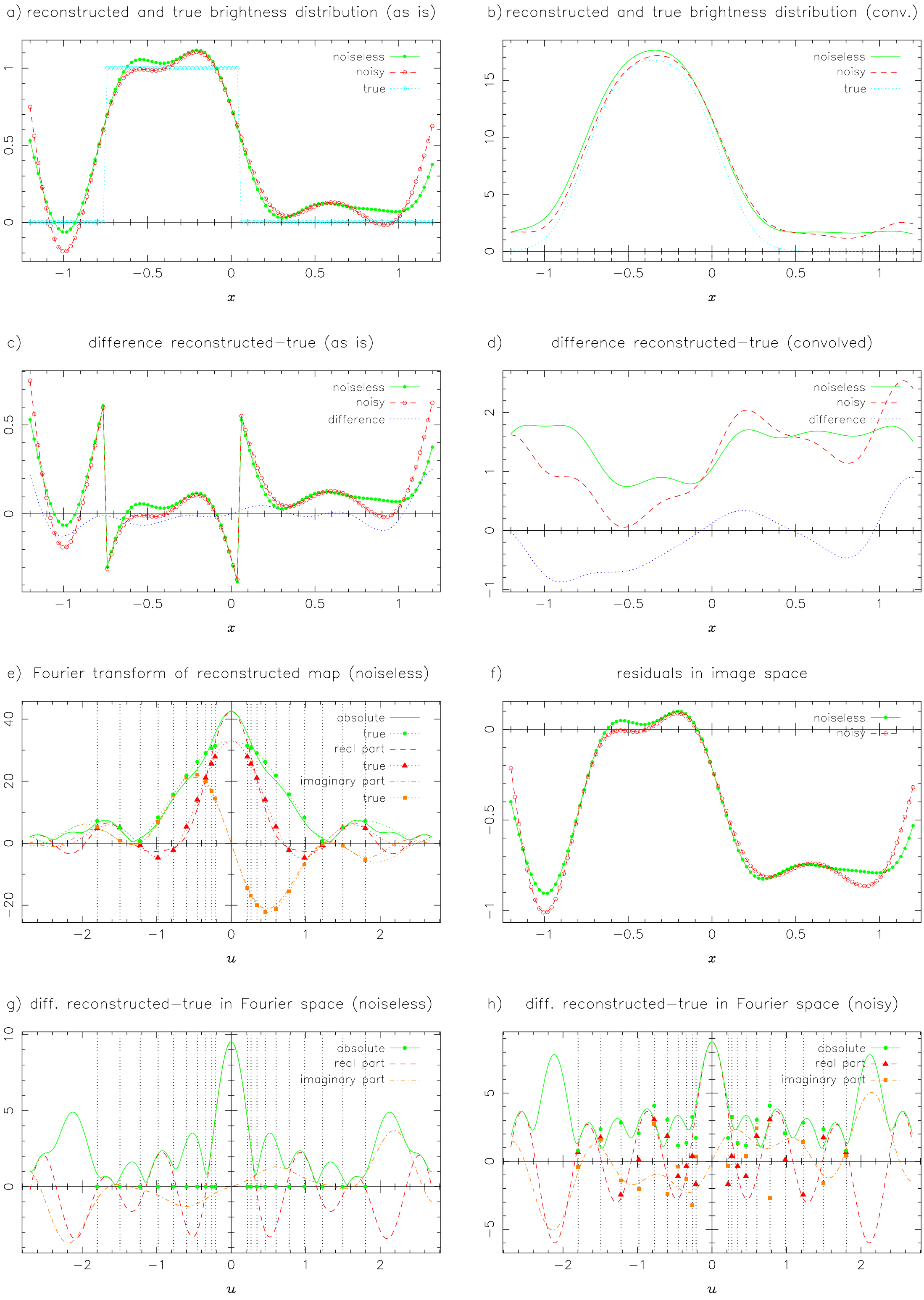


Figure 8.8: One-dimensional experiment using the strongly regularized ($\lambda = 0.85$) solution around constant 1. For details, see section 8.9, especially page 108 and section 8.9.2.

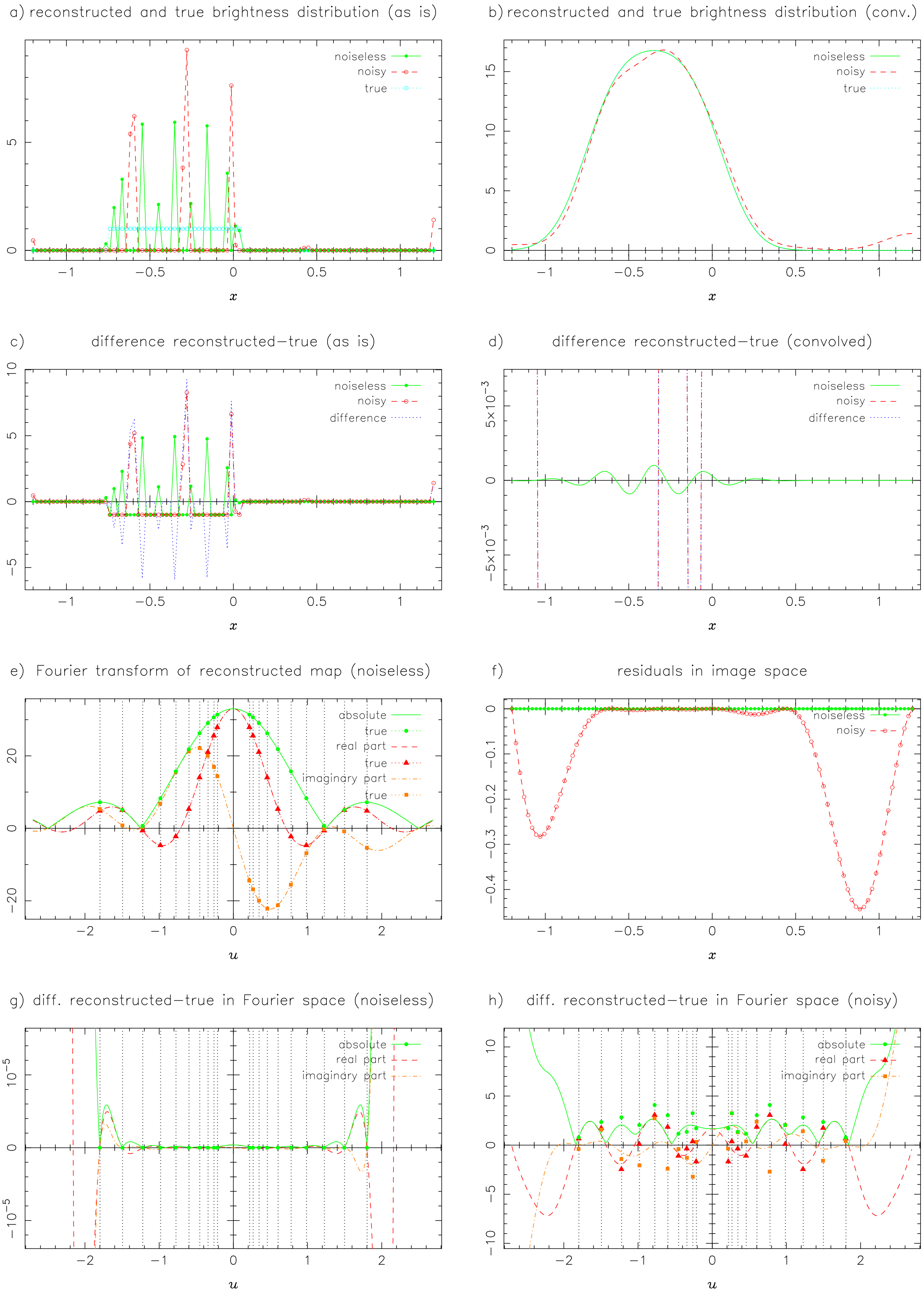


Figure 8.9: One-dimensional experiment using NNLS without regularization. For details, see section 8.9, especially page 108 and section 8.9.3.

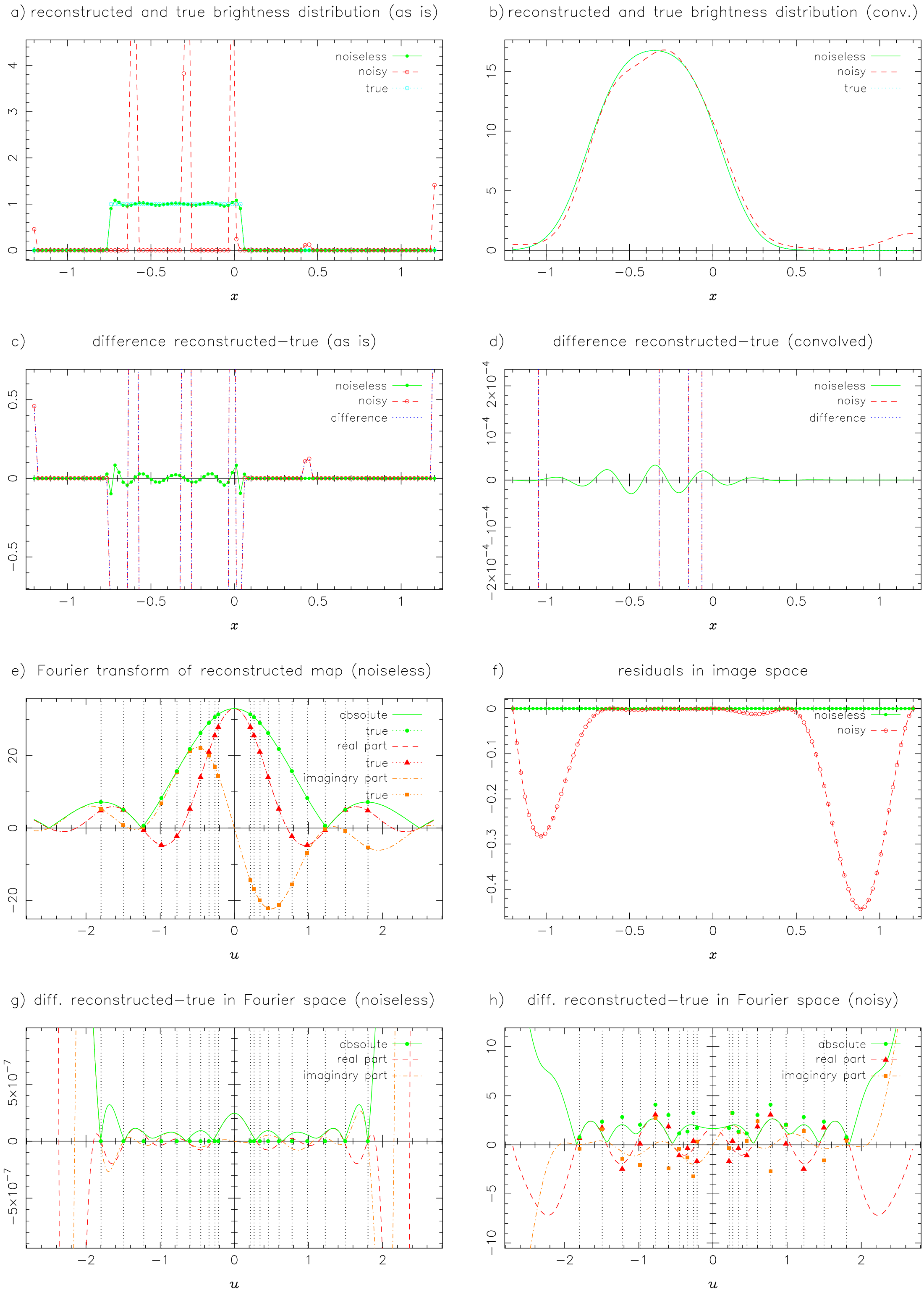


Figure 8.10: One-dimensional experiment using NNLS, regularized with $\lambda = 10^{-12}$ (principal NNLS solution). For details, see section 8.9, especially page 108 and section 8.9.3.

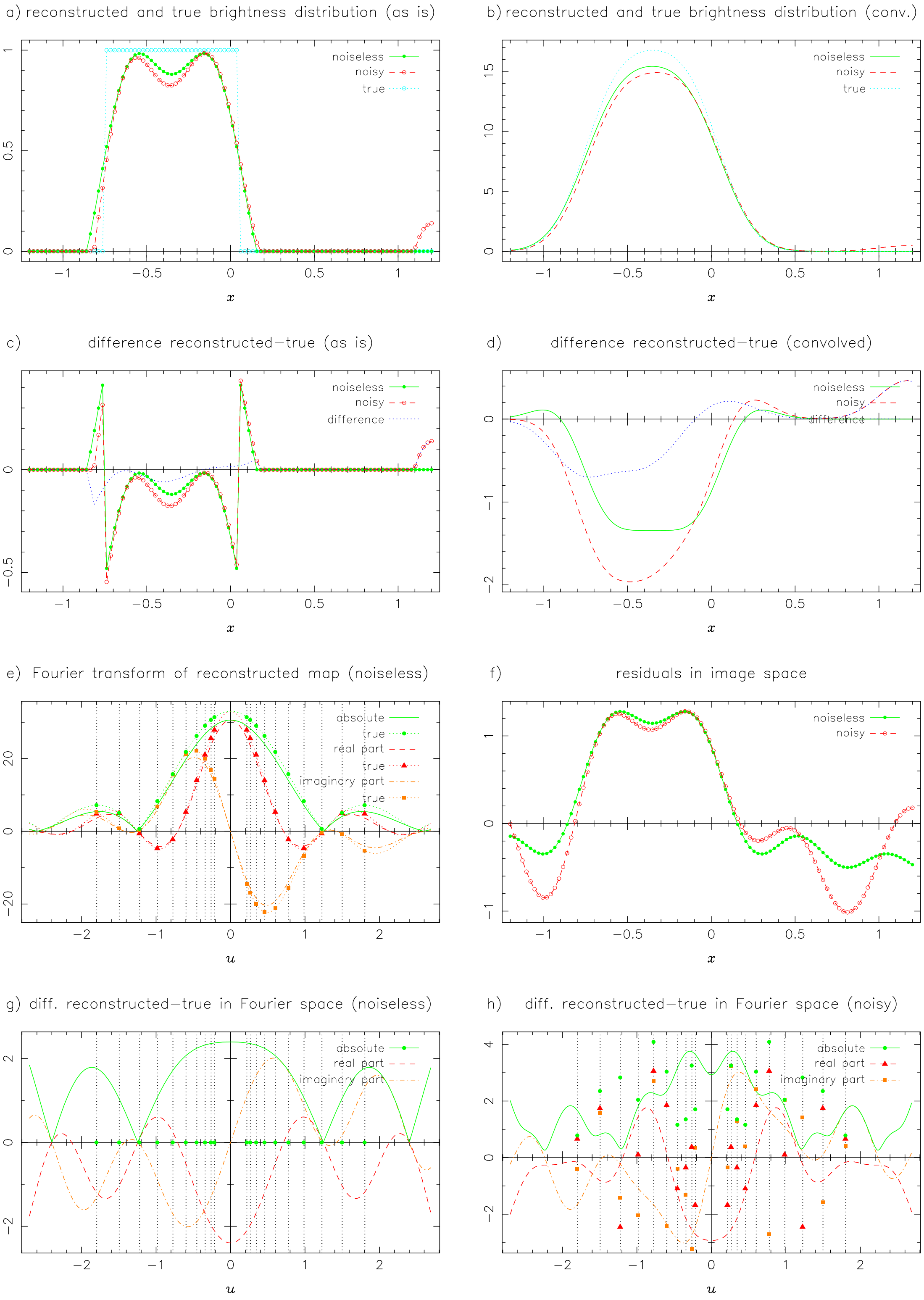


Figure 8.11: One-dimensional experiment using NNLS, regularized with $\lambda = 1.3$. For details, see section 8.9, especially page 108 and section 8.9.3.

For the reconstruction of the source in a lensing scenario (section 9.24), we would like a method that works without explicit convolution with a CLEAN beam. Strongly regularized direct inversion and strongly regularized NNLS can both be used then. Although the errors with NNLS are slightly smaller, the statistical properties of direct linear inversion are much simpler and the method might thus provide the more robust results.

8.10 Technical details of CLEAN

8.10.1 Gridding and FFT

The calculation of the dirty map and dirty beam with equations (8.14) and (8.15) (direct Fourier transform) is computationally very expensive. If the number of visibilities is not very small, another strategy can be faster by orders of magnitude, although it is conceptually more complicated. The idea is to apply the method of 'fast Fourier transform' (FFT), which is very efficient in the case of many components. While direct Fourier transform (DFT[7]) needs of the order n^2 steps to compute a complete transform of n components, FFT needs only $\propto n \log n$ steps. In typical applications, FFT is almost always much faster. For details about Fourier transforms, see appendix B and references therein.

The disadvantage of the FFT algorithm is that it only works on a regular grid in both Fourier and image space. We therefore have to put the measured visibilities on a grid and cope with the symmetry properties of regular grids. The Fourier transform of a regularly sampled function always is periodic. The same is true in the opposite direction; Fourier transforms on regular grids can only describe periodic functions.

We now have to find a regular sampled approximation to the true collection of irregularly distributed visibilities. The most simple gridding consists in overlaying the visibilities with the grid and assign the (weighted) mean of all visibilities lying inside a pixel to this grid point. This approach is not only difficult to analyse but also has some disadvantageous properties.

For continuous visibilities, an optimal gridding procedure would be the sampling at the regularly spaced grid points. Unfortunately, the measured visibilities are discrete. They first have to be approximated by a continuous function, which can be done by convolution. The standard recipe is the following.

- Convolve the measured visibilities (interpreted as a collection of δ-functions) with a 'gridding convolution function' $\tilde{C}$.
- Sample the convolved visibilities with a regular spacing of Δu on n points ($\Delta u \times \Delta u$ on $n \times n$ points in the two-dimensional case).
- Apply the FFT to generate a first approximation of the dirty map and beam.
- Divide dirty beam and map by C, the inverse Fourier transform of $\tilde{C}$, to compensate for the uv-space convolution with $\tilde{C}$. This is called the grid-correction, although it does not correct for the sampling step but for the convolution. 'Convolution correction' would be a more appropriate term.

Without the sampling step, the division at the end would exactly cancel with the convolution in the first step, provided that C has no zeros. The choice of $\tilde{C}$ would therefore not be critical.

---

[7] The acronym DFT is also used for 'discrete Fourier transform'. In our context, all numerical Fourier transforms are discrete anyway.

Several points deserve attention here. First, the discrete grid has to be large enough to cover all measured visibilities to avoid loosing some of the data. If the maximal absolute u-spacing is called u_{max}, this condition can be written as the equation[8]

$$n\,\Delta u \geq 2\,u_{\mathrm{max}} \tag{8.69}$$

or equivalently

$$\Delta x \leq \frac{1}{2\,u_{\mathrm{max}}} \quad , \tag{8.70}$$

where the factor of 2 is due to the fact that the visibilities are symmetric around $u=0$ and the grid has an extension of only $n/2$ in each direction. The last equation can be interpreted in terms of the Nyquist theorem. The grid has to be fine enough to allow at least two samples even for the highest frequencies. Especially for high dynamic range data with unresolved components, a finer sampling can improve the results considerably. For compact components lying between the grid points, the gridded representation would consist of a large number of alternating positive and negative components extending far beyond the true position. These oscillations can deteriorate the deconvolution results significantly and should be avoided.

The regular sampling in uv space has the consequence that the image always is periodic. Emission at a position x would have the same effect on the visibilities as emission at another position $x + kn\Delta x$ for arbitrary k. This emission outside the lm grid would therefore be folded (‘aliased’) into the mapping area. Aliased sources can be detected as such by changing the gridding parameters, which leads to a shift of these components but leaves the emission inside the map unaffected. Very strong confusing sources near the limits of the primary antenna beams can even be more problematic, because their effective antenna gains would differ from station to station, leading to incorrect visibility measurements.

Let I and $\tilde{I}$ represent the true brightness distribution, I_{D} the correct dirty map and B and $\tilde{B}$ the dirty beam and uv coverage function. The uv sampling function is called $\tilde{\text{Ш}}$ and has a period Δu. Its inverse transform Ш has an image space period of $n\Delta x$. The approximated dirty map can then be written as follows.

$$I_{\mathrm{D}}^{\mathrm{FFT}} = \frac{\mathcal{FT}^{-1}\left(\left((\tilde{I}\tilde{B}) * \tilde{C}\right)\tilde{\text{Ш}}\right)}{C} \tag{8.71}$$

$$= \frac{\left((I * B)C\right) * \text{Ш}}{C} \tag{8.72}$$

$$= \frac{(I_{\mathrm{D}}C) * \text{Ш}}{C} \tag{8.73}$$

The last equation shows that the correct dirty map (consisting of the sources *and* their sidelobes) is multiplied with C, periodically replicated and then divided by C, *not* by the periodic replicate of C. The function C can therefore be chosen in a way to suppress the aliased copies of I_{D}. Ideally, C shifted by one period and divided by the unshifted C should be as close to zero as possible. Unfortunately, a rapidly declining C would imply a very extended convolution function $\tilde{C}$. On the other hand, the convolution function should be very compact to allow an efficient numeric convolution. This trade-off is mathematically equivalent to one aspect of the Heisenberg uncertainty relation. C and $\tilde{C}$ cannot both be arbitrarily compact at the same time.

---

[8] The same equations hold for the v or y direction.

For very extended $\tilde{C}$, the convolution would need more CPU time than is saved by the FFT algorithm. Tabulated and highly compact (total size of only a few pixels) convolution functions are normally used. They can still be optimized for maximal alias reduction. Figure 8.12 shows three possible convolution functions together with their Fourier transform. The pillbox function is most simple to realize but has the worst alias reduction. The truncated Gaussian is used by the DIFMAP software. Near the field edges, it is inferior to an optimally truncated Gaussian multiplied with a sinc function. Other, more efficient, convolution functions (e.g. spheroidal functions) are commonly used.

For details of the gridding procedure and other aspects of the imaging process (before CLEANing), see Sramek & Schwab (1989) and Briggs, Schwab & Sramek (1999).

8.10.2 Field size limitation

When using FFT to obtain the dirty beam and dirty map, both have the same size. In each iteration of CLEAN, the shifted beam must be subtracted from the residual map. To be able to cover the complete map with the shifted beam, the 'active' area of the residual map and therefore the area of the CLEAN map has to be reduced to the central quarter (half the side length) of the total map. Even stricter reductions are often useful to reduce the effect of aliasing, which is strongest near the field edges.

8.10.3 The Clark algorithm

A variant of CLEAN was proposed by Clark (1980). In its minor cycle, it consists of Högbom's classical CLEAN, but uses only a small patch of the dirty beam which includes the centre and the highest sidelobe. When a certain limit is reached in the minor cycle, a major cycle is performed, in which all components from the minor cycle are collected, transformed with a FFT, multiplied with the gridded weights (equivalent to convolution with the dirty beam), transformed back and subtracted from the residual map. This major cycle corrects for the errors introduced by the approximation of the dirty beam by using only a small patch. The only advantage of the Clark algorithm is speed, because the subtraction in the minor cycles is much faster than with the complete beam. The convolution in Fourier space is so efficient, that the total time needed for a certain number of iterations is still smaller than with standard CLEAN.

8.10.4 The Cotton-Schwab algorithm

In this algorithm (Schwab, 1984), the subtraction of CLEAN components in the major Clark cycle is performed on the ungridded visibilities using a direct Fourier transform or FFT with interpolation. The residual visibilities are then again gridded. This process corrects for the errors introduced by using regular grids to compute the dirty maps. Without this correction, the convolution equation (8.30) can often not be satisfied exactly, because of the gridding errors, which leads to diverging CLEAN results. Depending on the number of visibilities, the Cotton-Schwab algorithm can sometimes be significantly slower than the simpler Clark CLEAN.

The direct algebraic inversion methods, which have been discussed before, should usually not be applied to data obtained with FFT of gridded visibilities. Even without noise, the dirty map is in this case not exactly given by the true brightness distribution convolved with the dirty beam. Strategies similar to Cotton-Schwab must be used in cases where the direct Fourier transform is prohibitively expensive.

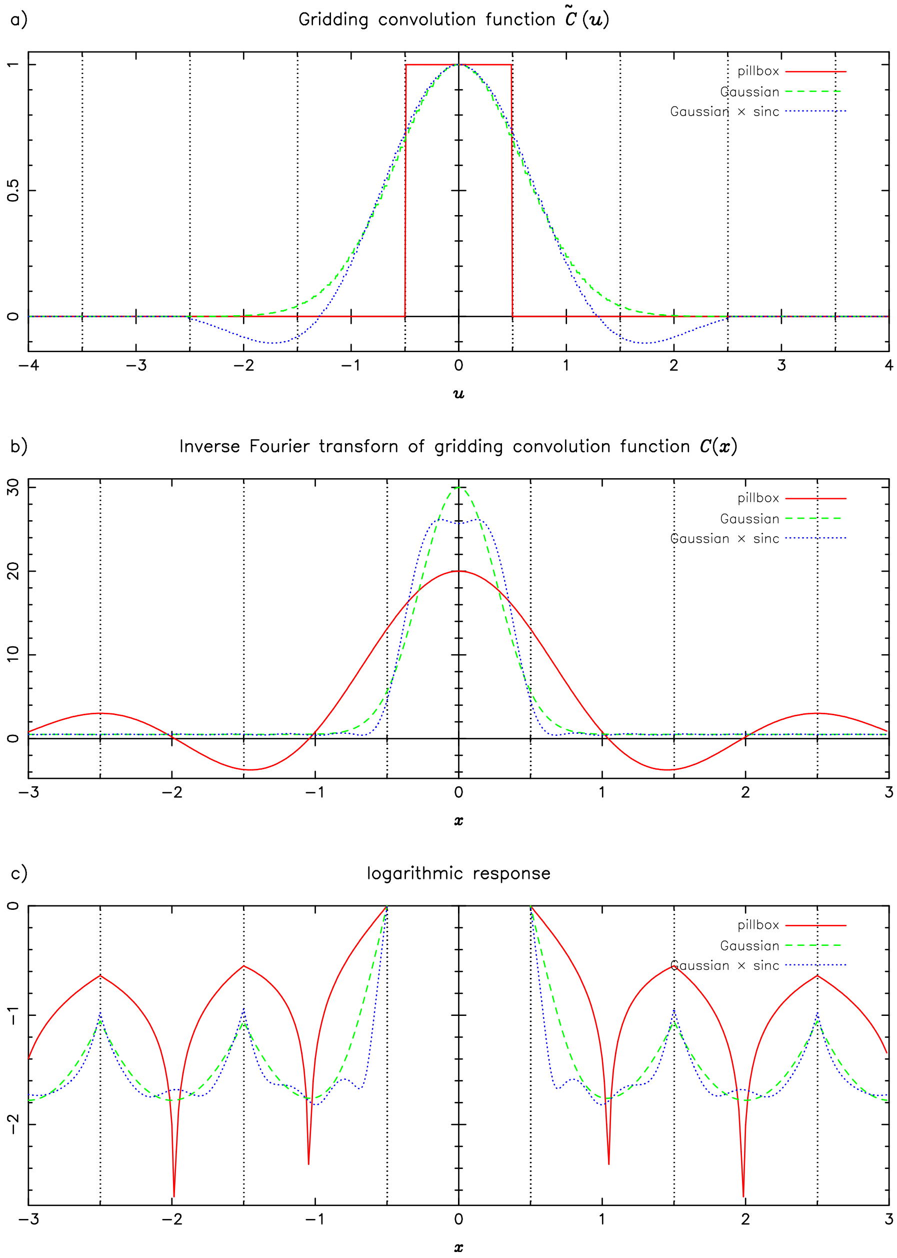


Figure 8.12: Three possible gridding convolution functions. (a) The functions in *uv*-space, (b) the inverse Fourier transform, (c) the logarithmic response (alias rejection). This is (b) with included 'gridding correction', i.e. division by the central period of $C(x)$. This function should be as low as possible outside the central period.

8.10.5 Weighting with CLEAN

Strong statistical arguments vote for natural weighting, because it minimizes the noise in the dirty map. There is, however, no reason to use the same weighting scheme during the whole CLEAN process. In the beginning, the S/N ratio is usually very high. In the limit of noiseless data, no weighting scheme is preferred for the others for statistical reasons. It is therefore appropriate to start with a non-natural weighting, which can improve the resolution and go to natural weighting only later, when the S/N comes close to unity. This is in fact done by many experienced observers. The mapping process is started with uniform weighting or a variant thereof and switches to natural weighting later, if the theoretical limit of the dynamic range should be reached. Maps produced with this strategy seem to be superior to ones with constant natural weighting. Thinking conservative, the naturally weighted CLEAN beam should be used at the end to produce the CLEAN map. This is, however, a matter of debate, since the definition of the CLEAN beam is rather arbitrary anyway.

8.10.6 Using CLEAN windows and non-negativity constraints

If it is known that the emission is confined to a certain region of the map, this a priori information should be used in the mapping process to reduce the freedom of the fit. This is especially true in combined mapping/self-calibration iterations.

Another physically motivated constraint is to disallow negative fluxes in the map. Sometimes this is implemented by stopping the CLEAN iteration when the first negative component is reached or by deleting negative components after some iterations. There is no theoretical foundation for this approach. Especially when changing the weighting or if strong compact components are present, negative components are often essential to fit the data, either to correct for too strong components added earlier or to represent components located not exactly on a grid point.

Another approach is to use only the positive components and disregard the others. All these approaches do *not* find the best non-negative solution. Other methods like NNLS have to be used to include the constraint in a rigorous way.

8.11 Conventional CLEAN beam convolution and alternatives

The normal procedure after producing a list of CLEAN components is to convolve it with a Gaussian CLEAN beam which is fitted to the central part of the dirty beam (see section 8.6.2). This beam is scaled to a peak of 1, so that the peaks in the CLEAN map are direct measurements of the total flux densities of unresolved components. The units of the map are therefore measured per beam. The level for resolved emission depends on the size of the beam. Another convention would be to scale the CLEAN beam to an integrated flux density of unity. The units would then be emission per unit area and the CLEAN map would provide a convolved surface brightness. Total flux densities could be derived by integrating the map. For point sources, the conventionally used former approach is more convenient, while the latter might be more useful for heavily resolved sources and for varying beams as in LENSCLEAN (see discussion in section 9.24).

8.11.1 Normalized beams

The integral over a beam following equation (8.21) can be calculated by writing it with coordinates $\mathbf{x}'$ in the principal axes system of G so that $\mathbf{x}^\dagger \mathsf{G} \mathbf{x} = \mathbf{x}'^\dagger \mathbf{x}'$. Without the irrelevant rotation, the transformation reads $x=\lambda_{\min}^{-1/2}x'$ and $y=\lambda_{\max}^{-1/2}y'$, where λ are the Eigenvalues of G.

$$\int \mathrm{d}^2x\, \mathrm{e}^{-\mathbf{x}^\dagger \mathsf{G}\mathbf{x}/2} = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}'}\right| \int \mathrm{d}^2x'\, \mathrm{e}^{-\mathbf{x}'^\dagger \mathbf{x}'/2} \tag{8.74}$$

$$= \frac{1}{\sqrt{|\mathsf{G}|}} \iint \mathrm{d}x\,\mathrm{d}y\, \mathrm{e}^{-(x^2+y^2)/2} \tag{8.75}$$

$$= \frac{2\pi}{\sqrt{|\mathsf{G}|}} \tag{8.76}$$

The normalized CLEAN beam can then be written as

$$G(\mathbf{x}) = \frac{\sqrt{|\mathsf{G}|}}{2\pi} \mathrm{e}^{-\mathbf{x}^\dagger \mathsf{G}\mathbf{x}/2} \quad . \tag{8.77}$$

If varying (either with position or CLEAN iteration number) beams are to be used, only the normalized beams produce meaningful results. The maps are then smoothed surface brightness maps with conservation of total flux. This would not be the case with the conventional beams with unity peak. Total flux densities would then depend on the beam size and constant surface brightness regions would not be shown as such, because the intensity in the CLEAN map would depend on the beam size.

8.11.2 Positional accuracy

The size of the beam should somehow resemble the accuracy of the corresponding CLEAN component position. To justify convention or develop a better approach, it is necessary to understand this accuracy. A very simple analysis is possible for a toy model which consists of only one component. The residuals after subtraction of a component with optimal flux at position $\mathbf{x}_0$ are (see e.g. section 8.8.4):

$$R^2 = R_0^2 - \sum_j w_j I_\mathrm{D}^2(\mathbf{x}_0) \tag{8.78}$$

To estimate the accuracy of the position $\mathbf{x}_0$, the residual difference ΔR^2 relative to the optimum has to be analysed.

$$\Delta R^2 = \sum_j w_j \left(I_\mathrm{D}^2(\mathbf{x}_0) - I_\mathrm{D}^2(\mathbf{x}) \right) \tag{8.79}$$

For natural weighting, the $1\,\sigma$ confidence limit can be calculated from the χ^2 distribution. It is at $\Delta R^2 = 1$ for one coordinate and $\Delta R^2 = 2.3$ for the combination of both coordinates. If the data consist of only one point source and if the S/N is not too low, the dirty map is approximately equal to the dirty beam scaled with the source flux S. If the beam is parametrized with a Gaussian, the equation is especially simple to solve for $\Delta x \ll$ FWHM:

$$\Delta R^2 = S^2 \frac{8\ln 2\, \Delta x^2}{\mathrm{FWHM}^2} \sum_j w_j \tag{8.80}$$

Remembering equation (8.18), which gave the noise as $N = 1/\sqrt{\sum w_j}$, the positional accuracy can be written as

$$\Delta x = \sqrt{\frac{\Delta R^2}{8\ln 2}}\, \frac{\mathrm{FWHM}}{S/N} \quad . \tag{8.81}$$

The first factor is of the order 1. We recover the well known fact that the positional accuracy is the size of the beam divided by the signal to noise ratio.

We learn that the normal CLEAN beam, which is fitted to the dirty beam, is appropriate for low signal to noise but actually obstructs information from high dynamic range observations of well separated point sources by smoothing the map far to strongly. No better approach has been developed yet, because the situation is more complicated in the case of extended emission or multiple very close point sources.

For well separated point sources, a simple idea for improvement would be to resize the CLEAN beam according to the magnitude of the peak in the dirty *map* and convolve the component with that beam. For a point source, these beams would grow larger in the course of CLEANing, because the peak would decrease and the positional accuracy become worse. This leads to the alternative of using a deconvolution function that resembles the resized beam only after summation over all CLEAN components at this position. This function can easily be calculated analytically.

Unfortunately, this approach is only sensible for well separated point sources but not for extended emission where the accuracy of positions does not only depend on the S/N but also on the shape of the dirty *map* around the CLEAN component. Even without noise (and therefore infinitely small resized beams), CLEAN beam convolution is absolutely necessary for extended source. This was already shown in the numerical experiments in section 8.9.1, see Figures 8.4 and 8.5.

8.11.3 Size of CLEAN beam from the dirty map

Another more consistent approach is to resize the CLEAN beam according directly to the dirty *map* instead of the *beam* and use (8.79). With the same formalism as in section 8.6.2, slightly modified to include the position and the measured visibilities, it is easy to fit a Gaussian to the dirty map at the peak position directly using the visibilities. Alternatively, the fit can also be done in the image plane. For separated point sources, this Gaussian would resemble the CLEAN beam scaled to the peak flux, but can be much more extended for extended sources. In fact, it can grow without limit which is clearly not very useful in practical mapping procedures. The beam is fitted to the map only locally and should therefore not become very large only because the local curvature is small.

It is known from experience that the CLEAN beam fitted to the dirty beam works well in most cases. Hence, this beam should be taken as a limit for the modified beam. For bright point sources, the beam should become very compact, and approach the limit of the classical CLEAN beam for very extended emission. The simplest idea to achieve this, is to add the G matrices for both beams to obtain the final beam. For circular beams, this is equivalent to

$$\mathrm{FWHM}_{\mathrm{final}} = \frac{\mathrm{FWHM}_{\mathrm{beam}}\mathrm{FWHM}_{\mathrm{map}}}{\sqrt{\mathrm{FWHM}_{\mathrm{beam}}^2 + \mathrm{FWHM}_{\mathrm{map}}^2}} \quad , \tag{8.82}$$

where $\mathrm{FWHM}_{\mathrm{beam}}$ stands for the normal CLEAN beam fitted to the dirty beam and $\mathrm{FWHM}_{\mathrm{map}}$ for the beam derived locally from the dirty map. The latter becomes smaller for higher flux levels. This approach uses approximately the smaller of the two beams, with a smooth transition. Maybe a correction has to be applied to assure that the final beam is never smaller than any of the two beams from which it was built. This can be achieved by taking the arithmetical mean of the G matrices instead of the sum. The expression in equation (8.82) is then modified by a factor of $\sqrt{2}$.

8.11.4 Using the mean instead of the best solution

Another idea that might be worthwhile following, is independent from the CLEAN method. Usually, the goal is to produce the *best* map consistent with the data and subject to some regularization constraint. Alternatively, one might think of finding the *mean* of all possible maps, weighted with their Bayesian probability. If the map is parametrized by the values of all its pixels, the probability density for a flat prior would be proportional to $\mathrm{e}^{-\chi^2/2}$ (see e.g. section 2.8). This is a multidimensional Gaussian. Hence, the mean would be exactly the same as the minimum, because of the symmetry. The method becomes more interesting, if alternative priors are used. Even very simple approaches, like exponential distributions for non-negative fluxes, might lead to improved results, when compared with the normal CLEAN solution. We have already started working on the integration, but the analytical difficulties are considerable.

8.12 Self-calibration and hybrid mapping

The details of preparing and calibrating interferometric data are far beyond the scope of this thesis. We nevertheless want to touch the topic of self-calibration, because it will be important for LENSCLEAN later.

To understand the problem, we have to go back to equation (8.6), which tells us how the visibilities are calculated in the correlator. If we now include possible phase shifts and gain errors of the atmosphere/receiver system, we can easily calculate how this changes the visibilities. Writing $\tilde{I}^{\mathrm{uncal}}$ for the uncalibrated visibilities and g_i for the complex gain factor of telescope i at a certain time, we can write

$$\tilde{I}_{ij}^{\mathrm{uncal}} = g_i\, g_j^{\star}\, \tilde{I}_{ij} \quad . \tag{8.83}$$

Phase and amplitude errors can be caused by the atmosphere and the receiver system. Amplitude errors are usually only varying slowly, while phases can shift on shorter timescales. With the observation of reference sources, only the slow variations can be corrected to a certain degree. For high dynamic range imaging, another method has to be used to reach the theoretical noise level.

One possible approach is to use only certain combinations of visibilities, which are independent of gain errors, the so-called closure phases and amplitudes. One example is the expression

$$\tilde{I}_{ij}^{\mathrm{uncal}}\, \tilde{I}_{jk}^{\mathrm{uncal}}\, \tilde{I}_{ki}^{\mathrm{uncal}} = |g_i|^2\, |g_j|^2\, |g_k|^2\, \tilde{I}_{ij}\, \tilde{I}_{jk}\, \tilde{I}_{ki} \quad , \tag{8.84}$$

whose phase is independent of calibration errors. Unfortunately, all closure quantities are non-linear combinations of several visibilities, which makes direct mapping very difficult. Fits of simple models are nevertheless possible.

A better approach is to include the gain corrections explicitly in the calculations and fit both the brightness model *and* the gain factors to the data. For each integration interval, $n(n-1)/2$ independent complex visibilities are available for an array of n stations. The n complex gain factors consist of $2n-1$ independent real parameters, because a global phase shift in all g_i does not change anything. Without calibration errors, $n(n-1)$ real equations would be provided by the data. After (implicitly or explicitly) solving for the gain factors, we are left with $n(n-3)+1$ real equations to fit the brightness model. For $n \gg 1$, we therefore do not loose much information in this process. In reality, the situation is even better, because gain factors do not change arbitrarily fast. For m integrations, the number of visibilities scales with m, but the number of gain factors scales more weakly.

The method of fitting the gain factors to a brightness model of the object in question itself is called ‘self-calibration’, because the data are used to calibrated themselves. Practically, the complete method works like follows:

- Start with an initial calibration using reference sources or self-calibration with a simple a priori model.
- Mapping: Build a brightness model from these data with a method like CLEAN, NNLS or MEM (maximum entropy).
- Self-calibrating: Fit the gain factors numerically to minimize the residuals between model and measured data. Possible variants include fixing the amplitudes or using a smoothed version of the best fits.
- Continue with mapping until the changes become insignificant.

In most situations, this mapping/self-calibrating loop converges to a sensible solution. When convergence is reached, the solution is automatically an optimal fit of the gain factors *and* brightness model to the data. For details, see Cornwell & Fomalont (1989, 1999) and Fomalont & Perley (1999).

Correction of phase errors is equivalent to adaptive optics in optical observations, where phase shifts are corrected by deforming optical elements of the instrument. The advantage in radio astronomy is, that the complete data, including the phases, of all stations (corresponding to different parts of the mirror in optical telescopes) can be recorded electronically. The phase corrections can thus be applied later in combination with the deconvolution process.

Chapter 9

LENSCLEAN

In the last chapter we discussed the reconstruction of arbitrary sources from radio interferometry data. We now want to modify one of the methods discussed before to adapt it for lensed images. We will start with a given lens model and try to find the 'best' brightness distribution which is exactly compatible with the given lens model. Secondary results will be a model of the unlensed source and, much more important for the following step, the residuals itself. We will then vary the lens model to minimize these residuals. In this way we find the best model for the mass distribution of the lens which is our main goal. The produced maps of the lensed and unlensed source are also presented but will not be discussed in detail. They are nevertheless interesting results that will help in the understanding of radio sources.

Our method of choice, LENSCLEAN, was first proposed by Kochanek & Narayan (1992) and Ellithorpe, Kochanek & Hewitt (1996). We use a slightly different formalism, which is more general and covers arbitrary weighting schemes. In contrast to our derivations, the treatment in the aforementioned publications is only valid for uniform weighting. Surprisingly, several approximations used in these papers for non-uniform weighting cancel each other, so that our exact result does not differ from the previous ones in the end.

In the original publication (Kochanek & Narayan, 1992) the computations were done with a CLEAN map as basis, which was then deconvolved with the CLEAN beam. This introduced bias effects, which are corrected by the 'visibility LENSCLEAN' method proposed by Ellithorpe et al. (1996). We follow this approach by doing all the statistics in *uv* space, where the noise properties are much better understood than in image space. Image space CLEAN is only used as approximation for *uv* space calculations. Corrections for the errors introduced by this are applied by different means.

Our developments on LENSCLEAN have been done over a long period and improved the algorithm and the implementation successively, until the quality proved sufficient to use it for lens model fitting for the system B0218+357. We started with the original algorithm, which was *not* successful with our test case B0218+357. The improvements introduced by several changes of the algorithm were each very small, which made the development difficult. Only the combination of many modifications improved LENSCLEAN significantly and made it useful for B0218+357. We do not try to describe this evolution here and omit most of the unsuccessful blind alleys we reached during this course. To illustrate the effects of the improvements, we use an artificial reference data set, the details of which will be described later. It was built using the best LENSCLEAN results for a VLA data set of B0218+357. Comparisons will not be made with the original algorithm but with the optimal version. This helps in interpreting the properties of all changes but hides most of the development difficulties. The value of a small improvement is always more significant when compared with the optimal algorithm than when compared with a very imperfect version.

9.1 Previous work

Only a small number of articles has been published about the LensClean method and results. The publications mentioned before were for the systems PKS 1830–211 (Kochanek & Narayan, 1992) and MG 0414+0534 and MG 1654+1346 (Ellithorpe et al., 1996).

Chen, Kochanek & Hewitt (1995) used LensClean with Clean maps of the system MG 1131+0456. They worked with smoothed power-law and de Vaucouleurs models. For the power-law models, the best exponent is $\beta = 0.6 \pm 0.2$, significantly different from isothermal. The core radius of the best model is very high, $r_{\mathrm{c}} = 0\overset{\prime\prime}{.}19 \pm 0\overset{\prime\prime}{.}007$. The large core radius is required to explain emission close to the lens position which could be interpreted as a central image. The authors take the result that the best models in fact *do* produce a central image as evidence that this interpretation is correct. The same result would, however, be expected if the observed emission is not a central image but instead caused by the lens galaxy. It would be interesting to repeat the calculations with subtraction of the central emission. We expect that the exponent β would shift in the direction of isothermal models while r_{c} would become compatible with zero. It is very suspicious that the small β can be interpreted as a correction for the large core radius in the outer regions. Chen et al. (1995) perform such calculations only with a fixed value of β. Without allowing changes of β simultaneously with r_{c}, the results are not very meaningful. It is not surprising, that the best core radius is again very high, because it may be needed as compensation for the possibly incorrect β.

The error statistics is calculated in image space and takes into account the regions of different image multiplicities. We prefer the *uv* space statistics for different reasons (see section 8.8.2 for a comparison of image space and *uv* residuals). The expected residuals can be calculated numerically with mock data sets which consist of noise only. We can thus avoid the rather arbitrary calculations of the effective number of parameters of the emission models.

Kochanek (1995) presented results of the image plane algorithm for MG 1654+134. Different models (de Vaucouleur, different types of power-law models) are tried. The power-law models are generally better fits to the data than de Vaucouleur, which is a clear hint for dark matter in the lensing galaxy. Small values of β need relatively large core radii, while isothermal models are happy with small core radii, even compatible with zero. The best fits are ellipsoidal almost isothermal models. If the lensing galaxies in MG 1654+134 and MG 1131+0456 are similar, these results are at least compatible with our aforementioned interpretation of the results from Chen et al. (1995).

An alternative method to the LensClean inversion, based on the maximum entropy method (MEM), was proposed by Wallington, Kochanek & Narayan (1996) under the name LensMEM. This method is complementary to LensClean in the same sense as MEM is complementary to Clean in the unlensed situation. MEM generally works better with smooth extended emission, while Clean is superior for ensembles of point sources. These properties are easy to understand. MEM and LensMEM optimize the solution by adding an entropy term to the residuals. This can be interpreted as a very specific kind of regularization, which is highly inappropriate for compact sources. The LensMEM algorithm calculates the residuals in the image plane, which is equivalent to *uv* space residuals only in certain cases (see section 8.8.2). Because of this and the expected problems with a system like B0218+357, which is dominated by two bright point-like images, we prefer LensClean before LensMEM.

9.2 Point sources

We start with the discussion of a lensed point source. Arbitrary sources will later be approximated by a collection of (very many) such sources at different positions. For a point source at position $\mathbf{z}_\mathrm{s}$ with (unlensed) flux density S, the given lens model predicts the number of images n, the positions of the images $\mathbf{z}_k$ and their (absolute) amplifications μ_k. All these parameters, including n, depend on the source position. For the given source, we therefore expect n images with the given positions and apparent flux densities $S_k = \mu_k S$. The measured visibilities for this source would be

$$\tilde{I}^\mathrm{m}(\mathbf{u}) = S \sum_{k=1}^{n} \mu_k \mathrm{e}^{2\pi\mathrm{i}\,\mathbf{u}\cdot\mathbf{z}_k} \quad . \tag{9.1}$$

If the fitting method puts a component at a position $\mathbf{z}_k$, this corresponds to a source component at $\mathbf{z}_\mathrm{s}$. If this source component may be multiply imaged, the method has to include the other images of the same source component as well. Otherwise, the brightness model would be incompatible with the given lens model.

9.3 LENSCLEAN in a nutshell

Explained in a short and simplified way, LENSCLEAN works like follows.

- Start with the dirty map and iterate the following loop, just as in standard CLEAN.
  - Find the peak[1] in the residual map.
  - Use the given lens model to determine the positions and amplifications of all images corresponding to the same source position as the given peak.
  - Shift the dirty beam to all these image positions and scale it proportionally to the amplification.
  - Subtract the scaled and shifted beams from the dirty map.
  - Add the positions and flux densities to the list of CLEAN components. If needed, also build a list of components in the source plane.
- Build the CLEAN map from the list of components by adding the CLEAN beam at the given positions with the given flux densities. This is equivalent to a convolution with the CLEAN beam.
- Add the final residual map to the CLEAN map, if the residuals are still not negligible. This last step is only useful if the residuals consist of noise only. Residuals which are incompatible with the lens model should never be included in the CLEAN map.

The residuals, quantified in an appropriate way, can be used to judge the lens model itself. The model parameters are then varied to minimize the residuals. In this way the best fitting lens model is found. We now want to describe the details and give a mathematical justification for the algorithm.

---

[1] This is a very crude simplification. Details will be explained later.

9.4 The lens model and our test case

The main reason for our LENSCLEAN work was the lens system B0218+357, which was discussed before. Without LENSCLEAN, it has not been possible to constrain the position of the galaxy centre with sufficient accuracy to determine the Hubble constant. We therefore developed a new version of LENSCLEAN, working on the available data for B0218+357. To keep things simple, most of this work was done with the singular isothermal ellipsoidal potential lens models (‘SIEP’).

We used these models for two reasons. The first is a practical one; the lens equation can be inverted analytically for SIEP models (see section 4.5), which greatly reduces the burden of coding and computation. The numerical LENTIL algorithm, which allows the use of arbitrary lens models, was developed only after the main work for LENSCLEAN was finished. We will present some preliminary results for other lens models later.The second reason is the fact that the main problem of lens models for B0218+357, the unknown lens position, has its full impact already for this family of models. Effects of more general mass models can only be relatively small corrections to the SIEP results.

We already discussed the problems with B0218+357 earlier (section 7.6). If we only use the two compact components as constraints, we have two two-dimensional positions and two flux densities to constrain the lens models. The position and flux density of the unlensed source have to be included in the models implicitly or explicitly, so that the compact images can provide only three effective constraining equations for the mass models. This is true regardless of the method we use to utilize the constraints. LENSCLEAN without the ring would therefore just as classical model fitting not be able to determine all five parameters (total mass scale α_0, ellipticity ε, and the lens centre’s position) of the model. Here the ring comes into play. It provides a number of additional constraints which make the unique definition of the best lens model possible. The flux density from the ring is much smaller than from the compact components, however. We therefore have to take extreme care to avoid any errors or bias effects introduced by the compact images, because even small residual errors from these could easily dominate the total residuals. The effects of the ring would then be obscured by numerical noise.

9.5 Strategy to fit lens models for B0218+357

Numerical noise and bias effects, which were still very strong in the original LENSCLEAN algorithm, made it impossible to perform a straightforward fit of all five model parameters simultaneously. The three constraints from the compact images are so strong compared to the ones from the ring, that the eigenvalues of the curvature matrix of the residual function have very different orders of magnitude. In other words the residual function has very narrow valleys in the five-dimensional parameter space. In two directions, the gradients are very steep, while they are shallower by several orders of magnitude in other directions. This, in combination with the fact that these valleys are not straight but curved, poses extreme difficulties for any minimization algorithm. Even very small numerical noise can be much stronger than the shallow slopes locally, making it impossible to follow the curved valleys to finally find the global minimum. For most calculations we used the downhill-simplex minimization algorithm (Press et al., 1992). More sophisticated methods like Powell’s algorithm do not work well with LENSCLEAN residuals because of the numerical noise. Downhill-simplex is quite robust and proved to be at least as efficient and reliable as other methods in our case.

To overcome the difficulties, a very primitive but powerful approach was used. We scanned the plane of possible lens centre positions z_0 with a regular grid and performed fits of the other parameters for all these positions. In the LENSCLEAN model fitting, the z_0 are fixed, so that only three parameters have to be determined by LENSCLEAN. The information from the bright compact images alone is sufficient to constrain these parameters very well, so that the minimization algorithm now has much less problems to find the minimum. Interpreted in the picture of finding the lowest point of a narrow and curved valley, we perform cuts through the valley. We can then easily find the centre in each of these cuts. Later we plot the values of all this central points to find the globally lowest point.

The plots of residuals of the best fitting lens models for different centre positions proved to be valuable not only for finding the global minimum, but also to detect numerical difficulties and other fundamental problems. A two-dimensional plot provides much more information than its minimum alone.

The disadvantage of this method is obvious; we have to perform LENSCLEAN fits for each z_0, which typically means some 100 fits for each plot, each fit taking about 100 iterations of LENSCLEAN, each iteration with several thousand CLEAN loops or up to about 10^8 LENSCLEAN loops for one plot. Each of these loops is much more costly than a non-lensed CLEAN loop, one taking approximately 0.1 sec. In one example, we needed 41 days of CPU time on modern PCs (Intel Pentium II-III with 200–800 MHz) to produce one plot of 212 points with about 70 iterations and 2000 LENSCLEAN loops each.

Very many of such plots as well as other calculations had to be done, mainly to improve the algorithm and perform extensive tests of all its aspects. In total ca. 14 years of CPU time have been used. Almost all of the calculations were done on normal PCs. Parallelization was easy for most of the calculations, so that up to about 30 PCs at a time could be used for one run. This use of normal PCs, mainly during the night time or in otherwise idle periods, provided enough flexibility to do all the development work. With the finished algorithm, final ‘production runs’ should better be done on faster machines. For these runs, it does not matter if the jobs have to be queued for several days before execution, which would be deadly in the development phase.

9.6 Tests of different LENSCLEAN variants

To study the effects of several LENSCLEAN modifications, we produced an artificial data set with similar properties as the 15 GHz VLA data set (Stokes LL part). It was built from an emission model which is the result of LENSCLEANing the real data with one of the best lens models. The lens model parameters are the following:

$$z_0 = \left(0\overset{\prime\prime}{.}264\,174, 0\overset{\prime\prime}{.}121\,697\right) \tag{9.2}$$

$$\alpha_0 = 0\overset{\prime\prime}{.}158\,213 \tag{9.3}$$

$$\varepsilon = (-0.009\,231, -0.083\,508) \tag{9.4}$$

One version of the data set was used directly, while Gaussian noise with the same properties as in the real data was added to the second version. No calibration errors were included for most of the calculations.

To keep the numerical effort reasonable, only a one-dimensional family of lens models was tested with this data set. Comparison of the results is also easier with one-dimensional curves than with two-dimensional maps. The lens centres of this models were chosen to be located on a straight line connecting the best fitting spherical model from equation (9.21)–(9.23) and the correct model (9.2). The range was extended in both directions beyond these two models. We

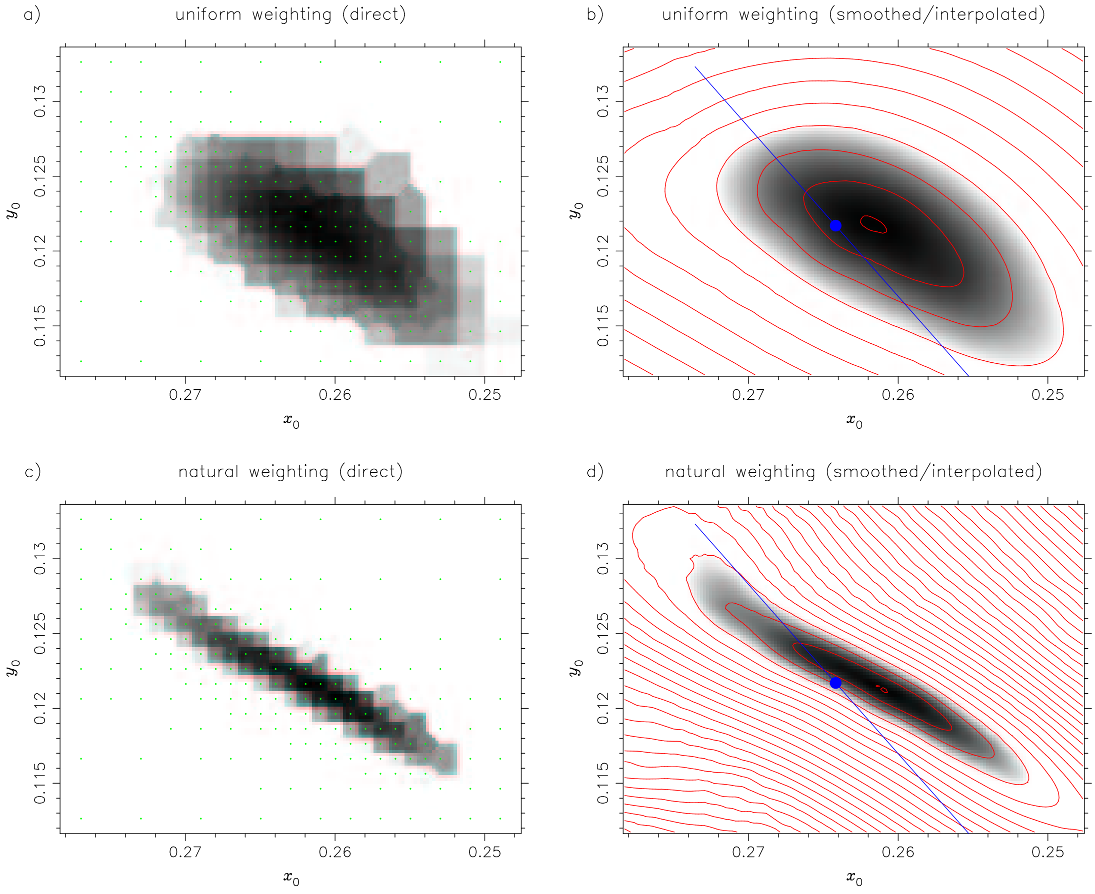


Figure 9.1: Residuals for the test data set. (a) and (b) (top) show uniformly weighted results for the fitted models (5000 iterations), (c) and (d) (bottom) are naturally weighted residuals for the same lens models ($10\,000 + 5000$ iterations). (a) and (c) (left) are direct plots. The positions for the fits are marked by dots. (b) and (d) (right) are a smoothed/interpolated version of the same data with contour lines (confidence limits of $0.1, 1, 2, 3, \ldots \cdot \sigma$). The straight line marks the one-dimensional family of lens models used for the numerical tests. The circle marks the correct model.

then performed a complete LENSCLEAN lens model fit of the other parameters for 128 positions on this line. The best version of LENSCLEAN as described later was used for this purpose.

Figure 9.1 shows a two-dimensional map of residuals with the one-dimensional family of models and the correct lens model marked. The smoothed version was calculated by performing a local quadratic fit. This was done by fitting a quadratic polynomial to the complete data set with Gaussian weights depending on the distance to the position for which the interpolation is calculated. The Gaussian's σ was chosen as the radius of the smallest circle which encloses at least 6 data points. 6 is the number of parameters of the quadratic form. This form of smoothing proved to be superior to other approaches, especially to global polynomial fits. Any kind of simple convolution smoothing would smear the minimum. This is not the case for the local quadratic fit.

With uniform weighting, the minimum of residuals is displaced from the correct lens model by about 2.5 mas. This is the remaining error of even the best LENSCLEAN variant. On the line of test models, the minimum is located more or less at the correct model.

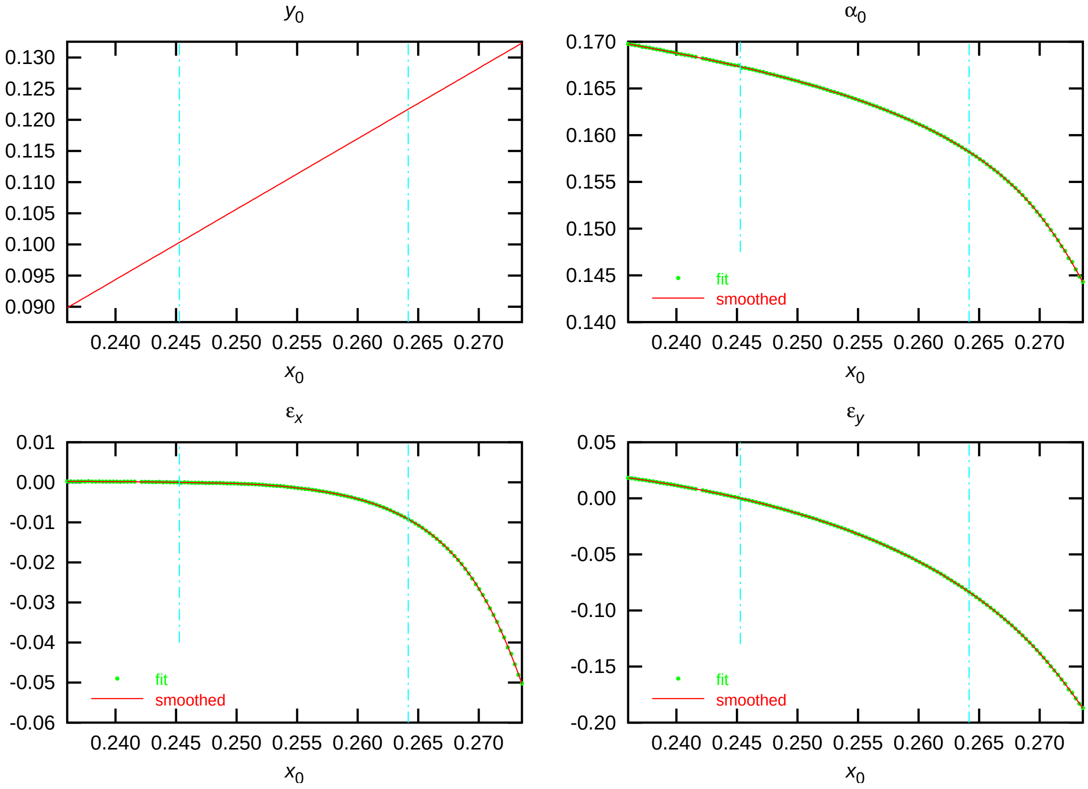


Figure 9.2: Lens models for the numerical tests of different LENSCLEAN variants. x_0 was used as independent variable, z_0 is located on a straight line. The parameters α_0 and ε were fitted for each z_0 to minimize the residuals. Dots show these fits, lines show a smoothed interpolation. The two are indistinguishable in these plots. The vertical lines mark the spherical (left) and correct (right) model.

With natural weighting (LENSCLEAN performed with 10 000 uniformly weighted iterations followed by another 5000 with natural weighting), the error is slightly larger, ca. 3.5 mas.

The residuals as well as the model parameters are rather smooth functions of x_0, which was used as independent variable. To prevent the remaining fluctuations from affecting the results, we also calculated a smoothed version of the model parameters (see Figure 9.2). A local cubic fit was used for this purpose[2]. With modified LENSCLEAN versions, we then calculated residuals for these fixed models. In this way, we could avoid very expensive complete LENSCLEAN fits for each test. Of course, this approach relies on the assumption, that the LENSCLEAN version used to determine the models is close to optimal. Extensive numerical tests with real and artificial data confirm this assumption.

The fits were performed with 5000 iterations and uniform weighting. To illustrate the effect of smoothing the lens model parameters, we also calculated residuals with exactly the same LENSCLEAN algorithm as in the fits but with the smoothed lens models. This was done with uniform weighting for 10 000 iterations before switching to natural weighting and performing another 5000 iterations. Figure 9.3 shows the residuals for the fits and for the smoothed models.

The most prominent feature in the residual plots is the minimum very close to the correct

---

[2] Numerically, we performed a global cubic fit to the data for each x_0 position with Gaussian weights for all data points. A Gaussian σ of 2 mas was used.

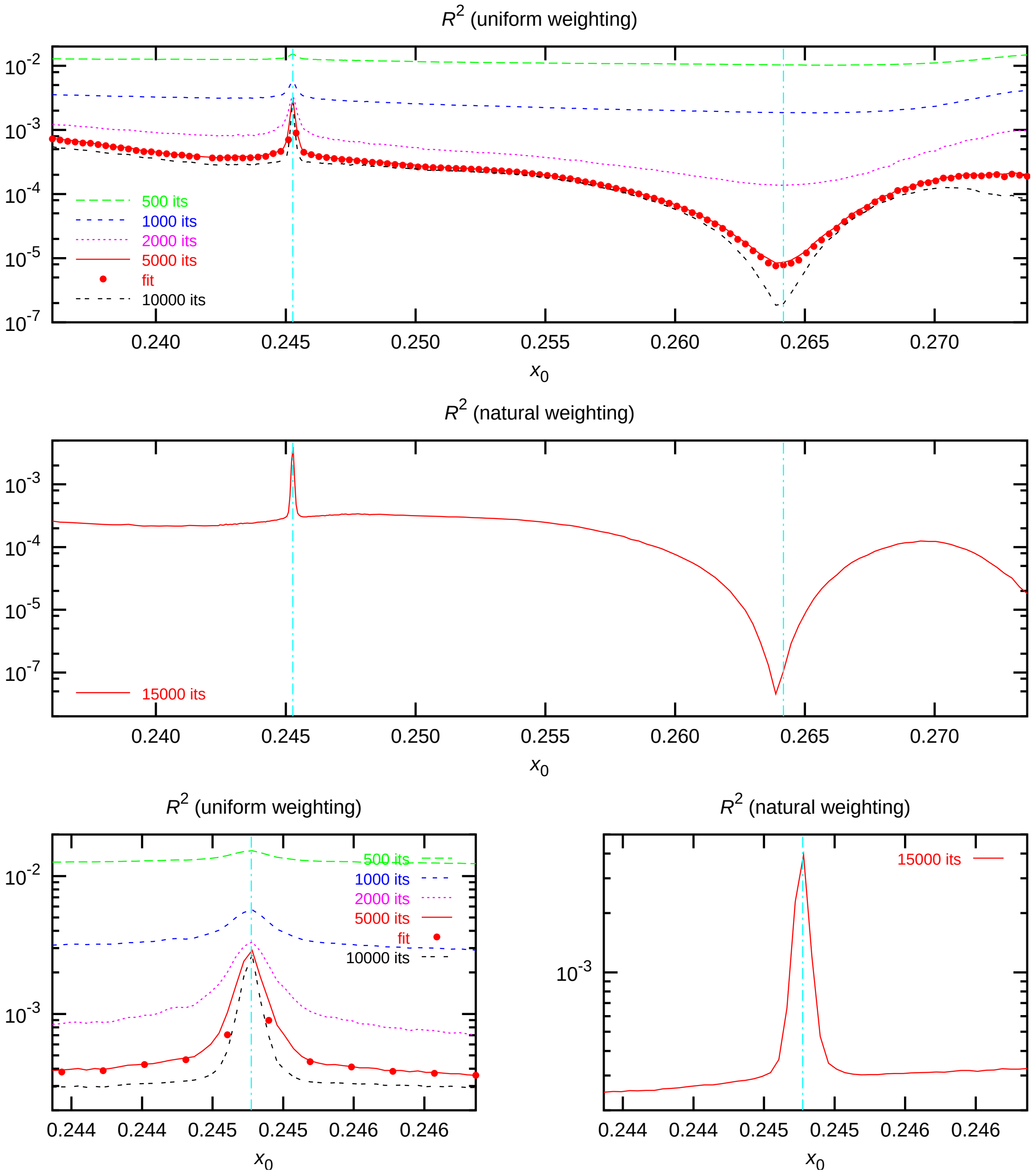


Figure 9.3: LENSCLEAN residuals for the test models, calculated with the optimal LENSCLEAN version and noiseless data (logarithmic). The nearly spherical models are shown magnified in the bottom panels. The maximal ε in this range is very small (ca. 0.004). The vertical lines mark the spherical (left) and correct (right) model.

model. The accuracy in x_0 is much better than 1 mas. The difference between the residuals of the fits and the residuals for the smoothed parameters is small and (most important) shows no systematic effect near the minimum. The minimum becomes more prominent for more iterations. Qualitatively, the behaviour is the same for natural and uniform weighting, apart from a decline for very high x_0 with natural weighting.

We also notice strange peaks near the spherical model. For models very close to the spherical case, the residuals are much higher than for more elliptical models. After more iterations the effect becomes more prominent, but the region in which the results are affected becomes smaller. This behaviour is suspicious, because spherical models are more simple than elliptical ones. We would therefore prefer to trust the results for spherical models and interpret the lower residuals for other models as an artifact of the algorithm. This simple picture is not correct, however. We will see later in section 9.13 that an amplification limit is responsible for this behaviour. Without the limit, the strange peak disappears, but the residuals become slightly more noisy then.

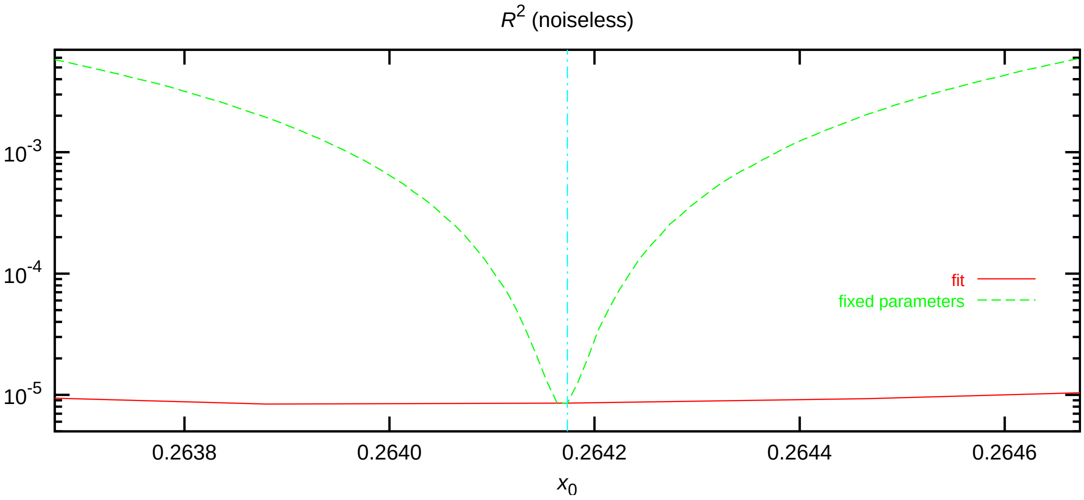


Figure 9.4: LENSCLEAN residuals for the test models with fixed parameters α and ε calculated for noiseless data. The result for the fitted parameters is also shown for comparison (see Fig. 9.3). The vertical line marks the correct model. Note the highly magnified scale of x_0.

The residuals in Figure 9.3 are mainly determined by the ring, because the two compact components can always be fitted exactly. The accuracy of LENSCLEAN, of the order a few mas is impressive, given the low resolution of the data (ca. 90 mas) and the absence of sharp features in the ring. To illustrate the potentially even higher accuracy in the presence of compact emission, we present results for the same lens centres with the other parameters (α and ε) fixed in Figure 9.4. With these lens models, the two compact bright components themselves also contribute to the residuals and produce a very sharp minimum. The accuracy in this case is about 0.01 mas in the noiseless as well as in the noisy case.

The effect of noise can be seen in Figure 9.5, which was calculated with the same lens models but with noisy *uv* data. We notice that the uniformly weighted residuals with enough iterations show the same behaviour as in the noiseless case. The minimum is still very well defined and located at the same correct position. The direct comparison with the shifted noiseless residuals for 5000 iterations shows that the slope of the curve becomes steeper in the noisy case, which appears unnatural at first sight. Naively, we expect deeper minima without noise. A possible explanation is, that the efficiency of LENSCLEAN is reduced by noise, and this effect is stronger for incorrect lens models where many more iterations are needed to achieve a low residual level.

For natural weighting, we notice a significant shift of the minimum in the one-dimensional cut. This is mainly due to the change of shape in Figure 9.1. In this plot, we see that the confidence region shrinks quite drastically in one direction for natural weighting, while it is more or less unaffected in the other direction, which is more important for H_0, see Figure 7.8 on page 82. Other numerical experiments showed that uniform weighting is generally superior

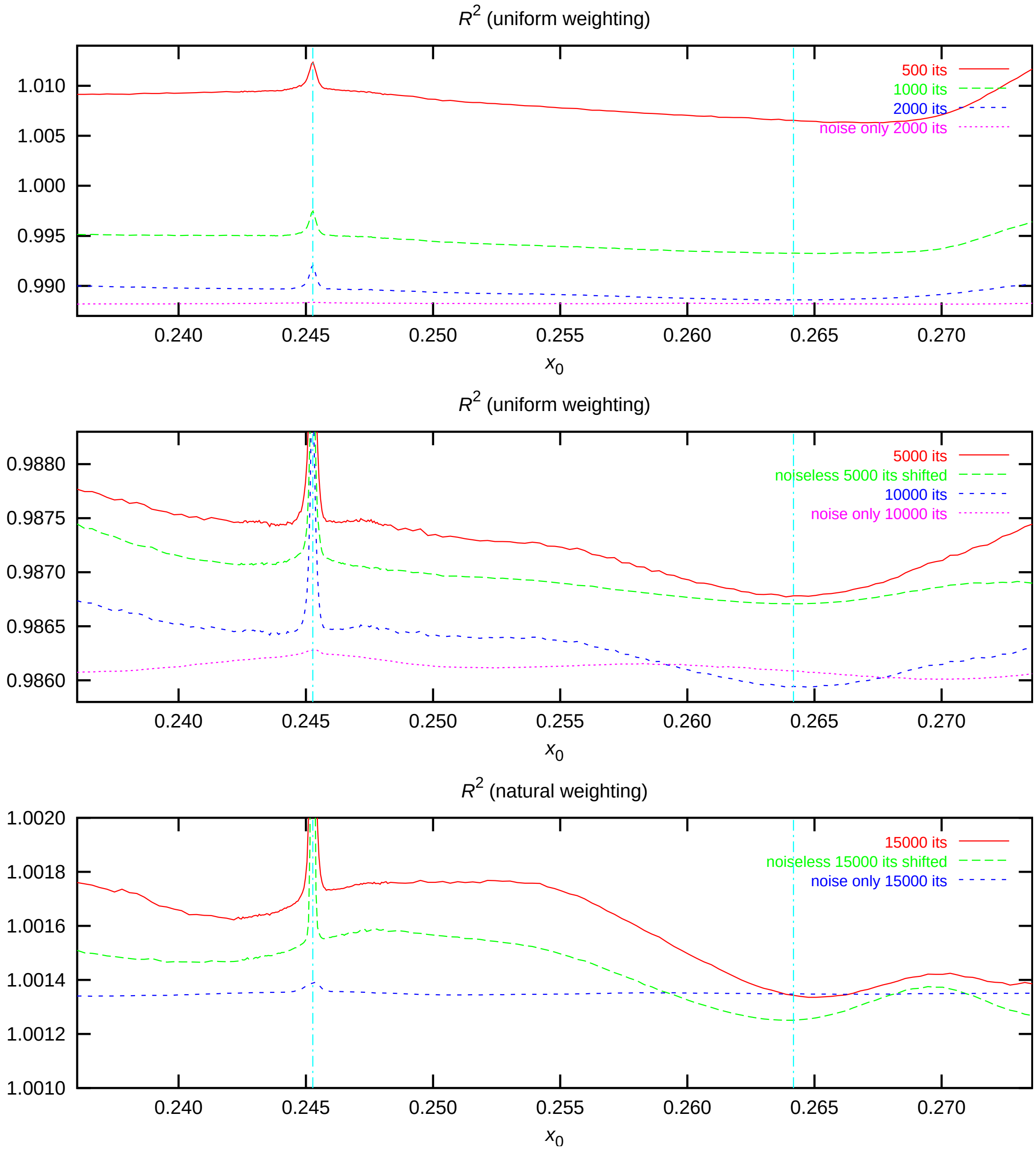


Figure 9.5: LENSCLEAN residuals for the test models, calculated with the optimal LENSCLEAN version and noisy data (5000 iterations, uniform weighting). A shifted version of the noiseless case is included for comparison. The 'noise only' curve was calculated with a data set consisting of noise only without a signal. The vertical lines mark the spherical (left) and correct (right) model.

to natural weighting regarding the accuracy of determined lens models. The formal statistical errors may be smaller for natural weighting, but possible bias effects are often larger. In some cases, they can then dominate the errors, making formal statistical error measures meaningless. To some extent, this can even be seen in Figure 9.1. For uniform weighting, the true error is well inside the 1σ-limit, while it is almost at the 2σ-limit for natural weighting. The true error is slightly smaller for uniform weighting in this case.

With changing lens model parameters, the effective number of emission parameters will generally change as well, because the regions of different image multiplicities depend on the lens model. Singly imaged regions leave more freedom for the emission model than multiply imaged regions. The limit of maximal freedom is reached with no lens at all. To estimate the bias effects from the varying effective number of model components, we included results for a noise only data set in Figure 9.5. For very many iterations, these residuals are surprisingly larger than the total ones near the correct model. This is an effect of the non-linearity of LENSCLEAN. The slope of the noise only curve is shallow enough to be neglected in the determination of the best fit. A correction can to first order be applied by subtracting the noise-only residuals from the real residuals. The shift in x_0 would be smaller than 1 mas.

9.7 Accuracy

With natural weighting, the accuracy of the results can be estimated easily, using the results from section 2.7. The fit of M parameters leads to a residual minimum lower by $\Delta\chi^2$ than the residuals for the correct lens model. The difference $\Delta\chi^2$ follows a χ^2-distribution with $\nu = M$, which has a mean of M and a variance of $2M$. For the normalized residuals, the expected difference is M/N. Our data have 104 007 good complex visibilities, which means $N = 208\,014$. For the lens centre's position, the relevant parameters are x_0 and y_0, which means $M = 2$. The other parameters are not relevant for the residual difference here (see section 2.7). The limiting values for the 68.4 %, 95.4 %, and 99.73 % confidence intervals (corresponding to 1, 2, and $3\,\sigma$ of a Gaussian distribution) for $M = 2$ are $\Delta\chi^2 = 2.30$, 6.17, and 11.8.

We show the confidence limits for our test data set in Figure 9.6. Please note that y_0 is not the optimal one but chosen to be located on a straight line. We merely present this plot to give a crude estimate of the accuracy. The error limits of the real results will be discussed in section 9.21.2.

For uniform weighting, we use the approximation explained in section 2.5, especially equation (2.67). We simply scale the expected residual differences $\Delta\chi^2$ from above with the expected $\Delta R^2/R^2$ for one parameter, which in the case of our data set is $3.355 \cdot 10^{-5}$ (compared with $\Delta\chi^2/\chi^2 = 1/N = 4.807 \cdot 10^{-6}$). We note that the real error is much smaller for uniform weighting, whereas the expected errors are smaller for natural weighting. For natural weighting, the true error is at the $1\,\sigma$ limit. For uniform weighting, it is much smaller than the expected error. This might indicate that the error limit estimate for uniform weighting is too conservative. This is, of course, small number statistics of the worst kind ($n = 1$).

Monte Carlo simulations were used to count the effective number of parameters of the emission model. We performed 100 LENSCLEAN runs with the (fixed) correct lens model and the artificial *uv* data set with different noise realizations. We compared the residuals χ_0^2 for the true emission model, as it was used in building the data set, and the resulting LENSCLEAN residuals (5000 iterations for uniform, $10\,000 + 5000$ iterations for natural weighting). The result of the former is $\overline{\chi_0^2} = 0.9982$, the expected value is $\overline{\chi_0^2} = 1 \pm 0.003$. The mean difference of $\Delta\chi^2 = 416$ correspondsto $M = 416$ parameters for the emission model. The expected scatter of this value is then 28.8, almost equal to the numerical value of 29.2. This confirms that LENSCLEAN residuals have the same statistical properties as real fits of the emission model.

If we assume that the number of free parameters M can be approximated by the number of beams filling the total map area, the diameter of these beams would be about 71 mas, which is smaller than the FWHM of even the uniformly weighted beam (about 83 mas) but of the same order of magnitude.

The uniformly weighted residuals are $\overline{R^2} = 1.0047 \pm 0.0084$ (mean and RMS scatter) com-

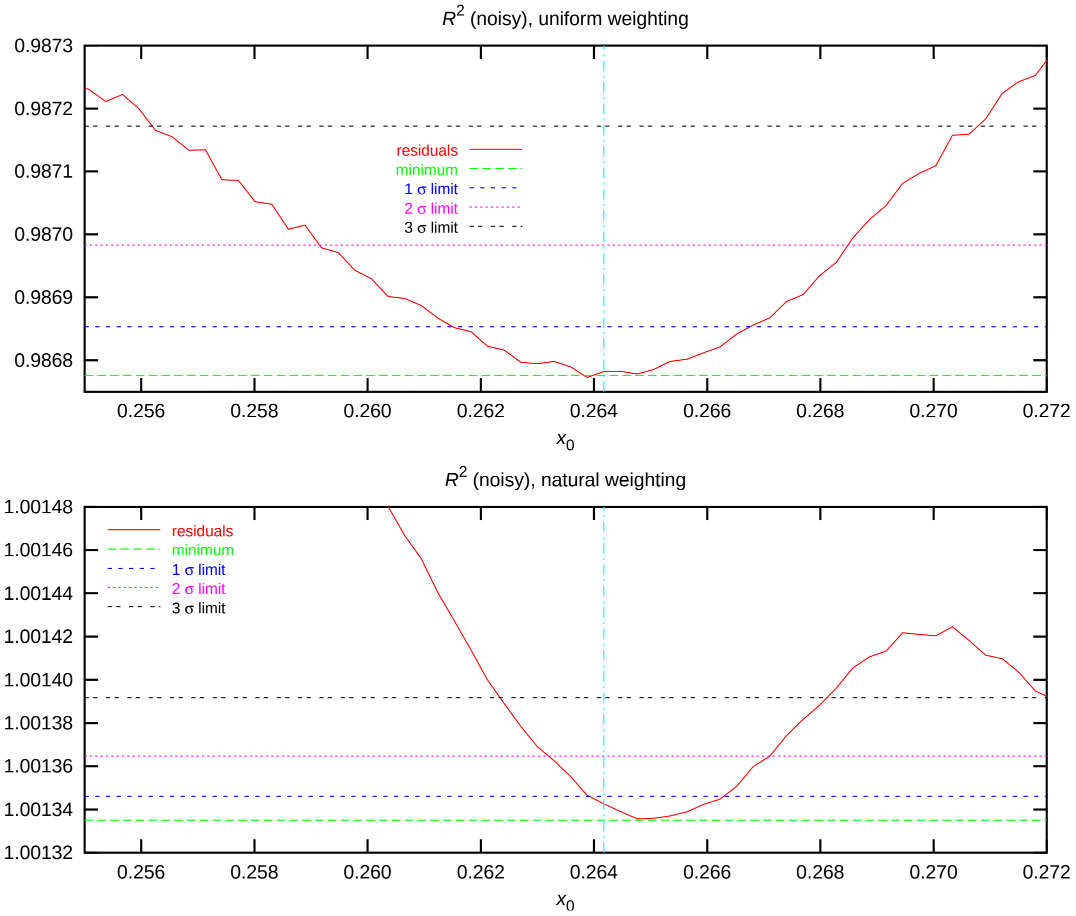


Figure 9.6: LENSCLEAN residuals for the test models, calculated with the optimal LENSCLEAN version and noisy data. The upper plot is uniformly weighted (5000 iterations), the lower one naturally (15 000 iterations in total). Error limits (1, 2, 3 σ) are included in both plots, the limits for uniform weighting are estimates. The vertical line marks the correct model.

pared to the theoretical expectation of $\overline{R^2} = 1 \pm 0.0091$. The numerical result for the difference is $\Delta\overline{R^2} = 0.0179 \pm 0.0023$. This is in moderately good agreement with the approximated expectation of $\Delta\overline{R^2} = 0.014$ from $M = 416$ with equation (2.67).

As expected, χ_0^2 and $\Delta\chi^2$ are uncorrelated for natural weighting. For uniform weighting, there is a significant correlation between R_0^2 and ΔR^2.

9.8 The original LENSCLEAN component selection

In unlensed CLEAN, the next component to subtract is always the peak in the residual dirty map. We have to generalize this selection scheme for the lensed case. To do this, we go back to the mathematical foundations of standard CLEAN (section 8.8.4) and try to apply the same concept for lensed sources. The idea in standard CLEAN was to select as next component the one which minimizes the residuals *in this step* and multiply the optimal flux with a small loop gain γ (section 8.8.3) to stabilize the algorithm and improve the resulting maps.

Analogously to (8.56), we calculate the residuals after the subtraction of a source component as described in section 9.2.

$$R^2 = \sum_j w_j \left| \tilde{I}_j - S \sum_{k=1}^{n} \mu_k \mathrm{e}^{2\pi \mathrm{i}\, \mathbf{u}_j \cdot \boldsymbol{z}_k} \right|^2 \tag{9.5}$$

$$= \sum_j w_j \, |\tilde{I}_j|^2 - 2S \sum_j w_j \tilde{I}_j \sum_{k=1}^{n} \mathrm{e}^{-2\pi \mathrm{i}\, \mathbf{u}_j \cdot \boldsymbol{z}_k} + S^2 \sum_j w_j \left| \sum_{k=1}^{n} \mathrm{e}^{2\pi \mathrm{i}\, \mathbf{u}_j \cdot \boldsymbol{z}_k} \right|^2 \tag{9.6}$$

We can split this in the constant and variable parts:

$$R^2 = R_0^2 - \sum w_j \, \Delta R^2 \tag{9.7}$$

$$R_0^2 = \sum_j w_j \, |\tilde{I}_j|^2 \tag{9.8}$$

$$\Delta R^2 = 2S \sum_{k=1}^{n} \mu_k I_k - S^2 \sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'} \tag{9.9}$$

Here we used the definitions

$$I_k = I_{\mathrm{D}}(\boldsymbol{z}_k) \quad , \tag{9.10}$$

$$B_{kk'} = B(\boldsymbol{z}_k - \boldsymbol{z}_{k'}) \tag{9.11}$$

for the dirty map and dirty beam. Please remember that n as well as the $\boldsymbol{z}_k$ and μ_k depend on the source position. For given $\boldsymbol{z}_{\mathrm{s}}$, the optimal source plane flux is

$$S = \frac{\sum_{k=1}^{n} \mu_k I_k}{\sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'}} \quad , \tag{9.12}$$

with resulting residual difference

$$\Delta R^2 = \frac{\left(\sum_{k=1}^{n} \mu_k I_k \right)^2}{\sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'}} \quad . \tag{9.13}$$

The position $\boldsymbol{z}_{\mathrm{s}}$ now has to be chosen in a way to maximize ΔR^2. This is the generalization of finding the peak in the dirty map, which is used in standard CLEAN. If we apply a loop gain γ to the optimal flux S, the residual reduction in (9.9) scales with a factor $\gamma(2-\gamma)$.

LENSCLEAN now continues by subtracting the shifted beam at all positions $\boldsymbol{z}_k$ simultaneously with flux densities $\gamma \mu_k S$. In contrast to standard CLEAN, the image plane residuals do not usually converge to zero, because an exact fit will not always be possible. Several stopping criteria for the algorithm have been discussed. The most simple one, which we regularly use, is a constant number of iterations. A limiting value for ΔR^2 or other criteria can also be used.

9.9 Unbiased LENSCLEAN

The recipe of the last section can be used to find one of the best fitting solutions and determine the residuals, if very many iterations can be performed. There are, however, several problems with this approach. First, some regions of the lens plane are CLEANed with higher priorities than others, depending on the lens model. This certainly leads to bias effects, when compared to unlensed CLEAN results with the same data. In some cases, they may improve the results and are thus advantageous, but usually they deteriorate the produced maps significantly.

Another aspect, which is even more important in our context, is the fact that CLEAN and LENSCLEAN usually converge extremely slowly in the later stages (remember the numerical experiments of section 8.9.1). We therefore have to worry, if the remaining residuals after stopping the CLEANing, say after a fixed number of iterations, are really representative for the residuals of the converged result. We later want to use these residuals to judge the ability of the lens model itself to fit the data and have to assure that this can be done reliably with the numerically calculated residuals. The residuals depend only very weakly on the lens model parameters, and even weak bias effects have to be avoided.

To improve the selection of CLEAN component positions and fluxes, we try to achieve equal reductions of the residuals for equivalent lens models in each step. We can then be sure that, after a fixed number of iterations, equivalent lens models lead to the same residuals. This can not be done for arbitrary lens models and source brightness distributions in a simple way. Since compact emissions are CLEANed away very quickly and most bias problems we encountered in our work were related to more or less smooth extended emission, we will concentrate on the latter.

The most simple source in this context consists of a constant surface brightness I in the region of interest. As lensing conserves surface brightness, *all* lens models are equivalent in this case. This includes the case of no lens at all. The residual reduction in one iteration can then be calculated following (9.13). It depends on the dirty beam and (via the image positions and amplifications) on the lens model and the source component position. To understand this dependence better, we can approximate ΔR^2 for well separated images, for which the dirty beams do not overlap significantly and $B_{kk'} \approx \delta_{kk'}$. If the amplifications of the n images are furthermore all equal, the residual reduction is

$$\Delta R^2 \approx \gamma(2-\gamma)\, n I \quad , \tag{9.14}$$

if we include a loop gain γ. In this case LENSCLEAN will preferably remove images of higher multiplicities to reduce the residuals more quickly. Since the success of CLEAN relies on the uniformity of component selection, the resulting maps will not only be worse than without a lens, but they will also depend on the lens model parameters, even though all models are equivalent. Even more important for us is the fact that the residuals will depend on the lens model. They are therefore no good measure for the goodness of fit of the lens model itself. This problem made it impossible to obtain useful results for B0218+357 with the originally published LENSCLEAN. We call this old version 'KNE-LENSCLEAN' from now on (standing for the authors of Kochanek & Narayan, 1992 and Ellithorpe et al., 1996).

We tried several methods to correct for this bias effects. The most simple idea was to adjust the loop gain γ to correct for the factor n in equation (9.14). The best solution so far exactly corrects for the lens model dependence for constant I. This is done by applying a correction factor to ΔR^2 which only depends on the lens model and image position, but not on the residual map:

$$\Delta R^2_{\text{corr}} = g \, \frac{\sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'}}{\left(\sum_{k=1}^{n} \mu_k \right)^2} \Delta R^2 \tag{9.15}$$

$$= g \left(\frac{\sum_{k=1}^{n} \mu_k I_k}{\sum_{k=1}^{n} \mu_k} \right)^2 \tag{9.16}$$

Here we have applied a factor g that plays the role of a loop gain. We now have to adjust the source plane flux S to achieve exactly this ΔR^2_{corr} after subtracting the components with $\gamma \mu_k S$. Since the residuals are quadratic in S, two solutions are possible. We choose the smaller one which converges to zero if g gets very small.

$$S_{\text{corr}} = \gamma S \tag{9.17}$$

$$\gamma = 1 - \sqrt{1 - g \frac{\sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'}}{\left(\sum_{k=1}^{n} \mu_k \right)^2}} \tag{9.18}$$

This expression cannot be interpreted easily. Because CLEAN works best for small loop gains anyway, we also want to present the limiting case for small g:

$$\gamma = \frac{g}{2} \, \frac{\sum_{k=1}^{n} \sum_{k'=1}^{n} \mu_k \mu_{k'} B_{kk'}}{\left(\sum_{k=1}^{n} \mu_k \right)^2} \tag{9.19}$$

$$S_{\text{corr}} = \frac{g}{2} \, \frac{\sum_{k=1}^{n} \mu_k I_k}{\left(\sum_{k=1}^{n} \mu_k \right)^2} \tag{9.20}$$

The last expression has a simple interpretation; the source plane flux is a weighted (with μ_k^2) mean of the individual source plane fluxes I_k/μ_k scaled with the loop gain factor. This method to reduce the bias effects is a rigorous and uniquely defined choice for constant surface brightness sources. In other cases it should at least reduce the errors and seems to be the optimal way to handle this problem in the framework of LENSCLEAN. Please note that both manifestations of the bias problem are corrected by the adaptive loop gain: the uneven CLEANing and the biased residuals.

Another derivation of the same component selection scheme can be obtained by using the subtraction of a fixed small total flux density in the lens plane. The optimal position is then the same as in the maximum of equation (9.16). The source plane flux is then given by the fixed lens plane total flux divided by the sum of amplifications. The adaptive loop gain γ for this approach differs from our unbiased method, of course.

9.9.1 Effects on the quality of maps

To test the effect on the quality of the produced maps, we performed tests with a noiseless artificial data set. For the lens, we used a SIEP model without external shear. It is centred on $z_0 = (0,0)$ and has $\alpha_0 = 0.158$ and $\varepsilon = (-0.009, -0.084)$, which is very similar to the best LENSCLEAN models for B0218+357. For the source, we used a constant surface brightness disc around z_0 with radius 80 mas to produce a wide smooth ring. For the numerical calculations, the disc was approximated by about 5000 components, leading to about 12 000 lens plane components. We performed runs with 5000 LENSCLEAN iterations and a gain of $\gamma = 0.05$ for KNE-LENSCLEAN and $g = 0.2$ for the unbiased algorithm. These values are equivalent for $n = 2$ equally bright and well separated images, cf. (9.20). The *uv* coverage and weights were taken from the 15 GHz VLA data set.

Figure 9.7 shows the image multiplicities (number of images corresponding to the same source position), the bias factor, which is the reciprocal of the correction factor in (9.15), the amplification of the images, and the sum of amplifications for all images corresponding to the same source position. We see that KNE-LENSCLEAN especially favours CLEANing in the quadruple image regions.

Figures 9.8 and 9.9 (page 148 and 149) show the ΔR^2 and squared residual map after 0, 200, 1000, and 5000 iterations with the noiseless artificial data set. We realize that ΔR^2 is initially dominated by the bias factor (compare with Figure 9.7b). The quadruply imaged region is preferably CLEANed, leading to a negative imprint of the bias factor on the residual map after 200 iterations. This shape is still noticeable at later stages. The non-uniform CLEANing can potentially deteriorate the result at the end.

Results of analogous calculations are shown for our new unbiased LENSCLEAN technique in Figures 9.10 and 9.11 (page 150 and 151). The CLEANing is much more uniform now. Hints of effects of the lens model are only noticeable at later stages at a far smaller residual level.

The results for the final CLEAN map are compared in Figure 9.12 on page 152. We present the real model which was used to build the artificial *uv* data set, a non-lensed CLEAN result produced with DIFMAP with 10 000 iterations and $\gamma = 0.05$ to achieve a residual level very similar to the LENSCLEAN results, and finally the results from KNE-LENSCLEAN and the unbiased modification. The latter is a far better approximation to the true model, as can be seen by the smaller residuals in the background. Surprisingly, the unlensed CLEAN map is even slightly superior to the best LENSCLEAN result in this case. One should keep in mind, however, that all residuals seen in Figure 9.12 are very small and would be hidden in the noise in most cases. The bias effects in standard LENSCLEAN are therefore not a major problem for the produced maps. The more important effect on the fitted lens model will be investigated next.

9.9.2 Effects on the residuals

The main goal in this work is to use LENSCLEAN to constrain mass models of gravitational lenses. To be able to judge the goodness of fit of the lens model, we have to calculate meaningful residuals. The final residuals after convergence of LENSCLEAN are not affected by the bias effect and could therefore be used to find the best model. Unfortunately, convergence is extremely slow, and small bias effects on the rate of residual decline are a matter of concern. If we stop the iteration after a fixed number of loops or use any other simple stopping criterion, the residual differences can be dominated by the bias effect.

To compare our unbiased LENSCLEAN with the KNE algorithm, we use the models from section 9.6. The results presented there were calculated with the best version of unbiased

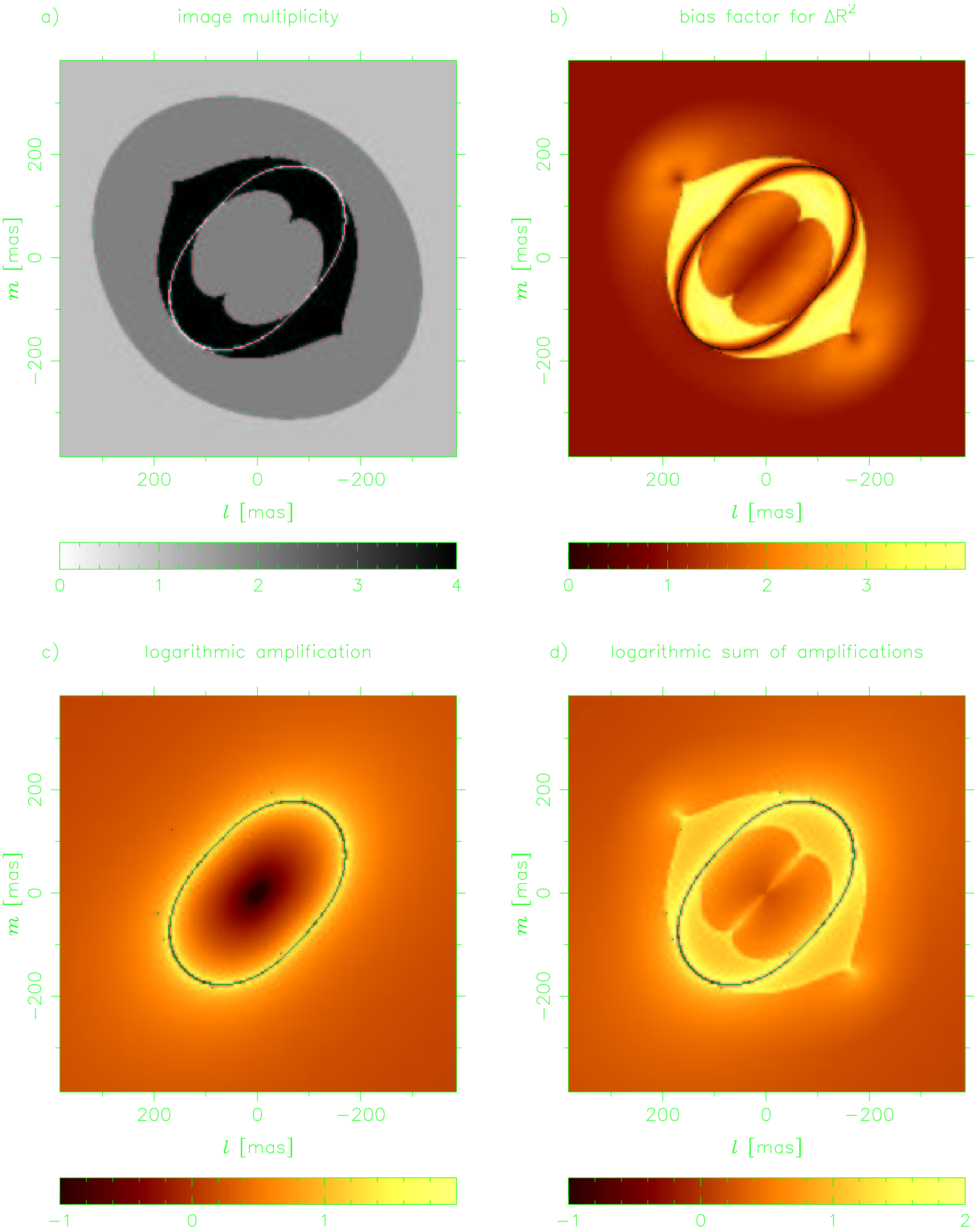


Figure 9.7: Data for the test lens model, which is very similar to the best model for B0218+357. (a) image multiplicities, (b) the bias factor in standard LENSCLEAN, (c) amplification, (d) sum of amplifications of all images corresponding to the same source position. Only pixels with amplifications less than 100 were used. The pixels near the critical curve are the ones rejected by this criterion.

LENSCLEAN. Residuals calculated for the same lens models but with KNE-LENSCLEAN are shown in Figure 9.13 and 9.14 on page 153 and 154.

In the noiseless case, three effects are visible for the old algorithm. First, the residuals are

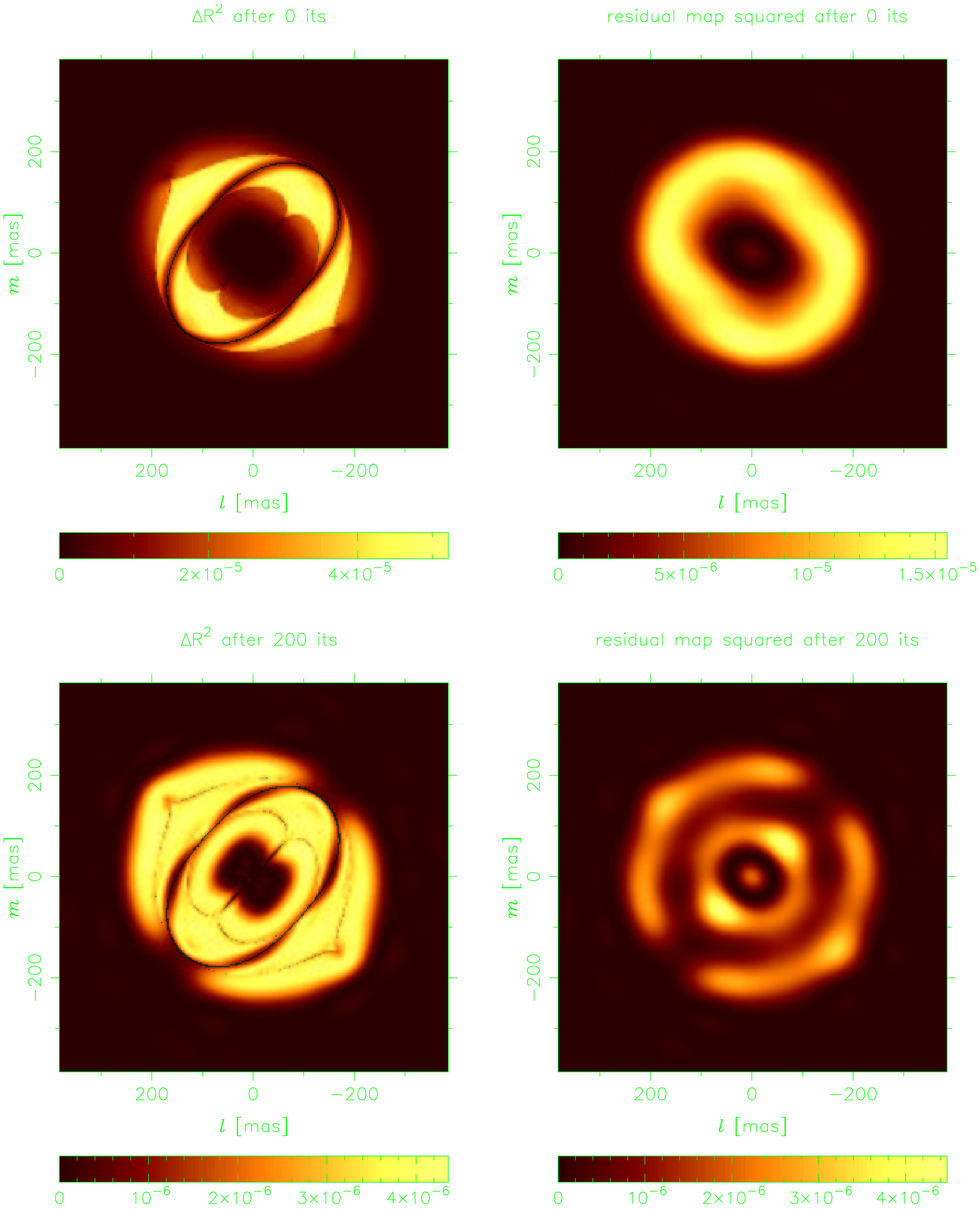


Figure 9.8: Progress in a typical KNE-LENSCLEAN run. The left side shows ΔR^2, the right side the squared residual dirty map. Top: before the first iteration, bottom: after 200 iterations.

much higher even for a large number of iterations. Second, the minimum shows a small but nevertheless significant shift. The KNE-LENSCLEAN method is therefore not able to recover the correct lens model with the same accuracy as unbiased LENSCLEAN. This effect would be shown even stronger in a true lens model fit with the KNE algorithm. Finally, the curve shows bad fluctuations near the peak of the spherical models. Even far from the peak, the curve is

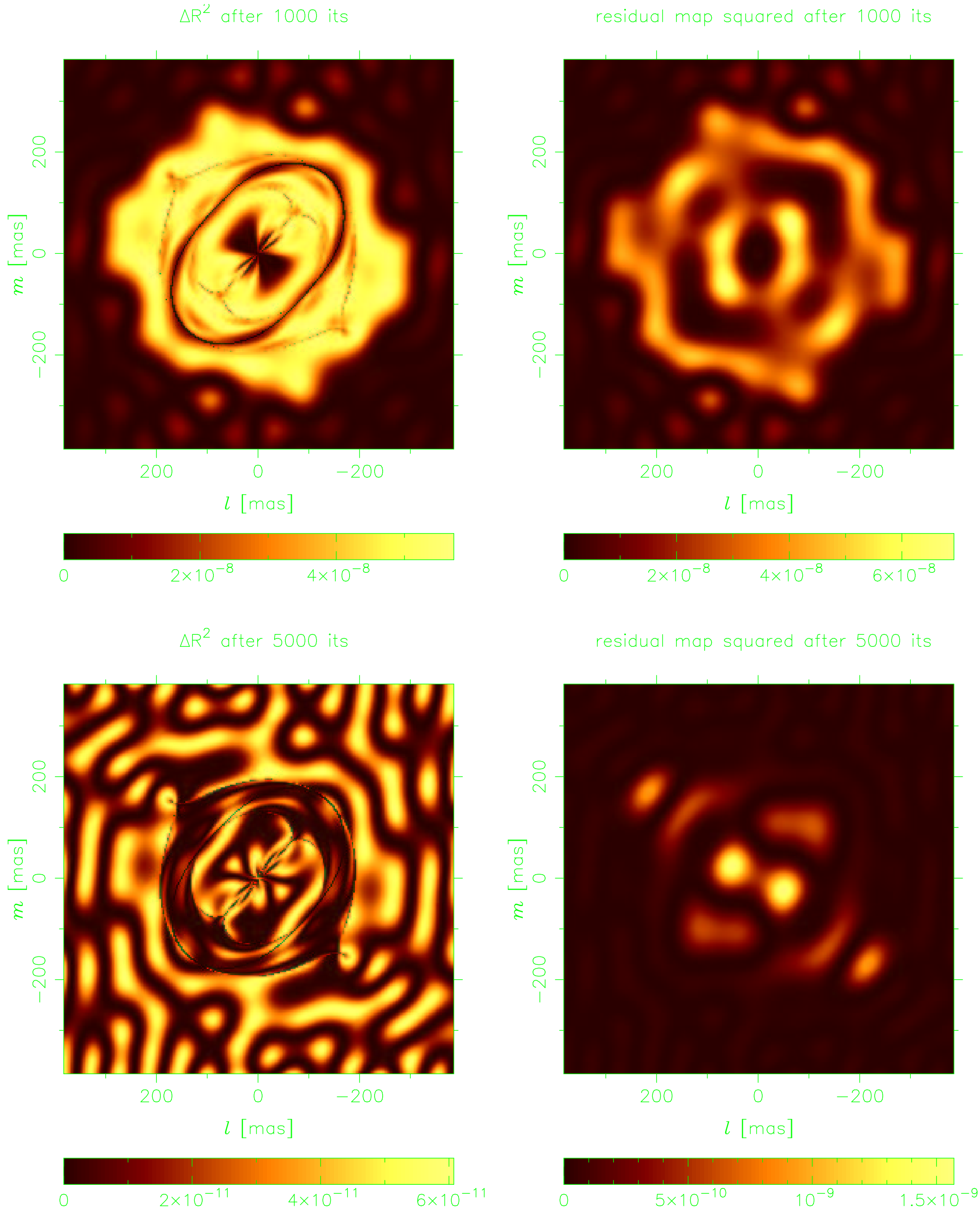


Figure 9.9: Progress in a typical KNE-LENSCLEAN run. The left side shows ΔR^2, the right side the squared residual dirty map. Top: after 1000, bottom: after 5000 iterations.

much more bumpy than the unbiased one. For very symmetrical lenses, this makes a reliable fit impossible, because the minimization routine would get stuck at one of the local minima. The plots show only a one dimensional cut through the five-dimensional parameter space of the residual function. In the complete space, the fluctuations would be even more harmful. Qualitatively the same effects are visible in the results for noisy data (Figure 9.14).

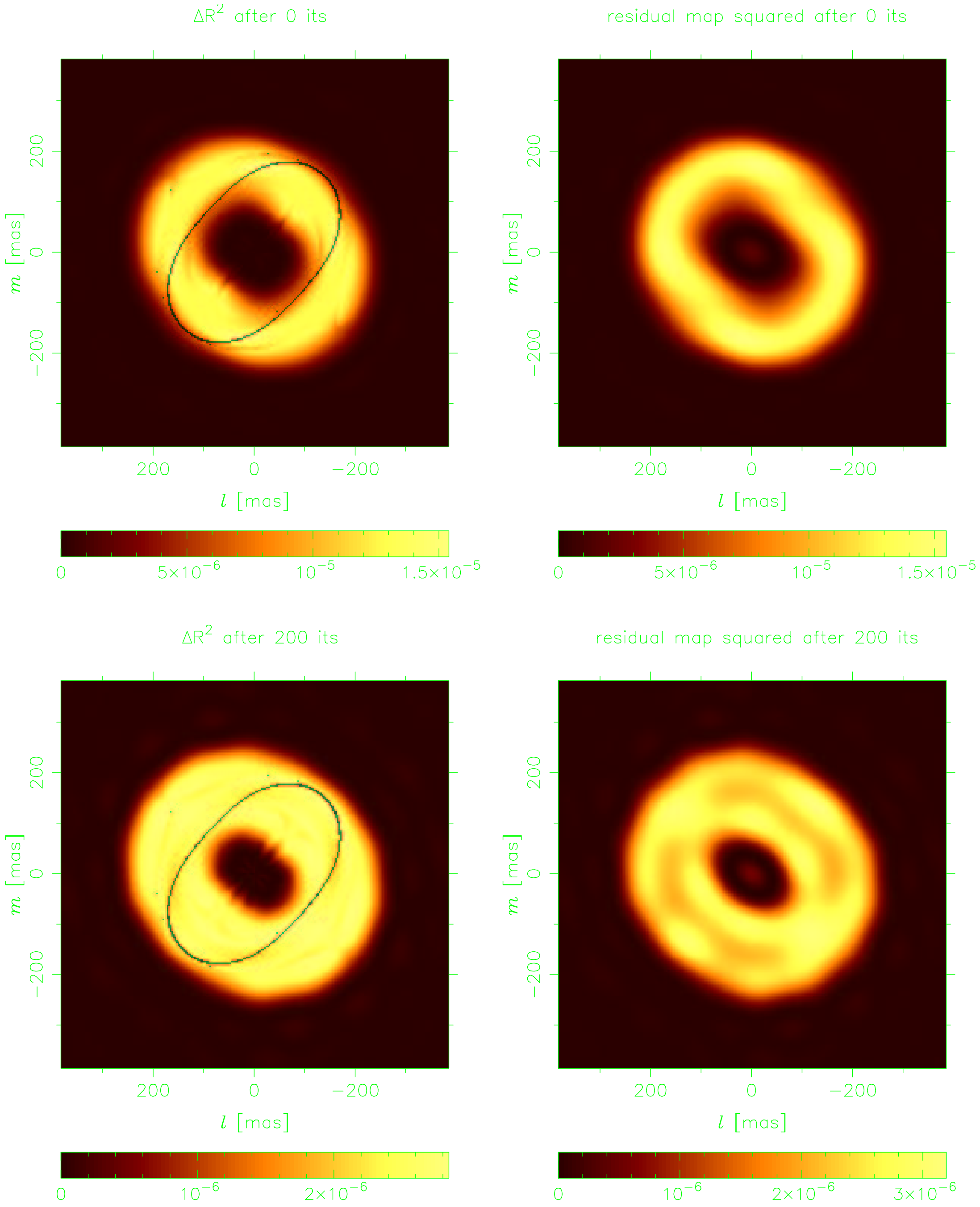


Figure 9.10: Progress in a typical unbiased LENSCLEAN run. The left side shows ΔR^2, the right side the squared residual dirty map. Top: before the first iteration, bottom: after 200 iterations.

We mentioned before that both variants of LENSCLEAN should lead to the same results in the limit of convergence. Even a few thousand iterations are difficult to perform for fits because of the huge amount of CPU time needed. To get a feeling of the numerical behaviour for very many iterations, we calculated residuals for up to 10^6 iterations for two lens models with both variants. The results in Figure 9.15 (page 155) show that the residuals decrease much faster

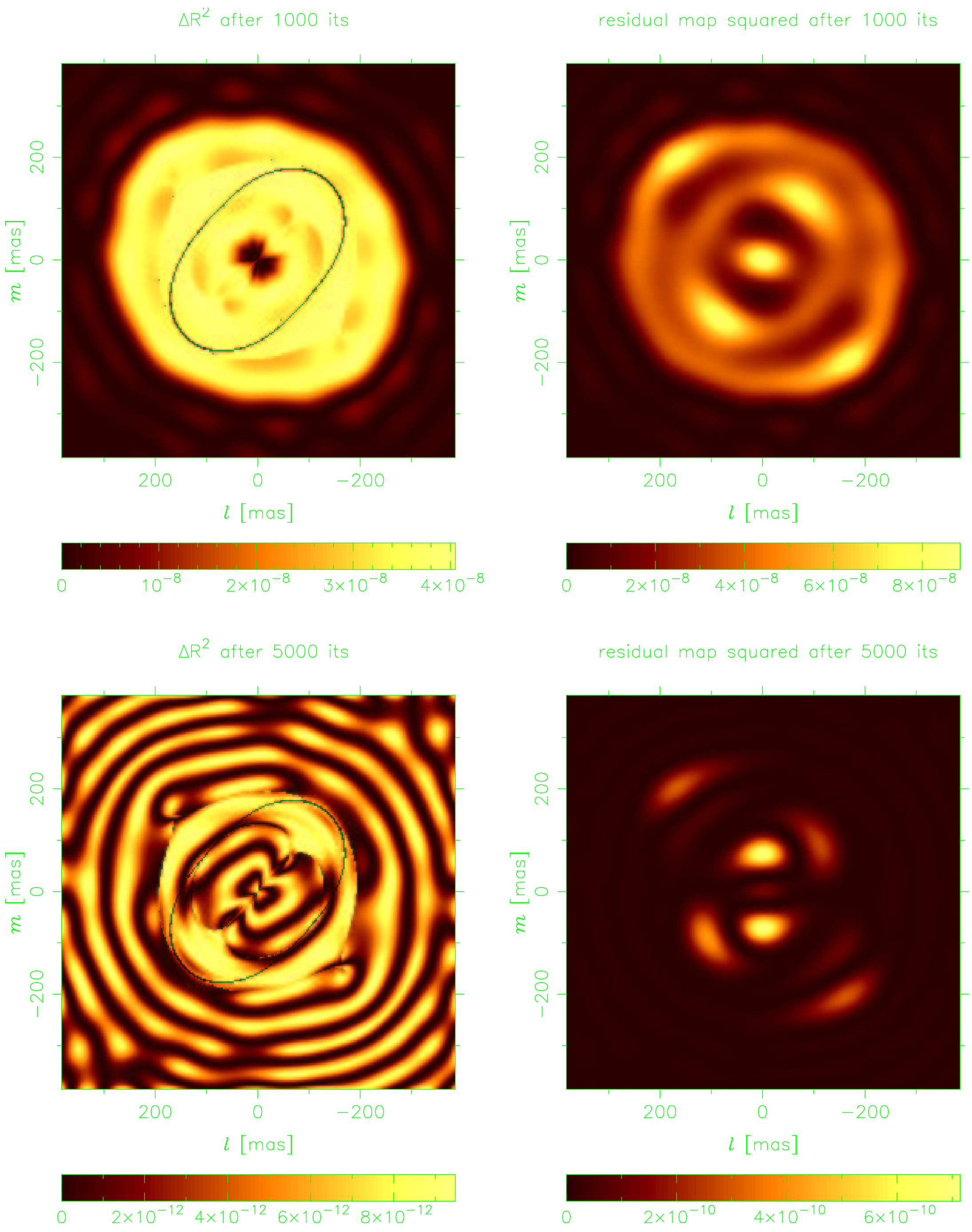


Figure 9.11: Progress in a typical unbiased LENSCLEAN run. The left side shows ΔR^2, the right side the squared residual dirty map. Top: after 1000, bottom: after 5000 iterations

.

with the unbiased algorithm than with KNE-LENSCLEAN. For an incorrect lens model, both results become very similar from about 10 000 iterations on. For the correct model, on the other hand, the difference is very significant. To reach the level of unbiased LENSCLEAN with 5000 iterations ($R^2 = 2.7 \cdot 10^{-6}$), the classical KNE algorithm needs about 240 000 iterations!

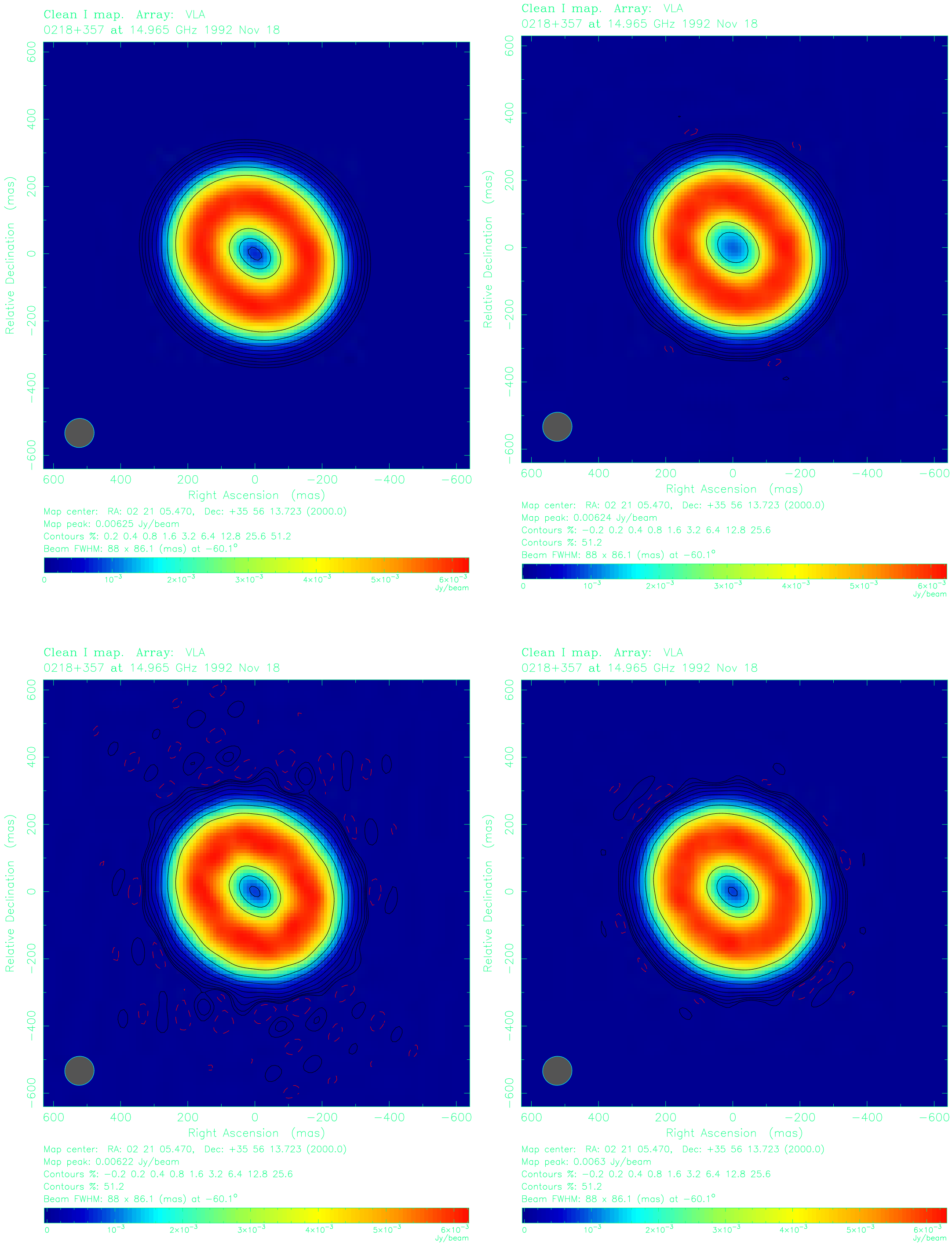


Figure 9.12: Comparisons of the final CLEAN maps. (a) The true model data set (ideal solution), (b) CLEAN map produced with DIFMAP, (c) KNE-LENSCLEAN, (d) our new unbiased LENSCLEAN method

The simulations were calculated with $\gamma = 0.1$ for KNE, $g = 0.38$ for unbiased LENSCLEAN and $\gamma = 0.21$ for CLEAN without a lens. These gains are equivalent in the case of two equally bright and well separated images, leading to very similar residuals in the beginning. The appear-

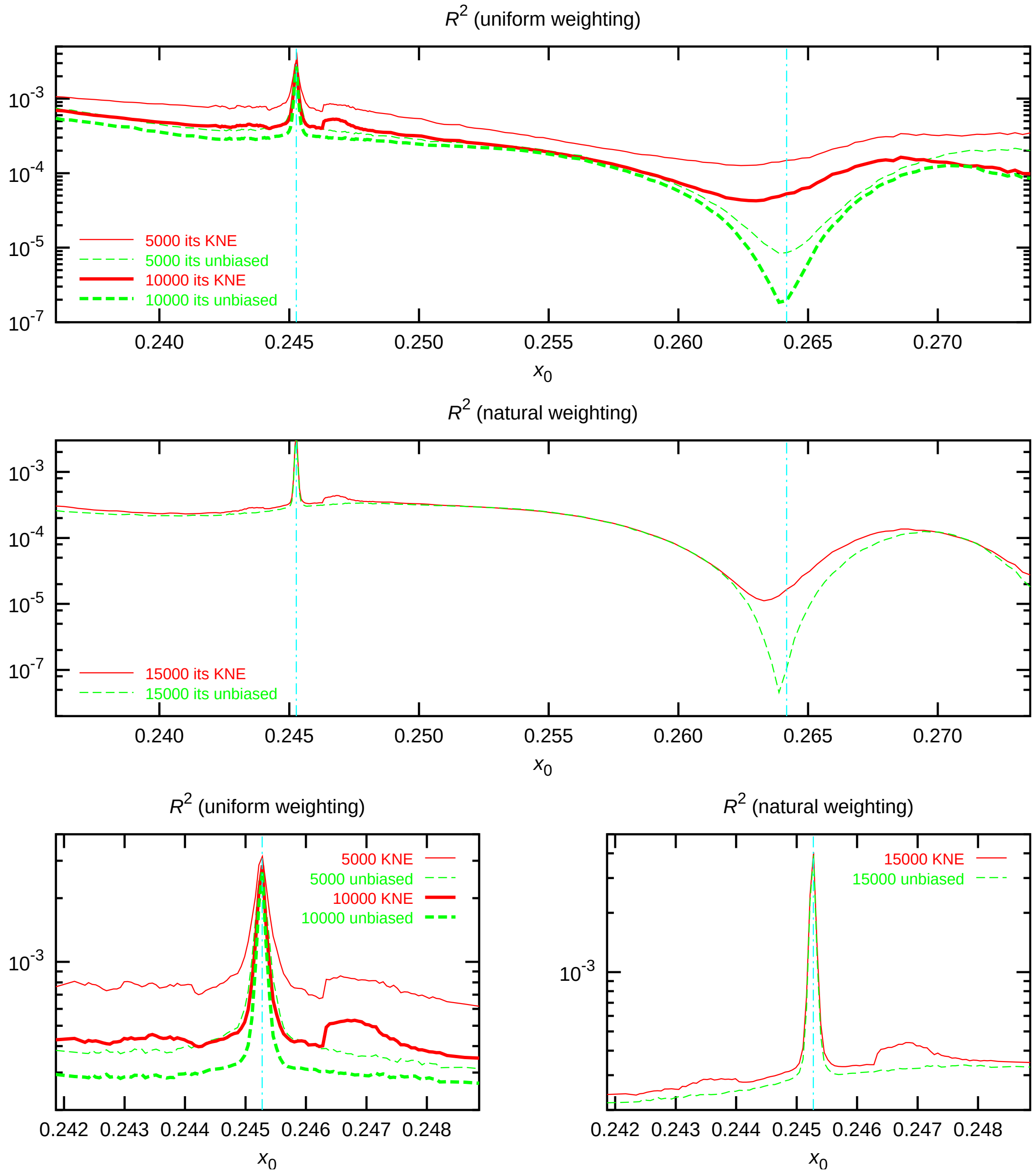


Figure 9.13: LENSCLEAN residuals for the test models, calculated with the KNE (biased) LENSCLEAN version (solid lines) and noiseless data (logarithmic). The unbiased results are shown dashed for comparison. The nearly spherical models are shown magnified in the bottom panels. The vertical lines mark the spherical (left) and correct (right) model.

ance of the double-logarithmic plot would not change considerably with different gains.

The peak near spherical models and the numerical problems of the KNE-algorithm near this peak are not of real concern here, because the best model is sufficiently far away to be influenced only insignificantly. For systems with very small real ellipticities, more serious problems have to be expected. To investigate this subject, we built another artificial data set as before, this time

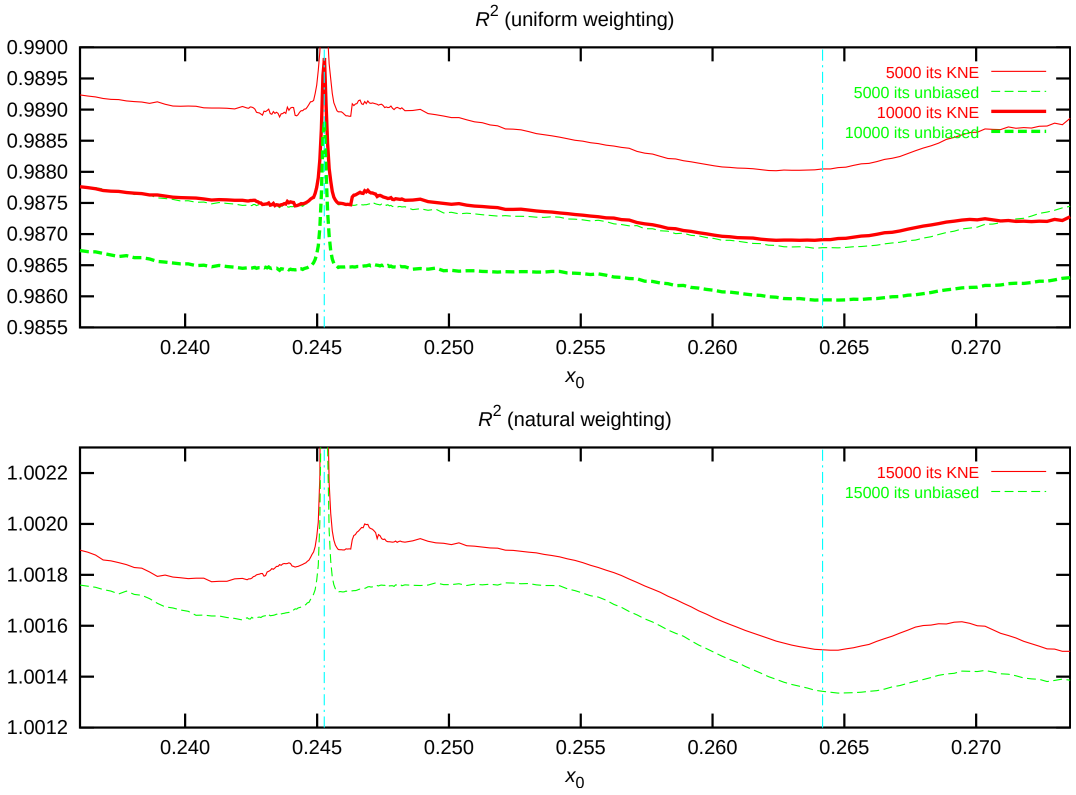


Figure 9.14: LENSCLEAN residuals for the test models, calculated with the KNE (biased) LENSCLEAN version (solid lines) and noisy data. The unbiased results are shown dashed for comparison. The vertical lines mark the spherical (left) and correct (right) model.

using the best fitting spherical lens model which has the following parameters:

$$z_0 = \left(0.^{\prime\prime}245\,489, 0.^{\prime\prime}100\,476\right) \tag{9.21}$$

$$\alpha_0 = 0.^{\prime\prime}167\,418 \tag{9.22}$$

$$\varepsilon = 0 \tag{9.23}$$

With these data, we again fitted the parameters α_0 and ε, using z_0 on a straight line as before. The fit was performed with noiseless data. We then smoothed the parameter functions as before and recalculated the residuals for these models, using noiseless and noisy data and both variants of LENSCLEAN. Some of the results are shown in Figure 9.16 on page 156.

In the noiseless case, the minimum of the residuals is very close to the correct position with unbiased LENSCLEAN. The peak for spherical models has vanished, but the residual function still is more noisy in this region than for elliptical models. We also note relatively low residuals for larger x_0 which are probably a result of the data set which was built from the real data. The real data are best fitted with elliptical models lying exactly in this region. Bias effects are therefore not unexpected. The curve for the KNE-algorithm has strong discontinuities near the correct model and shows no nicely defined parabolic minimum at all.

With the noisy data, strange effects become apparent. The global minimum now jumps to very elliptical models close to the best results for B0218+357. This is not an effect of the noisy alone. In Figure 9.16 we also show residuals for a data set consisting of noisy only (exactly the same noise as in the data+noise set). This curve has a peak near spherical models, but does

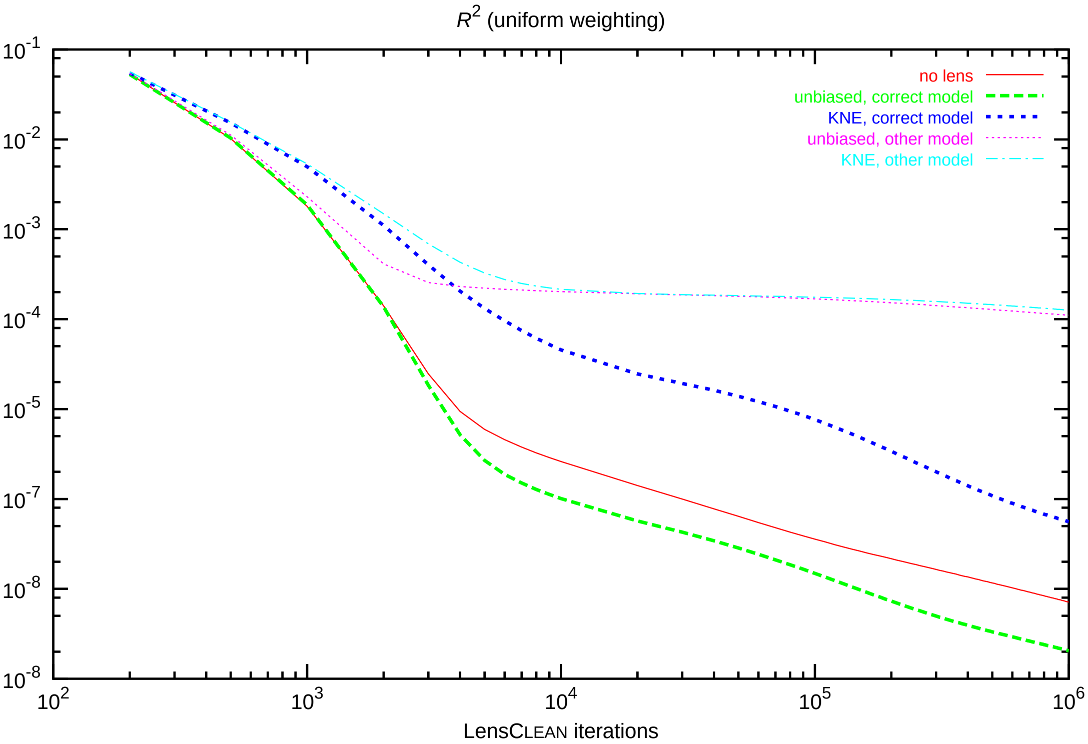


Figure 9.15: Residuals for up to 10^6 LENSCLEAN iterations for the correct model (two thick dashed lines) and for an alternative model from the one dimensional family used before ($x_0 = 0.254$, two thin dashed/dotted lines). The unbiased LENSCLEAN variant is by far superior to KNE-LENSCLEAN for the correct model and in discriminating between the models. CLEAN without a lens is also included for comparison (solid line). The noiseless artificial data set was used for this simulation.

not decrease as strong for larger x_0 as the other data set. Even more confusing is the fact that the residuals are larger for noise only than for the complete data set for some ranges of x_0. The contributions from noise and signal seem to combine in a strange and non-linear way. As the noise is statistically independent from the signal, we expect additivity of the contributions in the limit of convergence. Unfortunately, LENSCLEAN converges much too slow to test this hypothesis. This problem therefore has to wait for a direct algebraic deconvolution method to be solved. For the moment, we should keep in mind possible problems with all variants of LENSCLEAN for nearly spherical lenses.

9.10 Working in the lens plane

A regular grid in the source plane is not an appropriate approach for possible emission components because of the non-homogeneous magnification effects of the lens. Near the caustics, the magnification is very high (infinite in the limit) and the grid would have to be very dense.

We used a regular grid in the lens plane to overcome this difficulty. The resolution is then the same for the whole map, just as in the measured data. For each pixel, the corresponding source position can be calculated from the lens equation. The other solutions of the lens equation are the secondary images which are normally not located exactly on the grid. The multiply imaged regions of the source plane are also multiply mapped with this approach, which introduces some

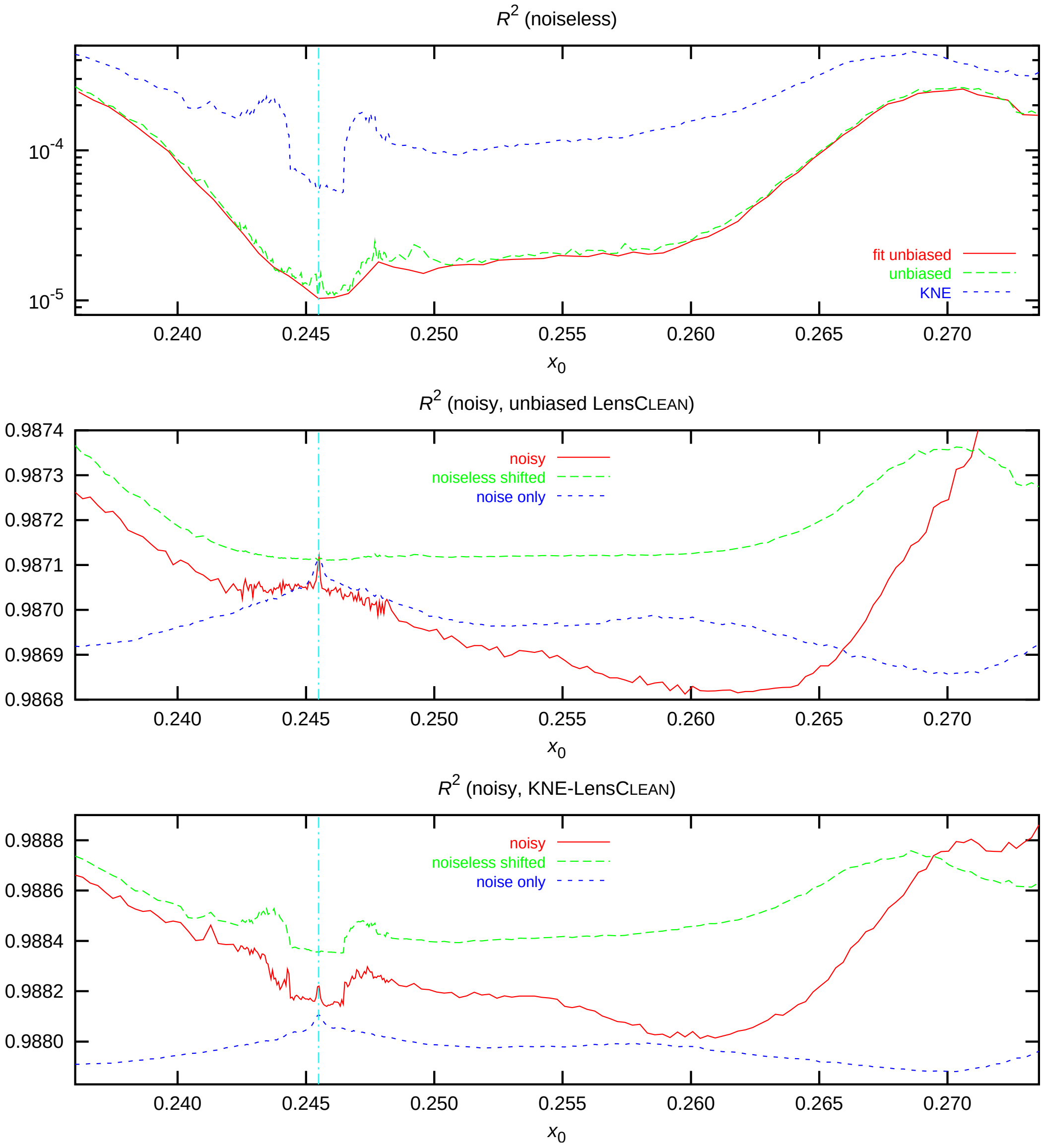


Figure 9.16: LENSCLEAN residuals for the spherical test data set, calculated with 5000 iterations and uniform weighting. (a) noiseless, (b) noisy, unbiased LENSCLEAN (including a noise only result), (c) noisy, KNE-LENSCLEAN (including a noise only result). The vertical line marks the spherical model used to build the data set.

computational overhead. An adaptive grid in the source plane would be more efficient but is conceptually more complicated. It may be tried in the future.

With non-lensed CLEAN, the grid must be fine enough to sample the image plane with the Nyquist frequency. Even finer grids are superior for very bright and compact components. For LENSCLEAN, the grid must also be sufficiently fine to resolve all features of the lens model, i.e. the critical curves and amplification gradients. As usual, there is a trade-off between the accuracy of the calculations and the numerical effort. For our B0218+357 data, we used a standard pixel size of 5 mas and a map size of 512×512 pixels[3]. Only the central quarter of this field can be

---

[3] Nyquist sampling requires only a pixel size five times larger than this, ca. 25 mas.

used for CLEAN components, because of the limited size of the dirty beam. This active area has a size of $1\farcs28 \times 1\farcs28$ and is large enough to cover the emission of the ring, but does not include the outer parts of the jet. These outer parts do not provide information about the lens model, because they are only singly imaged.

9.11 Continuous fits of the compact components

The gridding, which is used in CLEAN and LENSCLEAN, can cause serious numerical problems. Most of these are greatly reduced by the modifications of LENSCLEAN that are discussed below. The remaining errors can still be of concern, however. The problem is not so much the shift of the minimum of residuals but the bumps in the residual function, which produce many local minima. The algorithm to find the best model can easily be fooled by these local minima and produce incorrect results.

With B0218+357, most of this numerical noise is caused by the two bright compact components, which are much stronger than the ring. Therefore, the noise can be reduced by removing these compact components with a continuous algorithm. We do this by starting with one gridless LENSCLEAN step with a very high loop gain to remove the compact components almost completely. To determine the best position and flux for this first step, we use a direct Fourier transform to calculate samples of the dirty map and beam at arbitrary positions. A numerical algorithm is then used to optimize the positions. This first pair of components must be subtracted from the ungridded visibilities to avoid gridding errors.

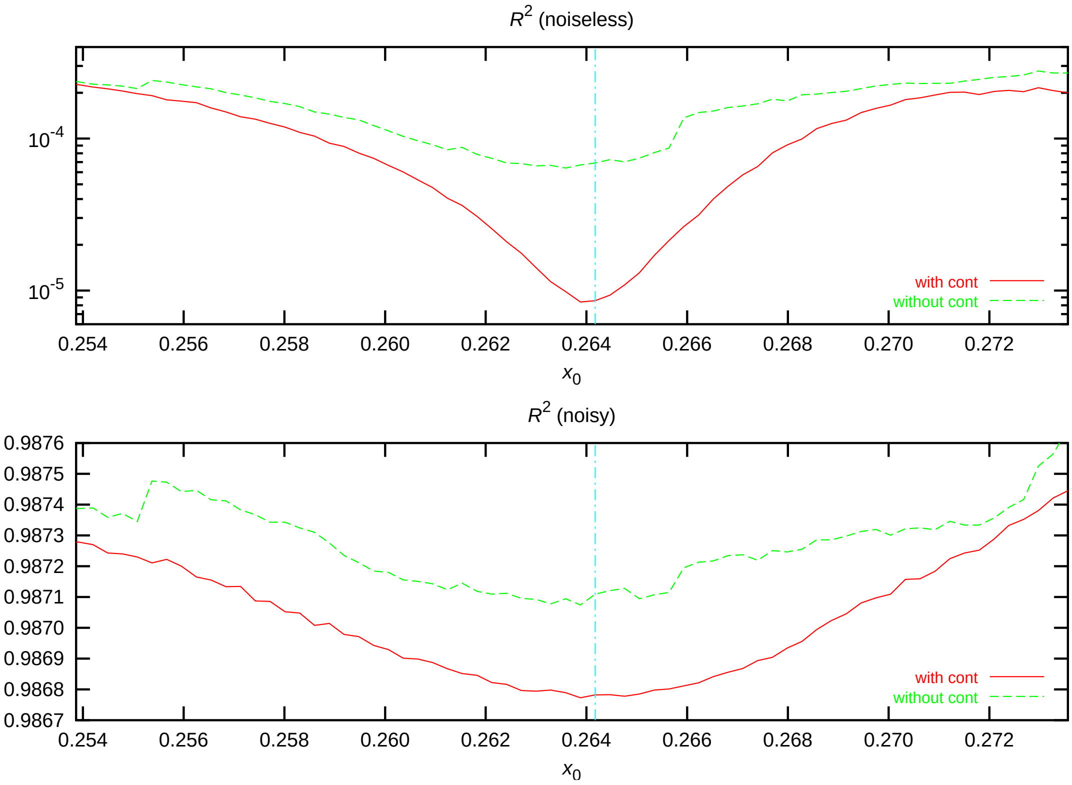


Figure 9.17: LENSCLEAN with and without the first continuous step (5000 iterations, uniform weighting). The vertical lines mark the correct lens model.

The compact components are more or less point-like and can therefore be removed in one step with a high gain. We use a loop gain of $\gamma = 0.98$ for the VLA data set and 0.95 with the MERLIN data (see section 9.22). With this gain, we avoid removing parts of the ring with this first LENSCLEAN step. Numerical experiments show that this is the optimal approach. Figure 9.17 compares the results with and without the first continuous step. The bumps in the residual curves are greatly reduced by the modification and the minimum becomes more prominent.

Experiments with the real data set showed the same behaviour. The improvement is therefore not an artifact of using one algorithm with the artificial data produced with the other one.

9.12 The inclusion of negative components

Real radio sources are not expected to have components with negative fluxes. One might therefore try to reduce the freedom of the models by imposing a non-negativity constraint. This can be done in a natural way in algebraic deconvolution with the NNLS method. CLEAN itself can be modified by allowing only positive components. Unfortunately, this non-negative CLEAN does *not* find an optimal non-negative solution of the deconvolution problem (see section 8.10.6).

With non-lensed CLEAN, negative components are sometimes needed to represent compact components which are not located exactly on grid points. With LENSCLEAN, these negative and positive components are then mapped into the source plane differently. Especially in regions of large amplification gradients, this is a problem. The combination of components might describe the data very well, but the structures in the source plane can become unphysical and be a very bad representation of the real source structure.

The best way to avoid this would be a real lensed NNLS version. This approach is, however, numerically very problematic as we will discuss in section 9.23.3. For the moment, we have to use variants of LENSCLEAN. Extensive numerical experiments have shown that the non-negativity constraint can in some cases improve the sensitivity of LENSCLEAN in the sense of more significant differences between the residuals of correct and incorrect lens models. In other cases, the constraint introduces serious bias effects and produces bumps in the residuals function. For the final modelling, we used the unconstrained version for this reason.

We compare LENSCLEAN runs with and without including negative components in Figure 9.18. We see that a small but significant shift of the minimum and an asymmetry of the curve is introduced by the non-negativity constraint. This is due to the fact that LENSCLEAN cannot be used to find the best non-negative solution.

9.13 Amplification limits

The gridding in the lens plane leads to discontinuities of the residual function. Especially pixels with very high amplification can produce significant jumps for values of model parameters where they appear or vanish. To smoothen the residual function, it sometimes helps to exclude all pixels with amplifications above a certain limit. We used a limit of 100 and excluded not only the pixels themselves but also ones for which the *secondary* images have amplifications above the limit. In the case of the B0218+357 data, the residual function does not change very much near the minimum. The limit thus reduces numerical problems without altering the results. A disadvantage of excluding high amplifications arises when real compact emission is present in these areas. This does not seem to be the case for B0218+357.

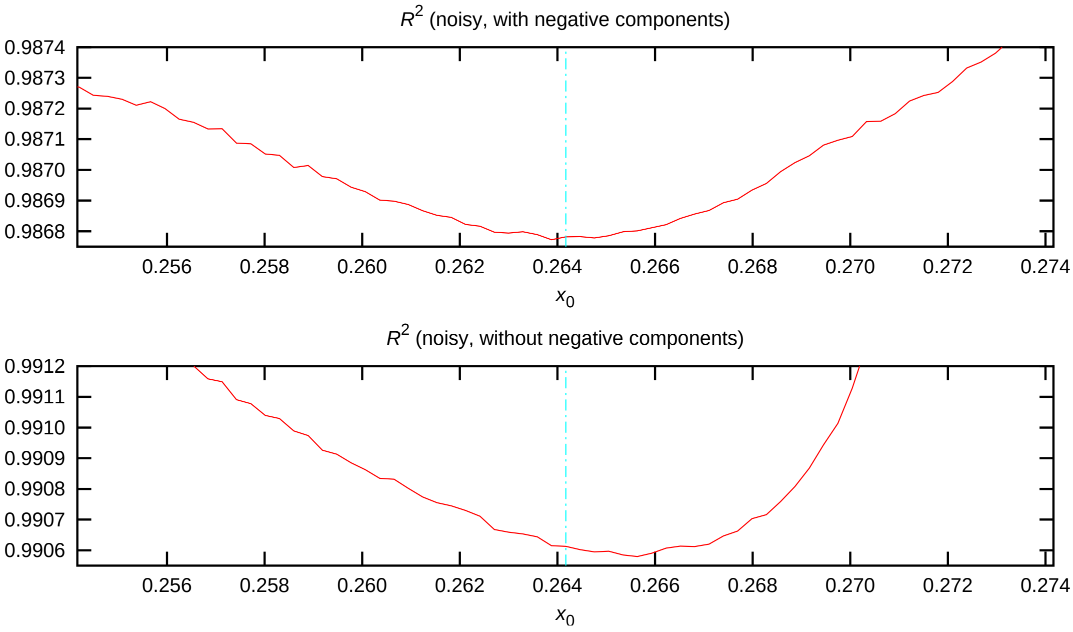


Figure 9.18: The reference version of LENSCLEAN (above) compared with the non-negative variant (below). 5000 iterations with uniform weighting were used. The vertical line marks the correct lens model.

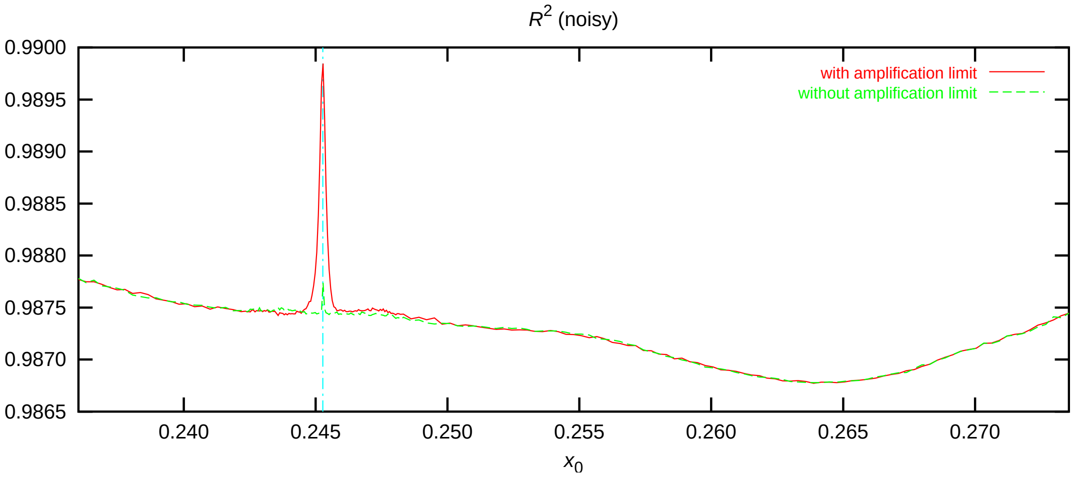


Figure 9.19: LENSCLEAN with and without amplification limit (5000 iterations, uniform weighting). The peak is an artifact of the limit and almost vanishes without it. The vertical line marks the spherical model.

Near the spherical lens models, the amplification limit excludes all pixels with image multiplicities of four. This reduces the model's freedom so much that the peak in the residual functions, which was already observed in section 9.6, is produced. Without the limit, the peak shrinks and becomes very narrow. See Figure 9.19 for a comparison. No systematical changes can be seen near the best models.

9.14 Interpolation and *uv*-space subtraction

With non-lensed CLEAN, the only errors introduced by using FFT on gridded data are the wraparound error to produce a periodic function and the aliasing. With LENSCLEAN, the problem is much more serious. Even if we place the primary images exactly on a regular grid, secondary images of the same source positions do not normally coincide with grid points. This means that the dirty beam and dirty map have to be interpolated to select the optimal pixel. Some kind of correction also has to be used when subtracting the shifted beam for non-grid positions. We combined two ideas to solve the problem. The most important one is to use a modified Cotton-Schwab algorithm and subtract the components at their true (ungridded) positions from the ungridded visibilities at certain intervals and recompute the residual maps from these corrected visibilities. Extensive tests were performed to analyse how many iterations can be performed between these subtractions without deteriorating the results. Subtractions after 20, 50, 200, 1000, and 5000 iterations were normally performed. Smaller intervals changed the results only insignificantly.

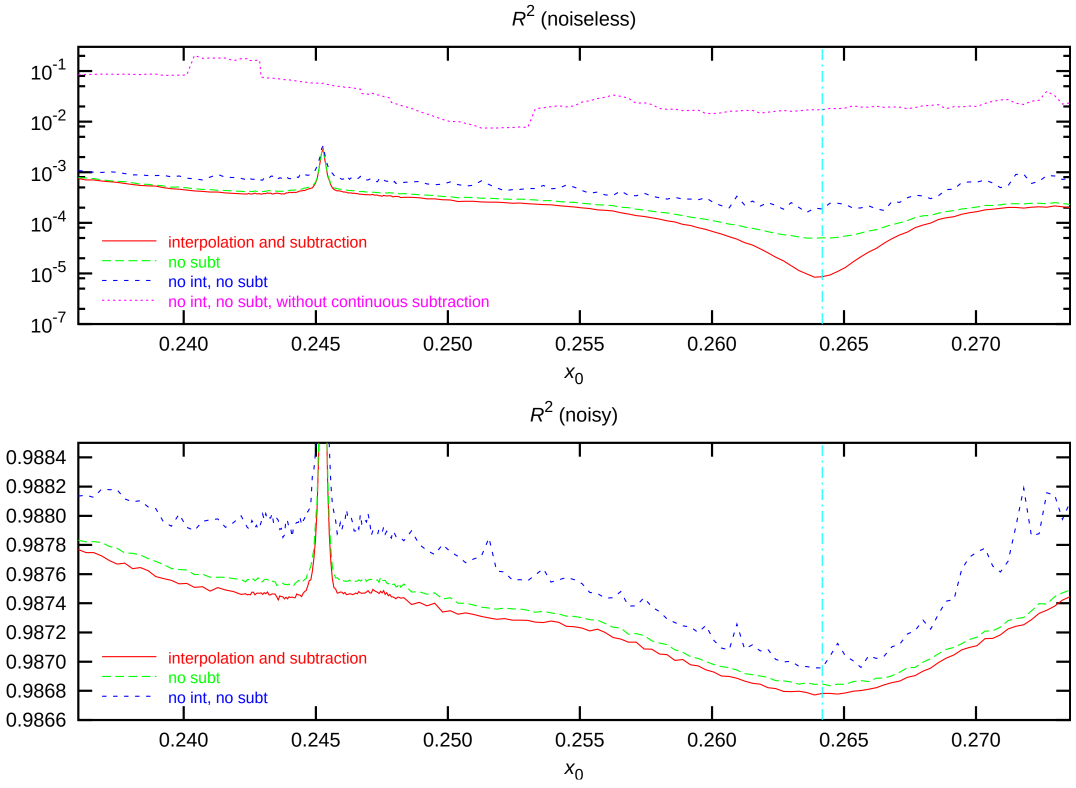


Figure 9.20: LENSCLEAN with interpolation (int) and *uv* space subtraction (subt), without *uv* space subtraction and without both (5000 iterations, uniform weighting). The upper plot also includes results without interpolation and *uv* space subtraction and without the first continuous step. The vertical line marks the correct lens model.

To select the optimal pixels, we used an implicit interpolation scheme. The primary images are always located exactly on grid points and are kept as they are. The secondary images are shared between the four nearest grid points. The amplification is distributed over the pixels in a way to achieve an effective bilinear interpolation. For n true images, we therefore perform the computations for $4n-3$ virtual gridded images. Subtraction in image space is also done with these virtual images. This interpolation scheme transforms the point-like images into small but nevertheless extended approximations. The residual errors left by the algorithm are corrected with the aforementioned *uv* space subtraction. Without interpolation, this had to be done much more often, leading to very slow computations.

Figure 9.20 shows the residuals with interpolation *and uv* space subtraction, with interpolation but without subtraction and without both, using the nearest grid position to select the pixel and to subtract the dirty beam. The largest part of the errors is introduced by the very bright compact components, which are already subtracted gridless (see section 9.11). Without this continuous subtraction, interpolation and *uv* space subtraction are absolutely essential to produce meaningful results at all. The uppermost curve in Fig. 9.20 shows results without interpolation, *uv* space subtraction and the first continous step. The numerical noise is so dominant in this case, that the curve ist useless to find the best model. The example of these modifications illustrates how some of the improvements of LENSCLEAN reach their full impact only in combination.

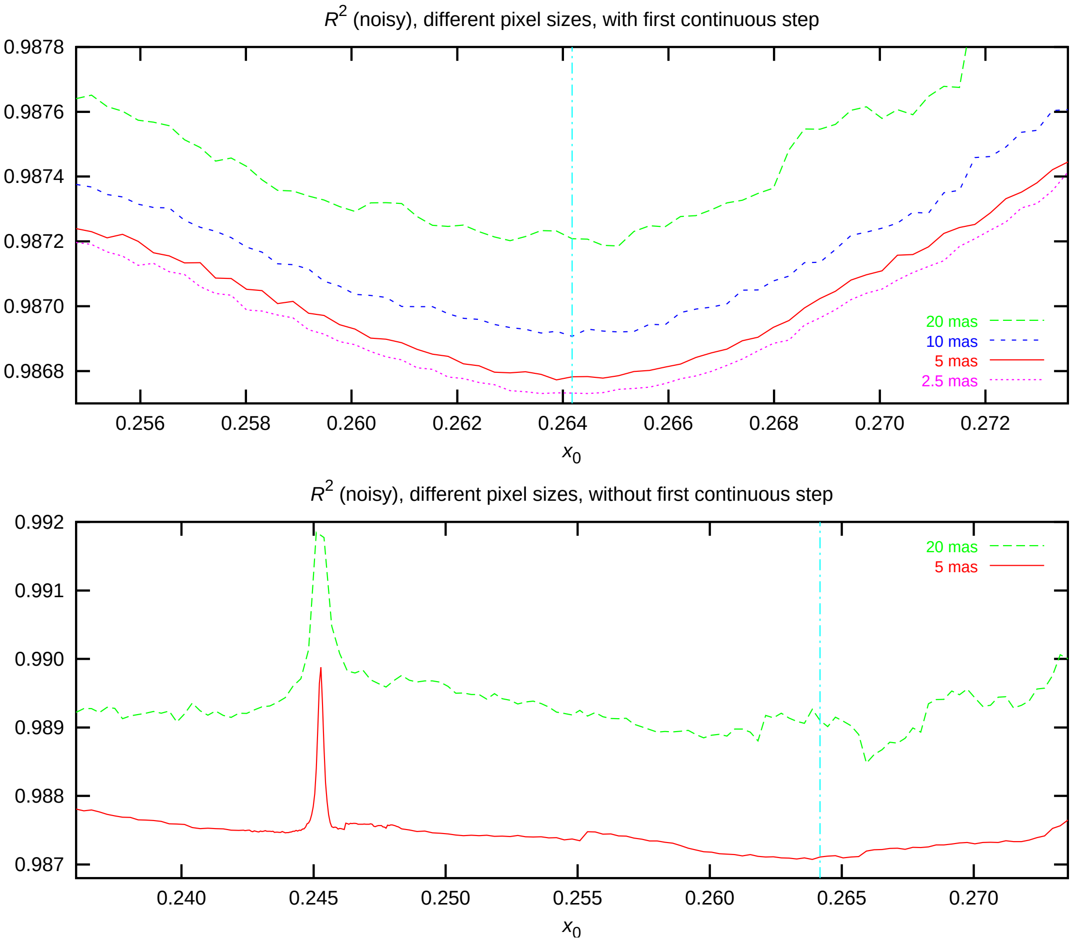


Figure 9.21: LENSCLEAN with different pixel sizes between 20 and 2.5 mas (5000 iterations, uniform weighting). The standard pixel size is 5 mas. The upper panel shows results with, the lower one without the first continuous step. In the latter case, small pixels are even more important than in the former one. Note the different scale of x_0 in the plots. The vertical line marks the correct lens model.

9.15 Choosing the pixel size

The discontinuities in the numerical residual function have their reasons in the discrete nature of the CLEAN algorithm. This is partly due to the gridding. Some modifications of the algorithm to reduce this effect have been discussed in the previous sections. To improve the results further, a fine grid with small pixels has to be used.

The Nyquist theorem gives an upper limit of 25 mas for possible pixel sizes with the 15 GHz VLA data, if we do not want to loose parts of the high-resolution visibilities. As discussed before, the grid must also be fine enough to resolve all important features of the lens. Figure 9.21 shows a comparison of results for different pixel sizes between 2.5 and 20 mas. The 20 mas pixels are definitely too large, but there is not much improvement for sizes much smaller than a few mas, where other sources of numerical noise dominate. Our standard of 5 mas seems to be a good compromise. Larger pixels do not lead to systematic errors, but the increased numerical noise of the residual function makes reliable fits increasingly difficult.

Without the first continuous LENSCLEAN step, even the compact components would have to be represented by the regular grid. In this case, too large pixels have an even stronger effect. 5 mas is at the limit of useful values then, the results with 20 mas are absolutely meaningless.

9.16 Choosing the gain

The second source of discontinuities in the residual function is the finite loop gain in the discrete iterations. To reduce the errors, a small loop gain with very many iterations should be used. Unfortunately, LENSCLEAN is numerically very expensive and small gains can not be used on a regular basis. Our standard loop gain parameter for unbiased LENSCLEAN is $g = 0.38$. For n well separated equally bright images, this corresponds to

$$\gamma = 1 - \sqrt{1 - \frac{g}{n}} \quad . \tag{9.24}$$

For $n = 1/2/4$ images, the loop gain is therefore $\gamma \approx 0.2/0.1/0.05$. About 2000 to 5000 iterations are usually needed with this gain to obtain useful residuals.

A comparison of 5000 iterations at this standard value with 25 000 iterations at $g = 0.076$ ($\gamma \approx 0.04/0.02/0.01$) is shown in Figure 9.22. The total level of the residuals is comparable, as expected. The much smoother curve for the smaller gain shows, how important the effect is. Unfortunately, we cannot do all the calculations with a gain much smaller than $g = 0.38$ with the available equipment. This large gain does not introduce systematic effects besides the numerical noise.

9.17 Classical pre-fits

The calculation of best-fit residuals for a given lens position z_0 is numerically very expensive with the full LENSCLEAN-algorithm. A fit of the remaining three lens model parameters has to be performed, which typically takes 150 iterations. We tried several ideas to avoid these full fits, without much success.

The remaining parameters for fixed z_0 can be determined easily with LENSCLEAN, because the two bright components provide enough constraints for them. The ring modifies the model only to a small extent but is the main factor for the best fit residual level. Its effect becomes

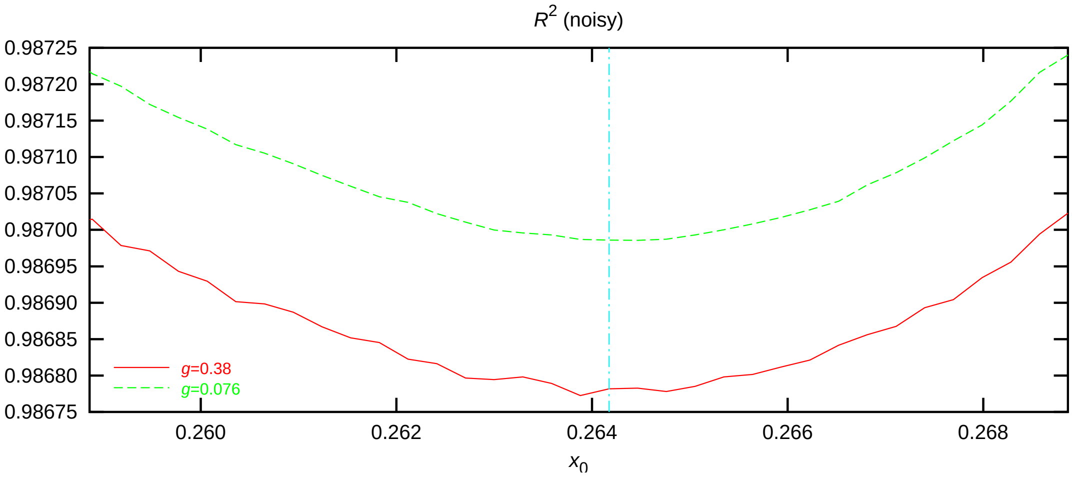


Figure 9.22: LENSCLEAN with the standard loop gain parameter $g = 0.38$ (5000 iterations) and with $g = 0.076$ (25 000 iterations) for comparison. Uniform weighting was used for both calculations. The higher gain produces more numerical noise but no systematic shift of the minimum position. The vertical line marks the correct lens model.

important when the residuals for different z_0 are compared to find the best lens position. These differences are defined only by the ring.

The most simple idea was to do a classical lens model fit for a fixed lens position, using the positions and flux densities of the compact components alone. This method is in the standard algorithm only used to find the starting values for the numerical LENSCLEAN fit.

The classically fitted models are a good approximation for the full LENSCLEAN fits and might thus be used as a substitute[4]. Of course LENSCLEAN still has to be used to calculate the final residuals for the classically fitted lens model. The recipe of an accelerated algorithm is the following: Choose a lens position, do a classical model fit and perform *one* LENSCLEAN run for the best fitting lens model. This reduces the number of LENSCLEAN runs needed for each z_0 from about 150 to 1. Unfortunately, LENSCLEAN is so sensitive to small deviations of the lens model, that the residuals are not a very good measure for the goodness of fit of the lens model. This problem is partly due to the difficulties of determining the positions and fluxes of the compact components with high accuracy. This leads to small but nevertheless significant errors in the classical lens model fit. In a fit of the components, the parts of the ring near the components are always included in the fit, leading to small errors in the positions and flux densities. Since these errors depend on the weighting, we tried to reduce them by using uniform weighting to reduce the effect of small *uv* spacings where the ring is stronger than at larger spacings[5]. We still expect contributions from the ring at the level of some mJy. From our noiseless artificial data set, we obtained a relative position $\mathrm{B}-\mathrm{A}$ of $(310.552, 127.095)$ mas and flux densities of 0.76234 and 0.20156. With natural weighting, the positions shift by about 0.5 mas, and the fluxes increase by about 15 mJy each. The flux ratio decreases from 3.78 to 3.56.

Figure 9.23 shows residuals for the same z_0 values as before but with classically fitted model parameters using the data given above. We also included the results for slightly different values

---

[4] The deviations are typically less than 1 mas in α_0 and much less than 0.001 in ε.

[5] DIFMAP does all model fitting with natural weighting, regardless of the assigned `uvweight`. Uniformly weighted fitting was performed with our own software and with reweighted data in DIFMAP.

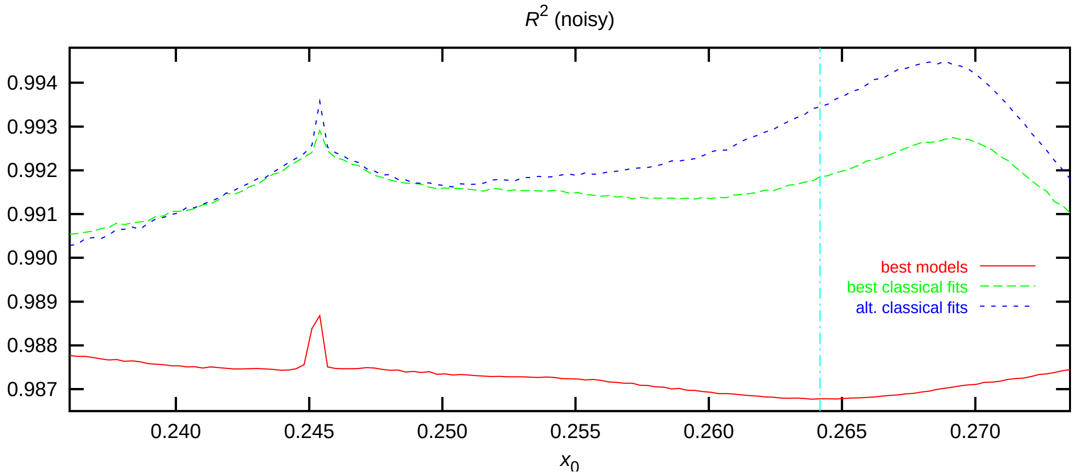


Figure 9.23: LENSCLEAN residuals calculated for classically fitted lens models (5000 iterations, uniform weighting). Two data sets for the compact components where used, which differed by only about 0.1 mas in position and 0.02 % in flux ratio. The vertical line marks the correct lens model.

(shift of about 0.1 mas and change of flux ratio of about 0.02 %) for comparison. We see that even these small changes produce very different results, and (even worse) both classically fitted results do not represent the residual curve for the LENSCLEAN models.

This high sensitivity to small changes of the models is mainly a result of the compact components themselves. Since the information from them was already used in the classical fit, we should exclude them from the following LENSCLEAN somehow.

9.17.1 Correction by free fit of compact components

One possibility to exclude the compact components from the LENSCLEAN process is to replace the continuous first LENSCLEAN step by a free (without a lens model) fit and subtraction of the two bright components. With this modification, small deviations in the lens model do not change the residuals caused by the compact components. They only influence the CLEANing of the ring. As the ring is much weaker and very resolved, the influence should be reduced by a large amount. A serious problem remains; subtracting the fitted compact components also removes parts of the ring near the images and thus changes its structure significantly. How strongly this changes the results is difficult to estimate in advance.

Figure 9.24 shows results with free fits and total (gain=1) subtraction of the compact components. Using the classically fitted lens models instead of the best LENSCLEAN results does not change the residuals significantly, which proves that the free fits reduce the influence of small model changes as intended. However, the perturbations of the ring are so strong that the residual curve cannot be used to determine the best lens model. The similarity of the curves in Figure 9.24 with those in Figure 9.23 for the classically fitted models are no coincidence. The free fit removes the compact components at the same positions and with the same flux densities that were used for the classical fits. In the case of the latter, the compact components are then removed by the lensed fit of the compact components, leading to approximately the same residuals.

The failure of this approach is partly due to the difficulties in the determination of the 'true'

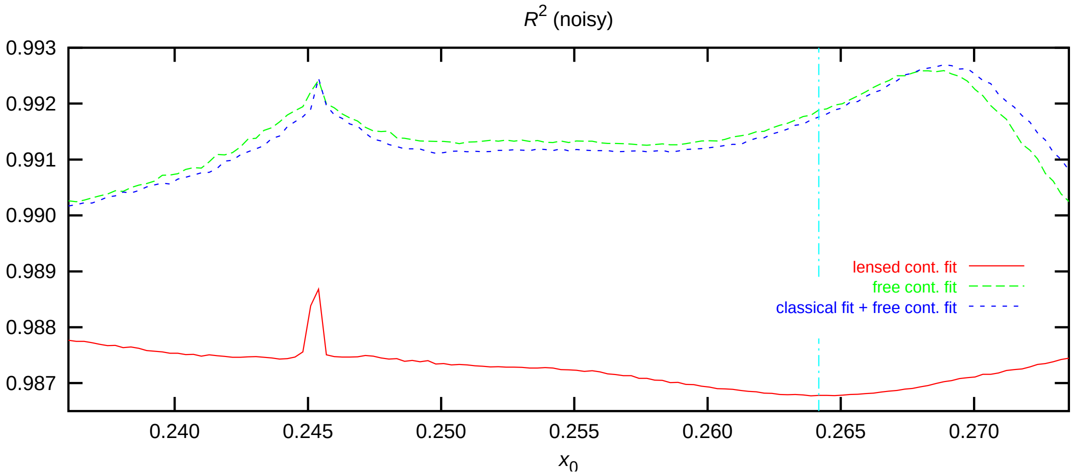


Figure 9.24: LENSCLEAN residuals with free continuous fit and subtraction of the compact components (gain=1) performed for the best models and classically fitted models (5000 iterations, uniform weighting). The vertical line marks the correct lens model.

positions and flux densities of the compact components and partly due to the fact that calculating the residuals for models which fit the compact images optimally is not equivalent to optimize the models for the compact images and the ring *simultaneously* and use these residuals to determine the best lens model.

9.17.2 Correction by cutting out the compact components

Another idea is to switch off the lens for small regions around the compact components and allow free (unlensed) CLEANing in these regions. Formally, this can be done by deleting all secondary images for primary images inside of small circles around the bright images. With this change, the regions around A and B can be CLEANed independently. Small changes in the lens model should then not influence the residuals caused by the compact images but by the ring only. The problems of the method discussed above can be avoided with this variant, because any modifications of the ring by free fits of the compact components can now be corrected at later stages.

We tried different sizes of circles and present the results in Figure 9.25. The upper panel shows results for the best LENSCLEAN models, the lower one for the classically fitted models. From the former, we see that cut radii larger than 20 mas change the shape of the residual function and begin to shift the minimum. A lower limit for the radii can be determined from the lower plot for classically fitted models. For 5 mas, the residuals are too high, because components confined to this small circle cannot reproduce the compact emission very well. Even the 20 mas curve shows a minimum shift of about 2 mas. The results for 50 mas radius seem to reproduce the direct fit results quite well, but the significantly changed shape of the residual function for the fitted models (upper panel in Fig. 9.25) tells us to interpret the results with caution.

That LENSCLEAN before reaching convergence is not a true fit to the data, can be seen from the fact that the residuals in the upper plots are generally larger for the modified LENSCLEAN version, although the emission model has more freedom in this case, which should reduce the residuals after convergence. This is caused by the decrease in efficiency of the modified algorithm. It takes very many iterations for LENSCLEAN to actually take advantage of the added

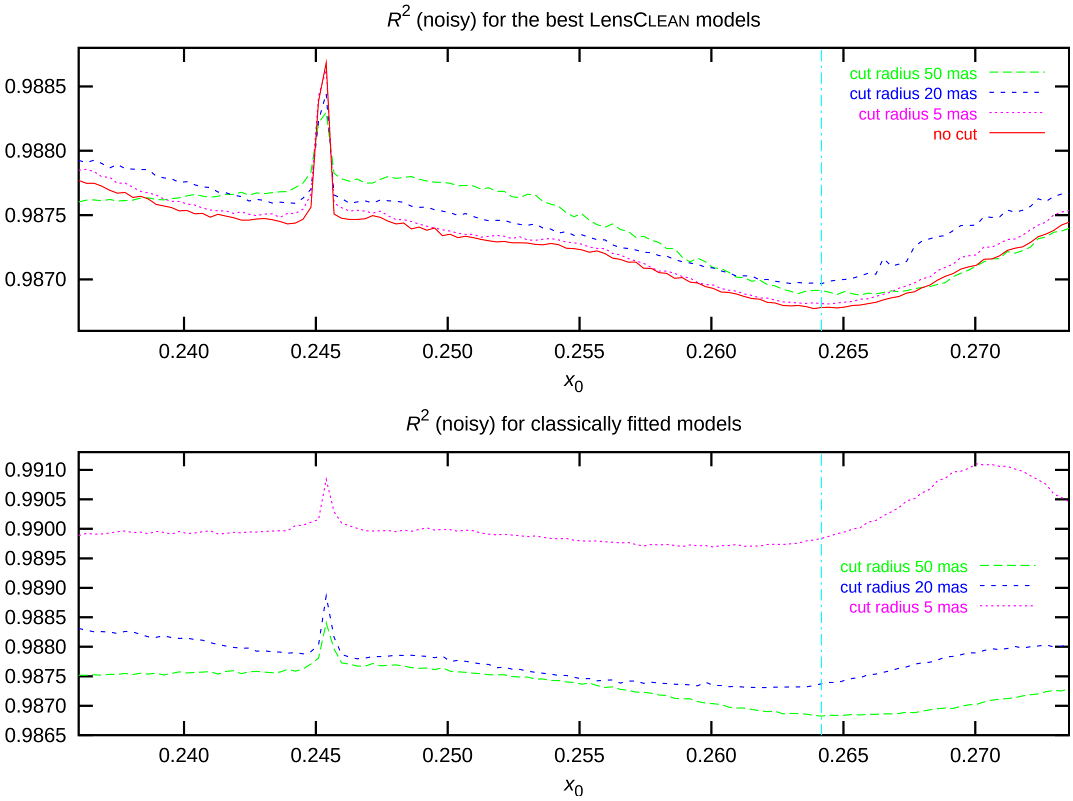


Figure 9.25: LENSCLEAN with allowing free CLEAN inside of circles around the compact images (5000 iterations, uniform weighting). The upper panel shows results for the best LENSCLEAN models, the lower one for classically fitted models. The vertical line marks the correct lens model.

freedom.

The fact that the optimal cut radius can only be determined from numerical experiments and comparison with the full LENSCLEAN fit results and the difficulty of estimating the errors rules out the use of the modified algorithm on a regular basis. Maybe a better way to exclude the compact emission from the LENSCLEAN process can be found in the future. This would result in an acceleration of LENSCLEAN fits by up to two orders of magnitude in the case of B0218+357.

9.18 Windows

In unlensed CLEAN, windows are normally used, outside of which no emission is allowed. These windows accelerate the CLEAN process and can improve the results considerably, especially when used in self-calibration loops. If any dubious features are present in the initial maps, the observer can try to exclude this emission from the windows and try to produce a map that is consistent with the self-calibrated data and does not show the dubious emission. If this is not possible, the features are either real or caused by other problems with the data, which can not be removed with self-calibration.

The purpose of LENSCLEAN is to find the best lens model. The decision to use windows should therefore depend on the influence on the residuals. Confining the emission to regions where it is expected, gives the emission models less freedom and should therefore increase the sensitivity for small changes of the lens models. On the other hand, bias effects may be introduced by the windows. For B0218+357, the advantages are very small, while potential bias effects can be strong.

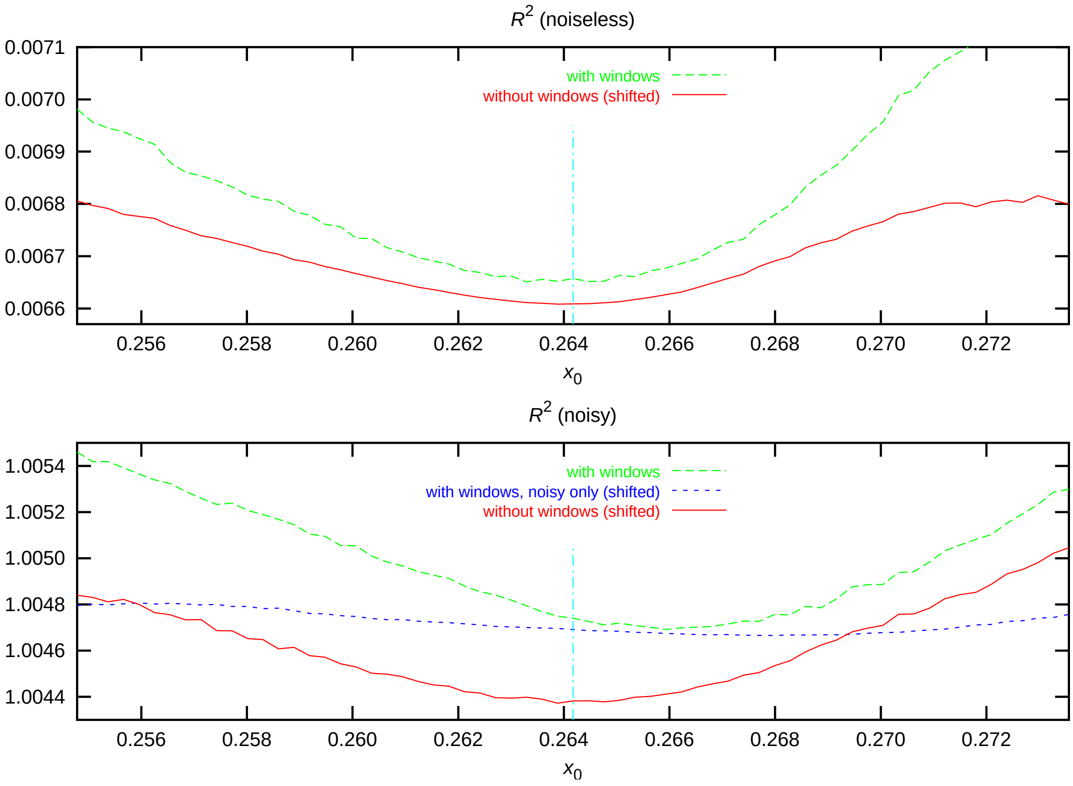


Figure 9.26: LENSCLEAN with and without windows (5000 its, uniform weighting). In the upper plot (noiseless data), the windowless residuals were shifted by 0.0066 to help comparing the two. In the lower plot (noisy data), the shift is 0.0176. We also included results for a noise-only data set for comparison. The bias is only partly explained by the behaviour of this data. The vertical lines mark the correct lens model.

Figure 9.26 shows the results with rather conservatively large windows, which cover the emission from the ring and the compact images. The windows are given by two overlapping circles with radius $0\farcs35$ around $(0\farcs179, 0\farcs103)$ and $(0\farcs309, 0\farcs143)$. Emission is allowed, if the primary image is inside of the windows, regardless of the position of the secondary images. Without noise, the residual curve becomes a bit steeper with the windows, but this does not help in determining the minimum, because the numerical noisy increases, too. No shift of the minimum is noticeable. With noise, things are less promising. We notice a significant shift of the minimum, which is caused by the combination of noise and windows. When the lens models change, the effective area of the windows also changes as a result of shifts of the secondary images. The fraction of noise that can be removed by LENSCLEAN therefore depends on the lens model, leading to a bias. To test this interpretation, we also include residuals for a noise-only data set with the same lens models and the same windows. When subtracting these noise-only residuals, the minimum still is displaced from the correct position. This is partly due to the non-linearity of LENSCLEAN before convergence.

We conclude that windows are not helpful in the case of B0218+357. No windows are therefore used in our standard algorithm and for the final results.

9.19 Extinction

The flux ratio A/B in B0218+357 shows significant variations with wavelength (see section 7.8). One possible explanation is an effective (possibly scattering induced) extinction of the A component at lower frequencies. If LENSCLEAN is used to find the best lens model, this effect has to be taken into account. Otherwise the lens model will be adjusted to have the same effect on the flux ratio as the extinction, leading to significant errors. Two ideas were tested to accomplish this. One is to compensate for this effect by adding emission to A (or subtracting from B) before LENSCLEAN is started. A difficulty with this approach is the determination of the correct position of the component. Even small errors can influence the results. The other idea may be more accurate when the extinction is not confined to a very small region in component A but to a larger area, so that the ring is also affected. We can take the effect into account by artificially reducing the amplifications given by the lens model near the A component. The size and shape of the absorbing region is not known, of course.

Even without real extinction, small corrections may be needed to correct for flux ratio errors due to flux variations in combination with the time-delay. For the 15 GHz data set, we performed the fits with both corrections and with different values for the extinction. The position of the best lens model shifts with the extinction, but the best value for the correction seems to be compatible with 0. Corrections of the order 5 % increase the residuals significantly. We conclude that extinction can be neglected at 15 GHz, just as expected. We will come back to this subject later, when we discuss the 5 GHz MERLIN data.

9.20 Self-calibration

Tests show that calibration errors can shift the residual minimum by a few mas. This effect is not suprising given the high dynamic range of the B0218+357 data. Calibration errors scale with the very high total flux, while the relevant signal is only proportional to the flux density of the ring. To take into account possible miscalibration, a loop of alternating LENSCLEAN and self-calibration with the LENSCLEAN brightness model should be used for each tested lens model. Unfortunately, even small changes in how the self-calibration is performed can change the residuals quite drastically. We nevertheless performed some tests with this approach, but learned that the whole process becomes very unstable. The residuals for data sets self-calibrated with different lens models cannot be compared directly with each other for several reasons. One reason is the fixing of the flux scale that is necessary to prevent the solutions to converge to zero with zero residuals. Every lens model would be compatible with no emission at all. This step has the effect that the visibilities after self-calibration are proportional to the best solution (with smallest residuals) but are not exactly the same and that the residuals are therefore *not* minimized. Different flux scale corrections for different lens models can then lead to strong variations of the residuals which are not related to the ability of the lens models to explain the observations.

As a compromise, we performed self-calibration not for each test lens model, but only for the best fitting models (including the fit of z_0). In this way, we only compare residuals obtained from one and the same self-calibration procedure in each iteration, avoiding the aforementioned problems. The process works like follows:

- Map and self-calibrate the data in a standard way with AIPS or DIFMAP with initial self-calibration against a point source model to produce an initially self-calibrated *uv* data set. This step is important, because it can in exactly the same way also be performed with artificial data. The initial self-calibration with a point source model destroys all previous calibration steps, so that the real and artificial data start in the same state. Only then can the results be compared.

- Change the weights of the data to obtain effective uniform weighting when natural weighting is chosen in DIFMAP. Self-calibration in DIFMAP always used natural weighting, which in combination with the uniform weighting in LENSCLEAN would not converge to a self-consistent solution. Uniform is superior to natural weighting in LENSCLEAN, as was shown before.

- Do LENSCLEAN lens model fits for a range of fixed lens positions. Determine the model parameters of the minimum by fitting a polynomial to the central part (around the minimum) of the grid of lens positions. This procedure reduces the effects of numerical noise on the resulting lens model drastically.

- Perform a LENSCLEAN/self-calibration loop with this fixed lens model by iterating the following loop 50 to 100 times:

    - Use LENSCLEAN to build an emission model with the given lens model and the latest version of the *uv* data. The non-negative variant is generally used for this, as allowing negative components is known to give self-calibration to much freedom.
    - Use this brightness model to self-calibrate the *initial* data set with DIFMAP[6]. Phase and amplitude are corrected with a solution interval of 0 with fixing the flux scale (`selfcal true,false,0` in DIFMAP). The integration bins of the 15 GHz VLA-data are 1 minute. This gives the process much more freedom than most observers would say is necessary. We nevertheless used this massive self-calibration to be sure to remove all calibration errors. Comparison with results for artificial data can be used to estimate the errors. Regarding the lens models, this is a very conservative approach.

      We always used the initial data set as basis for the self-calibration to prevent rounding errors from accumulating. The solutions would diverge after many iterations otherwise.

- Repeat the loop with the new *uv* data set. Find a new best lens model and continue until the shape of the residual function (and especially the position of the minimum) does not change any more.

To be sure, that this process does not introduce strong bias effects, we performed the same process with artificial data. After a few steps, the minima do not shift any more, proving the stability of the approach. We also performed tests with a data set which was initially LENSCLEAN–self-calibrated with an *incorrect* lens model. With this data set, the minimum shifted slightly in the direction of the incorrect model. After some iterations, however, it moved back to the correct model, which proves the reliability of this approach.

---

[6] We used a patched version of the DIFMAP software, which does not change the weights in the course of self-calibration. Otherwise the weights would decrease in each iteration to minimize the residuals, leading to diverging solutions.

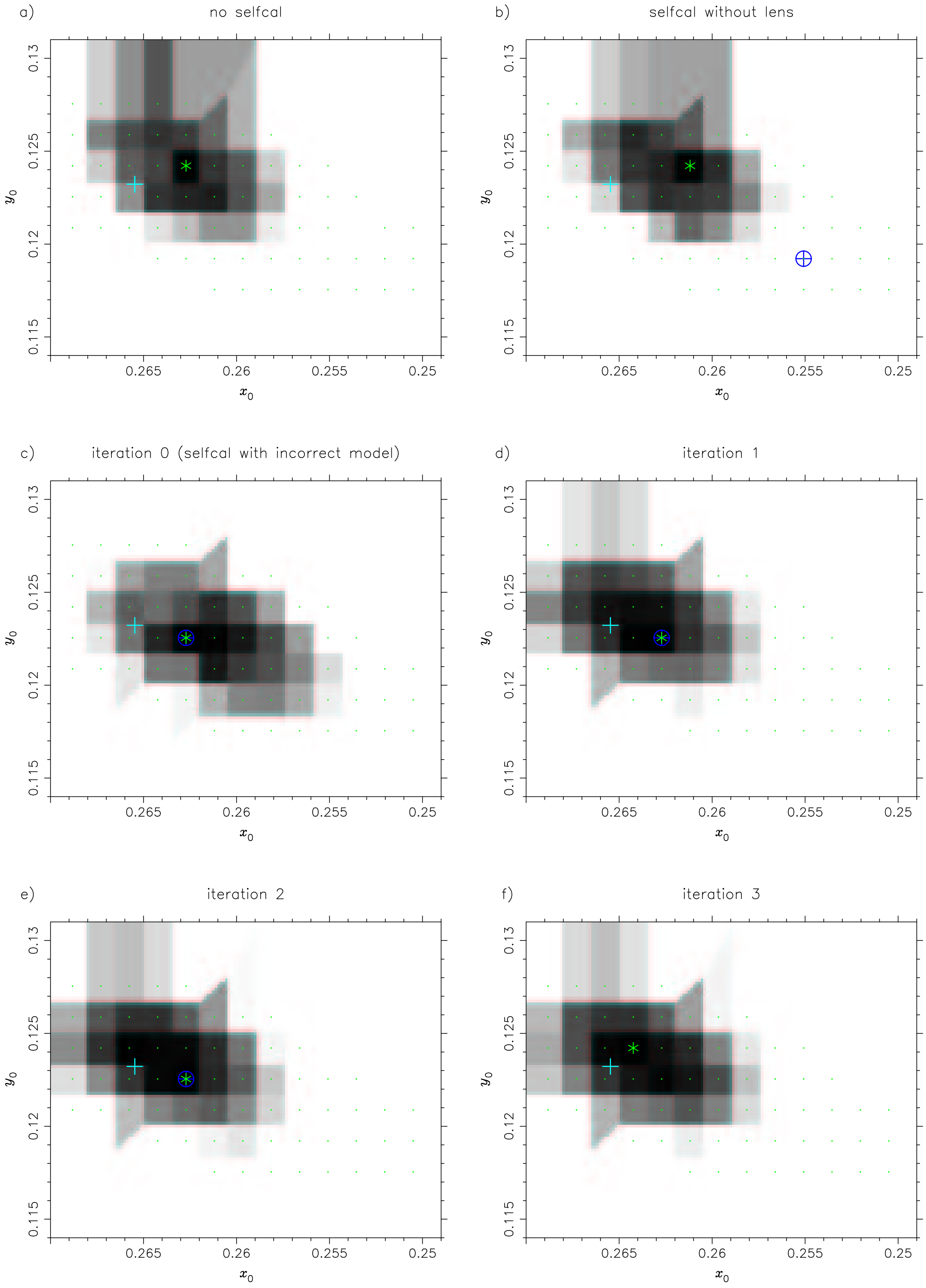


Figure 9.27: Residuals at all stages of the LENSCLEAN/self-calibration iteration for an artificial data set (direct plot of the data). (a) no self-calibration, (b) self-calibration without lens, (c)–(f) self-calibration with LENSCLEAN. The plus signs mark the model used to produce the artificial data, the asterisk the residual minimum, and the cross-hair the lens model used for the next self-calibration. In (b) this is an arbitrary incorrect model, in the other cases, it is the residual minimum.

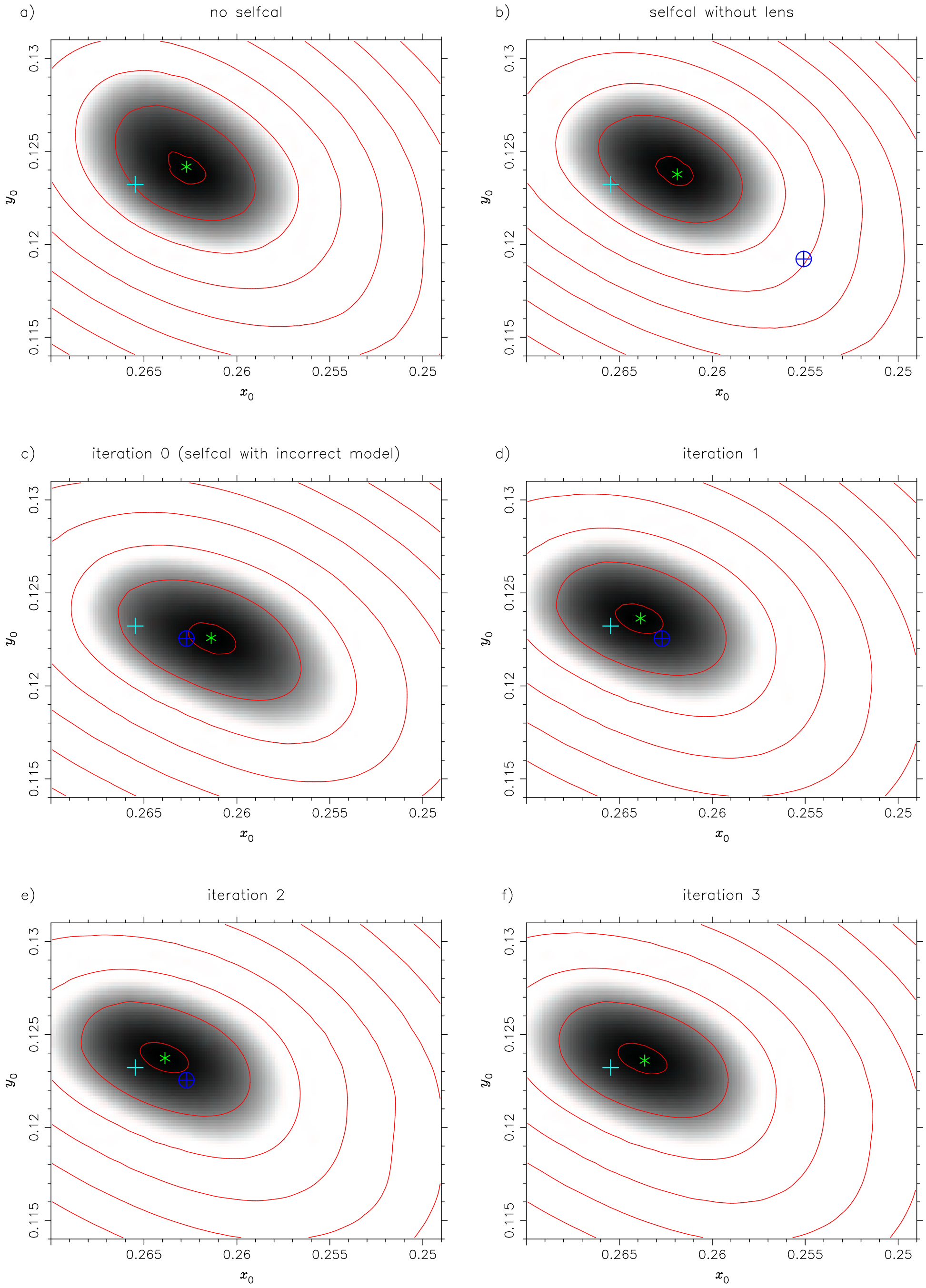


Figure 9.28: Residuals at all stages of the LENSCLEAN/self-calibration iteration for an artificial data set (interpolated and smoothed plot of the data with confidence limits of $0.1, 1, 2, 3, \ldots \cdot \sigma$). (a) no self-calibration, (b) self-calibration without lens, (c)–(f) self-calibration with LENSCLEAN. The plus signs mark the model used to produce the artificial data, the asterisk the residual minimum (smoothed data), and the cross-hair the lens model used for the next self-calibration. In (b) this is an arbitrary incorrect model, in the other cases, it is the residual minimum of the direct data.

In Figures 9.27 and 9.28 we show the results for an artificial data set. We started the loop with a self-calibrated (without lens) data set and used an incorrect lens model for the next step. We then performed a few iterations, in which the best lens model from the previous steps were used for the next self-calibration step. We used a lens centre of $(0\overset{\prime\prime}{.}265\,466, 0\overset{\prime\prime}{.}123\,222)$ to build the data set. The residual minimum at the end (smoothed version) is at $(0\overset{\prime\prime}{.}263\,636, 0\overset{\prime\prime}{.}123\,600)$. The error in the result is only 1.9 mas.

2000 iterations with included negative components (also for the self-calibration) were used for these test runs. As LENSCLEAN is not fully converged at 2000 iterations, the confidence limits in Figure 9.28 may be a bit optimistic.

The shifts induced by incorrect self-calibration were only small (especially when comparing the final result with the result without self-calibration), so that the whole self-calibration is only a small correction. Otherwise the residuals for the final *uv* data set could not be used to determine confidence limits, because the are only fits of the lens model parameters but not of the gain factors in self-calibration.

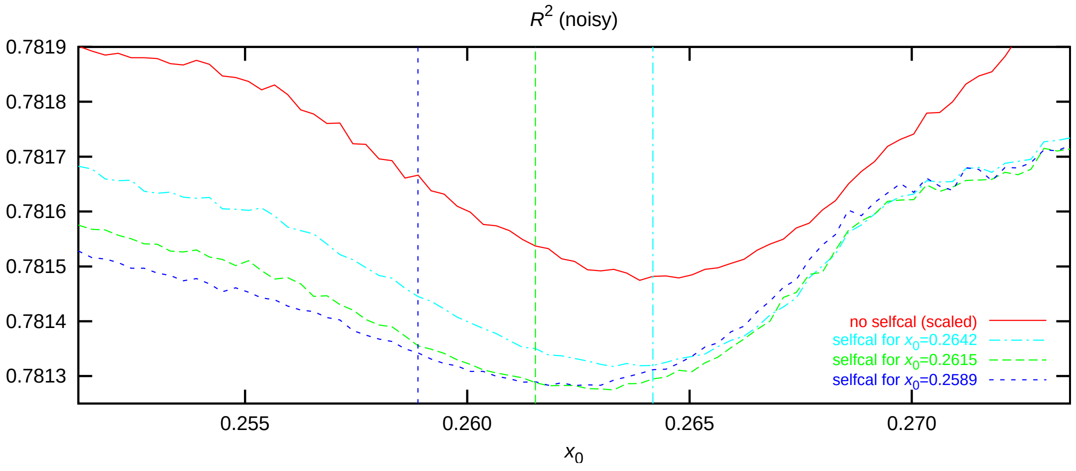


Figure 9.29: LENSCLEAN results with self-calibrated data sets. The upper line shows the residuals without self-calibration (5000 iterations, uniform weighting). The other three curves are calculated with data sets self-calibrated with three different lens models (marked by vertical lines). The right one of these is the correct lens model used to build the synthetic data set.

To make the collection of one-dimensional calculations complete, we also present self-calibration results for the same models as in the last sections in Figure 9.29. Self-calibration was performed with 2000 iterations without negative components with the uniformly reweighted data set. The shift of the minimum is of the order 1 mas when the correct model is used for self-calibration and much smaller than 50 % of the displacement of the lens centre for incorrect lens models. This leads to very rapid convergence of the LENSCLEAN/self-calibration iterations. We also notice that the curvature of the residual function, which is used to estimate confidence limits, is not reduced significantly with self-calibration. The final function is therefore a very good representation for the residuals without calibration errors and can be used to determine the final model parameter uncertainties.

The stable behaviour of self-calibration is mainly due to the very good *uv*-coverage of the long-track VLA data. With MERLIN data, self-calibration introduces many problems.

9.21 Results for the 15 GHz VLA data for B0218+357

9.21.1 Stokes LL and RR separately

We performed several LENSCLEAN/self-calibration loops as described in the last section, to obtain results for the lens models which are not influenced by initial calibration errors. The LL and RR polarization data sets were used independently to cross-check the results and to get a better feeling for the accuracy. The results for LL polarization are shown in Figures 9.30 and 9.31. On the first page, we plotted the fit results directly (grey values chosen with respect to the fit closest to each point), while the results were interpolated and smoothed for the second page. The same algorithm as for Figure 9.1 was used for the smoothing (see explanation on page 136). The results for RR polarization can be found in Figures 9.32 and 9.33 (page 176 and 177).

As described before, we started with a conventionally prepared and self-calibrated data set. Even this initial step was done independently for LL and RR. To produce these results LENSCLEAN was used in our ‘standard version’:

- FFT for dirty beam and maps, truncated $\exp\cdot\mathrm{sinc}$ as gridding convolution function (see section 8.10.1)
- unbiased LENSCLEAN component selection (see section 9.9)
- continuous fits of the compact components with gain $\gamma = 0.95$ (see section 9.11)
- negative components allowed (see section 9.12)
- amplification limit of 100 (see section 9.13)
- interpolation for secondary images (see section 9.14)
- exact subtraction from ungridded visibilities after 20, 50, 200, 1000 and 5000 iterations (see section 9.14)
- 5000 iterations with a loop gain of $g = 0.38$ (see section 9.16)
- 512×5 mas field size (see section 9.15)
- classical model fits as starting values, full fit with LENSCLEAN (see section 9.17)
- no CLEAN windows (see section 9.18)
- no extinction correction (see section 9.19)

In iterations 0 to 3, the lens model used for self-calibration was simply the one with the smallest residuals from the initial data set and the previous results. For iteration 4, we performed a third-order polynomial fit of the residuals in an area close to the minimum and used the minimum of this polynomial. In this way we reduced the effects of numerical noise on the results. The lens model parameters can be found in Table 9.1 on page 178.

In the LENSCLEAN runs which produced the emission models for the self-calibration steps we used almost the same variant as described above. The only modifications are a smaller number of iterations (2000) and the exclusion of negative components. A number of 50 alternating LENSCLEAN/self-calibration iterations were used.

For the final result (Figure 9.34 on page 179), we used the best-fitting models from the fifth self-calibration iteration, fitted a fourth-order polynomial to the parameters and calculated

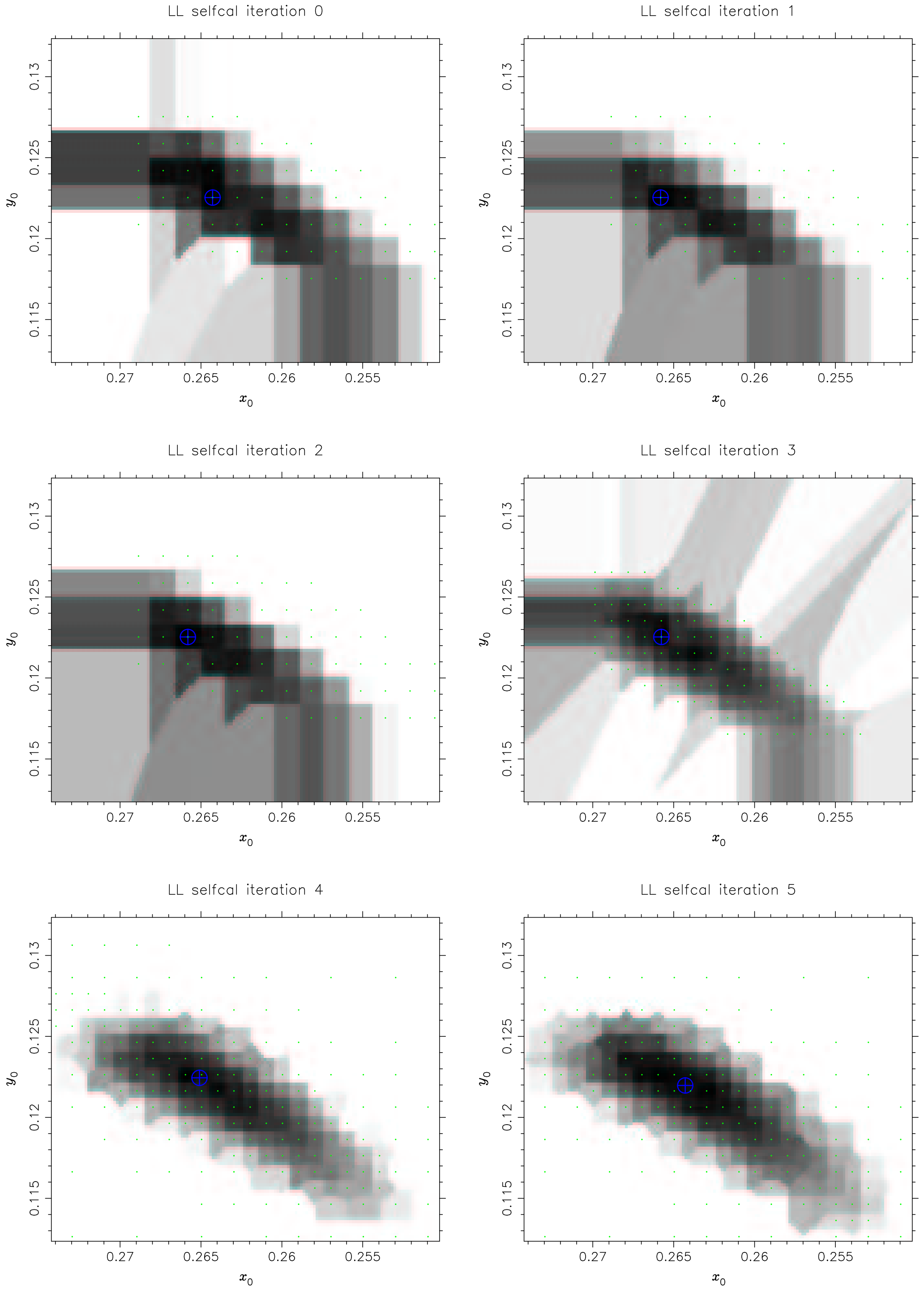


Figure 9.30: Residuals at all stages of the LENSCLEAN/self-calibration iteration for the LL polarization 15 GHz B0218+357 VLA data set (direct plot of the data). In the first four plots, the marked models (used for the next self-calibration steps) are the best fits. In the lower ones, they are minima of polynomial fits.

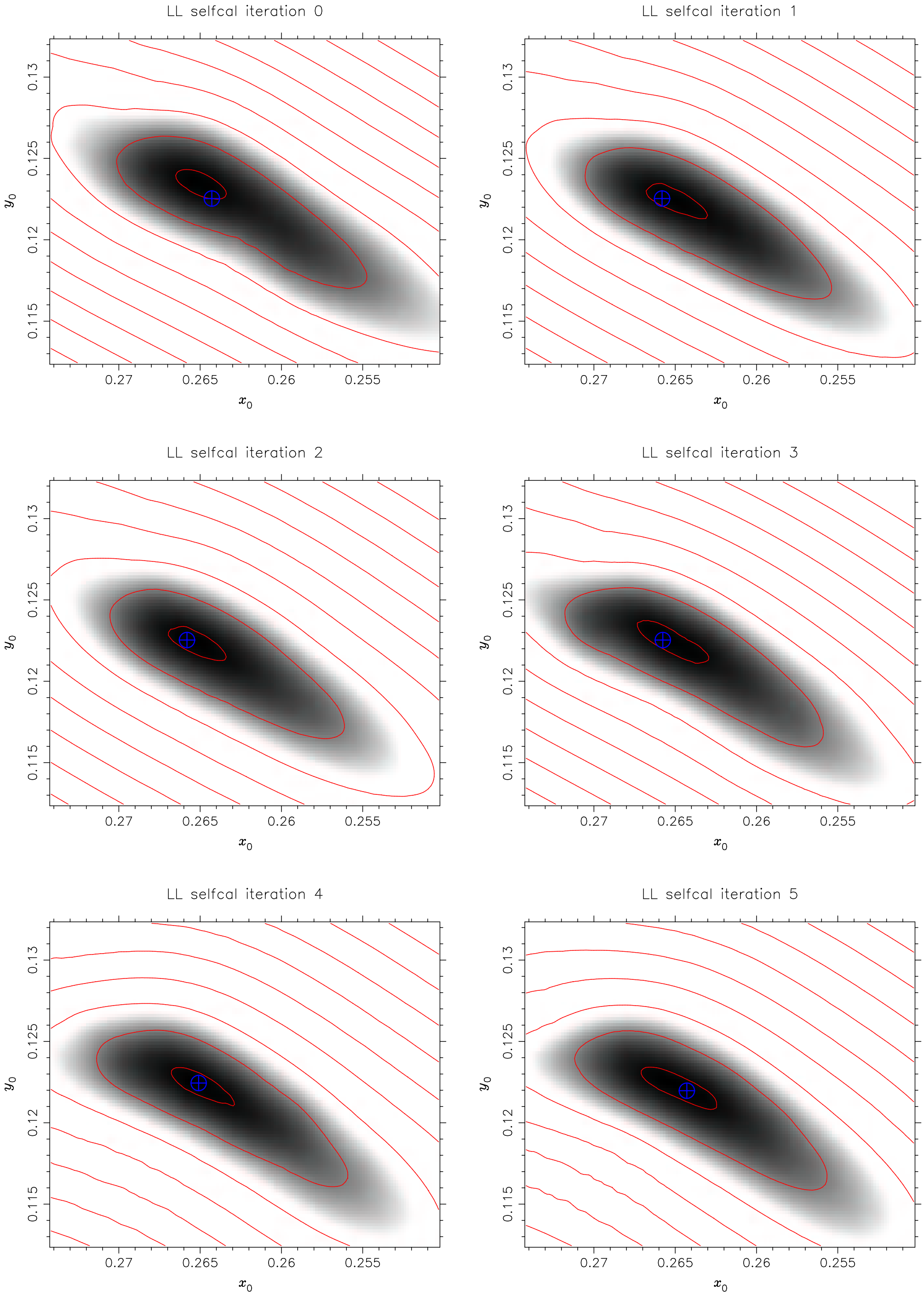


Figure 9.31: Residuals at all stages of the LENSCLEAN/self-calibration iteration for the LL polarization 15 GHz B0218+357 VLA data set (interpolated and smoothed plot of the data, confidence limits of $0.1, 1, 2, 3, \ldots \cdot \sigma$). In the first four plots, the marked models (used for the next self-calibration steps) are the best fits. In the lower ones, they are minima of polynomial fits.

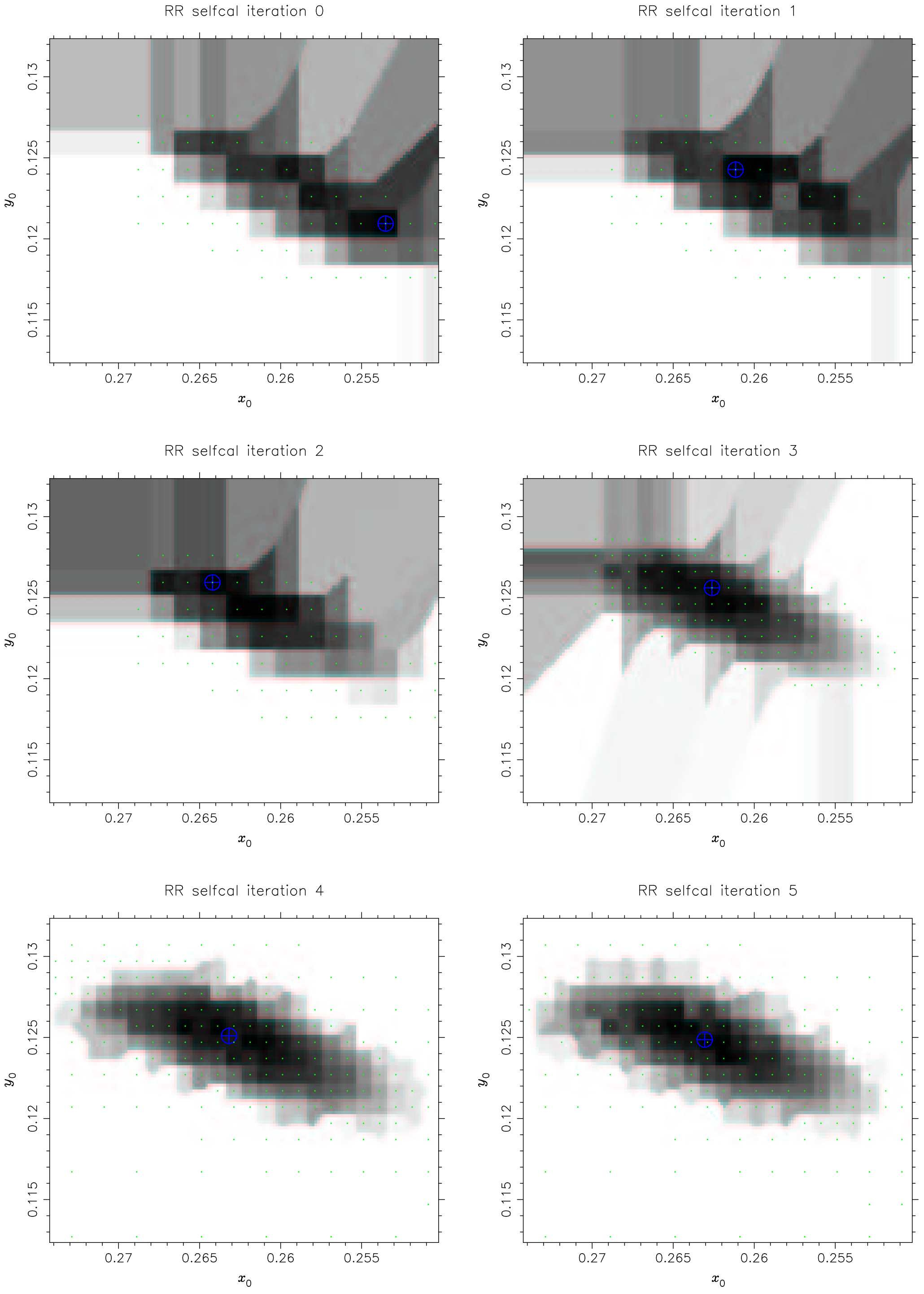


Figure 9.32: Residuals at all stages of the LENSCLEAN/self-calibration iteration for the RR polarization 15 GHz B0218+357 VLA data set (direct plot of the data). In the first four plots, the marked models (used for the next self-calibration steps) are the best fits. In the lower ones, they are minima of polynomial fits.

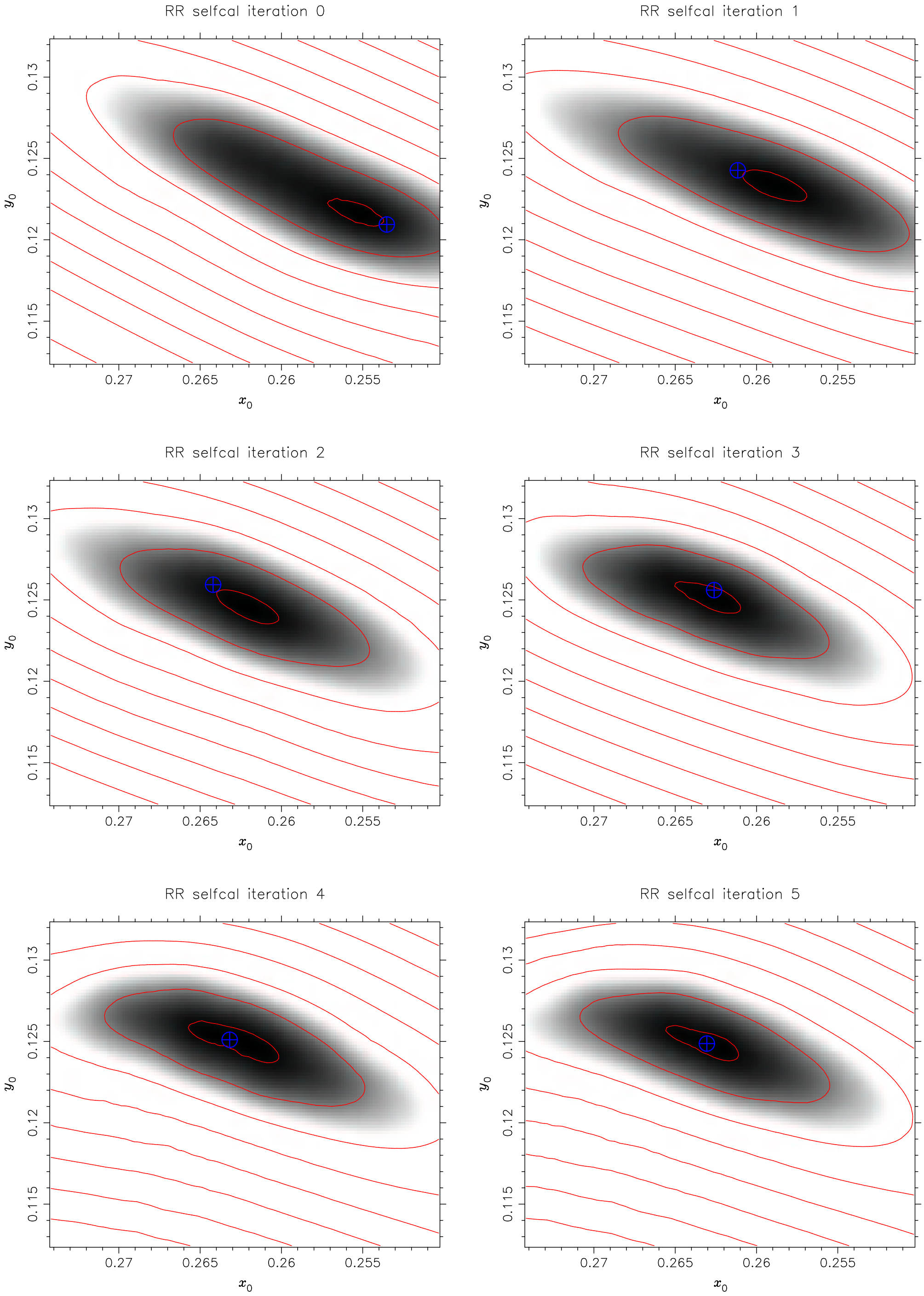


Figure 9.33: Residuals at all stages of the LENSCLEAN/self-calibration iteration for the RR polarization 15 GHz B0218+357 VLA data set (interpolated and smoothed plot of the data, confidence limits of $0.1, 1, 2, 3, \ldots \cdot \sigma$). In the first four plots, the marked models (used for the next self-calibration steps) are the best fits. In the lower ones, they are minima of polynomial fits.

Stokes	iteration	model	z_0	α_0	ε
LL	initial	best	0.″264 283 0.″122 534	0.″157 992	−0.008 729 −0.088 092
LL	1	best	0.″265 809 0.″122 534	0.″157 064	−0.012 809 −0.090 813
LL	2	best	0.″265 809 0.″122 534	0.″157 044	−0.012 861 −0.090 751
LL	3	best	0.″265 762 0.″122 534	0.″157 061	−0.012 743 −0.090 694
LL	4	fit	0.″265 092 0.″122 443	0.″157 506	−0.010 974 −0.088 910
LL	5	fit	0.″264 288 0.″121 972	0.″158 117	−0.009 165 −0.084 942
RR	initial	best	0.″253 520 0.″120 938	0.″163 704	+0.012 713 −0.063 850
RR	1	best	0.″261 151 0.″124 272	0.″159 380	−0.000 436 −0.091 045
RR	2	best	0.″264 204 0.″125 938	0.″156 975	−0.007 711 −0.106 740
RR	3	best	0.″262 608 0.″125 605	0.″158 197	−0.003 322 −0.101 163
RR	4	fit	0.″263 185 0.″125 101	0.″157 962	−0.004 882 −0.099 760
RR	5	fit	0.″263 055 0.″124 870	0.″158 117	−0.004 541 −0.098 240

Table 9.1: Lens models used in the LENSCLEAN/self-calibration iterations for the 15 GHz B0218+357 VLA data. See Figures 9.30 to 9.33.

LENSCLEAN residuals on a finer grid, using the polynomial fits for the parameters. The minima for LL and RR are not exactly at the same position, but are both compatible with the mean of the two. A preliminary version of these results was presented in Wucknitz (2002b).

The normalized residuals in the last self-calibration iteration are 0.7708 and 0.7988 for LL and RR polarization. The expected value without self-calibration would be 1 ± 0.009,[7] as can be calculated with equation (2.54). The scatter is of no practical value here, because the weights are never known with an accuracy better than a few per cent. The values are significantly smaller than 1, because the self-calibration is able to remove parts of the noise. Tests with CLEAN and self-calibration in DIFMAP (without a lens) resulted in residuals of ca. 0.78, very similar to (and within the $1\,\sigma$ limit compatible with) our results. The LENSCLEAN/self-calibration loop with artificial data also led to residuals of 0.78. These results show that the data are consistent with the assumed lens model, producing residuals exactly in the expected range.

We nevertheless have to keep in mind that this test is not very sensitive, because of the

---

[7] We neglected the number of free parameters of the emission model here, which is approximately the number of independent beams in the map or the area of the field divided by area of the beam, which is of the order a few 10^2 and therefore much smaller than the number of measurements (ca. $2 \cdot 10^5$). A more accurate estimate of this number of parameters is therefore not necessary, as long as we can be sure that it does not vary much with the lens model.

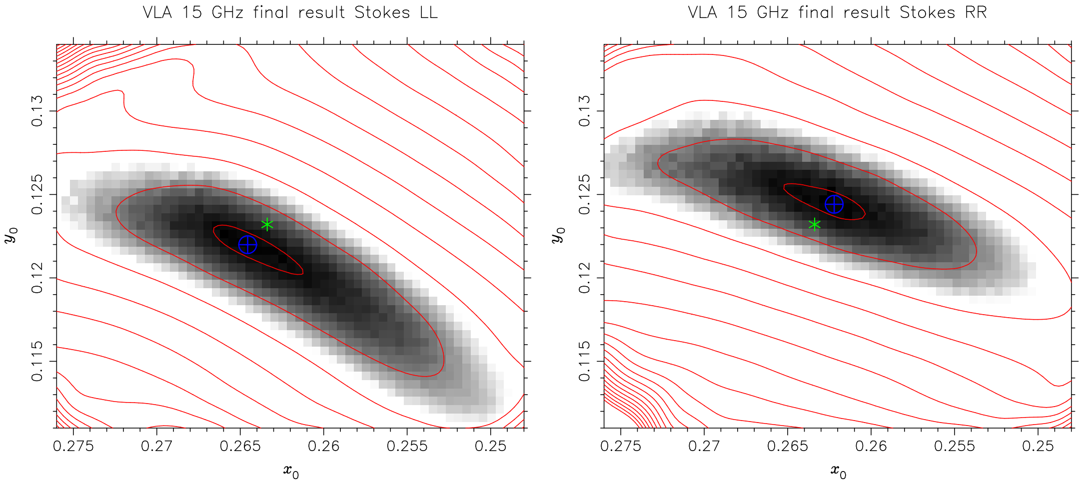


Figure 9.34: Final residuals for the 15 GHz B0218+357 VLA data set for LL and RR polarization. The results are shown directly as greyscale and are smoothed for the contour lines (confidence limits of $0.1, 1, 2, 3, \ldots \cdot \sigma$). The cross hairs mark the residual minima for LL and RR, the asterisk the mean of the two. The increased density of contour lines in the edges is an artifact of the interpolation.

inaccuracy of the weights, which makes the normalization of residuals difficult. We will see later that the residual map shows no significant structure correlated to the lens model. This is probably a better measure for the goodness of fit in this case than the formal error statistics.

9.21.2 Stokes I final result

To reach the highest possible accuracy, we returned to the combined Stokes I data set. The emission model in the self-calibration loop was built with LENSCLEAN for Stokes I (mean of LL and RR), but the self-calibration with DIFMAP was performed for LL and RR separately. Differences in the calibration errors of different Stokes parameters could be corrected in this way. Only positive components were allowed in building the emission model for self-calibration of the uniformly weighted data. For natural weighting, we started with 3000 iterations with uniform weighting and switched to natural for another 2000 iterations. Because of this combination, negative components were also allowed.

For the LENSCLEAN model fitting, negative components were allowed in both cases for reasons discussed before (section 9.12). 5000 iterations were used for uniform and $3000+3000$ for natural weighting. This brings the algorithm so close to convergence, that the shape of the residual function would not change significantly with more iterations.

As a starting model, we used the mean of the best z_0 for LL and RR, selected the other parameters for this lens position separately from the fits of both polarizations and used again the mean of LL and RR. This provides a better starting model than taking the mean of the best LL and RR models directly. Convergence of the LENSCLEAN/self-calibration iteration was very fast. After three loops, the lens model did not change any more within small fractions of the accuracy.

Figure 9.35 shows the final results for both weighting schemes. The confidence regions are significantly smaller than in the separate LL and RR results. The results for uniform and natural weighting are not exactly equal, but there exists a model between the two which is consistent

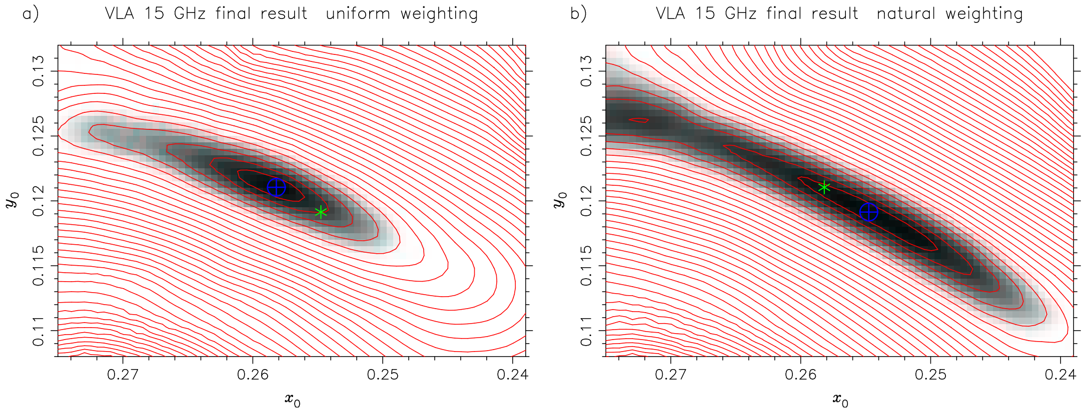


Figure 9.35: Final residuals for the 15 GHz B0218+357 VLA data set for Stokes I. The results are shown directly as greyscale and are smoothed for the contour lines (confidence limits of $1, 2, 3, \ldots \cdot \sigma$). (a) uniform, (b) natural weighting. The cross hairs mark the respective residual minima, the asterisk marks the minimum for the other weighting scheme.

with both weightings within the $1\,\sigma$ confidence region. The formal uncertainty is smaller for uniform weighting and we also expect smaller systematic errors in this case. The secondary local residual minimum near $z_0 = (0\farcs273, 0\farcs126)$ in the naturally weighted results can be ruled out as a realistic lens model by using the uniformly weighted results. The lens model parameters together with confidence limits are presented in Table 9.2 and 9.3.

parameter	best	$1\,\sigma$	$2\,\sigma$	$3\,\sigma$
x_0	$0\farcs259\,335$	$+0\farcs0018$ $-0\farcs0007$	$+0\farcs0033$ $-0\farcs0022$	$+0\farcs0063$ $-0\farcs0047$
y_0	$0\farcs121\,048$	$+0\farcs0007$ $-0\farcs0003$	$+0\farcs0017$ $-0\farcs0013$	$+0\farcs0027$ $-0\farcs0023$
α_0	$0\farcs161\,631$	$+0\farcs0004$ $-0\farcs0010$	$+0\farcs0012$ $-0\farcs0020$	$+0\farcs0023$ $-0\farcs0039$
ε_x	$+0.004\,318$	$+0.0013$ -0.0032	$+0.0038$ -0.0066	$+0.0073$ -0.0130
ε_y	$-0.070\,718$	$+0.0022$ -0.0060	$+0.0083$ -0.0132	$+0.0152$ -0.0230
$\lvert\varepsilon\rvert$	$0.070\,850$			
θ	$-43\fdg253$			
ΔT [arcsec^2]	$0.039\,260$	$+0.0006$ -0.0003	$+0.0012$ -0.0009	$+0.0022$ -0.0018

Table 9.2: Parameters of the best fitting lens models for B0218+357 (final result for VLA 15 GHz Stokes I data, uniform weighting).

Note that the confidence limits are calculated for one parameter each and not for the combination of all parameters. The limits (especially $1\,\sigma$) are not expected to be extremely accurate, because they were calculated by taking the range of all allowed lens models on a $0.5 \times 0.5\,\mathrm{mas}^2$ grid for z_0. For uniform weighting, this included only 7 lens models. The $3\,\sigma$ limits for natural

parameter	best	1σ	2σ	3σ
x_0	0.″254764	+0.″0022 −0.″0038	+0.″0062 −0.″0063	+0.″0222 −0.″0093
y_0	0.″119141	+0.″0011 −0.″0024	+0.″0031 −0.″0039	+0.″0081 −0.″0054
α_0	0.″163402	+0.″0016 −0.″0011	+0.″0026 −0.″0034	+0.″0035 −0.″0171
ε_x	+0.009143	+0.0039 −0.0030	+0.0066 −0.0098	+0.0084 −0.0611
ε_y	−0.058528	+0.0126 −0.0072	+0.0196 −0.0221	+0.0261 −0.0889
$\lvert\varepsilon\rvert$	0.059238			
θ	−40.°561			
ΔT [arcsec^2]	0.037961	+0.0008 −0.0015	+0.0023 −0.0025	+0.0078 −0.0036

Table 9.3: Parameters of the best fitting lens models for B0218+357 (final result for VLA 15 GHz Stokes I data, natural weighting).

weighting include the secondary residual minimum.

We also calculated residuals for a noise-only data set with the same lens models to estimate the effect of the varying effective number of parameters. If the real residuals are corrected for this bias effect, the minimum shifts by less than 1 mas, which is well within the expected error bars. The correction was thus not included in the results presented here.

To check the accuracy of our error estimates for uniform weighting (it is straight χ^2 statistics for natural weighting), we performed Monte Carlo simulations with a known lens model and artificial data with the same noise characteristics as the real data. The results of 21 runs are shown in Figure 9.36. 17 (81 %) of the results lie within the 1σ limit and all within 2σ. The small number of realizations is not sufficient for a very robust statistics, but the error limits from the residual statistics seem to be a very realistic estimate. They are at least not excessively optimistic.

It is now possible, to calculate the amplification ratio for the best lens models. This depends slightly on the exact image positions; possible values are between ca. 3.77 and 3.83. This result is not affected by the ring emission and is thus more reliable than direct measurements of the flux ratio.

The relatively high ellipticity of the mass model might be a reason to treat this results with caution. From equation (6.51), we know that the ellipticity of the mass distribution is approximately three times as high as that of the potential. If this is then converted to an axial ratio of the mass, the result is $a/b = 1.54$ for the uniformly weighted result and 1.43 for the naturally weighted. The optical images (see Figures 7.5 and 7.6) look more spherical, but they can only represent the luminous matter and not all matter responsible for the lensing effect. We also have to keep in mind that the accuracy of the ellipticity is not very high, and that the data are also compatible with less elliptical mass distributions.

If the ellipticity is due to a tilted symmetrical disc, the inclination angle has to be 45–50° (0° would be face-on, 90° edge-on). But this model is not very realistic, given the expectation that most of the mass is hidden in the dark matter halo of the galaxy.

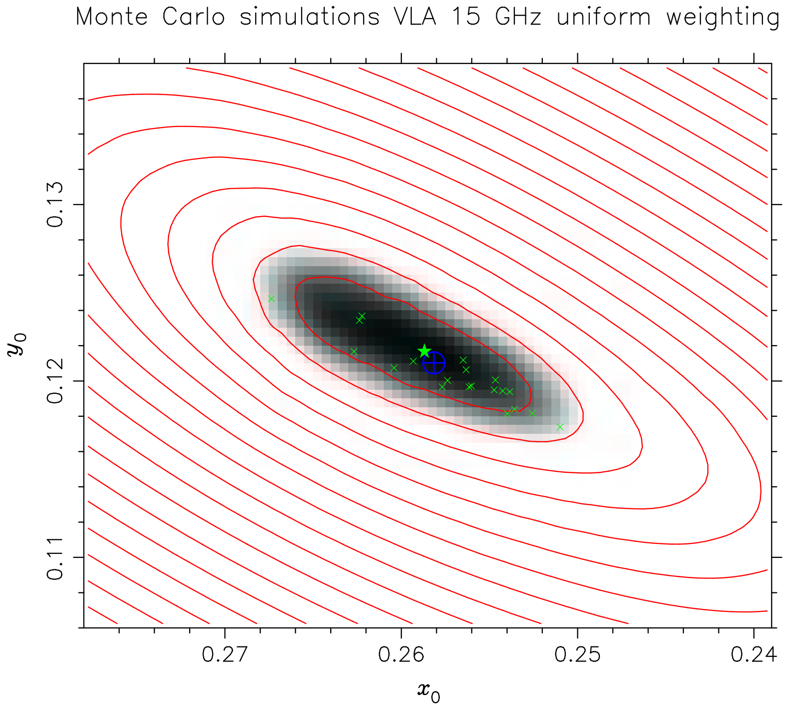


Figure 9.36: Results of Monte Carlo simulations for the 15 GHz B0218+357 VLA data (Stokes I). We picked one of the runs with a minimum (star) close to the model used to build the data (crosshair) for the greyscale and contour (confidence limits of $1, 2, 3, \ldots \cdot \sigma$) plot. The Monte Carlo results are shown as crosses.

The position angle of the major axis is between $-43°$ and $-41°$ in a mathematical sense and between $-47°$ and $-49°$ as an astronomical p.a. This should be compared with the p.a. of the image separation ($67\overset{\circ}{.}8$ astronomical). The projected axis of the galaxy is close to perpendicular to the image separation, which would lead to only small differences in the radial velocities of absorption lines in the spectra. Measurements for both components have not been possible yet but are expected for the near future (see e.g. Menten & Reid, 1996). Such measurements could help in constraining the circular velocity of the disc and the inclination angle. It would then be possible to separate the effect of the disc and the halo to some degree and break the disc/halo degeneracy.

9.21.3 The Hubble constant

The ΔT value for the LENSCLEAN models are included in Tables 9.2 and 9.3. A very conservative approach to estimate the errors is to use the minimum from natural and the maximum from uniform weighting. This leads to a $2\,\sigma$ region of $\Delta T = 0.035461\text{–}0.04046\,\mathrm{arcsec}^2$, e.g. equation (7.4). The accuracy is thus 6.6 %. The $2\,\sigma$ uncertainty of the uniformly weighted result alone is only 2.7 %.

The time-delay of $\Delta t = (10.5 \pm 0.4)$ days from section 7.5 and the distance parameters from Table 7.7 for the Einstein-de Sitter cosmological model ($d_{\mathrm{eff}} = 19\,609\,\mathrm{km\,s^{-1}\,Mpc^{-1}\,days\,arcsec^{-2}}$)

can be used to translate this to

$$H_0 = (70.9 \pm 5.4)\,\mathrm{km\,s^{-1}\,Mpc^{-1}} \quad , \tag{9.25}$$

see equation (3.43). The errors are the formal $2\,\sigma$ confidence limits from the uncertainty in the lens model *and* the time-delay. This value is in very good agreement with the results from the HST key project (Mould et al., 2000) of 71 ± 6 ($1\,\sigma$) but is incompatible with Parodi et al. (2000) and Sandage (1999), who determine values of 58.5 ± 6.3 (90 %) and 53 ± 7 ($1\,\sigma$), respectively. See also Tammann et al. (2002) for an up-to-date overview of the SN Ia results of this group (60 ± 5).

One should keep in mind, however, that this result is only valid for the singular isothermal ellipsoid lens model. We will discuss the effect of a different radial mass distribution in section 9.26. The preliminary results for different models give some evidence that the additional error will only be of the order 1 % for B0218+357.

The Hubble constant will be about 6 % higher (75 ± 6) for a flat Λ cosmological model (see Table 7.7). The uncertainties from the time-delay, the unknown cosmological parameters and the lens model are thus all of the same order of magnitude. To improve the results, all three factors have to be constrained with better accuracy.

We can now determine the age of the universe T_{U} from this result. For the Einstein-de Sitter universe it can be calculated by

$$T_{\mathrm{U}}(\mathrm{EdS}) = \frac{2}{3}\frac{1}{H_0} \tag{9.26}$$

$$= 9.2 \cdot 10^9\,\mathrm{years} \quad . \tag{9.27}$$

The higher value of H_0 for the flat Λ model is more than compensated by the change of the former factor $2/3$:

$$T_{\mathrm{U}}(\mathrm{flat}\,\Lambda) = 0.964\,\frac{1}{H_0} \tag{9.28}$$

$$= 12.5 \cdot 10^9\,\mathrm{years} \tag{9.29}$$

We omit the details of how these ages are calculated. They can be found in any textbook on classical cosmology.

A preprint published after this work was finished (Kochanek, 2002) claims that time delays of lenses with reliable models are compatible with the rather high values of H_0 from the HST key project only if the mass concentration is as compact as the light distribution ($\beta < 1$ in the picture of power-law models). Our result for B0218+357 does not confirm this view. The isothermal models lead exactly to the value preferred in that publication. The method of comparing values of H_0 from other methods with lensing results to learn more about the lens models, is in our opinion only of limited value because the results depend on so many astrophysical assumptions. We prefer to constrain the mass models with the lensing effect or by direct measurements. Systems with extended sources provide the possibility to measure the mass distribution relatively accurate. For B0218+357, we refer to the discussion in sections 7.6.2 and 10.2. The data are not compatible with $\beta < 1$ but favour values of $\beta = 1.04$. This is in contradiction with the interpretation from Kochanek (2002) that mass follows light more than it follows standard dark matter models.

9.22 The 5 GHz MERLIN data set

9.22.1 Self-calibration

This data set and its problems are explained in more detail in section 7.10. Another problem concerns self-calibration. Although the *uv* coverage is quite good in the combined data set, the number of telescopes of the MERLIN part is only $n = 6$, which gives self-calibration too much freedom. We nevertheless tried the same approach as with the 15 GHz VLA data set, alternating LENSCLEAN model fitting and self-calibration with the best lens models, see section 9.20. Unfortunately, it showed that the shift of the best model introduced by self-calibration is of the same order of magnitude as the error in the lens model used. This means that the LENSCLEAN/self-calibration iteration does not converge very well and that there is at least the possibility that it converges to an incorrect limit. We therefore did not use self-calibration with this data set for the results we want to present below. The 5 GHz VLA data were not included in the analysis to avoid the calibration problems of the combination.

9.22.2 Fits without self-calibration

Because self-calibration does not work well with LENSCLEAN and the MERLIN data, we used the data set directly as it was prepared by Andy Biggs when he produced the high-quality maps.

Experiments with different parameter sets for the LENSCLEAN algorithm showed that the problems become worse when more iterations are used. This means that the problems (calibration errors, absorption in parts of the system, emission from the lens galaxy itself) are significant only on small flux scales and do not show in earlier stages of the LENSCLEAN iteration. The best results were produced with 2000 iterations and the same algorithm as with the 15 GHz VLA data. The only difference is a smaller gain of 0.95 (compared to 0.98 for VLA) for the subtraction of the compact components in the first continuous step. The gain was decreased, because the ring is relatively stronger at 5 GHz, so that a smaller part of the bright images can be subtracted before emission from the ring is affected.

Results for Stokes LL, RR and I and for the three frequencies separately and combined are shown in Figure 9.37. No extinction correction was applied for these results. The residuals were smoothed with the usual locally weighted quadratic fit with a Gaussian σ of 5 mas to produce the plots. We notice several problems with the data. No nice quadratic minima are present and the results for the sub-data sets are not compatible with each other. The formal error bars are extremely small but cannot be believed. The behaviour of the combined data set, where a secondary minimum begins to appear, is not understood yet. It probably is related to small differences in the calibration of the different frequencies. The small errors do not affect the classical mapping considerably, but they seem to confuse LENSCLEAN.

9.22.3 Artificial MERLIN data

Figure 9.38 (page 186) shows results for an artificial data set, which was built to represent the complete MERLIN data set (all three frequencies, Stokes I). The emission model of the source was a best fit of the real data, the same extinction of about 20 % was used in LENSCLEAN and in producing the data set. We show results for uniform and natural weighting. The models used for the plot are again a locally weighted quadratical fit to results from LENSCLEAN model fitting. The same models were used for uniform and natural weighting. This procedure does not change the results significantly, but reduces the computing time to produce the plots. Because of numerical noise, the residuals have to be smoothed and interpolated to determine the minimum.

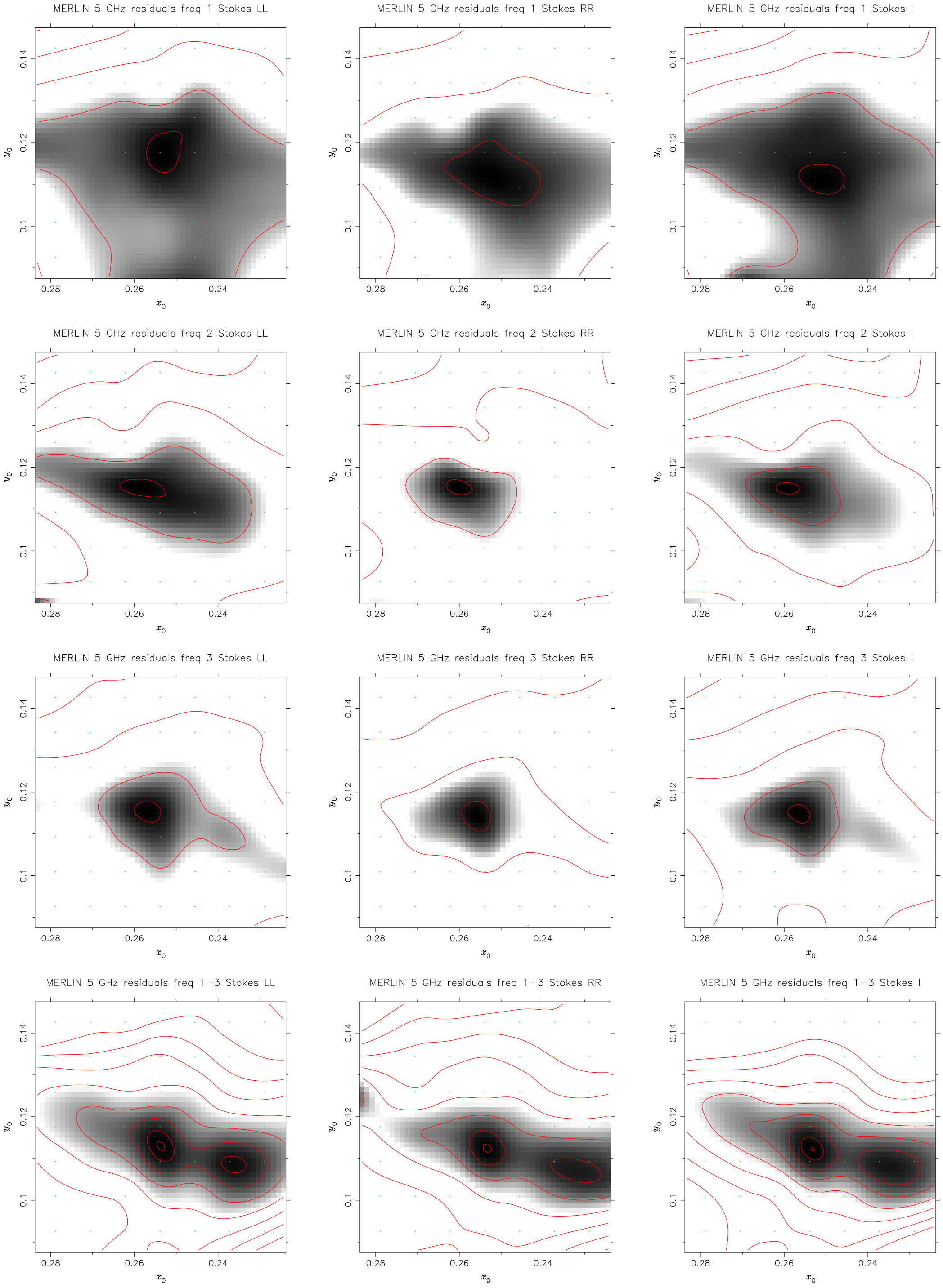


Figure 9.37: LensClean residuals for the 5 GHz B0218+357 MERLIN data set. From left to right, we show Stokes LL, RR, I. The first three rows are calculated for frequencies 1, 2, 3, the bottom row for the combined MFS data set. Fits were performed at the position marked by points. The greyscale and contour lines (formal significance levels of $1,5,15,20,25,30,35\,\sigma$) use a smoothed and interpolated version.

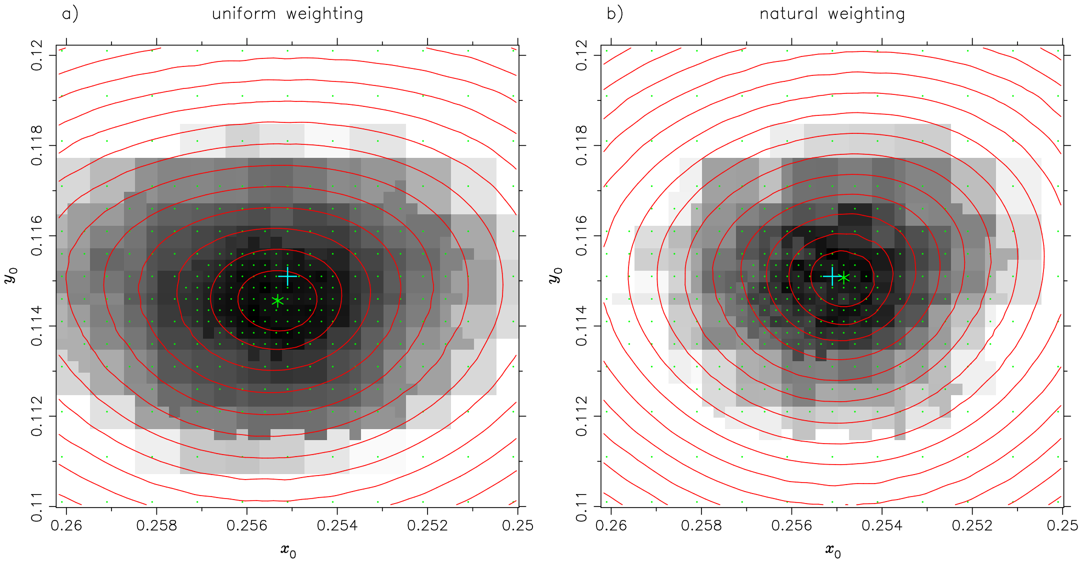


Figure 9.38: LENSCLEAN residuals for an artificial data set resembling B0218+357 at 5 GHz with MERLIN (all three frequencies, Stokes I). (a) uniform weighting 2000 iterations, (b) natural weighting $2500 + 2500$ iterations. The results are shown directly as greyscale and are smoothed for the contour lines (confidence limits of $1, 2, 3, \ldots \cdot \sigma$). The asterisk marks the residual minimum (smoothed version), the cross marks the model used to build the data set $z_0 = (0\overset{\prime\prime}{.}2551, 0\overset{\prime\prime}{.}1151)$.

This was done with a local Gaussian weighted quadratical fit ($\sigma = 1$ mas). The error in the result for the lens centre is 0.585 mas for uniform and 0.255 mas for natural weighting. Both results are well within the formal $1\,\sigma$ limit. The errors for α_0 are 0.065 and 0.093 mas, for the ellipticity they are 0.0020 and 0.0006.

These impressively accurate results show that the MERLIN data set *can* provide very strong constraints, once the calibration problems are solved. We plan to do a new reduction of the data from the beginning and perform the self-calibration and correction for spectral index differences in a way appropriate for LENSCLEAN. The small number of telescopes will, however, always introduce too much freedom for the models if self-calibration is used, even if the complete MFS data set is used.

Other calculations also showed that, in the absence of calibration errors, the extinction can be determined from the MERLIN data alone with a precision of a few per cent. The extinction also shifts the lens centre by about 3 mas for 10 % for simulated data.

9.22.4 Determining the source shift and extinction

The fact that the same lens model must be responsible for the apparent images at all frequencies (provided that no significant refraction in the lens plane is present) can be used to estimate the possible extinction and shift at lower frequencies. Less information is needed for this purpose than for a complete lens model fitting, so that the 5 GHz MERLIN data can be used for this purpose, despite their problems with free LENSCLEAN.

As substitute for a real simultaneous fit of the lens models to the VLA and MERLIN data, we used best fitting lens models from LENSCLEANing the 15 GHz VLA data and applied the same lens models to the 5 GHz data, allowing for a possible shift and an extinction in the A component. Formally, we fixed the data but varied the lens centre freely. Since all coordinates are measured relative to the A component, a shift of z_0 relative to A is equivalent to the same shift of the lensed source relative to the then fixed position of the lensing galaxy, but in the opposite direction.

The VLA data are relative to the J2000 equinox, while the MERLIN data are in the B1950 system. To avoid changing the *uv* data set, we applied the precession to the VLA lens model before fitting it with the MERLIN data. The shift of the coordinate system is not relevant, because all positions are relative to A. The rotation, on the other hand, must be applied. In terms of the astronomical position angle, the lens model must be rotated like follows:

$$\text{p.a.}_{1950} = \text{p.a.}_{2000} - 0\overset{\circ}{.}1964 \tag{9.30}$$

This rotation corresponds to a shift of about 1 mas of the B component, which is highly significant and definitely has to be taken into account to achieve correct results.

Several lens models from LENSCLEANing the VLA data were used in this procedure to estimate the accuracy of the results. We also performed the fits with subsets of the MERLIN data, consisting of the three frequencies and LL/RR polarization separately. Calibration errors of the data act differently at different frequencies and can be detected in this way. We used a range of extinction values for the numerical fits. In the process of model fitting (only the parameter z_0), the extinction was used as a fixed parameter. Finally, the residuals are plotted against the extinction to find the best model and to detect possible problems.

The extinction was taken into account by reducing the amplifications given by the lens model near the A component. A circular Gaussian with FWHM of 20 mas was used for the shape. Other methods to apply the extinction lead to very similar results. The size of the absorbing region is also not critical.

Figure 9.39 shows the results for one lens model and different sub-data sets of the MERLIN data (500 iterations, uniform weighting). The curves look very similar and the minimum is almost at the same extinction in all curves. Only for frequency 1, the optimal extinction is somewhat smaller than for the others. This confirms the view that possible small calibration errors (maybe caused by the spectral index correction) do not change the result for the extinction. There are still noticeable problems, e.g. the very different residual levels. Residuals are very high for the combined data set, especially for RR polarization and total intensity I. This is a property of the data itself and not of LENSCLEAN. Compare with the discussion in section 7.10.1. The similarity of the residual curves gives some confidence in the result for the extinction when it is compared with the prominent differences for free LENSCLEAN model fitting in Figure 9.37.

The influence of the VLA lens model itself is shown in Figure 9.40 on page 189, were several different lens models were used. These models are results at different (partly preliminary) stages of the VLA model fitting and self-calibration process. The details are shown in Table 9.4 (page 190). The models marked with $+$ and $-$ were derived from the 'UNI pre' model by changing the corresponding parameters within their total $1\,\sigma$ error bars. The confidence limits were calculated for the total five free parameters. Concerning the shift of the minimum (one parameter), this is equivalent to more than $2\,\sigma$. For the $\alpha\pm$, $\varepsilon_x\pm$ and $\varepsilon_y\pm$, the other parameters were fixed. For $x_0\pm$ and $y_0\pm$, the error ellipse of real fits was used. This collection should cover all possible lens models. In the result the minimum does shift significantly for some model variants, but it is in no case consistent with no extinction.

Figure 9.41 (page 190) shows the relative shift of the lens centre (MERLIN$-$VLA). As noted before, this is equivalent to the opposite shift of the (lensed) source relative to A. If the real extinction should be small, the source has to be shifted eastwards at 5 GHz. The shift would

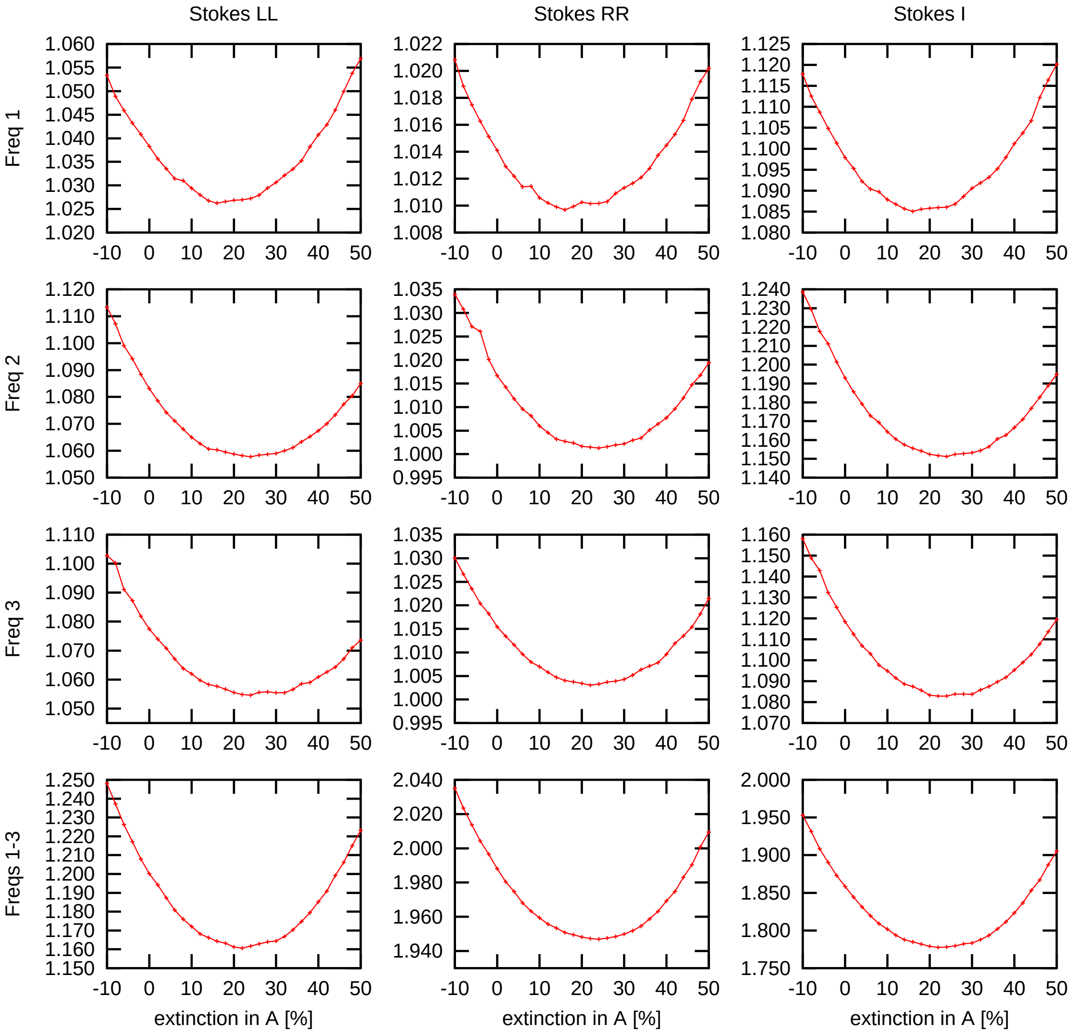


Figure 9.39: Residuals for 5 GHz MERLIN data vs. extinction in A. Lens models were fitted with free lens position and all other parameters fixed to one of the best fitting VLA models (called 'I UNI 1' in Fig. 9.40 and Tab. 9.4). From left to right: Stokes LL, RR and I. From top to bottom: Frequencies 1, 2, 3 and 1–3. 500 iterations with uniform weighting were used.

therefore approximately be down the jet (compare Figure 7.14). The fact that the zero positions of Δx_0 and Δy_0 are both at the same extinction level, which itself is compatible with the residual minima near 20 %, is a strong hint, that the frequency dependent flux ratios are not caused by an effective shift of the source but by extinction in the A component. If a shift was the reason for the observed effect, the zeros would not necessarily be at the same position and the minimum would be close to zero extinction.

Nevertheless, we should keep in mind that if the A component is significantly dimmed by extinction, similar effects have to be expected at other positions of the lens plane as well. This means that any LENSCLEAN results of the 5 GHz data have to be interpreted with caution. In any case, the results are clearly *not* compatible with no extinction at all.

The final decision about frequency dependent source positions in B0218+357 has to wait for the analysis of multi-frequency VLBI (VLBA+Effelsberg) observations that were taken in

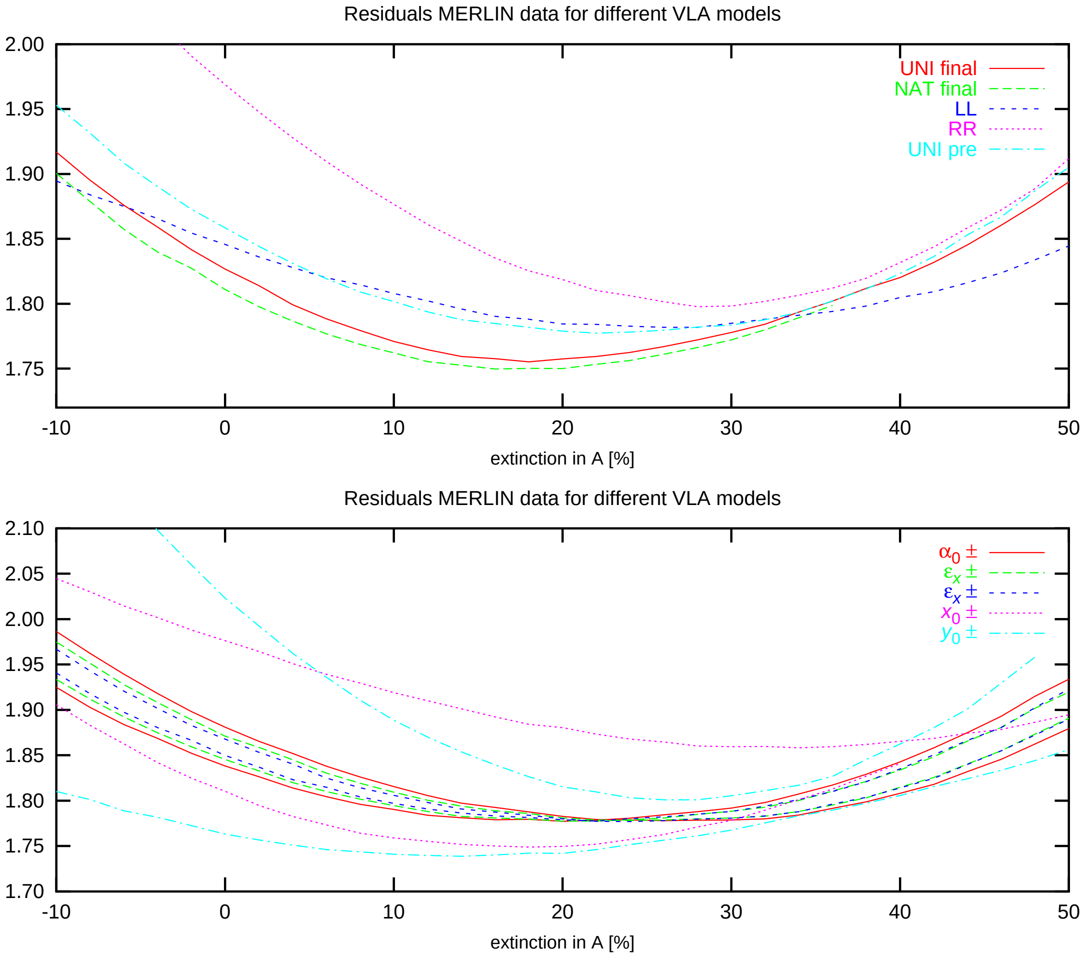


Figure 9.40: Residuals for 5 GHz MERLIN data vs. extinction in A. Lens models were fitted with free lens position and all other parameters fixed to some of the best fitting VLA models. For a description of the lens models, see Table 9.4. 500 iterations with uniform weighting were used with the complete MFS data set (Stokes I). The minima cluster around an extinction of about 20 %. None of the curves are compatible with zero extinction.

January 2002 (PI: Richard Porcas). These observations include phase referencing with several calibrator sources and should lead to more than sufficiently accurate absolute astrometry of both components at 1.6, 2.3, 5, 8.4, and 15 GHz. Any potential shift can then be measured directly. A uniform analysis of the data taken at one epoch can also provide better measurements of the frequency dependent flux densities of both components.

9.23 Algebraic LENSCLEAN

For future work and to reconstruct the source map, a more formal analysis of LENSCLEAN is required. The idea is to apply the general linear least squares theory (section 2.2) to the lensed scenario in the same way as it was done for the unlensed case in section 8.8.1. The lens model is kept fixed for the moment, and the relevant model parameters are now the emission components *in the source plane*. It is then possible to transform the concept of dirty map and dirty beam to

name	description
LL	final result Stokes LL, uniform weighting, see Fig. 9.34
RR	final result Stokes RR, uniform weighting, see Fig. 9.34
UNI pre	preliminary result Stokes I, uniform weighting
$\alpha_0\pm$	UNI pre with modified α_0 (± 0.05 mas)
$\varepsilon_x\pm$	UNI pre with modified ε_x ($\pm 0.000\,150$)
$\varepsilon_y\pm$	UNI pre with modified ε_y ($\pm 0.000\,400$)
$x_0\pm$	like UNI pre but max/min x_0 of error ellipse (y_0 free)
$y_0\pm$	like UNI pre but max/min y_0 of error ellipse (x_0 fixed)
UNI final	final result Stokes I, uniform weighting
NAT final	final result Stokes I, natural weighting

Table 9.4: VLA lens models used for the fits with MERLIN data (see Figure 9.40)

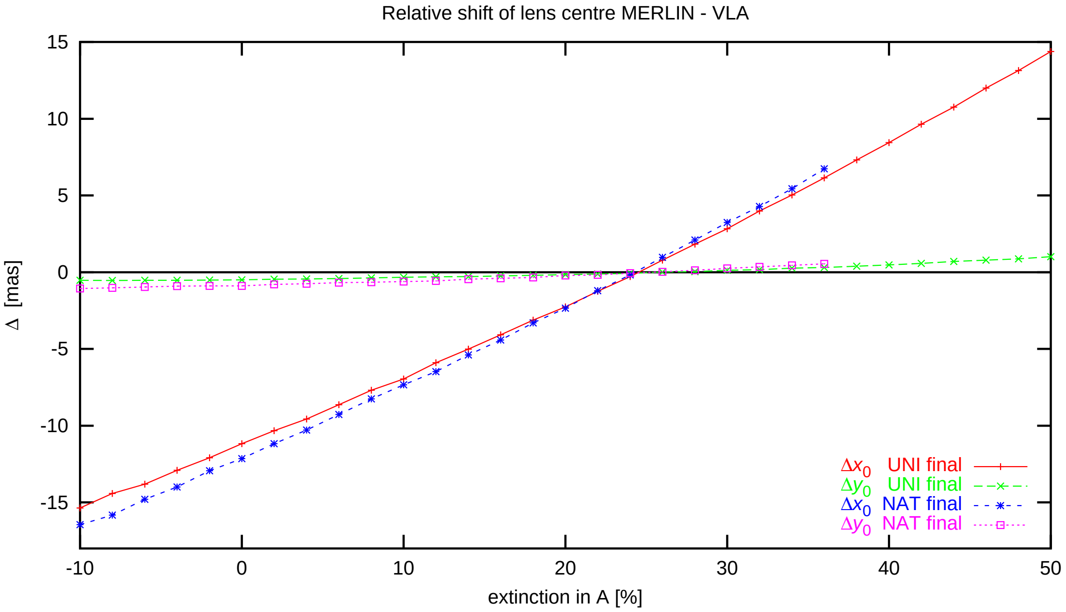


Figure 9.41: Relative shift of the lens centre (MERLIN−VLA) for different values of the extinction. The VLA-lens models 'UNI final' and 'NAT final' were used for the calculations with 500 iterations and uniform weighting with the complete MFS data set (Stokes I). The other lens models from Tab. 9.4 lead to very similar results. The shift of the source relative to the galaxy is the inverse of the values shown here.

the source plane and apply the known theory of CLEAN there.

9.23.1 Discrete formalism

In section 8.8.1 the emission components in the image plane (there was no lens then) acted as model parameters $\boldsymbol{x} \equiv \mathbf{I}$. This image plane brightness distribution (as collection of δ-components) $\mathbf{I}$ can now in the lensed case be derived from the brightness distribution in the source plane $\mathbf{I}_\mathrm{s}$ by means of the lens model. This can be formally written as

$$\mathbf{I} = \mathsf{L}\,\mathbf{I}_\mathrm{s} \quad . \tag{9.31}$$

If the positions in the lens and source planes are called z_ν and $z_{\mathrm{s}\nu_\mathrm{s}}$,[8] the matrix has non-vanishing elements only for those indices ν, ν_s for which the lens equation is satisfied:

$$z_{\mathrm{s}\nu_\mathrm{s}} = z_\nu - \alpha(z_\nu) \tag{9.32}$$

The values of L at these indices are the amplifications μ at the corresponding image positions z_ν. The parametrization must be complete, therefore exactly one ν_s with non-vanishing $L_{\nu\nu_\mathrm{s}}$ exists for each ν (one source position for each image). For fixed ν_s, the ν with non-vanishing matrix elements represent all the images for this source position (potentially several images for each source position). For each image position the corresponding source position and all secondary images of the same source position have to be included in the parametrization.

$$L_{\nu\nu_\mathrm{s}} = \begin{cases} \mu(z_\nu) & \text{if } z_{\mathrm{s}\nu_\mathrm{s}} = z_\nu - \alpha(z_\nu) \\ 0 & \text{otherwise} \end{cases} \tag{9.33}$$

If the model is now calculated in the source plane, a new matrix A_s has to be used, that transforms the model into a measurement.

$$\begin{aligned} \tilde{\mathbf{I}} &= \mathsf{A}\,\mathbf{I} && (9.34)\\ &= \mathsf{A}\,\mathsf{L}\,\mathbf{I}_\mathrm{s} && (9.35)\\ &= \mathsf{A}_\mathrm{s}\,\mathbf{I}_\mathrm{s} && (9.36)\\ \mathsf{A}_\mathrm{s} &= \mathsf{A}\mathsf{L} && (9.37) \end{aligned}$$

With this new matrix, the generalized dirty beam and dirty map in the source plane can be defined analogously to equations (8.34)–(8.37):

$$\begin{aligned} \mathsf{B}_\mathrm{s} &= \frac{1}{\operatorname{Tr}\mathsf{W}} \mathsf{A}_\mathrm{s}^\dagger \mathsf{W} \mathsf{A}_\mathrm{s} && (9.38)\\ &= \mathsf{L}^\dagger \mathsf{B} \mathsf{L} && (9.39)\\ B_\mathrm{s}(\nu_\mathrm{s}, \nu'_\mathrm{s}) &= \sum_{\substack{\nu(\nu_\mathrm{s})\\ \nu'(\nu'_\mathrm{s})}} \mu(\nu)\mu(\nu')B(\nu,\nu') && (9.40)\\ \mathbf{I}_{\mathrm{D}\,\mathrm{s}} &= \frac{1}{\operatorname{Tr}\mathsf{W}} \mathsf{A}_\mathrm{s}^\dagger \mathsf{W} \tilde{\mathbf{I}} && (9.41)\\ &= \mathsf{L}^\dagger \mathbf{I}_\mathrm{D} && (9.42)\\ I_{\mathrm{D}\,\mathrm{s}}(\nu_\mathrm{s}) &= \sum_{\nu(\nu_\mathrm{s})} \mu(\nu) I_\mathrm{D}(\nu) && (9.43) \end{aligned}$$

The sums in (9.40) and (9.43) are over all images ν of the source position ν_s and analogous for the primed indices. The indices are written as arguments to make the equations more readable. Contrary to the unlensed situation, the source plane dirty beam is not translation invariant anymore and multiplication with it is not a convolution.

As consistency check, the same procedure as in non-lensed CLEAN can now be performed: Find the position with the highest residual reduction if flux at this position is subtracted. We

---

[8] Indices ν in the lens plane and ν_s in the source plane.

start without a lens to derive the CLEAN recipe from the matrix formalism. The residuals can be calculated as follows:

$$\begin{aligned} R^2 &= (\tilde{\mathbf{I}} - \mathsf{A}\mathbf{I})^\dagger \mathsf{W} (\tilde{\mathbf{I}} - \mathsf{A}\mathbf{I}) && (9.44)\\ &= \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} - 2\mathbf{I}^\dagger \mathsf{A}^\dagger \mathsf{W} \tilde{\mathbf{I}} + \mathbf{I}^\dagger \mathsf{A}^\dagger \mathsf{W} \mathsf{A} \mathbf{I} && (9.45) \end{aligned}$$

CLEAN removes one component at a time. This corresponds to a $\mathbf{I}$ consisting of only of one component ν, leading to

$$R^2 = \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} - 2 I_\nu \, (\mathsf{A}^\dagger \mathsf{W} \tilde{\mathbf{I}})_\nu + I_\nu^2 \, (\mathsf{A}^\dagger \mathsf{W} \mathsf{A})_{\nu\nu} \quad . \qquad (9.46)$$

The optimal flux is

$$\begin{aligned} I_\nu &= \frac{(\mathsf{A}^\dagger \mathsf{W} \tilde{\mathbf{I}})_\nu}{(\mathsf{A}^\dagger \mathsf{W} \mathsf{A})_{\nu\nu}} && (9.47)\\ &= \frac{I_{\mathrm{D}\nu}}{B_{\nu\nu}} \quad . && (9.48) \end{aligned}$$

Dirty beam and map were normalized to a beam peak of $B_{\nu\nu} = 1$, so the optimal flux is given by the dirty map at this position. With this flux, the residuals are

$$\begin{aligned} R^2 &= \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} - \frac{{(\mathsf{A}^\dagger \mathsf{W} \tilde{\mathbf{I}})_\nu}^2}{(\mathsf{A}^\dagger \mathsf{W} \mathsf{A})_{\nu\nu}} && (9.49)\\ &= \tilde{\mathbf{I}}^\dagger \mathsf{W} \tilde{\mathbf{I}} - \mathrm{Tr}\, \mathsf{W} \frac{I_{\mathrm{D}\nu}^2}{B_{\nu\nu}} \quad , && (9.50) \end{aligned}$$

the optimum is again the peak of the dirty map. This formalism can without changes directly be transformed to LENSCLEAN. The optimal position is the maximum of $I_{\mathrm{D\,s}}^2(\nu_\mathrm{s})/B_\mathrm{s}(\nu_\mathrm{s}, \nu_\mathrm{s})$, the optimal flux is $I_{\mathrm{D\,s}}(\nu_\mathrm{s})/B_\mathrm{s}(\nu_\mathrm{s}, \nu_\mathrm{s})$. With (9.40) and (9.43) this is recognized as the result for KNE-LENSCLEAN given in equations (9.13) and (9.12).

9.23.2 Continuous formalism

In some contexts, it may be more convenient to use continuous functions for the brightness distribution in source and lens plane. The sum in the Fourier transform (8.32) now becomes an integral:

$$\begin{aligned} \tilde{I}_j^\mathrm{m} &= \int \mathrm{d}^2 z \, A(\mathbf{u}_j, \boldsymbol{z}) \, I(\boldsymbol{z}) && (9.51)\\ A(\mathbf{u}_j, \boldsymbol{z}) &= \mathrm{e}^{2\pi \mathrm{i} \mathbf{u}_j \cdot \boldsymbol{z}} && (9.52) \end{aligned}$$

In contrast to the amplifications of point sources, surface brightness is conserved in lensing, so that it is directly mapped with the lens equation:

$$\begin{aligned} I(\boldsymbol{z}) &= I_\mathrm{s}(\boldsymbol{z}_\mathrm{s}) && (9.53)\\ &= \int \mathrm{d}^2 z_\mathrm{s} \, L(\boldsymbol{z}, \boldsymbol{z}_\mathrm{s}) \, I_\mathrm{s}(\boldsymbol{z}_\mathrm{s}) && (9.54)\\ L(\boldsymbol{z}, \boldsymbol{z}_\mathrm{s}) &= \delta^2\big(\boldsymbol{z} - \boldsymbol{\alpha}(\boldsymbol{z}) - \boldsymbol{z}_\mathrm{s}\big) && (9.55) \end{aligned}$$

The dirty beam in the source plane can now be calculated analogously to (9.42).

$$I_{\mathrm{D\,s}}(z_{\mathrm{s}}) = \int \mathrm{d}^2 z\, L(z, z_{\mathrm{s}})\, I_{\mathrm{D}}(z) \tag{9.56}$$

To carry out the integration, we have to transform the δ-function from the source plane to the lens plane by using the following equation:

$$\delta^2\big(z - \alpha(z) - z_{\mathrm{s}}\big) = \sum_{k(z_{\mathrm{s}})} \left| \left| \frac{\partial z_{\mathrm{s}}}{\partial z}(z_k) \right|^{-1} \right| \delta^2(z - z_k) \tag{9.57}$$

The sum is taken over all images for the given source position. The result is the same as in the discrete case in (9.43).

$$I_{\mathrm{D\,s}}(z_{\mathrm{s}}) = \sum_{z(z_{\mathrm{s}})} \mu(z)\, I_{\mathrm{D}}(z) \tag{9.58}$$

The same formalism can be used to recover (9.40) in the continuous form from

$$B_{\mathrm{s}}(z_{\mathrm{s}}, z'_{\mathrm{s}}) = \int \mathrm{d}^2 z \int \mathrm{d}^2 z'\, L(z, z_{\mathrm{s}})\, B(z, z')\, L(z', z'_{\mathrm{s}}) \quad . \tag{9.59}$$

The result in the continuous formalism is

$$B_{\mathrm{s}}(z_{\mathrm{s}}, z'_{\mathrm{s}}) = \sum_{\substack{z(z_{\mathrm{s}}) \\ z'(z'_{\mathrm{s}})}} \mu(z)\, \mu(z')\, B(z, z') \quad . \tag{9.60}$$

Hence, the shape of the dirty beam and map do not depend on the sampling in the source plane. This fact is important when the dirty beam is used as model for the CLEAN beam.

9.23.3 Direct numerical inversion

Several methods of direct algebraic inversion of the unlensed convolution equations have been discussed in section 8.8. Some of them seem to be superior to CLEAN, so that similar approaches might also be useful in the lensed situation. The formalism to develop such algorithms is given in the last sections. No attempts to actually perform such calculations have been carried out so far because of the huge numerical burden. LENSCLEAN itself is at the limit of what can be done with normal computers these days and the direct methods even have greater demands on CPU time and memory. The problems are only partly the same as in the unlensed situation.

As discussed in section 9.14, the errors introduced by gridding and using FFT (see section 8.10.1) are much more serious in the lensed case, because it is not possible to use a regular grid for the emission model in the lens plane. This problem was overcome in LENSCLEAN by subtracting the components from the ungridded visibilities in certain intervals. This approach is appropriate for an iterational algorithm like LENSCLEAN. For direct methods, the same approach would mean to iterate alternating inversion and *uv* space subtraction, which is prohibitively expensive. Even only one direct inversion is at the very limit of today's computers capabilities.

As a result of the gridding problems, it is necessary to calculate the dirty beam and dirty map with the very inefficient DFT. This is possible for dirty maps with a moderate number of pixels *n*. For the dirty beam, however, the number of elements is of the order n^2 and thus extremely large. In the unlensed case, the beam is translation invariant, so that the number of independent elements on a regular grid is much smaller, of the order $4n$ in two dimensions (each difference of positions in the dirty map is a possible coordinate for the dirty beam). This is only true in the case of regular sampling of the image plane and therefore not applicable in a lensed situation.

Serious effort should be spent on this problem in the future, because a direct inversion of lensed sources is highly desirable. Especially the non-negativity constraint or other kinds of regularization have a potentially very high value in lens modelling. They will restrict the freedom of LENSCLEAN to produce unrealistic emission models without rejecting physically possible source structures and thus improve the capability to distinguish between good and bad lens models.

9.24 Reconstruction of the source

Even with unlensed CLEAN, the final CLEAN beam convolution is a somewhat arbitrary approach to reduce high frequencies that are not present in the measurements. This procedure nevertheless works very well in most cases and is conventionally used in almost all cases. It shall therefore also be used as basis for the reconstruction of the source here.

9.24.1 The source plane CLEAN beam

Things are simple in singly imaged regions with only slowly varying magnifications. The final CLEAN map of the source should be equal to the lens plane CLEAN map projected into the source plane. This is equivalent to projecting the CLEAN components and using projected CLEAN beams for the convolution. If the lens plane CLEAN beam is given by the matrix G (see section 8.6.2), the corresponding beam in the source plane can be calculated by applying equation (A.17) with the symmetric local magnification matrix M:

$$\mathsf{G}_{\mathrm{s}} = \mathsf{MGM} \tag{9.61}$$

Magnifications and source plane beam sizes will not be constant in realistic lenses. To assure flux conservation, the normalized beams (section 8.11.1) have to be used. The normalization factors of source and image plane Gaussians have a ratio of $\mu = |\mathsf{M}|$. Together with the amplification used to calculated the lens plane flux from the source plane flux, this assures surface brightness conservation.

Lenses used for LENSCLEAN will generally have multiply imaged regions. Otherwise LENSCLEAN would not make much sense. The task is now to construct a source plane beam from several projected lens plane beams of images belonging to one source position. These beams will usually have different sizes and shapes. A very conservative approach would be a large beam which covers all the single beams, but this would clearly hide information in the CLEAN maps. Kochanek & Narayan (1992) used circular beams and chose the size so that they just cover the intersection of all the projected single beams. With this approach the combined beam is in each direction larger or equal to the smallest of the single beams and does not introduce spurious small scale features, which should be smoothed out by the beams. This procedure is well justified for equal magnifications, but becomes very inaccurate for high amplification ratios. In this case the projected beams should somehow be weighted according to the lens plane flux densities which are proportional to the amplifications of the images. Images with very low amplification should not contribute significantly to the resulting source plane beam.

To find a better solution, we start with the source plane dirty beam as given in (9.40) and consider only a small neighbourhood of the source position z'_{s} as possible map positions z_{s}.

With this assumption, the number of images for both positions is equal and the image plane displacement of corresponding images $\Delta z = z - z'$ can be calculated linearly from the source plane displacement $\Delta z_\mathrm{s} = z_\mathrm{s} - z'_\mathrm{s}$:

$$\Delta z = \mathsf{M}\,\Delta z_\mathrm{s} \tag{9.62}$$

If ν and ν' count the images of z_s and z'_s (same index for corresponding images), the beam reads

$$B_\mathrm{s}(\Delta z_\mathrm{s}) = \sum_{\nu\nu'} \mu_\nu\,\mu_{\nu'}\,B(z_\nu - z_{\nu'} + \mathsf{M}_\nu\,\Delta z_\mathrm{s}) \quad . \tag{9.63}$$

Here it was assumed that μ and M are constant in the small regions of interest around each image. As a last approximation, we assume that the images are well separated, so that the lens plane dirty beams do not overlap for $\nu \neq \nu'$. All the approximations are well justified, if the images are located sufficiently far from the critical lines. Close to them, magnification gradients would become large and images would move very closely together. For the approximated result we obtain

$$B_\mathrm{s}(\Delta z_\mathrm{s}) = \frac{\sum\limits_\nu \mu_\nu^2\,B(\mathsf{M}_\nu \Delta z_\mathrm{s})}{\sum\limits_\nu \mu_\nu^2} \quad . \tag{9.64}$$

It is normalized to $B_\mathrm{s}(0) = 1$ for $B(0) = 1$. The source plane dirty beam thus is a weighted sum of the projected lens plane dirty beams. To fit a Gaussian to the centre of this beam, only the centres of the lens plane beams are needed. They are already approximated by a Gaussian G.

$$B_\mathrm{s}(\Delta z_\mathrm{s}) = \frac{\sum\limits_\nu \mu_\nu^2\,\mathrm{e}^{-z_\mathrm{s}{}^\dagger\,\mathsf{M}_\nu \mathsf{G} \mathsf{M}_\nu\, z_\mathrm{s}/2}}{\sum\limits_\nu \mu_\nu^2} \tag{9.65}$$

Close to the centre, this sum of Gaussians can again be fitted by one combined Gaussian. The coefficients can be calculated by comparing the second derivatives, which are given by the projected G and which add linearly:

$$\mathsf{G}_\mathrm{s} = \frac{\sum\limits_\nu \mu_\nu^2\,\mathsf{M}_\nu \mathsf{G} \mathsf{M}_\nu}{\sum\limits_\nu \mu_\nu^2} \tag{9.66}$$

In the case of equal magnifications, this approach is very similar to the one used by Kochanek & Narayan (1992). For circular beams (or in an arbitrary but fixed direction for arbitrary beams), the calculation of the mean in equation (9.66) approximately chooses the smallest of the beams. The same formalism as in equation (8.82) applies here.

Our choice of CLEAN beams in the source plane is arbitrary in some respects. It should work well for separated images, where the approximations are valid and where the linear transformation of beams is sensible. Problems are expected close to the caustics, where the linear theory should not be used. Convolution with any beam can spill emission across caustics to regions of different image multiplicity, even if the data rule out any emission there. This problem was already recognized by Kochanek & Narayan (1992). It can only be solved with an alternative reconstruction technique which is non-linear in the positions.

The modification that is introduced by using the unbiased LENSCLEAN variant (section 9.9) is not considered a problem here. In the framework of the approximations used, the correction factors are almost constant across the regions for which the beam was calculated and should therefore not introduce significant changes.

To solve the problem of source plane image reconstruction optimally, other methods should be developed in the future. Similar problems have been discussed for the unlensed situation in section 8.11. The idea of adaptive CLEAN beams, calculated from the residual dirty map is very promising (see section 8.11.3). The same approach can be used for LENSCLEAN.

9.24.2 Superresolution in the image plane

The combination of different beams in the source plane to form a combined beam usually improves the resolution, because in each direction the smallest of the beams dominates the combined beam. The improvement is thus achieved not by the magnification alone but by the combination of different projected beams. This improvement can be brought back to the lens plane by projecting the combined source plane beam (9.66) back to the image plane at the image positions and use these beams to produce the final CLEAN map.

This map can then also be used to suppress the spilling problem discussed above, which was mainly caused by projecting the beams linearly in regions where this is not appropriate. If we now project the final CLEAN map, built with the superresolved beam, back to the source plane (pixel by pixel), the resulting source plane map should be almost devoid of the disturbing spilled flux. In free interpretation of (9.58), a mean weighted with the amplifications should be used to reconstruct the source plane in multiply images regions.

9.25 Maps of the source and lens plane for B0218+357

9.25.1 15 GHz VLA data

We now apply the methods developed in the last sections to produce maps of the source and lens plane using the LENSCLEAN results. For the 15 GHz VLA data, the best fitting lens model for uniformly weighted data was used. To produce the final LENSCLEAN component list, 20 000 iterations with the self-calibrated data set on a grid of $512 \times 5\,\mathrm{mas}$ with the usual algorithm were performed to CLEAN sufficiently deep below the noise level. Figure 9.42 shows the lens plane maps made with unlensed CLEAN and LENSCLEAN. The two maps are almost indistinguishable. Hence LENSCLEAN with the best lens model is able to reproduce the data as well as unlensed CLEAN. The residual map would be zero after many more iterations in the singly imaged outer regions. In the presented map, small residuals are visible in the central part. These are not higher than expected for the theoretical noise limit. The lowest contour lines mark $2\,\sigma$ of the expected noise of $0.14\,\mathrm{mJy/beam}$ or $1.44 \cdot 10^{-8}\,\mathrm{Jy/mas^2}$.

We also present a superresolved lens plane map. This result is, however, difficult to interpret because of the varying resolution. The apparent small scale structure close to the critical curve is mainly a result of the small local restoring beam.

The source plane results are shown in Figure 9.43 on page 198. The superresolved version shows radially stretched details which are artifacts of the stretched restoration beams. The same is true for the ellipticity of the core component. The version reconstructed with circular beams is easier to interpret. The inner jet, defined by the inner subcomponents (not resolved in the maps), is directed eastwards. The jet then bends in a southern direction and crosses the caustic on its way. It is this part of the source which forms the Einstein ring in the lensed image.

A lower limit of 20 mas was used for the size of all beams. This does not change the results significantly but helps avoiding minor numerical problems.

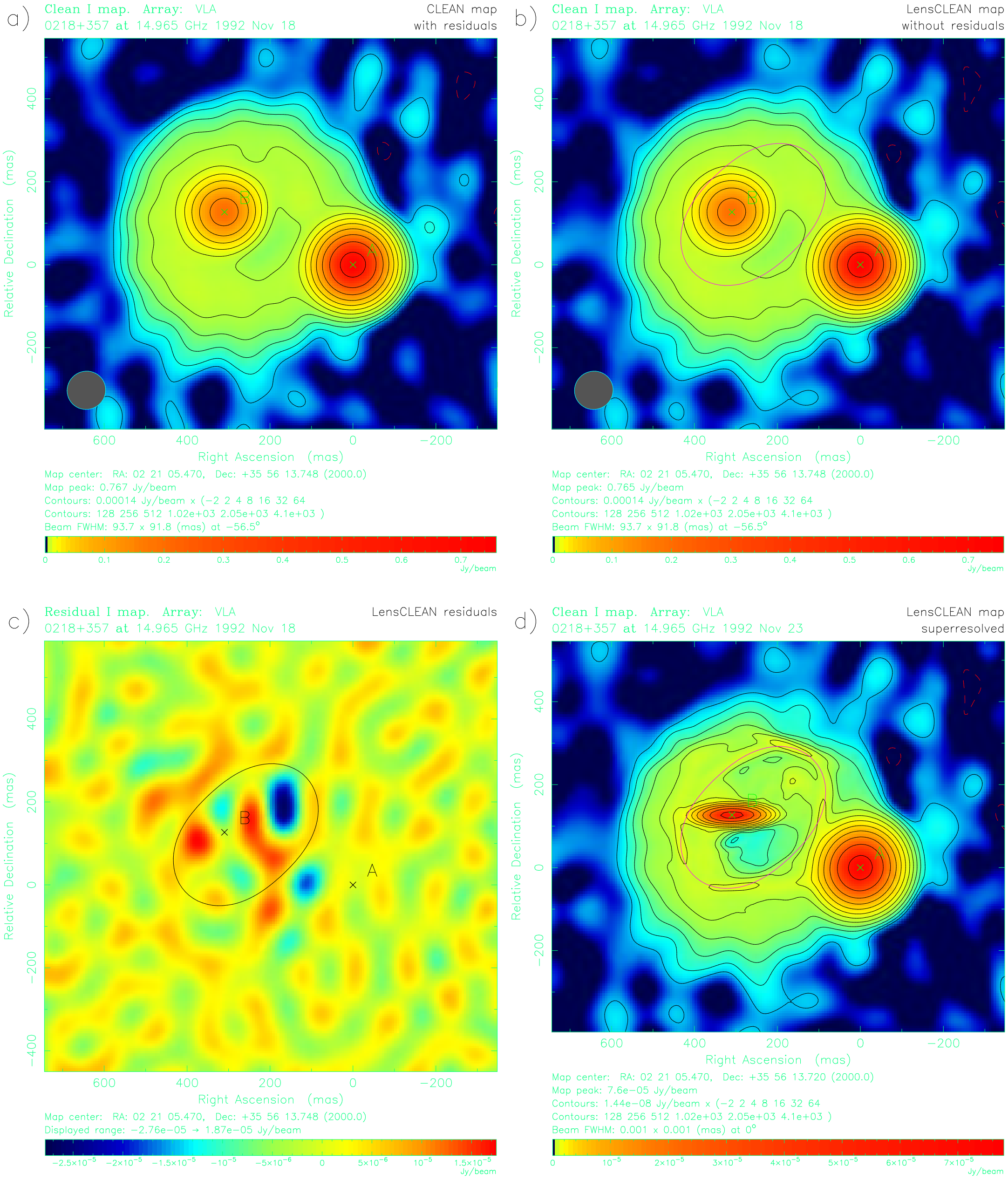


Figure 9.42: Resulting lens plane maps for the 15 GHz VLA data (uniform weighting). (a) Classical unlensed CLEAN map, produced with DIFMAP (with residuals). (b) LENSCLEAN map reconstructed with the same CLEAN beam (without residuals). (c) Residuals for the LENSCLEAN result. Significant residuals are only seen in the multiply imaged region near B. Even these are below the expected noise. Flux density units for (a), (b) and (c) are Jy/beam. (d) Superresolved LENSCLEAN map (without residuals) with varying elliptical beams. Units are not Jy/beam but Jy/mas^2. The scaling factor is 9751 mas^2/beam. Subplots (b) to (d) also show the (tangential) critical curve of the lens model. The logarithmic contour lines start with 2σ of the expected lens plane noise of 0.14 mJy/beam or $1.44 \cdot 10^{-8}$ Jy/mas^2.

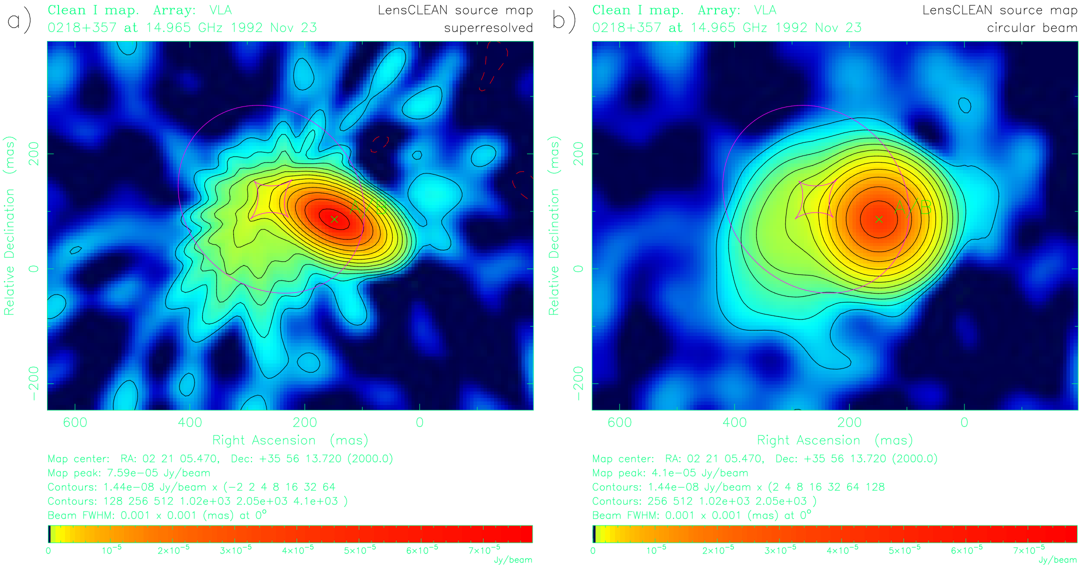


Figure 9.43: Source plane LENSCLEAN maps for the 15 GHz VLA data (uniform weighting). (a) Reconstructed with the local combined beams (superresolved). (b) Reconstructed with local circular beams (size of the major axis of the elliptical beam) to achieve isotropic, but only approximately homogeneous, resolution. Units are not Jy/beam but Jy/mas^2. The scaling factor is 9751 mas^2/beam. The logarithmic contour lines start with $2\,\sigma$ of the expected lens plane noise of $1.44 \cdot 10^{-8}$ Jy/mas^2. Also included is the diamond shaped (tangential) caustic and the elliptical 'cut' (Kovner, 1987) or 'pseudo-caustic'.

9.25.2 5 GHz MERLIN/VLA data

The problems with LENSCLEANing the MERLIN MFS data set have already been discussed before. Here we are not interested in finding the best lens model but only in the resulting lens and source plane maps. We can therefore use one of the best models from the VLA data and apply it to the MERLIN data set. It was shown in section 9.22.4 that the MERLIN data are only compatible with the VLA lens models if extinction of the order 20 % for the A component is taken into account. The exact value of the extinction if not very well defined, which makes using the MERLIN data more difficult. If the hypothesis of extinction in A is correct, no shift of the 5 GHz source relative to 15 GHz is needed to explain the differing flux density ratios. We therefore assume that the shift is exactly zero.

The results for the shift as a function of different extinction values can now be used to find the best extinction for a zero shift. An optimal extinction of 24 % can be estimated from Figure 9.41 for this case. The best lens model for this extinction (VLA reference model 'UNI final') was then fixed and used for fits to the MERLIN data. The other parameters were varied freely to minimize the residuals. A LENSCLEAN/self-calibration loop (2000 LENSCLEAN iterations without negative components, uniform weighting, 50 loops) was then performed with this fixed lens model to calibrate the data optimally. As for the final VLA results, the emission model was built for Stokes I and LL and RR were independently self-calibrated with this model. This process was repeated twice to correct for calibration errors as good as possible. The second iteration did not change the results significantly and was therefore not used for the maps.

We have to keep in mind that the data do have there problems with LENSCLEAN. The no-shift hypothesis might be incorrect and the model of a Gaussian area of 24 % extinction around the A component can only be a crude approximation of the real effects. We also noticed that

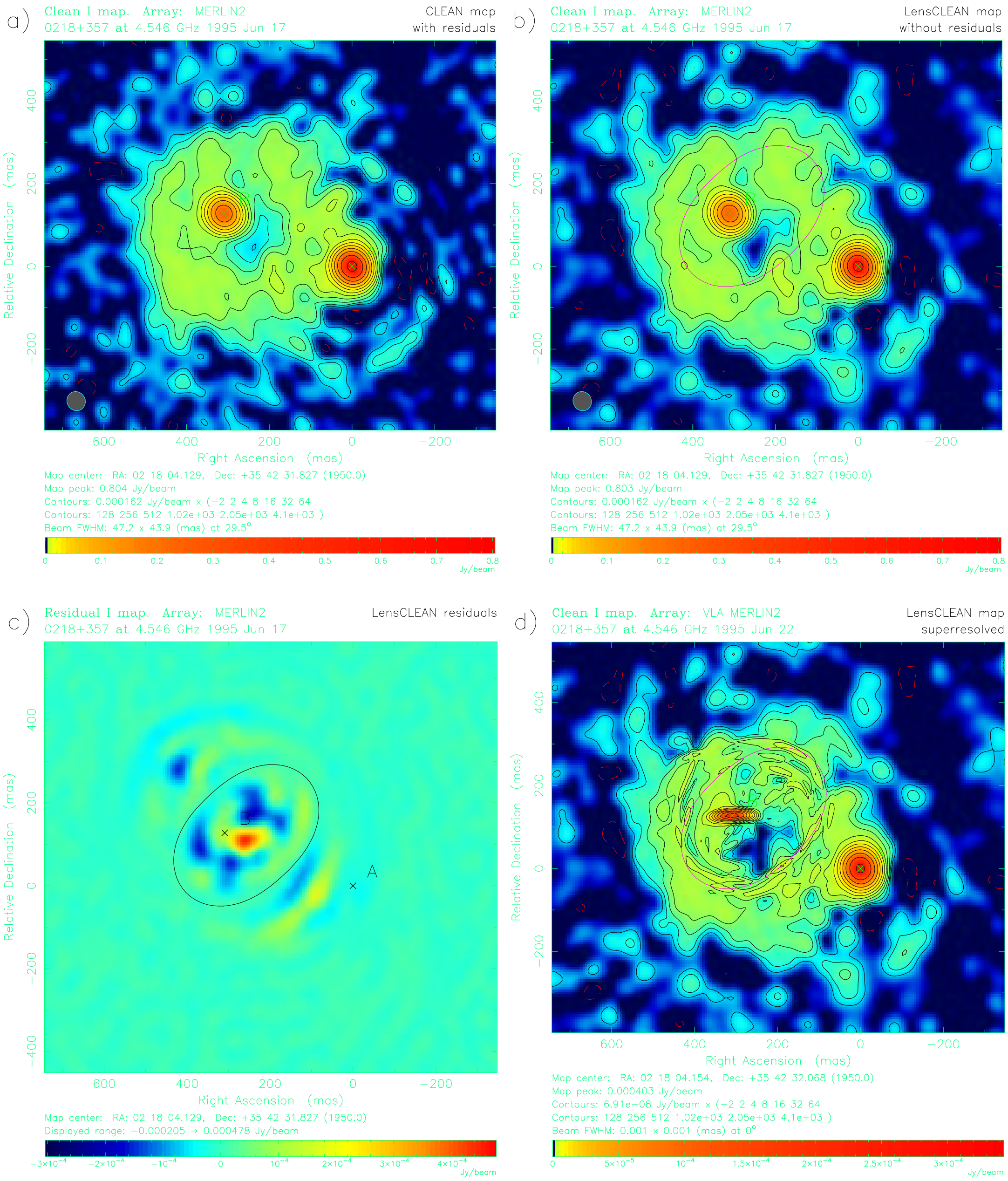


Figure 9.44: Resulting lens plane maps for the 5 GHz MERLIN data (uniform weighting). (a) Classical unlensed CLEAN map, produced with DIFMAP (with residuals). (b) LENSCLEAN map reconstructed with the same CLEAN beam (without residuals). (c) Residuals for the LENSCLEAN result. Flux density units for (a), (b) and (c) are Jy/beam. (d) Superresolved LENSCLEAN map (without residuals) with varying elliptical beams. Units are not Jy/beam but Jy/mas^2. The scaling factor is 2346 mas^2/beam. Subplots (b) to (d) also show the (tangential) critical curve of the lens model. The logarithmic contour lines start with 2σ of the expected lens plane noise of 0.16 mJy/beam or $6.91 \cdot 10^{-8}$ Jy/mas^2.

the self-calibration residuals for LL polarization are much smaller than for RR. This is probably related to the preprocessing of the data. The effect is only present for the complete MFS data set. For single-frequency sub-data sets, the residuals are almost equal. The resulting maps should in

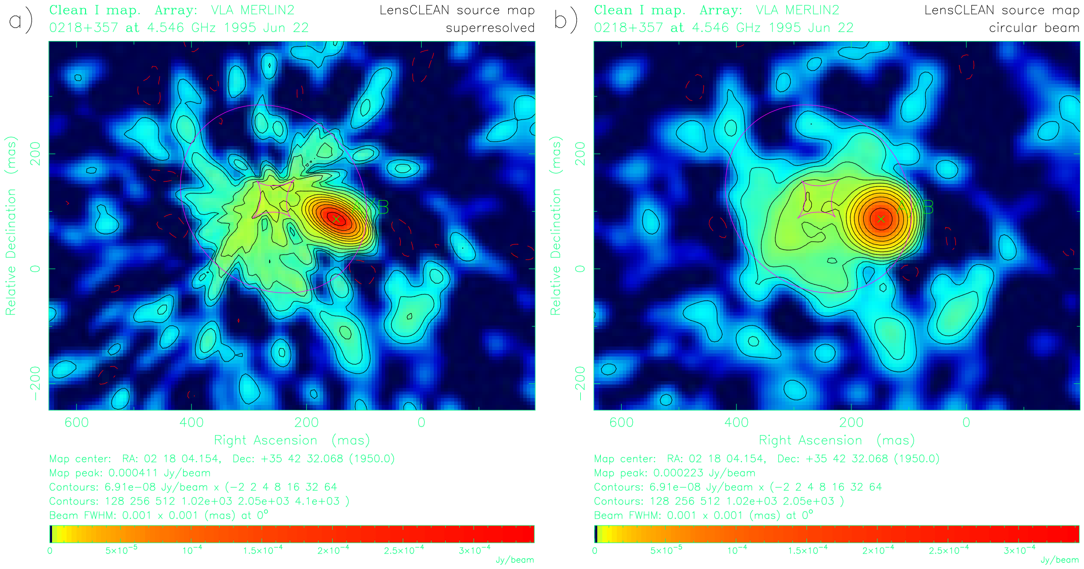


Figure 9.45: Source plane LENSCLEAN maps for the 5 GHz MERLIN data (uniform weighting). (a) Reconstructed with the local combined beams (superresolved). (b) Reconstructed with local circular beams (size of the major axis of the elliptical beam) to achieve isotropic, but only approximately homogeneous, resolution. Units are not Jy/beam but Jy/mas^2. The scaling factor is 2346 mas^2/beam. The logarithmic contour lines start with 2σ of the expected lens plane noise of $6.91 \cdot 10^{-8}$ Jy/mas^2. Also included is the diamond shaped (tangential) caustic and the elliptical 'cut'.

any case be interpreted with caution.

Results for the lens and source plane are shown in Figures 9.44 and 9.45, respectively. The results for classical CLEAN and LENSCLEAN are similar, but there are nevertheless some noticeable differences. We have to keep in mind, however, that the final calibration of the data was performed with emission models built with LENSCLEAN. The unlensed CLEAN is thus not completely independent of the LENSCLEAN results. A lower limit for the beam sizes of 10 mas was used for these plots.

In contrast to the VLA results (section 9.25.1), significant residuals above the noise level remain for the MERLIN data. No detailed analysis of the properties of LENSCLEAN residual maps has been done yet, but larger than average residuals are expected in the regions of very small amplification (in fact de-amplifications) close to the lens centre, because noise will partly be transferred from image positions corresponding to the same source positions, but with higher amplifications. One should therefore not naively interpret the positive residuals near the centre as evidence for a central image or emission from the galaxy itself. Monte Carlo simulations can be used in the future to obtain a statistical measure of the significance of these positive residuals. This has to wait for the recalibration of the data, however.

The source plane maps strengthen the interpretation of a curved jet emerging in the direction of the inner 1 and 2 components and then bending southwards. To investigate the jet's structure further outwards, the VLA part of the 5 GHz data set was used. The first calibration is sufficiently good, so that no LENSCLEAN/self-calibration is necessary. The CLEAN area was enlarged to enclose the outer parts of the jet. 20 000 LENSCLEAN iterations were used for the uniformly and 10 000 + 10 000 for the naturally weighted data. The results are shown in Figures 9.46 to 9.49 (pages 201–204). They were produced as before with a beam size limit of 20 mas.

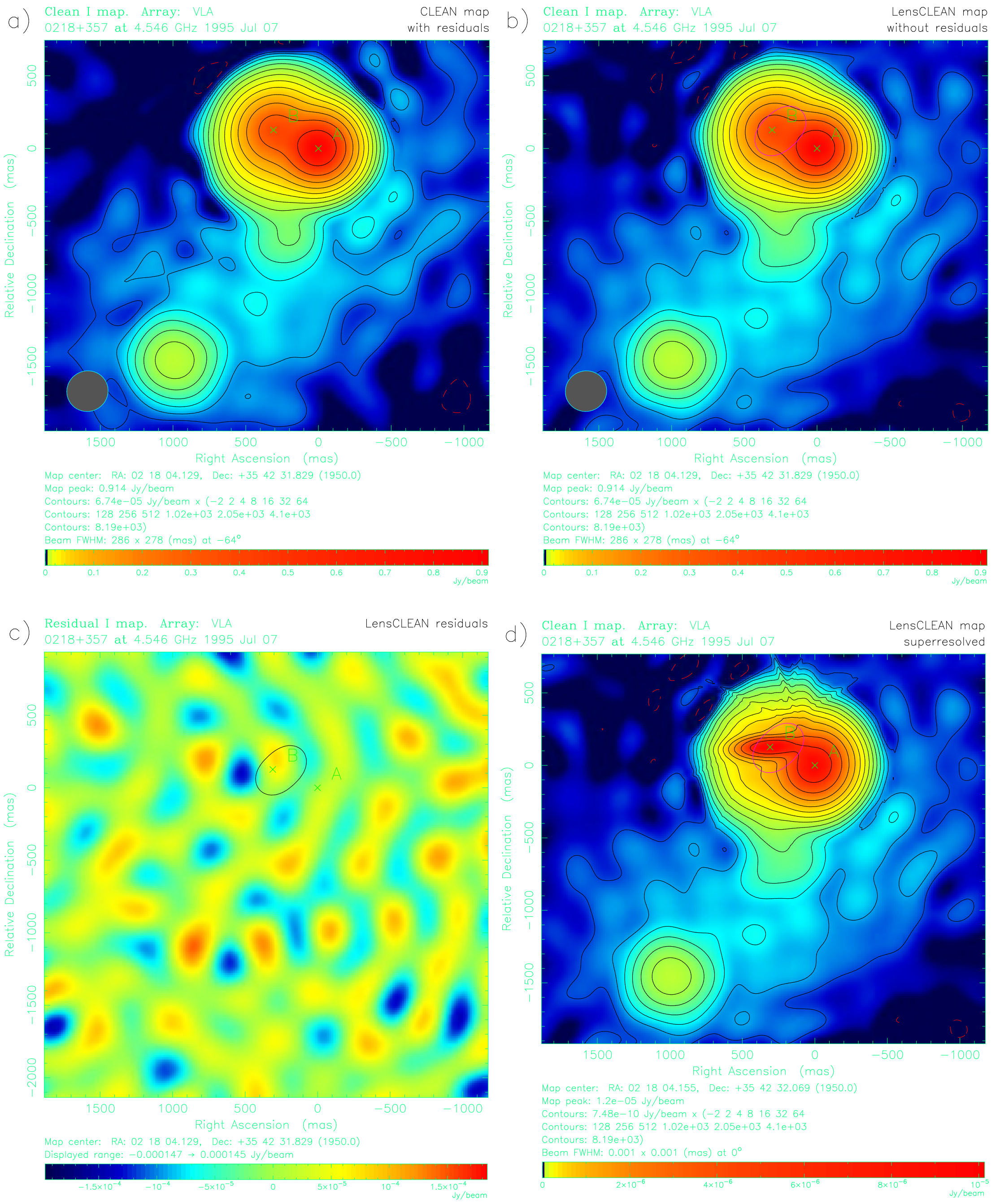


Figure 9.46: Resulting lens plane maps for the 5 GHz VLA data (uniform weighting). (a) Classical unlensed CLEAN map, produced with DIFMAP (with residuals). (b) LENSCLEAN map reconstructed with the same CLEAN beam (without residuals). (c) Residuals for the LENSCLEAN result. Flux density units for (a), (b) and (c) are Jy/beam. (d) Superresolved LENSCLEAN map (without residuals) with varying elliptical beams. The sharp elongated features at the top are artifacts of the mapping algorithm. Units are not Jy/beam but Jy/mas^2. The scaling factor is 90 160 mas^2/beam. Subplots (b) to (d) also show the (tangential) critical curve of the lens model. The logarithmic contour lines start with $2\,\sigma$ of the expected lens plane noise of 0.067 mJy/beam or $7.43 \cdot 10^{-10}$ Jy/mas^2.

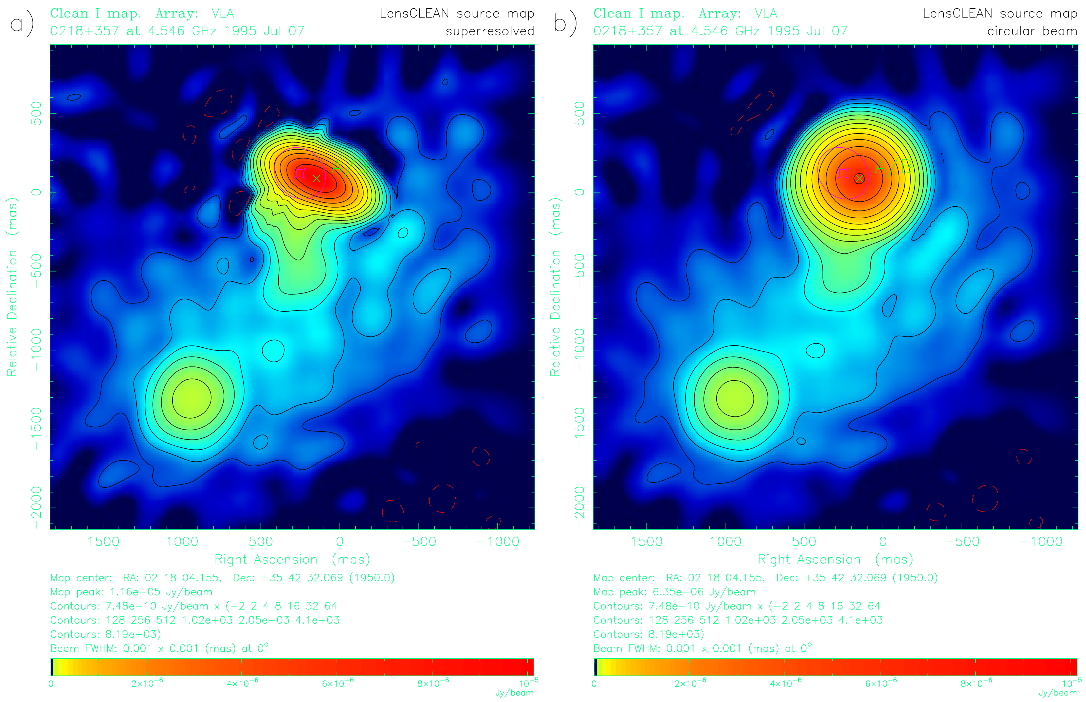


Figure 9.47: Source plane LENSCLEAN maps for the 5 GHz VLA data (uniform weighting). (a) Reconstructed with the local combined beams (superresolved). (b) Reconstructed with local circular beams (size of the major axis of the elliptical beam) to achieve isotropic, but only approximately homogeneous, resolution. Units are not Jy/beam but Jy/mas^2. The scaling factor is 90 160 mas^2/beam. The logarithmic contour lines start with 2σ of the expected lens plane noise of $7.43 \cdot 10^{-10}$ Jy/mas^2. Also included is the diamond shaped (tangential) caustic and the elliptical 'cut'.

9.26 Non-isothermal lens models

Serious indications for deviations from an isothermal potential were already discussed in section 7.6.2. Probable exponents for power-law ellipsoids are of the order $\beta \approx 1.05$. For a fixed lens position, the scaling of the Hubble constant would approximately conform to $H_0 \propto 2-\beta$ in B0218+357, leading to an overestimate of the Hubble constant by about 5 % if isothermal models are assumed. It is not sensible, however, to use a fixed lens position for an estimate of the influence of β, because the positional constraints derived from LENSCLEAN are only valid for isothermal models with $\beta = 1$. More complicated lens models, for which the inversion of the lens equation can not be done analytically, could not be used with LENSCLEAN until our recent development of a sufficiently reliable numerical algorithm (see chapter 5).

Only preliminary results are available at the moment because of the high computational burden. About 50–100 % more CPU time is needed with LENTIL and general lens models than with simpler lens models which are treatable analytically. As a first application, LENSCLEAN model fits were done for a number of fixed values of β close to 1. LENTIL works sufficiently reliable, so that almost no fits fail because of incorrect inversion of the lens equations. Unfortunately, the LENSCLEAN model fitting becomes much more unstable with $\beta > 1$ than with isothermal models. At the moment, about 5 % of the fits get stuck at local minima and do not converge to the global minimum. The reason for this bad performance is currently under investigation. It seems to be a generic problem of the specific lens models and not of the LENTIL algorithm.

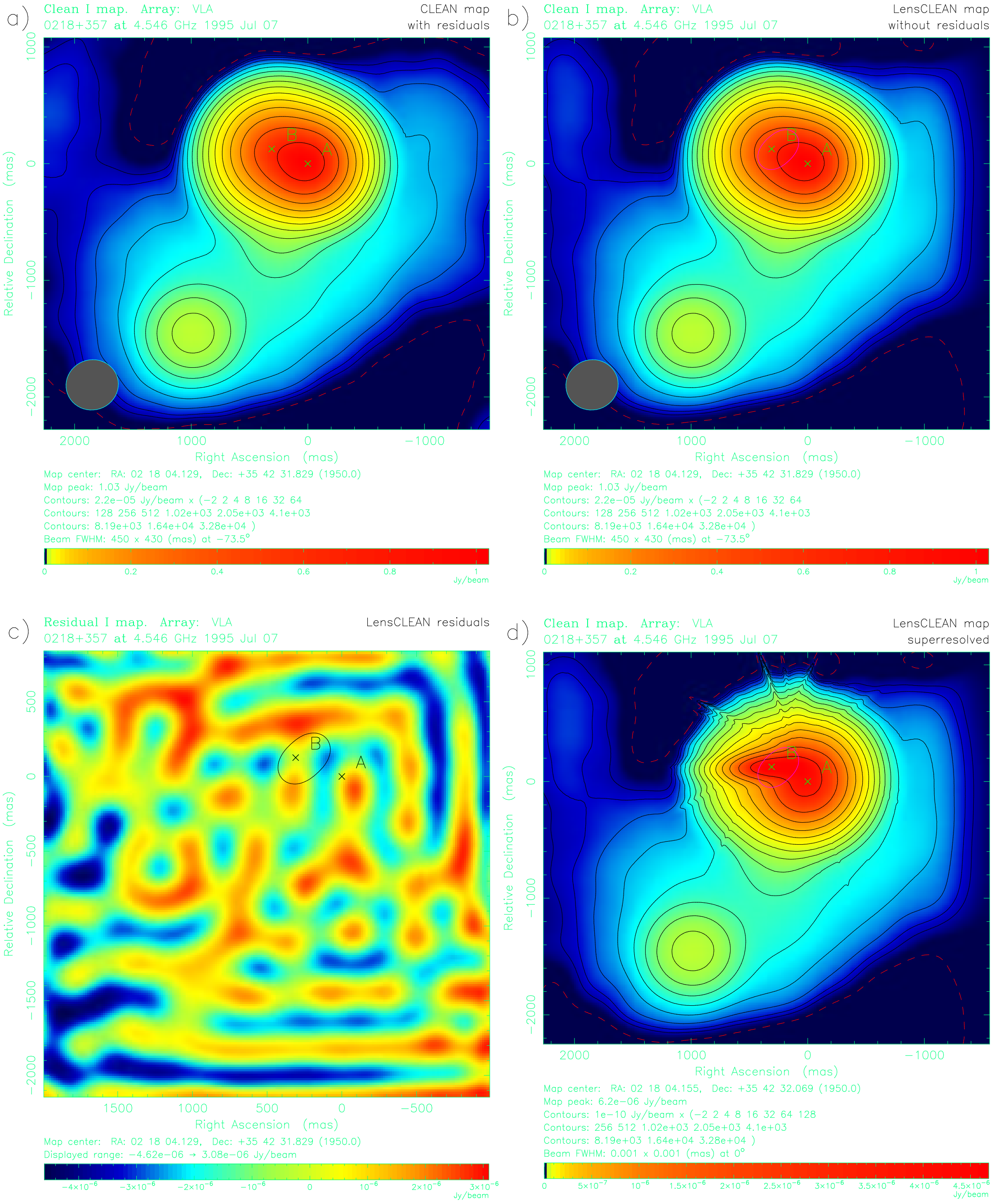


Figure 9.48: Resulting lens plane maps for the 5 GHz VLA data (natural weighting). (a) Classical unlensed CLEAN map, produced with DIFMAP (with residuals). (b) LENSCLEAN map reconstructed with the same CLEAN beam (without residuals). (c) Residuals for the LENSCLEAN result. Flux density units for (a), (b) and (c) are Jy/beam. (d) Superresolved LENSCLEAN map (without residuals) with varying elliptical beams. The sharp elongated features at the top are artifacts of the mapping algorithm. Units are not Jy/beam but Jy/mas^2. The scaling factor is 219 300 mas^2/beam. Subplots (b) to (d) also show the (tangential) critical curve of the lens model. The logarithmic contour lines start with $2\,\sigma$ of the expected lens plane noise of 0.022 mJy/beam or $1.00 \cdot 10^{-10}$ Jy/mas^2.

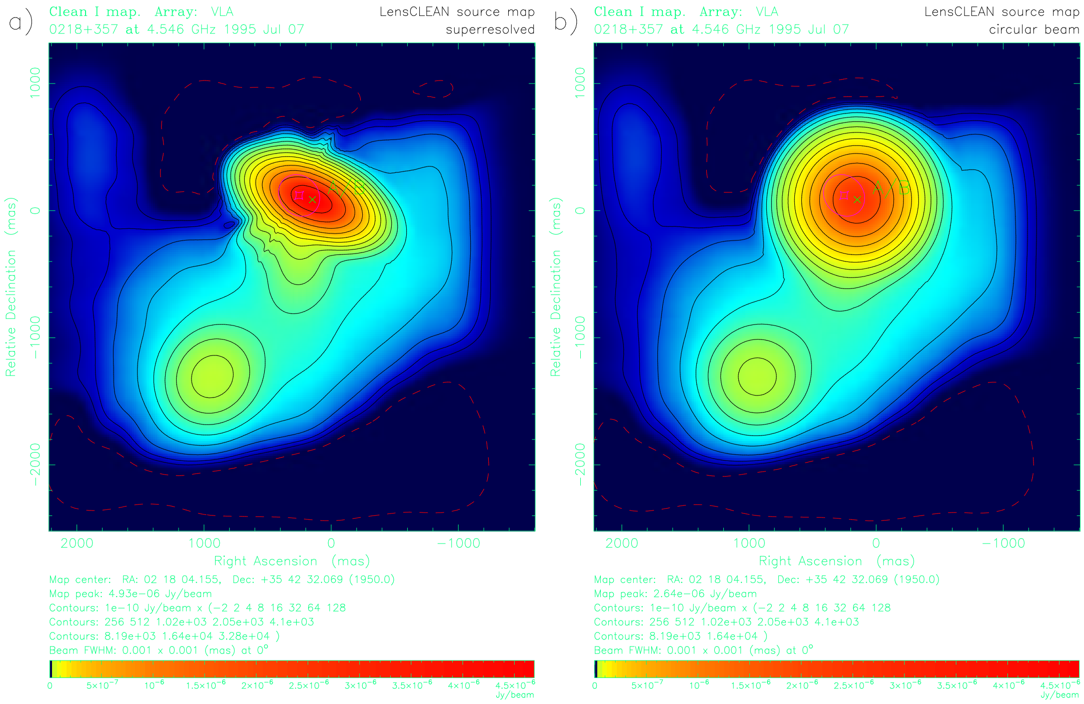


Figure 9.49: Source plane LENSCLEAN maps for the 5 GHz VLA data (natural weighting). (a) Reconstructed with the local combined beams (superresolved). (b) Reconstructed with local circular beams (size of the major axis of the elliptical beam) to achieve isotropic, but only approximately homogeneous, resolution. Units are not Jy/beam but $\mathrm{Jy}/\mathrm{mas}^2$. The scaling factor is $219\,300\,\mathrm{mas}^2/\mathrm{beam}$. The logarithmic contour lines start with 2σ of the expected lens plane noise of $1.00 \cdot 10^{-10}\,\mathrm{Jy}/\mathrm{mas}^2$. Also included is the diamond shaped (tangential) caustic and the elliptical 'cut'.

For future regular use of more general models, LENSCLEAN has to work more reliable in these cases. Because of this still unsolved problems, we only present very preliminary results from an older self-calibration data set. The results should only be compared with each other but not with results from different *uv* data sets.

β	x_0	y_0	relative H_0	$2-\beta$
0.98	$0.''2630$	$0.''1228$	1.010	1.02
1.00	$0.''2646$	$0.''1221$	1.000	1.00
1.02	$0.''2672$	$0.''1216$	0.997	0.98

Table 9.5: Preliminary LENSCLEAN results for different power-law lens models for B0218+357. H_0 is given relative to the result for $\beta = 1$. The last column presents the expected value for a fixed galaxy position.

Table 9.5 shows some of the results for the best galaxy position and the derived Hubble constant. We notice that the variations of the galaxy position cancel most of the direct effect of β, so that the deviations from the isothermal result are expected to be of the order 1 % or even less for $\beta \approx 1.05$. Whether this is just coincidence or a general property of power-law models for systems like B0218+357 has not been analysed yet. More serious effects are seen for the lens position, which shifts by about 1 mas for 1 % change in β.

The minimal residuals for the different power-law exponents between 0.98 and 1.02 vary only within the $1\,\sigma$ limits. The accuracy of β derived from these residuals would therefore be of the order a few per cent. Much tighter constraints are expected from detailed LENSCLEAN modelling of new 8.4 GHz VLBI data (see chapter 10). Together with the 15 GHz VLA data or other medium resolution data sets to determine the galaxy position, very good constraints for the lens model and the Hubble constant will be possible.

Chapter 10

New VLBI observations of B0218+357

10.1 The 8.4 GHz global VLBI data set

These data and their implications will be discussed in detail in Biggs et al. (2002). This short description is based on that publication.

Observations took place in November 2000 with a global VLBI network including the ten VLBA antennas (25 m each), six telescopes from the European VLBI network EVN (Effelsberg, Medicina, Noto, Onsala, Westerbork and Yebes), the phased inner nine telescopes of the VLA, which was in its widest A configuration and could thus not be used in total due to its small synthesized beam, and the two 70 m Deep Space Network (DSN) antennas Goldstone and Robledo. Observations started on 11 Nov. 20h UT for the European antennas and ended on 12 Nov. 9h UT for the VLBA.

The frequency of 8.4 GHz was chosen as an optimum between higher resolution and less effect of scattering at higher frequencies and higher sensitivity at lower frequencies. Two of the telescopes (Yebes and Owens Valley of the VLBA) did not deliver data because of technical problems. Apart from Noto and Onsala, which only recorded right-handed circular polarization, both senses were recorded at all telescopes, allowing full polarization data to be produced by the correlator.

As a result of serious amplitude calibration problems, the Westerbork data could not be used for the maps. The remaining data were mapped and self-calibrated in a standard way. For details, the reader is referred to Biggs et al. (2002). The resulting off-source noise of $30\,\mu\mathrm{Jy}/\mathrm{beam}$ is close to the theoretical expectations. The very good *uv* coverage for Stokes I is shown in Figure 10.1 for illustration. For uniform weighting the beam shape is almost Gaussian with only weak sidelobes.

The reduced data became available when the bulk of work for this thesis had already been finished. A detailed LENSCLEAN analysis will be done in the future. This will be more difficult than the analysis of the VLA and MERLIN data, because the number of visibilities is higher and two separated CLEAN windows have to be used to cover the regions of interest with the required resolution.

The sensitivity of the data is not sufficient to see the ring with its very low surface brightness at small scales. The main goal was to detect more jet components close to the core image, to obtain more constraints for the lens models. Another objective was to either detect a third central image or emission from the centre of the lensing galaxy itself. Both would provide extremely accurate constraints for the position of the galaxy. If, as was actually the case, no emission could be detected close to the expected lens centre, the data can at least give a strict upper limit for

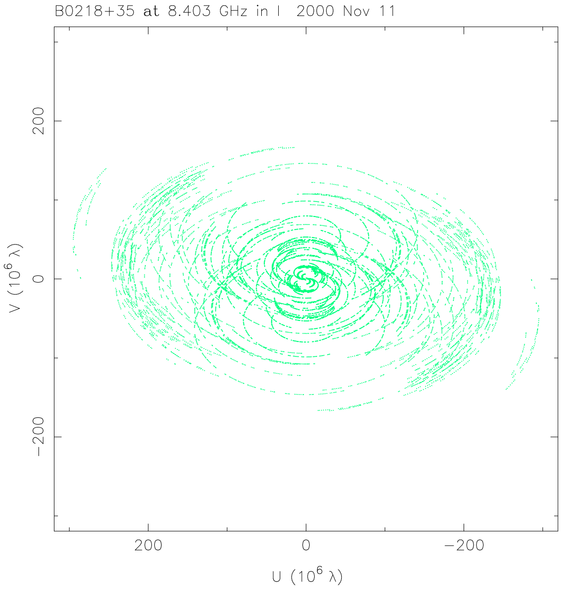


Figure 10.1: UV coverage of the 8.4 GHz VLBI observations of B0218+357

the flux of the central image. This would provide valuable information about the central density profile of the lens, in particular about the core radius.

Although some calibration problems had to be solved, the data are of a very good quality. Even preliminary maps showed a jet extending from the 1 and 2 central components both in A and B.

10.2 Projection into the source plane

Before doing the complete LENSCLEAN modelling, the data can at least be used as a consistency check of the existing lens models. In the first attempts, we used a classically fitted lens model with the lens centre from equation (7.7) to project the maps from the lens plane near A and B back to the source plane. These independent source maps can then be compared. As lens plane emission model we used a CLEAN component list kindly provided by Andy Biggs, which is the result of uniformly weighted mapping and self-calibrating the data with two windows, one around A and one around B.

Figure 10.2 (top) shows maps produced from this component lists and convolved with the beam of $1.36 \times 0.41\,\mathrm{mas}^2$ at p.a. $= -7\overset{\circ}{.}53$. The residuals were not included in these maps, because we are mainly interested in the projection in the source plane. The CLEANing was deep enough to produce meaningful maps even without the residuals. In the bottom part, all CLEAN components were projected into the source plane using the lens model. The CLEAN beam size and shape was transformed using equation (9.61) for each component position. The resulting beams are noticeable in the outer parts of the maps where only isolated components are present. Variations of the projected beams are strongest in the western part of the B map, close to the centre of the lensing galaxy. The logarithmic colour/greyscale encoding was chosen in a way to make structures near the noise level more apparent. The contour lines are the better indicator for the higher surface brightness regions.

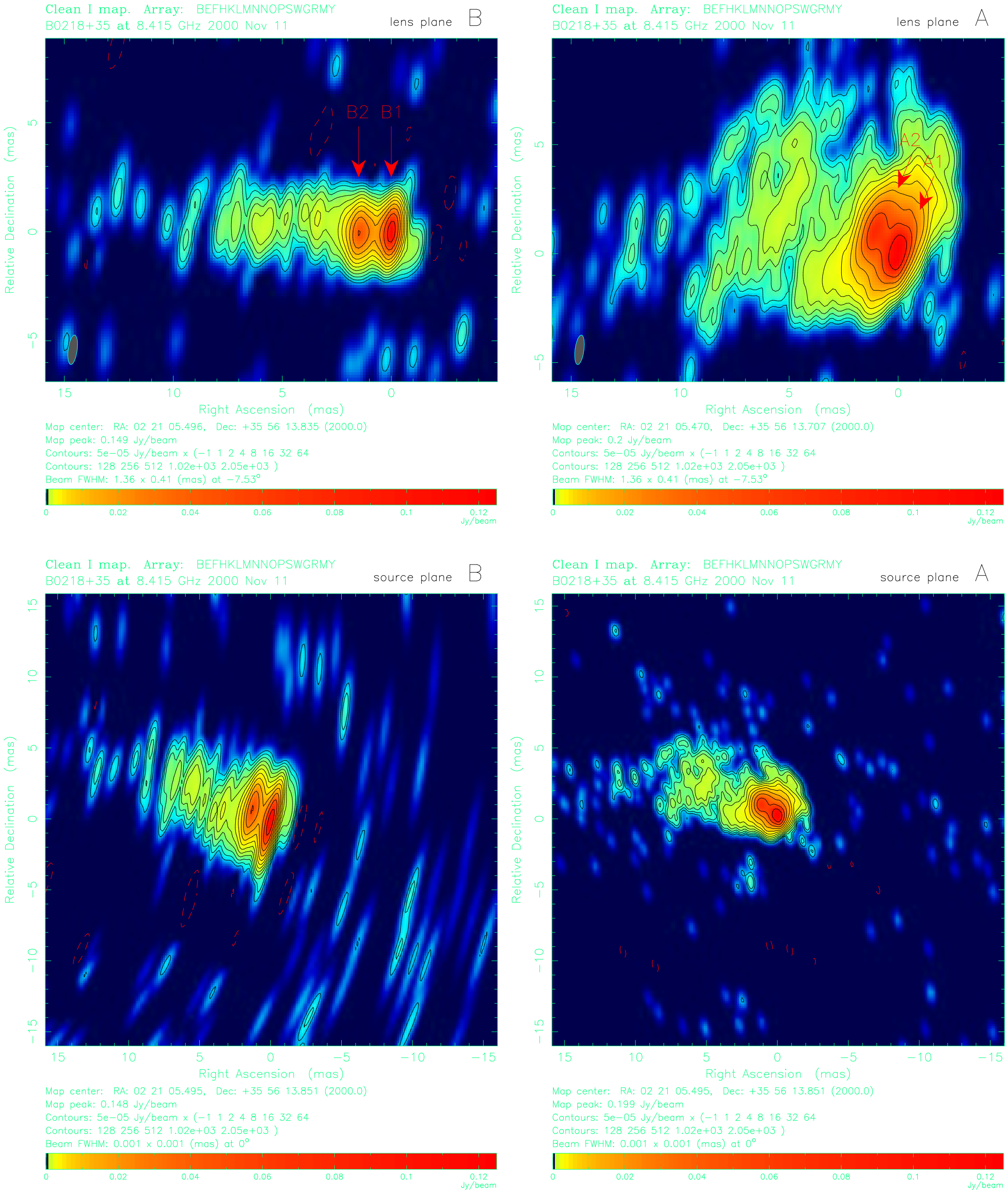


Figure 10.2: Top: Lens plane CLEAN maps without residuals (B and A). Bottom: Direct projection in the source plane. CLEAN beams are projected locally. Units are $\mathrm{Jy/mas^2}$ and not Jy/beam. The scaling factor is $0.6318\,\mathrm{mas^2/beam}$. Contour lines and colour/greyscale coding of the surface brightness is the same in all maps. Coordinate zeros are the positions of the (projected) A1 and B1 components.

Constant source plane beams were used to produce Figure 10.3. For the upper maps they where used directly as calculated for the core components of A and B independently. Comparison of the two is still difficult because of the very different resolution. For the lower maps, the same beam was used for A and B. It was chosen to be at least as large as both A and B in all directions but otherwise as small as possible. The resulting beam is very similar to the projected

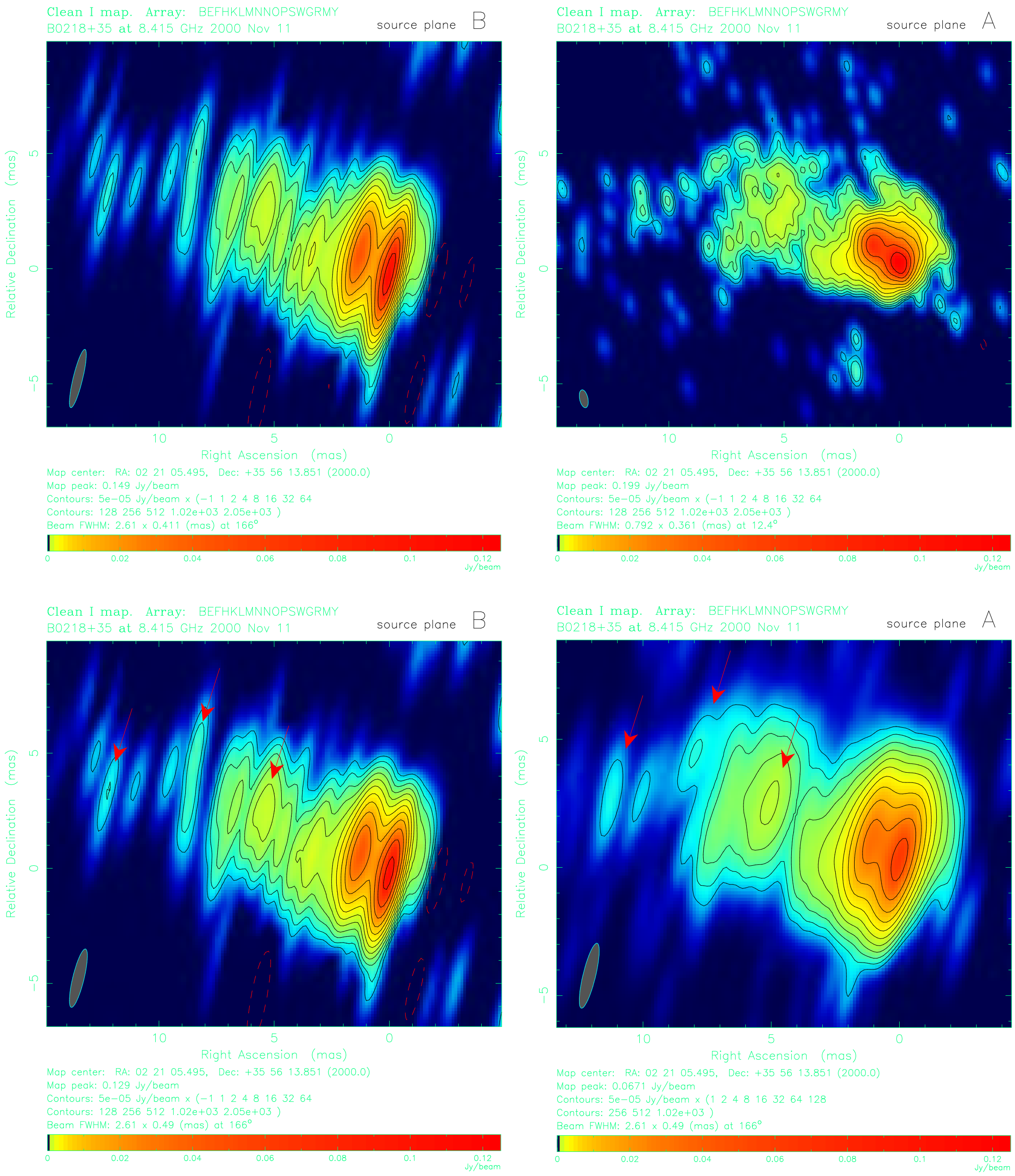


Figure 10.3: Projected source plane maps (B and A). Top: Beams projected at the core images. Bottom: Same beam for A and B. The arrows mark some corresponding features in A and B. Units are Jy/mas^2 and not Jy/beam. The scaling factor is 0.6318 mas^2/beam in the lens plane. Contour lines and colour/greyscale coding of the surface brightness is the same in all maps. Coordinate zeros are the positions of the (projected) A1 and B1 components.

B beam but slightly wider. The nominal resolution is now the same in A and B. Given the very different lens plane images (Figure 10.2 top), the source plane maps look remarkably similar. There are still some small but interesting differences in the details. The A component seems to be slightly smeared when compared with B. This can not be a residual effect of the different lens plane beams, because the resolution in A, which has the smaller projected beams, should be at least as high as that of B then. The opposite effect is observed. The most probable explanation is the presence of significant scatter broadening in A. This would be compatible with the frequency dependent flux ratios and other observed effects like peak fluxes smaller in A than in B at lower frequencies (see section 7.8).

A closer comparison of the A and B source plane maps shows that the B jet seems to be stretched by about 10 % relative to the A jet. This can only be explained with a different radial mass distribution, because the jet is directed more or less radially relative to the galaxy's centre. The 10 % effect can be explained by a power-law potential with $\beta \approx 1.04$. This is in very good agreement with the results from the central components 1 and 2 in section 7.6.2.

In the lower part of Figure 10.3, the B map was demagnified by 11 % to achieve the same visual appearance in A and B without repeating the projection with another lens model. The arrows mark some of the jet components that can be identified in A and B and which can be used as constraints for the lens models. A close comparison shows that some of the weak components are present in only one of the images and are probably noise, whilst others are very similar in both maps. The relative stretching seems to be the same for the whole visible jet. The two independent maps can also be used to study the source itself. A very interesting feature is the extension to the north west of the core, which is present in both maps and which may be a first hint of a counter jet.

We also repeated the same calculations with the best lens model from our LENSCLEAN analysis of the 15 GHz VLA data. The resulting source plane maps are very similar to the ones we have presented here.

No quantitative analysis has been done for the 8.4 GHz VLBI data yet. Model component fitting and classical lens modelling can be used to constrain the mass distribution much better than before. The accuracy of positions from model fitting are, however, difficult to estimate. A final analysis will therefore be based on LENSCLEAN, which can exploit the information without explicit assumptions for the source structure. We expect very precise and robust results for the radial mass distribution from this analysis.

Another goal of the observations was to detect a possible central image. No emission above the noise level is visible in the data near the expected position. This result can later be used to constrain the central mass concentration (core radius, cuspy density profile) with high accuracy.

Chapter 11

Discussion

The work presented here consisted of two main parts. The smaller one is chapter 6 which deals with model degeneracies in quadruple lens systems. We used a very general semi-parametric lens model approach to study the changes of time-delays and the determined Hubble constant with the assumed radial density slope, quantified by the radial mass index β. The use of a linear formalism made the study of the β dependence easy. For fixed external shear in quadruple lenses, this resulted in the simple scaling law $H_0 \propto (2-\beta)/\beta$, independent of the lens geometry, the time-delay ratios or the external shear. This means that a systematic error in the assumed β will have exactly the same effect on all lenses and will not show as scatter in the results. The good agreement between H_0 measurements from different lenses (Koopmans & Fassnacht, 1999) should therefore not be taken as evidence for an accurate determination of H_0. It merely shows that all lensing galaxies seem to have more or less the same β.

In nearly isothermal models, a systematic error of only 10 per cent in β will result in an error of about 20 per cent in the deduced Hubble constant. To compare the results from lensing not only with each other but also with results derived from other methods, this possible source of error has to be taken into account. Furthermore, it is important not only to be aware of this effect, but to try and obtain better constraints on the radial mass profile. Work in this direction constitutes the second part of this thesis.

We also quantified the effect of external shear by introducing the concept of a 'critical shear' $\boldsymbol{\gamma}_{\mathrm{c}}$. The effect of γ on H_0 is linear and strongest in the direction of $\boldsymbol{\gamma}_{\mathrm{c}}$. For a fixed direction, its amount is proportional to $\gamma/\gamma_{\mathrm{c}}$. The value of $\boldsymbol{\gamma}_{\mathrm{c}}$ can be found in a geometrical way. It is given by the ellipticity of the roundest ellipse passing through all images. For $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{\mathrm{c}}$ and non-vanishing β, the time-delays Δt (or alternatively the Hubble constant H_0) become zero. This is also true for a whole family of models which are represented by the less symmetric ellipses fitting the images. The models with vanishing time delays can be modified slightly to allow complete Einstein rings, the ellipticity of which is determined by the shear. Small perturbation of Einstein ring systems can lead to configurations with a very high number of images of one source.

The effect of shear is also the clue in understanding the compatibility of the general scaling law with the simpler one of $H_0 \propto 2-\beta$ for spherical models. In spherical models close to Einstein ring cases, the shear is constrained by the observational data to be $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{\mathrm{c}}(2-\beta)/2$. This special value of external shear changes the product of time delays and Hubble constant by a factor of $\beta/2$ relative to the shearless models.

Interestingly, the values of the critical shear and the Hubble constant in the general model (with arbitrary but fixed shear) do not depend on the position of the lensing galaxy. This may be of use for systems where this position cannot be determined accurately.

When using more general models or less constraints than in our calculations, the scaling laws still apply. They are only valid for a subset of the possible models then and one would have to expect even larger uncertainties when using the whole set. This also applies for lenses with less than four images. More special models on the other hand, like parametrized elliptical power-law models, may be able to constrain the range of possible results much better. Nevertheless, the scaling laws still apply for the range of models that are compatible with the constraints. Even in these cases, our results may be used to determine H_0 without explicit modelling.

The main part of our work is concerned with lens models for one of the most promising lens systems for cosmological applications, B0218+357. Classical model fits for this lens are presented in section 7.6. We showed that it is not possible to determine the position of the lensing galaxy because of a parameter degeneracy. The initial hope that the VLBI substructure can be used as constraint for this most important model parameter has not been fulfilled. This is a consequence of a special property of the isothermal models used. Because they do not magnify the source radially, the subcomponents, which are separated almost exactly in the radial direction, cannot constrain the position in this direction. Unfortunately, this is exactly the direction of the gradient of H_0, so that a very wide range of the Hubble constant is compatible with the observations. Even negative values cannot be ruled out from these data alone. In the direction perpendicular to this, the constraints are much tighter.

B0218+357 is the lens with the smallest image separation of all galaxy mass systems. The small separation of only $0\farcs33$, which is the same as the diameter of the Einstein ring, makes direct optical measurements of the galaxy position very difficult. Existing observations are far too inaccurate to provide useful constraints for H_0. This will hopefully change with scheduled HST observations in the near future.

A significant fraction of this thesis is devoted to the LENSCLEAN method. We started the development mainly to exploit the radio ring of B0218+357 to refine the lens models. It showed that this system is a particularly hard test case for LENSCLEAN because of the high dynamic range. Even small residuals from the bright compact images can dominate the effects from the ring unless extreme care is taken and the algorithm is used in its optimal version. We discussed a number of improvements which were necessary to make LENSCLEAN a useful method for lens models of B0218+357. The final version was used to determine the position of the lens galaxy from a 15 GHz VLA data set, which was possible with an accuracy of a few milliarcseconds for isothermal ellipsoidal lens models. The result for the Hubble constant is $H_0 = (71 \pm 5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for an Einstein-de Sitter universe or $H_0 = (75 \pm 6)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for the current standard model of a flat low-density universe with a cosmological constant. The error bars are $2\,\sigma$ confidence limits and include the uncertainty of the lens model and the time delay. Preliminary results for non-isothermal model provide evidence that deviations of the radial exponent of the order 5 % do not change the results for H_0 significantly. The direct effect from the lens model and the indirect effect from small induced shifts of the galaxy centre compensate to a large fraction.

A data set with higher resolution, which was taken in combination of the MERLIN array and the VLA at 5 GHz, could not be used to improve the results further for two reasons. One is given by serious calibration problems which affect LENSCLEAN much stronger than the normal mapping process. The data were indeed used to produce the best maps of the overall structure of B0218+357 yet (Biggs et al., 2001). Possible scattering in the lensing galaxy is another reason for the failure of LENSCLEAN model fits with these data. The effects are much more important at 5 GHz than at 15 GHz. For some time several theoretical ideas have been under discussion to explain the frequency-dependent flux ratios of the compact images. One is the strong effect of scattering in the A component which acts effectively as extinction. Another is a possible shift or change of structure of the source itself with frequency. Spectral index effects can produce small shifts of the visible ‘core’ components along the direction of the jet for decreasing frequencies. This shift, together with the strong differential amplification gradient of the lens, can produce an effect similar to the one observed. The comparison of models from 15 GHz data with data at 5 GHz including possible extinction and shifts made it possible to estimate these two effects. The results give strong evidence that the extinction of the A component is real and that no significant relative shift is present.

This result will be tested with a (yet unreduced) VLBI observation at five frequencies which can be used to measure the positions of the compact components with high accuracy in an absolute reference frame. Shifts much smaller than the ones expected without extinction can be detected with these data.

The probable extinction in the A component of ca. 24 % is, unfortunately, bad news for the application of LENSCLEAN for VLBI observations of B0218+357 at lower frequencies, which could otherwise provide very good constraints for the lens models. If such a high extinction is present at one position of the lens plane, it cannot be ruled out, that other regions with similarly strong effects exist. It is impossible to correct for this in a simple way.

Deep VLBI observations at 8.4 GHz, presented in chapter 10, are affected by the scattering to a lesser degree. Maps of the data show the jet at a scale of 10 mas for the first time. Many well defined features in the maps offer the possibility to determine the radial mass exponent with very high accuracy. As a preliminary analysis, we showed maps of the source plane calculated by projecting the maps of the A and B components. It shows that the source plane maps are in very good agreement only if the mass exponent is not exactly unity (isothermal) but $\beta \approx 1.04$. LENSCLEAN will be the method of choice for the final and quantitative analysis. This will provide not only the best-fit value but also reliable estimates of the uncertainty. The determination of the lens position will then be repeated with the best value of β, or the complete lens model will preferably be fitted simultaneously to the VLBI *and* VLA data. This will lead to the best constrained model of any time delay lens and to a very robust determination of the Hubble constant.

Further improvements of the results for B0218+357 will rely on progress in all three fields which contribute to the uncertainty. Possible errors from the time delay, the lens model and the cosmological parameters are of the same order of magnitude at the moment. All of these therefore have to be reduced to achieve significantly more accurate results. For the time delay, new monitoring data and a reanalysis of the old data can help in reducing the errors. It may also be possible to confirm the time delay in the optical or even at X-ray wavelengths where fast variations are common for BL Lac objects like B0218+357.

The lens models can be improved by better medium-resolution data to tighten the constraints for the galaxy position. These may be taken with the extended versions of the VLA or MERLIN or with the already existing VLA–Pie Town link. The results from observations with the new ACS detector on board the HST will provide a completely independent measurement of this parameter. Together, they can be used to cross-check the results or to constrain other parameters.

The lens galaxy is a gas-rich spiral which allows to study the kinematics of the disc with spectroscopic observations of absorption lines in the emission from the background source. Separate spectra for the two components are expected at radio wavelengths in the near future. It will also be investigated whether or not the spectral and spatial resolution of optical observations is sufficient to obtain useful information about the disc kinematics. Combined with shape parameters of the disc from optical observations, which can be used to infer the inclination and its position angle, an independent measure of the mass distribution will then be available. Together with the lens models, a partial breaking of the disc/halo degeneracy might even be possible. Such data will at the very least provide a very important cross-check of the lensing results.

Even more important than future improvements of the B0218+357 models will be the application of LENSCLEAN for as many radio systems with extended or multi-component sources as possible. Results of individual systems can always be impaired by several effects, e.g. small scale mass clumps like globular clusters in the lensing galaxy, unseen secondary galaxies close to the line of sight, cosmic shear or untypical properties of the individual lenses. To be able to detect such effects and to obtain reliable and representative results, detailed modelling efforts have to be spent for many systems. This should not only be done for time-delay lenses but also for other applicable systems. Most important for the Hubble constant is the radial mass distribution, which can only be constrained well with extended sources. LENSCLEAN is thus the method of choice for this problem. Our analytical work also showed that external shear can have significant effects in certain lens systems. Ellipticities of the main lenses and the external shear should therefore also be inspected precisely. The results will not only be important for cosmological applications but also for the study of high-redshift lensing galaxies, whose mass distributions cannot be investigated by other means.

We conclude that LENSCLEAN will probably play an important role in future lens work. It is currently by far the best method available to utilize the information from arbitrary, possibly extended, radio sources to determine mass models of gravitational lenses.

Appendix A

Parametrization and mapping of ellipses

A.1 General

Ellipses occur in several contexts in this work. They are used to describe mass distributions and potentials, components of lensed sources, Einstein rings and locations of images. A well-chosen parametrization of these curves can greatly help in the analytical and numerical calculations. An ellipse with major half axis a in the x direction and minor half axis b in the y direction is described by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad , \tag{A.1}$$

or in polar coordinates:

$$x = r\cos\phi \tag{A.2}$$
$$y = r\sin\phi \tag{A.3}$$
$$r_0 = ab\sqrt{\frac{2}{a^2+b^2}} \tag{A.4}$$
$$e = \frac{a^2-b^2}{a^2+b^2} \tag{A.5}$$
$$r(\phi) = r_0\,(1 - e\cos 2\phi)^{-1/2} \tag{A.6}$$

For elliptical potentials, a parametrization of the ellipticity with ε according to

$$a \propto 1+\varepsilon \quad , \tag{A.7}$$
$$b \propto 1-\varepsilon \tag{A.8}$$

is also used. The parametrization with the axial ratio $q = b/a$ (e.g. Kormann et al., 1994) has the disadvantage of breaking the symmetry of the equations. The relations between the different measures of ellipticity are as follows:

$$\begin{array}{ccccccc} q & = & \sqrt{\frac{1-e}{1+e}} & = & \frac{1-\varepsilon}{1+\varepsilon} & = & \frac{b}{a} \\ \frac{1-q^2}{1+q^2} & = & e & = & \frac{2\varepsilon}{1+\varepsilon^2} & = & \frac{a^2-b^2}{a^2+b^2} \\ \frac{1-q}{1+q} & = & \frac{1-\sqrt{1-e^2}}{e} & = & \varepsilon & = & \frac{a-b}{a+b} \end{array} \tag{A.9}$$

For the description of elliptical source components and to calculate how they project in the lensing process, we have to generalize for arbitrary orientation of the ellipses. The orientation is measured by the orientation angle θ of the major axis, in mathematical sense from the positive x direction through the positive y direction. With the approach

$$\boldsymbol{z}^{\dagger}\mathsf{E}^{-1}\boldsymbol{z} = 1 \quad , \tag{A.10}$$

the elements of the matrix E can be calculated either by starting from the polar form (A.6), subtracting θ from ϕ and using simple trigonometric formulas to separate the ϕ and θ parts, or by starting with the matrix E in the system of the main axis and apply a rotation matrix to $\boldsymbol{z}$. This rotation of $\boldsymbol{z}$ is equivalent to a transformation of E with the rotation matrix applied from the left and the inverse from the right side. The following parametrization of the matrix proves to be convenient:

$$\mathsf{E}^{-1} = \frac{1}{a^2b^2}\begin{pmatrix} R_+ - R_x & -R_y \\ -R_y & R_+ + R_x \end{pmatrix} \tag{A.11}$$

$$\mathsf{E} = \begin{pmatrix} R_+ + R_x & R_y \\ R_y & R_+ - R_x \end{pmatrix} \tag{A.12}$$

$$R_\pm = \frac{a^2 \pm b^2}{2} \tag{A.13}$$

$$R_x = R_- \cos 2\theta \tag{A.14}$$

$$R_y = R_- \sin 2\theta \tag{A.15}$$

The parameters R_+, R_x, R_y and the matrix E have the advantage that they are regular even for degenerated ellipses. In the case of elliptical Gaussian components, they are direct measures of the second order brightness moments and are thus related to observations in a direct and linear way. This is essential for a simple error propagation from the measurements to the ellipse parameters. The inverse matrix E^{-1} does not share this nice properties. If the measurement errors of the second order moments are of the same order of magnitude as the moments themselves, linear error propagation would not be possible for E^{-1}, because parameters would diverge within the allowed range. This is not a problem for E.

Please note that this formalism is the same as the one used to parametrize the CLEAN beam in equation (8.21). The matrix $\mathsf{E}^{-1} = \mathsf{G}$ describes the ellipse at which the beam has fallen to $1/\mathrm{e}$ of its peak.

A.2 Lens mapping

For lensing applications, the mapping of ellipses has to be investigated, e.g. the mapping from the source to the lens plane. If coordinates map like

$$\boldsymbol{z} = \mathsf{M}\,\boldsymbol{z}_\mathrm{s} \quad , \tag{A.16}$$

the ellipticity matrix E is correspondingly transformed like

$$\mathsf{E}_\mathrm{s}^{-1} = \mathsf{M}^\dagger\,\mathsf{E}^{-1}\,\mathsf{M} \tag{A.17}$$

or

$$\mathsf{E}_\mathrm{s} = \mathsf{M}^{-1}\,\mathsf{E}\,(\mathsf{M}^\dagger)^{-1} \quad . \tag{A.18}$$

This mapping becomes especially simple in the case of orthogonal mappings (rotations) or symmetric matrices M as they are given in the case of a lens mapping. For fixed M, the matrices E and E^{-1} transform linearly, which is another advantage of this parametrization. The relations to the parameters that were used before are given by

$$r_0 = \sqrt{\frac{R_+^2 - R_-^2}{R_+}} \quad , \tag{A.19}$$

$$e = \frac{R_-}{R_+} \quad . \tag{A.20}$$

For lens model fitting we need the linear transformation explicitly. With the normal inverse magnification matrix

$$\mathsf{M}^{-1} = \begin{pmatrix} 1-\kappa+\gamma_x & \gamma_y \\ \gamma_y & 1-\kappa-\gamma_x \end{pmatrix} \quad , \tag{A.21}$$

and the magnification $\mu = |\mathsf{M}| = \left((1-\kappa)^2-\gamma^2\right)^{-1}$, the transformation

$$\mathsf{E} = \mathsf{M}\,\mathsf{E}_\mathrm{s}\,\mathsf{M} \tag{A.22}$$

can be written as

$$\begin{pmatrix} R_+ \\ R_x \\ R_y \end{pmatrix} = \mathsf{A} \begin{pmatrix} R_+^\mathrm{s} \\ R_x^\mathrm{s} \\ R_y^\mathrm{s} \end{pmatrix} \quad , \tag{A.23}$$

with

$$\mathsf{A} = \mu^2 \begin{pmatrix} (1-\kappa)^2+\gamma_x^2+\gamma_y^2 & -2(1-\kappa)\gamma_x & -2(1-\kappa)\gamma_y \\ -2(1-\kappa)\gamma_x & (1-\kappa)^2+\gamma_x^2-\gamma_y^2 & 2\gamma_x\gamma_y \\ -2(1-\kappa)\gamma_y & 2\gamma_x\gamma_y & (1-\kappa)^2-\gamma_x^2+\gamma_y^2 \end{pmatrix} \quad . \tag{A.24}$$

This equation can be used with standard linear least-squares techniques (see section 2.2) to fit source component parameters to multiple image components and calculate χ^2 contributions for lens modelling purposes (section 4.3.5).

Appendix B

Fourier transforms and Fourier series

The properties of Fourier transforms and series are well known and can be found in a number of text books. The most important equations are summarized here to provide a reference list and to introduce the notations that is used in the main part of this thesis. We start with continuous Fourier transforms in $\mathbb{R}^n$ and interpret the Fourier series as a special case.

B.1 Continuous Fourier transforms

The Fourier transform (FTor $\mathfrak{FT}$) of an 'arbitrary'[1] function $f(\boldsymbol{x})$ ($\boldsymbol{x} \in \mathbb{R}^n$) is defined as follows:

$$F = \mathfrak{FT}\, f \tag{B.1}$$

$$F(\mathbf{u}) = \int \mathrm{d}^n x\, \mathrm{e}^{2\pi \mathrm{i}\, \mathbf{u}\cdot\boldsymbol{x}} f(\boldsymbol{x}) \tag{B.2}$$

The variable in Fourier space is also n-dimensional ($\mathbf{u} \in \mathbb{R}^n$). With this definition, the inverse transform is very similar to the transform itself:

$$f = \mathfrak{FT}^{-1} F \tag{B.3}$$

$$f(\boldsymbol{x}) = \int \mathrm{d}^n u\, \mathrm{e}^{-2\pi \mathrm{i}\, \mathbf{u}\cdot\boldsymbol{x}} F(\mathbf{u}) \tag{B.4}$$

Definitions without the 2π factor in the exponents are also possible, but they would introduce a scaling factor in either or both transforms. With the definitions for (one-dimensional) convolution and correlation,

$$(f * g)(x) = \int \mathrm{d}x'\; f(x - x')\, g(x') \quad , \tag{B.5}$$

$$\mathrm{Corr}(f, g)(x) = \int \mathrm{d}x'\; f^\star(x' - x)\, g(x') \quad , \tag{B.6}$$

---

[1] We do not claim true mathematical accuracy here. Most important is the integrability of $f(x)$. Generalizations can be treated with the formalism of distributions (eg. Forster, 1984; Wucknitz, 1996). The technical details can be very important in some contexts but shall not be copied from mathematical text books here.

the following relations can be derived:

function	Fourier transform	
$f(x)$	$F(u)$	(B.7)
$g(x)$	$G(u)$	(B.8)
$f(ax)$	$\frac{1}{\lvert a\rvert} F\left(\frac{u}{a}\right)$	(B.9)
$f(x-x_0)$	$F(u)\,\mathrm{e}^{2\pi \mathrm{i}\, u x_0}$	(B.10)
$(f * g)(x)$	$F(u)\,G(u)$	(B.11)
$\mathrm{Corr}(f,g)(x)$	$F^\star(u)\,G(u)$	(B.12)

The generalization for the n-dimensional case is straightforward. For real functions f, the FT shows the symmetry

$$F(-\mathbf{u}) = F^\star(\mathbf{u}) \quad . \tag{B.13}$$

The Wiener-Khinchin theorem for the autocorrelation

$$\mathcal{FT}\,\mathrm{Corr}(f,f) = |F|^2 \tag{B.14}$$

leads directly to the Parseval theorem:

$$\int \mathrm{d}x\, |f(x)|^2 = \int \mathrm{d}u\, |F(u)|^2 \tag{B.15}$$

A Gaussian transforms to a Gaussian:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\,\mathrm{e}^{-x^2/(2\sigma_x^2)} \tag{B.16}$$

$$F(u) = \mathrm{e}^{-u^2/(2\sigma_u^2)} \tag{B.17}$$

$$\sigma_u = \frac{1}{2\pi\,\sigma_x} \tag{B.18}$$

The full width at half maximum (FWHM) is defined as

$$f\left(\frac{\mathrm{FWHM}_x}{2}\right) = \frac{f(0)}{2} \quad , \tag{B.19}$$

$$F\left(\frac{\mathrm{FWHM}_u}{2}\right) = \frac{F(0)}{2} \quad , \tag{B.20}$$

and is related to σ via

$$\mathrm{FWHM} = \sqrt{8\ln 2}\,\sigma \quad . \tag{B.21}$$

FWHM in image and Fourier space have the following relation:

$$\mathrm{FWHM}_u = \frac{4\ln 2}{\pi}\,\frac{1}{\mathrm{FWHM}_x} \tag{B.22}$$

The transform of a pillbox function is a sinc:

$$f(x) = \begin{cases} 1 & \text{for } |x| < X \\ 0 & \text{for } |x| > X \end{cases} \tag{B.23}$$

$$F(u) = 2X\,\mathrm{sinc}(2uX) \tag{B.24}$$

$$\mathrm{sinc}\,x = \frac{\sin \pi x}{\pi x} \tag{B.25}$$

For generalized functions (distributions) the following pairS of functions/transforms follow trivially:

$$f(x) = \delta(x) \quad F(u) \equiv 1 \tag{B.26}$$

$$f(x) \equiv 1 \qquad F(u) = \delta(u) \tag{B.27}$$

For shifted δ-functions the translation theorem (B.10) can be applied. The ‘Sha’-function

$$Ш(x) = \sum_{k=-\infty}^{\infty} \delta(x-k) \tag{B.28}$$

is its own Fourier transform,

$$\tilde{Ш}(u) = \sum_{k=-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i} ku} \tag{B.29}$$

$$= \sum_{k=-\infty}^{\infty} \delta(u-k) \quad . \tag{B.30}$$

B.2 Fourier series and discrete Fourier transforms

As for the continuous Fourier transform, several different conventions are used for Fourier series. We use the one most appropriate in the radio interferometry context.

Let us assume that $f(x)$ is L-periodic, i.e. $f(x) = f(x+L)$. To calculate the Fourier transform of this function, we can start with the transform of one period of f, say $0 < x < L$. The complete function can be constructed from this period by convolving it with a modified Sha-function $Ш(x/L)/L$. Since this convolution in x-space is a multiplication in u-space with the transform of the modified $Ш$, the transform can be written as follows:

$$F(u) = Ш(uL) \int_0^L \mathrm{d}x\, \mathrm{e}^{2\pi \mathrm{i} ux} f(x) \tag{B.31}$$

The $Ш$ selects multiple frequencies of $1/L$. The transform and its inverse can therefore be written as

$$F(u) = \sum_{k=-\infty}^{\infty} F_k\, \delta\left(u - \frac{k}{L}\right) \quad , \tag{B.32}$$

$$f(x) = \sum_{k=-\infty}^{\infty} F_k\, \mathrm{e}^{-2\pi \mathrm{i} kx/L} \quad , \tag{B.33}$$

with

$$F_k = \frac{1}{L} \int_0^L \mathrm{d}x\, \mathrm{e}^{2\pi \mathrm{i} kx/L} f(x) \quad . \tag{B.34}$$

This transformation for periodic functions is called a Fourier series; the F_k are the Fourier coefficients. See e.g. Forster (1983) for details. Discrete functions f with n samples per period can be written as

$$f(x) = \sum_{j=0}^{n-1} f_j\, \delta\left(x - \frac{jL}{n}\right) \tag{B.35}$$

in the interval $0 < x < L$. This can be combined with equation (B.34) to obtain the Fourier coefficients

$$F_k = \frac{1}{L} \sum_{j=0}^{n-1} f_j\, \mathrm{e}^{2\pi \mathrm{i} kj/n} \quad . \tag{B.36}$$

These are periodic with $F_k = F_{k+n}$, analogously to f_j. The inverse transform in the discrete case is

$$f_j = \frac{L}{n} \sum_{k=0}^{n-1} F_k\, \mathrm{e}^{-2\pi \mathrm{i} kj/n} \quad . \tag{B.37}$$

Abstracting, from the continuous functions we can think of f_k as a discrete function with discrete Fourier transform LF_k. The regular sampling in x-space leads to periodicity in u-space and vice versa. Table B.1 summarizes the sampling intervals and periods in x and u-space.

	interval	period
x	$\Delta x = L/n$	$L = n\Delta x$
u	$\Delta u = 1/L$	$n/L = 1/\Delta x$

Table B.1: Sampling interval and period in image and Fourier space for the discrete Fourier transform

A FT of n terms can be split into the sum of two FTs with $n/2$ terms each. Splitting recursively until $n = 1$, leads to the very efficient FFT algorithm which needs of the order $\log_2 n$ steps to calculate one F_k, while the direct calculation would need n. For large n the saving in CPU time is remarkably high.

List of Figures

List of Tables

Bibliography

Abramowitz, M. & Stegun, I. A.: 1972, *Handbook of Mathematical Functions*, Dover, New York. 12

Bade, N., Siebert, J., Lopez, S., Voges, W. & Reimers, D.: 1997, RX J0911.4+0551: A new multiple QSO selected from the ROSAT All-Sky Survey, A&A **317**, L13. 64, 65

Barkana, R.: 1997, Analysis of time delays in the gravitational lens PG 1115+080, ApJ **489**, 21. 63

Biggs, A. D.: 1999, *Investigations of the Gravitational Lens System B0218+357*, PhD thesis, Jodrell Bank Observatory, Manchester University, UK. 71

Biggs, A. D., Browne, I. W. A., Helbig, P., Koopmans, L. V. E., Wilkinson, P. N. & Perley, R. A.: 1999, Time delay for the gravitational lens system B0218+357, MNRAS **304**, 349. 71, 73, 78, 79, 82, 83

Biggs, A. D., Browne, I. W. A., Muxlow, T. W. B. & Wilkinson, P. N.: 2001, MERLIN/VLA imaging of the gravitational lens system B0218+357, MNRAS **322**, 821. 69, 89, 214

Biggs, A. D., Wucknitz, O., Porcas, R. W., Browne, I. W. A., Jackson, N., Mao, S., Patnaik, A. R. & Wilkinson, P. N.: 2002, Global VLBI 8.4-GHz observations of JVAS B0218+357, MNRAS. submitted. 207

Blandford, R. D. & Narayan, R.: 1992, Cosmological applications of gravitational lensing, ARA&A **30**, 311. 17

Briggs, D. S.: 1995, *High Fidelity Deconvolution of Moderately Resolved Sources*, PhD thesis, The New Mexico Institute of Mining and Technology, Socorro, New Mexico. currently available from `ftp://ftp.aoc.nrao.edu/pub/dissertations/dbriggs/diss.html`. 107, 108, 112

Briggs, D. S., Schwab, F. R. & Sramek, R. A.: 1999, Imaging, *in* Taylor et al. (1999), p. 127. 123

Browne, I. W. A., Patnaik, A. R., Walsh, D. & Wilkinson, P. N.: 1993, The redshift of the lensing galaxy in the gravitationally lensed system B0218+35.7, MNRAS **263**, L32. 75

Burud, I.: 2002, *Gravitational lensing as a tool for determining the age of the Universe* , PhD thesis, Institut d'Astrophysique, Liege, Belgium. currently available from `http://vela.astro.ulg.ac.be/themes/dataproc/deconv/theses/burud/ibthese_e.html`. 64, 65

Burud, I., Courbin, F., Lidman, C., Jaunsen, A. O., Hjorth, J., Ostensen, R., Andersen, M. I., Clasen, J. W., Wucknitz, O., Meylan, G., Magain, P., Stabell, R. & Refsdal, S.: 1998, High-resolution optical and near-infrared imaging of the quadruple quasar RX J0911.4+0551, ApJ **501**, L5. 34, 64

Carilli, C. L., Radford, S. J. E., Menten, K. M. & Langston, G. I. (eds): 1999, *Highly Redshifted Radio Lines*, ASP Conf. Ser. 156. 236

Carilli, C. L., Rupen, M. P. & Yanny, B.: 1993, Neutral hydrogen 21 centimeter absorption at $z = 0.6847$ toward the 'smallest Einstein ring', ApJ **412**, L59. 75

Chang, K. & Refsdal, S.: 1976, Time delay effects for measuring cosmological distances, *L'Évolution des Galaxies et ses Implications Cosmologiques, Colloques Int. du CNRS*, Vol. 263, p. 369. 47

Chartas, G., Dai, X., Gallagher, S. C., Garmire, G. P., Bautz, M. W., Schechter, P. L. & Morgan, N. D.: 2001, Chandra detects a rapid flare in the gravitationally lensed mini-broad absorption line QSO RX J0911.4+0551, ApJ **558**, 119. 64

Chen, G. H., Kochanek, C. S. & Hewitt, J. N.: 1995, The mass distribution of the lens galaxy in MG 1131+0456, ApJ **447**, 62. 132

Clark, B. G.: 1980, An efficient implementation of the algorithm CLEAN, A&A **89**, 377. 123

Cohen, A. S., Hewitt, J. N., Moore, C. B. & Haarsma, D. B.: 2000, Further investigation of the time delay, magnification ratios, and variability in the gravitational lens 0218+357, ApJ **545**, 578. 73, 78, 79, 80

Corbett, E. A., Browne, I. W. A. & Wilkinson, P. N.: 1996, Radio measurement of the time delay in 0218+357, *in* Kochanek & Hewitt (1996), p. 37. 76

Cornwell, T. & Fomalont, E.: 1999, Self-calibration, *in* Taylor et al. (1999), p. 187. 129

Cornwell, T. & Fomalont, E. B.: 1989, Self-calibration, *in* Perley et al. (1989), p. 185. 129

Cornwell, T. J.: 1983, A method of stabilizing the CLEAN algorithm, A&A **121**, 281. 107

Courbin, F., Magain, P., Keeton, C. R., Kochanek, C. S., Vanderriest, C., Jaunsen, A. O. & Hjorth, J.: 1997, The geometry of the quadruply imaged quasar PG 1115+080: implications for H_0, A&A **324**, L1. 63

Crane, P., Albrecht, R., Barbieri, C., Blades, J. C., Bokensenberg, A., Deharveng, J. M., Disney, M. J., Jakobsen, P., Kamperman, T. M., King, I. R., Macchetto, F., Mackay, C. D., Paresce, F., Weigelt, G., Baxter, D., Greenfield, P., Jedrzejewski, R., Nota, A. & Sparks, W. B.: 1991, First results from the faint object camera — Images of the gravitational lens system G 2237+0305, ApJ **369**, L59. 61

Dyer, C. C. & Roeder, R. C.: 1972, The distance-redshift relation for universes with no intergalactic medium, ApJ **174**, L115. 19

Dyer, C. C. & Roeder, R. C.: 1973, Distance-redshift relations for universes with some intergalactic medium, ApJ **180**, L31. 19

Einstein, A.: 1915, Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie, *Sitzungsber. Preuß. Akad. Wissensch., erster Halbband*, p. 831. 22

Ekers, R. D., Fanti, C. & Padrielli, L. (eds): 1996, *Extragalactic Radio Sources*, IAU Symp. 175. 236

Ellithorpe, J. D., Kochanek, C. S. & Hewitt, J. N.: 1996, Visibility LensCLEAN and the reliability of deconvolved radio images, ApJ **464**, 556. 131, 132, 144

Evans, N. W. & Witt, H. J.: 2001, Are there sextuplet and octuplet image systems?, MNRAS **327**, 1260. 68

Falco, E. E., Gorenstein, M. V. & Shapiro, I. I.: 1985, On model-dependent bounds on H_0 from gravitational images: Application to Q0957+561A,B, ApJ **289**, L1. 49

Fassnacht, C. D., Pearson, T. J., Readhead, A. C. S., Browne, I. W. A., Koopmans, L. V. E., Myers, S. T. & Wilkinson, P. N.: 1999, A determination of H_0 with the CLASS gravitational lens B1608+656. I. Time delay measurements with the VLA, ApJ **527**, 498. 65

Fomalont, E. & Perley, R. A.: 1999, Calibration and editing, *in* Taylor et al. (1999), p. 79. 129

Forster, O.: 1983, *Analysis 1*, 4th edn, Vieweg. in german. 224

Forster, O.: 1984, *Analysis 3*, 3rd edn, Vieweg. in german. 12, 221

Gil-Merino, R., Wisotzki, L. & Wambsganß, J.: 2002, The Double Quasar HE 1104–1805: A case study for time delay determination with poorly sampled lightcurves, A&A **381**, 428. 34

Giovi, F. & Amendola, L.: 2001, The distance-redshift equation in quintessence cosmology and the estimation of H_0 through time delays, MNRAS **325**, 1097. 19

Gorenstein, M. V., Shapiro, I. I. & Falco, E. E.: 1988, Degeneracies in parameter estimates for models of gravitational lens systems, ApJ **327**, 693. 49, 56

Grundahl, F. & Hjorth, J.: 1995, The optical appearance of the gravitational lens system B0218+357, MNRAS **275**, L67. 76

Högbom, J. A.: 1974, Aperture synthesis with a non-regular distribution of interferometer baselines, A&AS **15**, 417. 104

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E. & Perley, R.: 1985, 2237+0305 — A new and unusual gravitational lens, AJ **90**, 691. 60

Impey, C. D., Falco, E. E., Kochanek, C. S., Lehár, J., McLeod, B. A., Rix, H.-W., Peng, C. Y. & Keeton, C. R.: 1998, An infrared Einstein ring in the gravitational lens PG 1115+080, ApJ **509**, 551. 63

Jackson, N., Helbig, P., Browne, I., Fassnacht, C. D., Koopmans, L., Marlow, D. & Wilkinson, P. N.: 1998, Lensing galaxies: light or dark?, A&A **334**, L33. 66

Jackson, N., Xanthopoulos, E. & Browne, I. W. A.: 2000, NICMOS images of JVAS/CLASS gravitational lens systems, MNRAS **311**, 389. 76

Kassiola, A. & Kovner, I.: 1993, Elliptic mass distributions versus elliptic potentials in gravitational lenses, ApJ **417**, 450. 29, 30

Kayser, R.: 1990, Macro- and micromodels for gravitational lenses, ApJ **357**, 309. 31

Kayser, R., Helbig, P. & Schramm, T.: 1997, A general and practical method for calculating cosmological distances, A&A **318**, 680. 19

Kayser, R. & Schramm, T.: 1988, Imaging procedures for gravitational lenses, A&A **191**, 39. 85

Kayser, R., Schramm, T. & Nieser, L. (eds): 1992, *Gravitational Lenses*, Proceedings of a Conference Held in Hamburg, Germany. Springer. Also Lecture Notes in Physics, volume 406. 236

Keeton, C. R.: 2001a, Computational methods for gravitational lensing, ApJ. submitted, also available as `astro-ph/0102340`. 38, 39

Keeton, C. R.: 2001b, Software for gravitational lensing. Software manual, available from Kochanek et al. (2002) or from `http://astro.uchicago.edu/~ckeeton/gravlens/`. 38

Keeton, C. R. & Kochanek, C. S.: 1997, Determining the Hubble constant from the gravitational lens PG 1115+080, ApJ **487**, 42. 63

Keeton, C. R., Kochanek, C. S. & Seljak, U.: 1997, Shear and ellipticity in gravitational lenses, ApJ **482**, 604. 63

Keeton, C. R., Mao, S. & Witt, H. J.: 2000, Gravitational lenses with more than four images. I. Classification of caustics, ApJ **537**, 697. 68

Kemball, A. J., Patnaik, A. R. & Porcas, R. W.: 2001, Polarization VLBI observations of the gravitational lens system B0218+357 at 8.4 GHz, ApJ **562**, 649. 72, 73, 74, 75, 82, 83, 85

Kneib, J.-P., Cohen, J. G. & Hjorth, J.: 2000, RX J0911+05: A massive cluster lens at z=0.769, ApJ **544**, L35. 35, 64, 65

Kochanek, C. S.: 1991, The implications of lenses for galaxy structure, ApJ **373**, 354. 31

Kochanek, C. S.: 1995, Evidence for dark matter in MG 1654+134, ApJ **445**, 559. 132

Kochanek, C. S.: 2002, Gravitational lenses, the distance ladder and the Hubble constant: A new dark matter problem, ApJ. submitted, also available as `astro-ph/0204043`. 183

Kochanek, C. S., Blandford, R. D., Lawrence, C. R. & Narayan, R.: 1989, The ring cycle — an iterative lens reconstruction technique applied to MG1131+0456, MNRAS **238**, 43. 85

Kochanek, C. S., Falco, E. E., Impey, C., Lehár, J., McLeod, B. & Rix, H.-W.: 2002, CASTLES website, `http://cfa-www.harvard.edu/castles/`. 35, 60, 63, 65, 66, 75, 77, 234

Kochanek, C. S. & Hewitt, J. N. (eds): 1996, *Astrophysical Applications of Gravitational Lensing*, IAU Symp. 173. 232, 235, 236

Kochanek, C. S. & Narayan, R.: 1992, LensCLEAN: An algorithm for inverting extended, gravitationally lensed images with application to the radio ring lens PKS 1830–211, ApJ **401**, 461. 4, 131, 132, 144, 194, 195

Koopmans, L. V. E. & Fassnacht, C. D.: 1999, A determination of H_0 with the CLASS gravitational lens B1608+656. II. Mass models and the Hubble constant from lensing, ApJ **527**, 513. 66, 67, 213

Kormann, R., Schneider, P. & Bartelmann, M.: 1994, Isothermal elliptical gravitational lens models, A&A **284**, 285. 30, 217

Kovner, I.: 1987, The quadrupole gravitational lens, ApJ **312**, 22. 45, 198

Kovner, I.: 1990, Fermat principle in arbitrary gravitational fields, ApJ **351**, 114. 22

Kundic, T., Turner, E. L., Colley, W. N., Gott, J. R. I., Rhoads, J. E., Wang, Y., Bergeron, L. E., Gloria, K. A., Long, D. C., Malhotra, S. & Wambsganß, J.: 1997, A robust determination of the time delay in 0957+561A, B and a measurement of the global value of Hubble's constant, ApJ **482**, 75. 76

Lawrence, C. R.: 1996, Observations of lens systems with Keck I, *in* Kochanek & Hewitt (1996), p. 299. 76

Lawson, C. L. & Hanson, R. J.: 1974, *Solving Least Squares Problems*, Prentice-Hall. 112

Lehár, J., Falco, E. E., Kochanek, C. S., McLeod, B. A., Muñoz, J. A., Impey, C. D., Rix, H.-W., Keeton, C. R. & Peng, C. Y.: 2000, Hubble Space Telescope observations of 10 two-image gravitational lenses, ApJ **536**, 584. 72, 74, 75, 76, 77, 80, 82, 83

Lopez, S., Wucknitz, O. & Wisotzki, L.: 1998, Detection of the lensing galaxy in HE 2149–2745, A&A **339**, L13. 35, 68

Mao, S. & Schneider, P.: 1998, Evidence for substructure in lens galaxies?, MNRAS **295**, 587. 66

Marsh, K. A. & Richardson, J. M.: 1987, The objective function implicit in the CLEAN algorithm, A&A **182**, 174. 106

Menten, K. M. & Reid, M. J.: 1996, Formaldehyde absorption at z=0.685 toward the ”Einstein ring” B0218+357, ApJ **465**, L99. 182

Metcalfe, N. & Shanks, T. (eds): 2002, *A New Era in Cosmology*, ASP Conf. Ser. in print. 237, 238

Mould, J. R., Huchra, J. P., Freedman, W. L., Kennicutt, R. C., Ferrarese, L., Ford, H. C., Gibson, B. K., Graham, J. A., Hughes, S. M. G., Illingworth, G. D., Kelson, D. D., Macri, L. M., Madore, B. F., Sakai, S., Sebo, K. M., Silbermann, N. A. & Stetson, P. B.: 2000, The Hubble Space Telescope key project on the extragalactic distance scale. XXVIII. Combining the constraints on the Hubble constant, ApJ **529**, 786. 1, 183

Narayan, R. & Bartelmann, M.: 1999, Gravitational lensing, *in* A. Dekel & J. P. Ostriker (eds), *Formation of Structure in the Universe, Proceedings of the 1995 Jerusalem Winter School*, p. 360. also available from `http://www.mpa-garching.mpg.de/Lenses/Preprints/JeruLect.html`. 17

Norbury, M. A., Rusin, D., Jackson, N. J. & Browne, I. W. A.: 2000, Limits on the core radii of JVAS/CLASS gravitational lenses, *Proceedings of a conference held in Granada, 17–20 of September*. 84

Parodi, B. R., Saha, A., Sandage, A. & Tammann, G. A.: 2000, Supernova type Ia luminosities, their dependence on second parameters, and the value of H_0, ApJ **540**, 634. 1, 183

Patnaik, A. R., Browne, I. W. A., King, L. J., Muxlow, T. W. B., Walsh, D. & Wilkinson, P. N.: 1992, 0218+357: The smallest separation lensed system, *in* Kayser, Schramm & Nieser (1992), p. 140. 70

Patnaik, A. R., Browne, I. W. A., King, L. J., Muxlow, T. W. B., Walsh, D. & Wilkinson, P. N.: 1993, B0218+35.7: A gravitationally lensed system with the smallest separation, MNRAS **261**, 435. 70, 71, 72, 73, 82

Patnaik, A. R. & Porcas, R. W.: 1999, Frequency dependent radio structure of the gravitational lens system B0218+357, *in* Carilli, Radford, Menten & Langston (1999), p. 247. 73

Patnaik, A. R., Porcas, R. W. & Browne, I. W. A.: 1995, VLBA observations of the gravitational lens system B0218+357, MNRAS **274**, L5. 70, 72, 73, 74, 75, 82, 83, 85

Pelt, J., Hoff, W., Kayser, R., Refsdal, S. & Schramm, T.: 1994, Time delay controversy on QSO 0957+561 not yet decided, A&A **286**, 775. 34

Pelt, J., Kayser, R., Refsdal, S. & Schramm, T.: 1996, The light curve and the time delay of QSO 0957+561, A&A **305**, 97. 34

Perley, R. A., Schwab, F. R. & Bridle, A. H. (eds): 1986, *Synthesis Imaging*. 93

Perley, R. A., Schwab, F. R. & Bridle, A. H. (eds): 1989, *Synthesis Imaging in Radio Astronomy*, ASP Conf. Ser. 6. 93, 232, 237

Porcas, R. W. & Patnaik, A. R.: 1996a, Multi-frequency VLBI observations of B0218+357, *in* Kochanek & Hewitt (1996), p. 311. 71, 88

Porcas, R. W. & Patnaik, A. R.: 1996b, Multi-frequency VLBI observations of the gravitational lens B0218+357, *in* Ekers, Fanti & Padrielli (1996), p. 115. 73

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P.: 1992, *Numerical Recipes in C. The Art of Scientific Computing*, 2nd edn, Cambridge University Press. 7, 30, 111, 134

Refsdal, S.: 1964, On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect, MNRAS **128**, 307. 1, 24

Refsdal, S.: 1966, On the possibility of testing cosmological theories from the gravitational lens effect, MNRAS **132**, 101. 24

Refsdal, S. & Surdej, J.: 1994, Gravitational lenses, *Rep. Prog. Phys.* **57**, 117. 17, 47

Remy, M., Claeskens, J.-F., Surdej, J., Hjorth, J., Refsdal, S., Wucknitz, O., Sørensen, A. N. & Grundahl, F.: 1998, Detection of the lensing galaxy for the double QSO HE 1104–1805, *New Astronomy* **3**, 379. 34

Saha, P.: 2000, Lensing degeneracies revisited, AJ **120**, 1654. 24

Saha, P. & Williams, L. L. R.: 1997, Non-parametric reconstruction of the galaxy lens in PG 1115+080, MNRAS **292**, 148. 27, 47, 63

Sandage, A.: 1999, Bias properties of extragalactic distance indicators. VIII. H_0 from distance-limited luminosity class and morphological type-specific luminosity functions for Sb, Sbc, and Sc galaxies calibrated using cepheids, ApJ **527**, 479. 1, 183

Scarpa, R., Urry, C. M., Falomo, R., Pesce, J. E., Webster, R., O'Dowd, M. & Treves, A.: 1999, The Hubble Space Telescope survey of BL Lacertae objects: Gravitational lens candidates and other unusual sources, ApJ **521**, 134. 35

Schechter, P. L., Bailyn, C. D., Barr, R., Barvainis, R., Becker, C. M., Bernstein, G. M., Blakeslee, J. P., Bus, S. J., Dressler, A., Falco, E. E., Fesen, R. A., Fischer, P., Gebhardt, K., Harmer, D., Hewitt, J. N., Hjorth, J., Hurt, T., Jaunsen, A. O., Mateo, M., Mehlert, D., Richstone, D. O., Sparke, L. S., Thorstensen, J. R., Tonry, J. L., Wegner, G., Willmarth, D. W. & Worthey, G.: 1997, The Quadruple Gravitational Lens PG 1115+080: Time Delays and Models, ApJ **475**, L85. 63

Schneider, P., Ehlers, J. & Falco, E. E.: 1992, *Gravitational Lenses*, Springer Verlag, A&A Library. 17, 21, 38

Schwab, F. R.: 1984, Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry, AJ **89**, 1076. 123

Schwarz, U. J.: 1978, Mathematical-statistical description of the iterative beam removing technique (method CLEAN), A&A **65**, 345. 105

Sereno, M., Covone, G., Piedipalumbo, E. & de Ritis, R.: 2001, Distances in inhomogeneous quintessence cosmology, MNRAS **327**, 517. 19

Sramek, R. A. & Schwab, F. R.: 1989, Imaging, *in* Perley et al. (1989), p. 83. 123

Tammann, G. A., Reindl, B., Thim, F., Saha, A. & Sandage, A.: 2002, Cepheids, supernovae, H_0, and the age of the universe, *in* Metcalfe & Shanks (2002). in print, also available as `astro-ph/0112489`. 183

Taylor, G. B., Carilli, C. L. & Perley, R. A. (eds): 1999, *Synthesis Imaging in Radio Astronomy II*, ASP Conf. Ser. 180. 93, 231, 232, 233

Thompson, A. R., Moran, J. M. & Swenson, G. W.: 1986, *Interferometry and Synthesis in Radio Astronomy*, Wiley-Interscience, New York. 93

Tonry, J. L.: 1998, Redshifts of the gravitational lenses B1422+231 and PG 1115+080, AJ **115**, 1. 64

Trotter, C. S., Winn, J. N. & Hewitt, J. N.: 2000, A multipole-Taylor expansion for the potential of the gravitational lens MG J0414+0534, ApJ **535**, 671. 32

Wallington, S., Kochanek, C. S. & Narayan, R.: 1996, LensMEM: A gravitational lens inversion algorithm using the Maximum Entropy Method, ApJ **465**, 64. 132

Wambsganß, J. & Paczyński, B.: 1994, Parameter degeneracy in models of the quadruple lens system Q2237+0305, AJ **108**, 1156. 47, 56, 60, 61

Wiklind, T. & Combes, F.: 1995, CO, HCO^+ and HCN absorption in the gravitational lens candidate B0218+357 at $z = 0.685$, A&A **299**, 382. 75

Williams, L. L. R. & Saha, P.: 2000, Pixelated lenses and H_0 from time-delay quasars, AJ **119**, 439. 27, 47

Wisotzki, L., Wucknitz, O., Lopez, S. & Sørensen, A. N.: 1998, First estimate of the time delay in HE 1104–1805, A&A **339**, L73. 34

Witt, H. J. & Mao, S.: 1997, Probing the structure of lensing galaxies with quadruple lenses: the effect of 'external' shear, MNRAS **291**, 211. 62

Witt, H. J. & Mao, S.: 2000, On the magnification relations in quadruple lenses: a moment approach, MNRAS **311**, 689. 68

Witt, H. J., Mao, S. & Keeton, C. R.: 2000, Analytic time delays and H_0 estimates for Gravitational lenses, ApJ **544**, 98. 47, 50, 57, 81

Witt, H. J., Mao, S. & Schechter, P. L.: 1995, On the universality of microlensing in quadruple gravitational lenses, ApJ **443**, 18. 47, 59

Wucknitz, O.: 1996, *Neuanalyse des Gravitationslinseneffektes unter besonderer Berücksichtigung kosmologischer Anwendungen*, Master's thesis, Hamburger Sternwarte, Universität Hamburg. in german, available from the author or from `http://www.astro.physik.uni-potsdam.de/~olaf/`. 17, 19, 22, 24, 30, 75, 221

Wucknitz, O.: 1997, Determining cosmological parameters from individual gravitational lens systems, *Proceedings of the 'Golden Lenses' workshop at Jodrell Bank*. available from the author or from `http://www.astro.physik.uni-potsdam.de/~olaf/`. 24

Wucknitz, O.: 2002a, Degeneracies and scaling relations in general power-law models for gravitational lenses, MNRAS **332**, 951. 47, 81

Wucknitz, O.: 2002b, LensCLEANing JVAS B0218+357 to determine H_0, *in* Metcalfe & Shanks (2002). in print, also available as `astro-ph/0112330`. 178

Wucknitz, O. & Refsdal, S.: 2001, Degeneracies of the radial mass profile in lens models, *in* T. G. Brainerd & C. S. Kochanek (eds), *"Gravitational Lensing: Recent Progress and Future Goals, ASP conf. series, Vol. 237"*, p. 157. 47, 56

Zhao, H. & Pronk, D.: 2001, Systematic uncertainties in gravitational lensing models: a semi-analytical study of PG1115+080, MNRAS **320**, 401. 63

Numbers at the end of each entry point to pages of this thesis where that work is cited. They are hyperlinks in the online-version. Most entries have links to the NASA Astrophysics Data System (ADS) at `http://adswww.harvard.edu/`. Abstracts and original articles can be retrieved from there. The `astro-ph` preprint links point to the e-print archive at `http://arXiv.org/`. External links may change in the future and are not kept up to date.

Acknowledgments

I would like to thank my supervisor Prof. Sjur Refsdal for the possibility to work in his group and for the complete freedom I had for my scientific projects. In numerous discussions he often surprised me by finding simple explanations for seemingly complicated problems. It may be this ability to separate the important and often simple core of a problem from irrelevant details, which makes up a great scientist.

Most of the work for this thesis has been done at the Hamburger Sternwarte. I want to thank all colleagues there who either contributed in a scientific way or by being around and keeping a good temper. This includes without exception all former and present members of the lensing group, especially Oliver Czoske, Achim v. Linde, Jochen Schramm, Bernd Neindorf and of course the chief 'Old Sjurhand' himself.

Besides the incredibly shrinking lensing group, many other persons helped to make the time spent at the institute a good one. I only want to mention a small number of representative friends and colleagues, being well aware of the fact that the list is highly incomplete. Hans Hagen and Dieter Engels are part of the inventory of the Observatory and have always been very helpful in many respects. Ana de la Varga proved that it is possible to finish a thesis even under less than easy circumstances, giving hope to all of us following her example. Volker-'Dr.'-Beckmann has always been a nice fellow to be with. Our observations on Calar Alto were very entertaining, although MOSCA was not. Dirk H. Lorenzen ('the press') helped in the final phase of this thesis by pretending to be convinced that I am doing a good job. Furthermore I want to thank Marc Hempel, Ansgar Reiners and Uwe Wolter for a number of serious conversations and (not to be forgotten) for the biscuits. Cora Fechner and Eckart Janknecht shall be named as representatives for all office mates I had during the last few years, although I was away most of the time since they moved into the Balcony Suite.

The colleagues and friends abroad, especially Olli Czoske and Uli Sperhake, were always available for questions and discussions about everything. Maybe we should have continued our meetings on general relativity as an e-seminar.

I worked at the Jodrell Bank Observatory (University of Manchester) for a few months in 1999 (then still called the Nuffield Radio Astronomy Laboratories) and for half a year in 2001. I am greatly indebted to all colleagues there for help in all respects. Especially Ian Browne, who made the CERES project (and CERES lunches at the Crown!) possible, and Neal Jackson, who was responsible for the funding of my second stay, deserve many thanks. The collaboration with Andy Biggs was also very important for my work. He taught me the practical reduction and analysis of radio data, which was of high value for my work on LENSCLEAN. Close cooperation with experienced and enthusiastic observers is extremely important for theoretical work, especially if the latter is somehow related to real data.

Shude Mao stressed the importance of my concept of a critical shear. This and the very helpful and constructive comments from Prasenjit Saha, who was the referee for that article, raised my motivation to find the simple geometrical interpretation. Discussions with Ester Piedi-

palumbo initiated my work on the subject of analytical quadruple lenses. I also thank G. A. Tammann for pointing out that my B0218+357 result for H_0 cannot be correct and that the exponent β should be calculated from my equations in combination with his results on H_0.

Too many people to mention them all made my visits in Jodrell a very nice experience. Without Martin Norbury and our shopping tours, I would have starved most of the time. Ian Browne deserves to be mentioned again, but I have been told to avoid the word ‘beer’ in a PhD thesis. I also want to thank (and apologize to) Anita Richards for the bike. May it rest in pieces.

There have also been bad times during this work and I want to thank everybody who helped when help was needed. Without the support of Sjur Refsdal, Hans Hagen, Thomas Berghöfer, everybody from Jodrell and especially Dieter Engels, it would have been even more difficult to defend against slanderous and absurd accusations. Without their help, I would not have been able to continue and finish my work.

Finally, I want to thank my family for support and sympathy for my morbid interest in astronomy and Ines Drescher for being there and for everything else.

This work was partially funded by

- Deutsche Forschungsgemeinschaft, reference no. Re 439/26–1 and 439/26–4,
- European Commission, Marie Curie Training Site programme, under contract no. HPMT-CT-2000-00069,
- European Commission, TMR Programme, Research Network Contract ERBFMRXCT96-0034 ‘CERES’.

Support from the Bundesanstalt für Arbeit is also appreciated. My work would have been much more difficult without the freely available software packages AIPS, CFITSIO, DIFMAP, EMACS, GNUPLOT, IRAF, (LA)TEX, PGPLOT and others. I thank everybody who contributed to these great tools. I pay my debt to society with the development of NOLO.

(Hamburg/Berlin/Potsdam, April 2002)