

9

Model fitting and image reconstruction

The OIFITS files produced at the end of the data-reduction process can be used to provide information about the object under study. There are two ways in which this information can be extracted, either in terms of parameters of a relatively simple model, or as a model-independent image. These two forms of information are related to one another and are often used in tandem. This chapter discusses the process of deriving these end products of an interferometric observation.

9.1 Bayesian inference

In Chapter 8 the overall inverse problem of interferometry was presented as the problem of determining the parameters of a model of the object being observed from the measured data values. The data-reduction process presented in that chapter does not fundamentally change that problem: the object model parameters still need to be determined, but, as the name implies, the data-reduction process reduces the volume of data that need to be considered.

Indeed, the remaining problem is not one of data reduction but of data interpretation, as the number of data points produced by the data-reduction process may be comparable to or even less than the number of model parameters. This kind of under-constrained problem is where a form of inference known as Bayesian inference is at its best.

Bayesian inference acknowledges the non-uniqueness of the model parameters in the majority of inverse problems. Instead of selecting a single model (for example a parameterised model together with a single set of values for the model parameters) which could have produced the data, it acknowledges that there may be many possible models and assigns a *probability* P between

0 and 1 to all of them in parallel. The probability assigned to a given model is used to denote the plausibility of the model based on all the available evidence, with 0 denoting impossible and 1 denoting certain. Bayesian probability extends Boolean logic, which can only deal with statements that are either true or false, to allow statements about which models are more likely to be true and which are more likely to be false.

Bayes' theorem is the fundamental 'engine' of Bayesian inference. It combines statements about previous knowledge with statements about experimental data to allow new statements to be made about the best models to explain the data. Bayes' theorem states that, given a set of data D , the probability $P(M|D)$ (the *posterior probability* or *posterior* for short) of a particular model M is given by

$$P(M|D) = \frac{P(M)P(D|M)}{P(D)}. \quad (9.1)$$

This theorem expresses the logic behind the scientific method in a mathematical formalism. It says that the plausibility of any given model after performing an experiment is a combination of the plausibility of the model before the experiment was performed $P(M)$ (the *prior probability* or *prior*) with an experimental update $P(D|M)$, known as the *likelihood* of the data, which expresses how well the data fits with the predictions of the model. The denominator $P(D)$ is a normalising factor, which ensures that the posterior probability sums to unity; in more advanced Bayesian inference it is called the 'Bayesian evidence' and can be used to distinguish between different classes of model, but it is not considered further here.

A full treatment of Bayesian inference in data analysis can be derived from a number of excellent introductions to the topic, for example that by Sivia and Skilling (2006). The following sections consider the likelihood and the prior probability in the context of interferometry, and show how these can be used together with Bayes' theorem for both model-fitting and imaging.

9.2 The interferometric likelihood

The likelihood $P(D|M)$ expresses how often a given model M will produce a given set of data D . In a perfect experiment, a single model can only produce a single possible set of measurements so $P(D|M)$ would be either unity if the data agree with the model or zero if the data disagree. In this case, all models which could not produce the observed data automatically de-select themselves via Bayes' theorem, as the posterior probability for such models would be zero.

In practice, all realistic models must include the effects of measurement noise, and so the agreement between the data and a given model is only probabilistic – a given model could possibly produce many different sets of measured values, some of these sets being more likely than others.

Determining the likelihood requires a process for producing the data from the model, in other words the forward problem discussed in Section 8.2. The forward problem introduced there starts with the model of the object with a set of parameters $\{\Theta_m\}$ and ends with the individual pixel intensity measurements in the fringe pattern $\{r_p\}$. The forward problem considered in this chapter is an abbreviated one as shown in Figure 9.1. This problem takes as its end point the OIFITS file produced by the data-reduction process described in Chapter 8. In other words, the data reduction moves the ‘data’ of the forward problem ‘backwards’ from a set of pixel intensities to an OIFITS file containing a set of calibrated averaged power-spectrum and bispectrum data, and/or a set of calibrated coherently averaged visibilities. The revised forward problem starts from the object model to produce object coherent fluxes $F(\mathbf{u}_{ij})$ as before. In order to cover the general case where the data may contain coherently integrated visibilities and not just bispectrum and power-spectrum data, the measurement model incorporates a set of antenna-based phase distortions $\{\epsilon_1, \dots, \epsilon_M\}$. These phase distortions are so-called ‘nuisance parameters’: parameters of the forward problem whose value we are not interested in but must be included in the measurement model in order to explain the data. In the full forward problem the set of nuisance parameters includes a complete description of the atmospheric wavefront distortions for every exposure; in this ‘data-reduced’ model there is only one set of piston phases for each coherently averaged data point, so the number of such nuisance parameters is greatly reduced.

The phase distortions combine with the object coherent fluxes to yield measured coherent fluxes

$$F_{ij} = e^{i[\epsilon_i - \epsilon_j]} F(\mathbf{u}_{ij}). \quad (9.2)$$

Note that the system visibility in this model is unity as any system-visibility variations are assumed to have been removed in the calibration process. The data are processed to produce power-spectrum and bispectrum data. The phase errors have no effect on these data, and so do not need to be included in the model if coherently averaged data are not present in the OIFITS file. Finally, the noiseless data are corrupted by additive Gaussian noise to produce the noisy but calibrated data in the OIFITS file. Importantly, it is assumed that

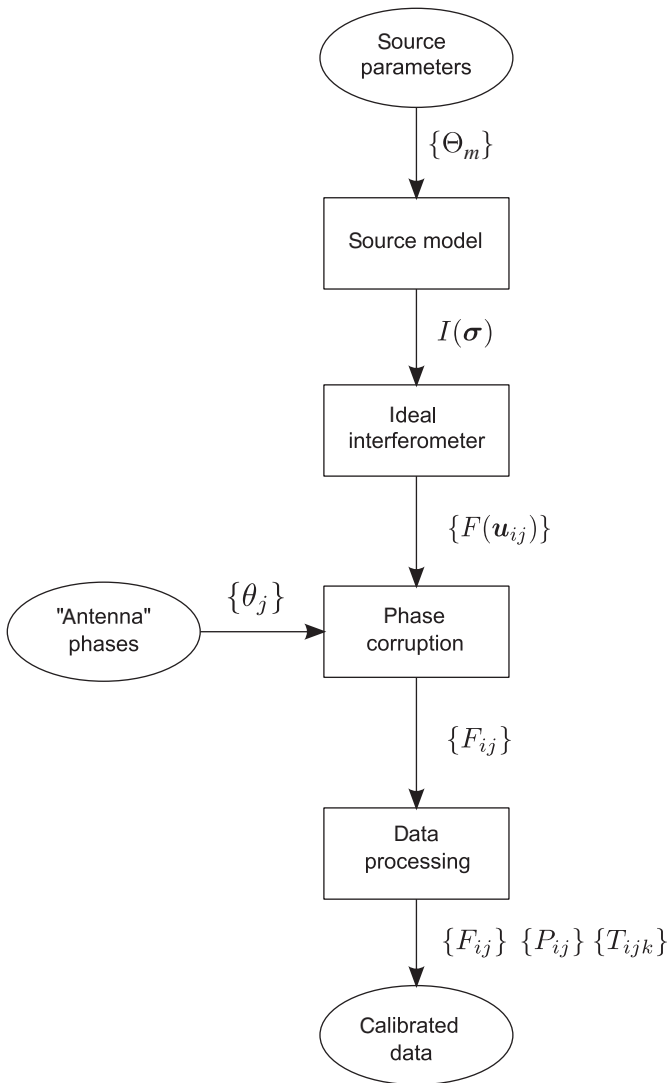


Figure 9.1 The abbreviated forward problem for model fitting.

the noise is uncorrelated between data samples, which is not always the case, as discussed in Section 9.6.4.

The likelihood of the data for a given model can be straightforwardly derived from the noiseless data if the noise is assumed to be Gaussian and

uncorrelated between the data points. If the noiseless data, whether bispectrum, power-spectrum and/or coherently averaged visibility data are denoted by the set $\{\mu_1, \dots, \mu_N\}$, then the likelihood of a given set of measured data $D = \{d_1, \dots, d_N\}$ will be given by

$$P(D|M) \propto \prod_{i=1}^N e^{-\frac{|d_i - \mu_i|^2}{2\sigma_i^2}}, \quad (9.3)$$

where σ_i is the standard deviation of the noise added to μ_i . This can be written as

$$P(D|M) \propto e^{-\chi^2}, \quad (9.4)$$

where χ^2 is the weighted sum of squared deviations:

$$\chi^2 = \sum_{i=1}^N \frac{|d_i - \mu_i|^2}{2\sigma_i^2}. \quad (9.5)$$

Thus, the maximum of the likelihood occurs for models which minimise the ‘distance’ between the noiseless data predicted by the model and the actual data, i.e. at the minimum of χ^2 . Choosing a model which minimises χ^2 is equivalent to least-squares fitting of models to data, and is a form of *maximum likelihood* estimation.

9.3 Model-dependent priors

Bayesian inference extends the ideas of maximum-likelihood estimation by considering the probabilities of the whole model space, not just the single point where the probability is maximised, and also by considering the prior probability as well as the likelihood. The prior depends, as might be expected from the name, on what prior information exists at the time of making the measurements. A specific example serves to illustrate the general process of constructing the prior.

In this example the object under study is a binary star system that is too small to be resolved by conventional telescopes. One possible model for this system is a physical model which has an orbit with a semi-major axis, eccentricity, inclination and so forth, and has masses, ages and metallicities of the constituent stars. This model would be appropriate for fitting to the data from observations at multiple epochs over the orbit of the binary, complemented with spectroscopic and photometric data.

When analysing data from a single epoch on its own, a model with the smallest number of parameters needed to model the expected brightness

distribution on the sky might be easier to work with. One such model might consist of a pair of stars, modelled as uniform disks with fluxes at the wavelength of observation of I_A and I_B and angular diameters θ_A and θ_B , respectively. The angular separation θ_{AB} of the pair of stars and the orientation ϕ_{AB} of this separation vector on the sky completes the set of model parameters.

Often the prior probability can be factorised into a set of independent probabilities

$$P(I_A, I_B, \theta_A, \theta_B, \theta_{AB}, \phi_{AB}) = P(I_A)P(I_B)P(\theta_A)P(\theta_B)P(\theta_{AB})P(\phi_{AB}) \quad (9.6)$$

but this is not always the case. For instance, if the total flux from the system $I_{\text{total}} = I_A + I_B$ is known a priori from single-telescope measurements then I_A and I_B are not independent variables. Instead, the prior probability distribution concentrates along a line in the (I_A, I_B) plane given by $I_A = I_{\text{total}} - I_B$.

Assuming that the total flux is not known and the rest of the prior is also factorisable, the problem of determining the model prior reduces to one of determining priors for each of the variables independently. The role of the prior is easiest to understand when it is used to exclude regions of model space which are unfeasible. For example, based on centuries of research we know that it is impossible (or at least extremely unlikely) for a star to emit negative flux so models with negative I_A or I_B would have zero prior probability.

Assigning values to the prior outside of the impossible regions of parameter space is less straightforward. A simple solution occurs in the case of maximum lack of knowledge: if pre-existing knowledge provides no reason to prefer any one from a number of different models, then all these models are assigned equal probabilities – the so-called *uniform prior*. In the absence of any information about the orientation ϕ_{AB} of the separation vector of the binary pair on the sky, any orientation is equally likely. Thus, a uniform prior over the range $-\pi$ to π can safely be chosen. For the same reason, uniform priors are good choices for the nuisance parameters ϵ_i in the measurement model.

It may be possible to constrain the separation of the binary pair θ_{AB} to be in some given range $\theta_0 \pm \sigma$ if there is some other information available, such as an approximate distance to the system and an orbital period from spectroscopy. In this case an appropriate prior for θ_{AB} might be a Gaussian with a mean of θ_0 and a standard deviation of σ .

If there is no observational evidence to favour any particular scale for θ_{AB} versus a scale which is ten times smaller or ten times larger than this then an appropriate prior would be one which is uniform in the logarithm of θ_{AB} . In other words, the prior $P(\theta_{AB}) \propto 1/\theta_{AB}$ would be appropriate (this is known in Bayesian inference as the ‘uninformed prior for a scale parameter’).

To avoid infinities in this distribution it is necessary to choose some upper and lower bounds for θ_{AB} outside which the probability tends rapidly towards zero. The upper bound could be the angular separation at which the binary pair would be resolvable with a single telescope while the lower bound could be the angular separation at which the stellar surfaces would be touching for any reasonable choice of system parameters.

Similar log-uniform priors could be used for the fluxes and angular diameters as these are also *scale parameters* – a key attribute these share is that they always take a value which is greater than zero. The exact choice of the cutoffs in such priors is usually unimportant, since if there are a reasonable amount of data then the likelihood will usually overpower these choices in determining the posterior.

9.4 Model fitting in practice

In principle, Bayesian solution of the inverse problem consists simply of computing the posterior probability using Equation (9.1) for every point in the model space. The resulting distribution can be summarised in terms of the location and extent of the region of highest probability. To exhaustively cover all of the multi-dimensional model space would take near-infinite amounts of computer time, but there exist computational techniques for minimising the time spent in regions of model space with low probability and interpolating the probability in higher-probability regions. The details of these techniques can be ignored for the purposes of this discussion and the problem can be treated as though we were in possession of an infinitely large computer, which can compute the probabilities for all points in parallel.

Model-fitting software is often purpose-written for a given astrophysical model, so that all of the physics of the problem can be incorporated. There exist also more general-purpose programs, for example LITpro (Tallon-Bosc *et al.*, 2008) and MFIT (see the supplementary online material referenced in Appendix B), which fit geometrical models. These models can consist of unresolved point sources, disks, shells and so forth.

Model-fitting programs read in calibrated data from an OIFITS file and typically perform either a least-squares fit to the data or find the location of the maximum posterior probability. The output therefore typically consists of a single-point estimate of the model parameters, and some indication of the region in model space around which χ^2 is within some bounds.

Using such programs is relatively straightforward providing only simple models are used. The main difficulty that can arise is if the posterior probability

has multiple local maxima, which can cause problems with many optimisation programs, as they can become ‘trapped’ in these minima. To diagnose this situation it is often useful to look at plots of the actual data against the noiseless data predicted by the model, to see if there are obvious regions of systematic misfit. Discarding the data in these regions or downweighting these and restarting the model fit can give insight into whether the data or the model is at fault. Another technique is to start the model fit from a number of different starting models to see if the solution converges to the same answer every time.

9.5 Model-independent priors

A Bayesian inference framework requires a model for the object being observed. If the physics of the object being studied is well understood, this model can be a relatively simple model with few parameters. Such models have the advantage that relatively little data are needed to adequately constrain the model parameters. Having fitted this model to the data, the best-fit model can be displayed as an image, hence the term *model-dependent imaging*.

The opposite end of the spectrum occurs when relatively little is known a priori about the object under study, such that a whole host of quite different physical and/or geometric models are consistent with existing knowledge. One approach in this situation would be to construct a set of parameterised models for all of these possibilities and then see which of these models the interferometric data are consistent with. Providing an exhaustive set of such models may prove infeasible and so an alternative approach is to try to reconstruct an image which is independent of any specific model of the object, just as would be possible in conventional imaging using a camera.

In Bayesian terms this can be expressed as a ‘model’ which consists of a number of pixels with unknown brightnesses. This model has the advantage of generality, but the downside is that the number of model parameters is then as large as the number of pixels, which could easily be in the millions. This poses two problems. There is a computational problem of dealing with models of this large dimensionality, but methods of dealing with this are relatively well understood. A more serious problem is that the model is likely to be underconstrained by the data. The number of independent coherent flux measurements in a typical optical interferometric experiment is typically less than 1000 or so, and so cannot possibly constrain a million-pixel image on its own.

Additional constraints are therefore needed in order to ‘regularise’ the image reconstruction process. In Bayesian image reconstruction, these constraints are embedded in the prior, and this form of regularisation will be assumed in the

following discussion. Even given this regularisation, the data need to be not only numerous but sufficiently diverse to avoid degeneracies in the image, that is to say to avoid the situation where many significantly different images fit the data. The following sections look at the interaction of the constraints from the data with the constraints from the a priori information in a quantitative (but not rigorous) fashion, before discussing how model-independent image reconstruction is achieved in practice.

9.6 Preconditions for imaging

9.6.1 Sampling the (u, v) plane

In Section 2.3 it was shown that reconstruction of an $N \times N$ pixel image was possible in principle if the object visibility function is sampled on an $N \times N$ square grid in the (u, v) plane. In practice, visibility samples are never taken on a regular grid, as even if the telescopes are arranged on the ground in a regular grid, baseline projection effects would distort the grid differently for different objects and for different times of night.

Most interferometers have only a few telescopes and so can only sample the visibility function at a few (u, v) locations at a time. Earth rotation can be used to ‘sweep out’ ellipses on the (u, v) plane, and, if enough time is available, different array configurations can be used on different nights. Even so, the resulting coverage will still be quite uneven as shown in Figure 9.2.

The question then becomes whether, given some quasi-random sampling in the (u, v) plane, there is adequate information to reconstruct a reliable image of the target. One way to address the problem is via the idea of ‘invisible distributions’.

An interferometer samples the object visibility function at a set of discrete locations $\{\mathbf{u}_i\}$ in the (u, v) plane. At all other locations the visibility function is unknown. Assuming there are no atmospheric phase errors and no other sources of noise affecting the measurements, the coherent flux $F(\mathbf{u})$ corresponding to the true brightness distribution $I(\boldsymbol{\sigma})$ will clearly fit the measured data exactly. However, there is an infinite set of other object brightness distributions, which will also fit the data exactly. These will be brightness distributions that are the sum of the true object brightness distribution and any distribution $I_{\text{invis}}(\boldsymbol{\sigma})$ which has the property that

$$F_{\text{invis}}(\mathbf{u}_i) = \mathcal{F}\{I_{\text{invis}}(\boldsymbol{\sigma})\}(\mathbf{u}_i) = 0 \quad (9.7)$$

for all the (u, v) points at which the data have been measured, but is non-zero elsewhere. The brightness distribution I_{invis} is known as an invisible

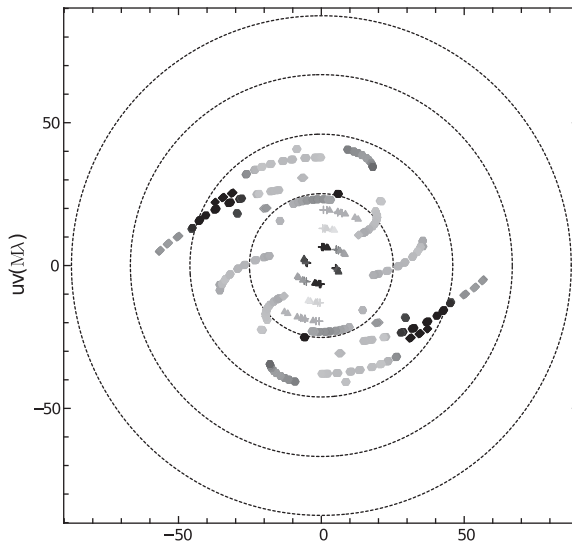


Figure 9.2 The (u, v) plane coverage of observations of VY CMa with the VLTI and the PIONIER instrument. From Monnier *et al.* (2014).

distribution because an object with this brightness distribution would be invisible to an interferometer with the (u, v) coverage given by $\{u_i\}$.

The existence this degeneracy in the image-reconstruction process may make the possibilities for image reconstruction look hopeless. However, there exist a-priori constraints on the invisible distributions, which can be used to ‘regularise’ the problem. The first of these is image positivity. It is known that no object can emit light with a negative intensity and so all plausible images must be non-negative everywhere. A large fraction of the set of invisible distributions will have negative regions, and if these correspond to regions where the true target is zero, this will lead to an unphysical image.

A second constraint is the finite angular extent of the target (also known as *finite support*). Most astronomical targets can be assumed to consist of a comparatively bright target surrounded by a black background. The angular extent of the illuminated region of the target (the ‘support’ of the target) can be estimated for many objects from other information about the object. For example, the temperature of an object in thermal equilibrium with the radiation from a star will fall with distance from the star, and so at infrared wavelengths this can be used to estimate the maximum extent of thermal emission from objects in the star’s environment. In addition, images of the object taken with single telescopes serve to provide an upper limit to the angular extent.

This constrains the invisible distributions to being ones which do not have any light emission outside a given region centred on the target. This in turn can be used to constrain the invisible distributions. An invisible distribution with a finite angular extent θ_d can be modelled as

$$I_{\text{invis}}(\sigma) = g(\sigma) \text{rect}(|\sigma|/\theta_d), \quad (9.8)$$

where $g(\sigma)$ is a function whose extent is not limited and $\text{rect}(x)$ is the unit ‘top-hat’ function defined in Equation (1.58), which limits the support of the invisible distribution. The coherent flux will therefore be given by

$$F_{\text{invis}}(\mathbf{u}) = G(\mathbf{u}) * \text{jinc}(\pi|\mathbf{u}|\theta_d), \quad (9.9)$$

where $G(\mathbf{u}) = \mathcal{F}\{g(\sigma)\}$. The effect of the convolution with the jinc function is to ‘smooth’ the Fourier-plane function G over a region approximately $1.22/\theta_d$ in size, and so the values of $F_{\text{invis}}(\mathbf{u})$ will be correlated over approximately this distance. The Fourier transform of the invisible functions must be zero at the sample locations $\{\mathbf{u}_i\}$, and this implies that F_{invis} will have a restricted magnitude over a set of circular regions of order $1/\theta_d$ in radius around the sample points.

If the sampling is such that these regions around the sample points cover most of the (u, v) plane out to some maximum spatial frequency then the invisible distribution will have ‘nowhere to hide’. Simulations confirm this general rule that images can be reconstructed from a relatively arbitrary sampling, provided that the samples are spaced at roughly $1/\theta_d$ intervals. In addition, these simulations show that metrics for the quality of image reconstructions from a given (u, v) sampling are typically related to the size of the largest ‘hole’ in sampling, i. e. the size of the largest region where the Fourier transforms of the invisible distributions are unconstrained.

9.6.2 Phase information

The image reconstruction results quoted above were assuming a phase-stable interferometer, i. e. ones for which both the visibility modulus and phase can be measured. In many early optical interferometers, beams from only two telescopes could be combined at any one time, so only the visibility modulus and not the closure phase could be measured. The question arose whether with sufficient modulus information images could be reconstructed. The answer seems to be that so-called ‘phaseless imaging’ is possible with high-quality data and/or simple scenes to image, but image reconstruction from phaseless data with scenes of unknown complexity and the levels of systematic errors, random errors and missing data encountered in real life has proven elusive.

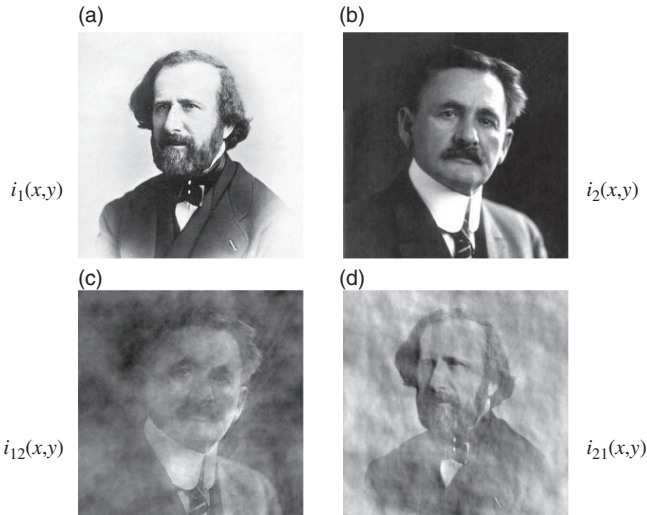


Figure 9.3 Upper row: images of two founders of interferometry – H. Fizeau (a) and A. A. Michelson (b). Lower row: hybrid images made from combining the Fourier amplitudes of Fizeau and the Fourier phases of Michelson (c) and from combining the Fourier phases of Fizeau and the Fourier amplitudes of Michelson (d).

This indicates that the phase contains information which is vital for imaging. A simple illustration of this can be derived with the aid of a numerical experiment. In this experiment we take two images such as the upper two images in Figure 9.3 and take their Fourier transforms. If the two image intensities are $i_1(x, y)$ and $i_2(x, y)$, respectively, we can denote the Fourier transforms as

$$A_1(u, v)e^{i\phi_1(u, v)} = \mathcal{F}\{i_1(x, y)\}$$

and

$$A_2(u, v)e^{i\phi_2(u, v)} = \mathcal{F}\{i_2(x, y)\}.$$

We can then make a hybrid of the two images by combining the Fourier phases from one image with the corresponding Fourier amplitudes from the second and taking the inverse Fourier transform:

$$i_{12}(x, y) = \mathcal{F}^{-1}\{A_1(u, v)e^{i\phi_2(u, v)}\}.$$

We can also make the complementary pairing i_{21} , which has the amplitudes from i_2 and the phases from i_1 . The surprising result can be seen in Figure 9.3, namely that the hybrid images strongly resemble the image from which their phase information is derived and bear no resemblance to the image from which

Table 9.1 *The fraction of the phase information captured by the closure phases in an array of M telescopes; this is the ratio of the number of independent closure phases N_{clos} to the number of independent object phases $N_{\phi} = M(M - 1)/2$ measured by a phase-stable interferometer with the same number of telescopes. The number of independent closure phases is given by $N_{\phi} - N_{\epsilon}$ where $N_{\epsilon} = M - 1$ is the number of independent antenna-based errors.*

M	N_{ϕ}	N_{ϵ}	N_{closure}	$N_{\text{closure}}/N_{\phi}$
2	1	1	0	0.00
3	3	2	1	0.33
4	6	3	3	0.50
5	10	4	6	0.60
10	45	9	36	0.80
15	105	14	91	0.87
20	190	19	171	0.90

their amplitude information is derived. This indicates that most of the useful information in an image is contained in its Fourier phases and not the Fourier amplitudes: the amplitudes give information about how *much* structure is in an image, but the phases tell us *where* that structure is.

9.6.3 Information loss in the closure phase

While the closure phase is dependent only on the object visibility phase and not on the atmospheric phase disturbance, it is not immediately clear how to make use of the object phase information contained in the closure phase. It is clear that because of the ‘antenna-based’ phase errors, some information has been lost compared to measuring the phases directly.

The amount of information loss can be estimated by comparing the number of independent closure phase measurements available from an array of M telescopes to the number of phase measurements from the equivalent phase-stable interferometer. The larger the number M of telescopes in the array, the more efficient the closure phase is at capturing object phase information, because the number of object phases rises as $M(M - 1)/2$ while the number of atmospheric phase unknowns rises as $N_{\epsilon} = M - 1$. It can be seen from Table 9.1 that the fraction of the object phase information which is retained in the closure phases tends rapidly towards unity as the number of telescopes increases.

One piece of information which cannot be recovered from closure phases is information about the absolute position of an object. This can be shown

by considering independent observations of an arbitrary object brightness distribution $I(\boldsymbol{\sigma})$ and the same distribution displaced by some vector offset $\boldsymbol{\sigma}_1$, given by

$$I'(\boldsymbol{\sigma}) = I(\boldsymbol{\sigma} - \boldsymbol{\sigma}_1) = I(\boldsymbol{\sigma}) * \delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}_1), \quad (9.10)$$

where the $*$ operator denotes convolution. From the convolution theorem for Fourier transforms, the corresponding object visibility functions $V(\mathbf{u})$ and $V'(\mathbf{u})$ are related through

$$V'(\mathbf{u}) = V(\mathbf{u})e^{-2\pi i \boldsymbol{\sigma}_1 \cdot \mathbf{u}}. \quad (9.11)$$

As a result, the phases of the visibility functions will be related to one another via

$$\phi'(\mathbf{u}) = \phi(\mathbf{u}) - 2\pi i \boldsymbol{\sigma}_1 \cdot \mathbf{u}. \quad (9.12)$$

The measured closure phases on a triangle of telescopes denoted by p, q and r are therefore related by

$$\begin{aligned} \Phi'_{pqr} &= \phi_{pq} - 2\pi \boldsymbol{\sigma}_1 \cdot \mathbf{u}_{pq} + \phi_{qr} - 2\pi \boldsymbol{\sigma}_1 \cdot \mathbf{u}_{qr} + \phi_{rp} - 2\pi \boldsymbol{\sigma}_1 \cdot \mathbf{u}_{rp} \\ &= \Phi_{pqr} - 2\pi \boldsymbol{\sigma}_1 (\mathbf{u}_{pq} + \mathbf{u}_{qr} + \mathbf{u}_{rp}), \end{aligned} \quad (9.13)$$

where \mathbf{u}_{pq} is the spatial frequency corresponding to the baseline between telescopes p and q . For a closed triangle of baselines, $\mathbf{u}_{pq} + \mathbf{u}_{qr} + \mathbf{u}_{rp} = 0$ and so $\Phi_{pqr} = \Phi'_{pqr}$. Thus, all the closure phases measured on a given object are the same as those measured on the same object displaced, and so the absolute position of the object cannot be measured using the closure phase alone.

9.6.4 Using all the closure phases

For an array of M telescopes, there are $\frac{1}{6}M(M-1)(M-2)$ different baseline triangles, but the number of linearly independent *object* closure phases (i.e the number of constraints on the phase of the object Fourier transform) that can be measured is only $\frac{1}{2}(M-1)(M-2)$. In other words, in a noise-free instrument there would be no point in computing all the possible closure phases since they could all be derived from a linearly independent subset. In the presence of noise, however, it would be worthwhile computing the full set of closure phases if the errors on the different closure phases were uncorrelated, since we would then gain in the number of independent constraints on the object closure phases.

The correlation of two closure phases is the ratio of their covariance to the geometric mean of their variances. The covariance of the closure phases can be calculated from the covariance of the noise on the bispectrum values resolved

along lines in the complex plane, which are perpendicular to their respective mean vectors. This calculation can be broken down into the calculations of two covariances:

$$\text{covar}_1[A, B] \equiv E(AB^*) - E(A)E(B^*) \quad (9.14)$$

and

$$\text{covar}_2[A, B] \equiv E(AB) - E(A)E(B), \quad (9.15)$$

where $E(X)$ denotes taking the expectation (mean) of the quantity X . The covariance along any two directions can then be calculated from

$$\begin{aligned} \text{covar}[A, B, \theta_A, \theta_B] = & \frac{1}{2} \text{Re}\{\text{covar}_1[A, B]e^{-i(\theta_A - \theta_B)} \\ & + \text{covar}_2[A, B]e^{-i(\theta_A + \theta_B)}\}, \end{aligned} \quad (9.16)$$

where $\text{covar}[A, B, \theta_A, \theta_B]$ is the covariance of two complex variables A and B along lines in the complex plane at angles of θ_A and θ_B to the real axis.

The calculations proceed along the same lines as that for computing the variance of the bispectrum in Section 5.5.3. The covariance is zero if the closure triangles do not share a common baseline, but if they do share a common baseline the closure phase covariance is given by

$$\frac{\text{covar}[T_{123}, T_{234}, \phi_{123} + \pi/2, \phi_{234} + \pi/2]}{|F_{12}F_{23}^2F_{31}F_{34}F_{42}|} = \gamma_{23}^{-2},$$

where γ_{23} is the SNR of the coherent flux on the shared baseline. The correlation coefficient is therefore

$$\begin{aligned} \mu_{123,234} &\equiv \frac{\text{covar}[T_{123}, T_{234}, \phi_{123} + \pi/2, \phi_{234} + \pi/2]}{(\text{var}[T_{123}, \phi_{123} + \pi/2]\text{var}[T_{234}, \phi_{234} + \pi/2])^{1/2}} \\ &= \frac{1}{3 + 3\gamma^{-2} + \gamma^{-4}}, \end{aligned}$$

where it has been assumed for simplicity that all the signal-to-noise ratios (SNRs) are the same, i.e. $\gamma_{ij} = \gamma$. Hence,

$$\mu_{123,234} \simeq \begin{cases} 1/3 & \text{at high SNRs} \\ \gamma^4 \simeq 0 & \text{at low SNRs} \end{cases}.$$

Thus, for data where the SNR of the single-exposure coherent flux is high, the noise on the closure phases is strongly correlated and therefore the full set of closure phases contains no new information compared to a linearly independent subset. When the SNR on each exposure is low, however, the noise on the closure phases becomes decorrelated and so using the full set of closure phases will provide a larger number of independent constraints on the object phase.

The above result holds for the average of many bispectrum measurements: for any one exposure, the closure-phase errors cannot all be independent from one another as they arise from the same set of phases. The resolution of this apparent contradiction is that a lack of *correlation* only implies *independence* when the variables involved have Gaussian distributions, and the noise on the bispectrum values in a single exposure do not have Gaussian distributions. The central limit theorem means that the average of many per-exposure bispectrum data results in Gaussian noise and therefore independent data.

9.6.5 Phase reconstruction from the closure phase

If the closure phase can be used to reconstruct the individual object visibility phases, then the equivalence between phase information and closure-phase information is clear. This can be done if the configuration of collectors in the interferometer is appropriately chosen. One such configuration is the linear chain of telescopes such as shown in the lower half of Figure 2.5. If the projected spacing between adjacent telescopes is given by $\mathbf{u}_0\lambda$ then the closure phase measured on the linear ‘triangle’ between telescopes 1, 2, and 3 will be given by

$$\Phi_{123} = \phi(\mathbf{u}_0) + \phi(\mathbf{u}_0) - \phi(2\mathbf{u}_0), \quad (9.17)$$

where $\phi(\mathbf{u})$ is the phase of the object visibility function at spatial frequency \mathbf{u} . Thus, if $\phi(\mathbf{u}_0)$ is set to some arbitrary value, for example zero (this is equivalent to choosing an arbitrary position for the object, which as shown above cannot be determined from the closure phases), then the value of $\phi(2\mathbf{u}_0)$ can be solved for using

$$\phi(2\mathbf{u}_0) = 2\phi(\mathbf{u}_0) - \Phi_{123}. \quad (9.18)$$

Similarly, the value of closure phase on the next-larger triangle is given by

$$\Phi_{134} = \phi(\mathbf{u}_0) + \phi(2\mathbf{u}_0) - \phi(3\mathbf{u}_0) \quad (9.19)$$

and so the value of $\phi(3\mathbf{u}_0)$ can be solved for, giving

$$\phi(3\mathbf{u}_0) = 3\phi(\mathbf{u}_0) - \Phi_{123} - \Phi_{134}. \quad (9.20)$$

For an array consisting of a longer chain of telescopes, this process can be repeated recursively to determine the phases $\phi(4\mathbf{u}_0)$, $\phi(5\mathbf{u}_0)$ and so on, up to the phase for the maximum baseline measured by the array. These phases can then be combined with the visibility moduli measured on the corresponding baselines to form a set of estimates of the visibility equivalent to those measured by a phase-stable interferometer.

This procedure, known as *redundant-spacing calibration* shows that closure-phase measurements can act as an effective substitute for phase measurements. However, it does require a special array configuration: as the name suggests, arrays which allow this kind of phase reconstruction must have a substantial degree of baseline redundancy; that is, there are repeated measurements of the visibility of the same spatial frequency by different pairs of telescopes. This means that the number of distinct spatial frequencies measured by the interferometer is less than that measured by a non-redundant interferometer where all telescope pairs measure a unique spatial frequency.

9.7 Model-independent image reconstruction

Given sufficient (u, v) coverage and phase information, it would seem tempting to reconstruct images by continuing with the ‘estimator’ style of data reduction, reconstructing the image from the calibrated Fourier data through a series of steps. Such processes have been tried, using the closure phase to reconstruct a set of phases and combining these with the visibilities estimated from the power spectrum, inverse Fourier transforming to get an image and then ‘cleaning’ the image to impose constraints such as non-negativity and finite extent. These processes have met with some success, but have generally fallen out of favour with respect to more Bayesian image reconstruction methods, which directly fit the image to the noisy data.

Bayesian methods are more robust to missing data, for example not requiring redundant telescope arrays in order to allow phase reconstruction and allowing the use of Fourier coverage with significant ‘holes’. There are a number of such Bayesian methods (or Bayesian-style methods) for image reconstruction in optical interferometry. Those in active use at the time of writing include (in alphabetical order) programs such as BSMEM (Buscher, 1993), IRBis (Hofmann *et al.*, 2014), MACIM (Ireland *et al.*, 2006), MIRA (Thiébaud, 2008), SQUEEZE (Baron *et al.*, 2010) and WISARD (Meimon *et al.*, 2005), but many more are under development and the state of the art is constantly improving.

9.7.1 BSMEM

The BSMEM program is described here, not to advocate it as the best solution but to provide an example to illustrate the process of image reconstruction. The BiSpectrum Maximum Entropy Method (BSMEM) program was designed, as

its name suggests, to reconstruct images from bispectrum (as well as power-spectrum data).

The image reconstruction process can be understood in terms of a standard Bayesian model-fitting where the model consists of a number of image pixels on a regular square grid. The model parameters are the brightnesses of these pixels and in a typical image reconstruction there will be thousands if not millions of such parameters. The number of calibrated data points from an optical interferometer will measure in the dozens to the hundreds at most, and so the problem is highly underconstrained.

In BSMEM this problem is ‘regularised’, that is additional constraints are introduced in order to make the problem well-posed, using the principle of maximum entropy. This principle will be introduced in this discussion in terms of a Bayesian prior; in reality, despite the fact that the maximum entropy principle was introduced in a Bayesian context, the exact link between maximum entropy methods and Bayesian analysis remains controversial. More sophisticated analyses suggest that maximum-entropy methods can be better thought of as having Bayesian analysis as a special case, rather than vice versa (Skilling, 1988a; Giffin, 2008).

These conceptual issues are glossed over here and instead a pseudo-Bayesian rationale for the image reconstruction method adopted is presented, which is plausible if not looked at too closely. The reader is asked to suspend disbelief on the grounds that more rigorous rationales require considerably more space to explain; the interested reader can consult the literature (Skilling, 1988b; Gull and Skilling, 1999). The proof of the maximum-entropy pudding is that it has been successfully used in many different fields to give results superior in many ways to other existing methods.

The maximum-entropy prior for a set of pixels with intensities $\{i_1, \dots, i_{N_{\text{pix}}}\}$ is given by

$$P(I, \alpha | M) \propto e^{\alpha S}, \quad (9.21)$$

where S is the relative ‘entropy’ of the image with respect to a prior model image $M = \{m_1, \dots, m_{N_{\text{pix}}}\}$ and given by

$$S = - \sum_{p=1}^{N_{\text{pix}}} \left[i_p - m_p - i_p \log(i_p/m_p) \right]. \quad (9.22)$$

The parameter α is a ‘hyperparameter’ of the prior and controls how strongly the entropy affects the posterior probabilities. The model image M encodes prior information about the image, for example the finite support constraint

can be encoded as low values of m_p outside some region, indicating that little or no flux resides in these ‘blank spaces’.

Given the maximum-entropy prior, Bayes’ theorem gives the posterior probability as

$$P(I, \alpha | M, D) \propto e^{\alpha S - \chi^2}. \quad (9.23)$$

A convenient representation for the multi-dimensional posterior is the single image I , which has the maximum posterior probability, in other words the image which maximises $\alpha S - \chi^2$.

For a given value of α the probability can be maximised by maximising the entropy S and by minimising the misfit with the data χ^2 . The former term comes from the prior and the latter comes from the likelihood. The relative importance of these two terms is given by the hyperparameter α . If α is small, the maximisation tends towards the minimum of the misfit χ^2 between the actual data and the noiseless data predicted from the image I . If α is large, the maximisation tends towards the maximum of the entropy S , which occurs when the image is identical to the prior image M . Using too small a value of α can over-fit the data at the expense of the prior while using too large a value of α tends to emphasise the prior at the expense of the data.

A balance between these two can be achieved by choosing the value for α such that the maximal value of $\alpha S - \chi^2$ occurs when $\chi^2 = N_{\text{dat}}$. This is equivalent to maximising the entropy under the constraint that $\chi^2 = N_{\text{dat}}$. This value of χ^2 occurs when the distance between the actual data and those which would be observed if I were the true image is of order of one standard deviation per data point.

If the model M is uniform over some region and zero elsewhere, the net effect of this choice for α is that the selected image I is the ‘smoothest’ image that fits the data, as the entropy penalises images which have more structure. In any interferometric experiment (and indeed in any imaging experiment), the measured Fourier components do not extend beyond some finite limit in the (u, v) plane, and so the object can have arbitrary amounts of small-scale structure without affecting the data. The maximum-entropy method automatically limits any attempts to ‘guess’ at this structure while at the same time providing an image that shows any structure for which there is good evidence.

The maximum-entropy prior also has the useful property of enforcing positivity in the image – the entropy contributed by any pixel tends towards $-\infty$ as the pixel intensity tends towards zero and so pixel values are ‘steered’ away from negative values by the entropy.

9.7.2 Running BSMEM

The BSMEM program implements the entropy-maximisation described above for data D , which consist of an arbitrary mixture of bispectrum and power-spectrum data – at present it cannot make use of coherently-integrated visibility data directly. The data, together with the errors on the data, are read in from an OIFITS file specified by the user.

The user also needs to specify a model image M . The default model is a broad Gaussian centred in the middle of the image, which can be used to signify that the object is thought to have an extent that is smaller than some characteristic contour (such as the half-maximum contour) of the Gaussian but where there is not a definite cut-off in the maximum size of the possible structure. This encourages the algorithm to find images which fit inside this contour but does not completely exclude slightly larger images.

An arbitrary model image can also be loaded from a user-provided file. This is typically derived from lower-resolution data – sometimes it is derived from a previous image reconstruction phase using the same data, which the user has adjusted and smoothed to indicate plausible areas for the flux in the object to reside. This allows the user to strengthen the limit-support constraint based on prior knowledge of what structures are likely to exist in the object under study, and allows images to be reconstructed from more sparsely sampled data than looser constraints. Needless to say, such user adjustment requires some care to avoid introducing spurious structure into the image.

The program works in terms of an $N \times N$ -pixel image I where each pixel represents a square region on the sky of a fixed angular size. This pixel size can be automatically selected by the program or chosen by the user. Typically, pixels which are a few times smaller than the theoretical maximum resolution of the data given by u_{\max}^{-1} are used. In other words, the pixels are smaller than the ‘resels’ (resolution elements) of the instrument. This allows for some ‘super-resolution’, where details can be recovered in the image which are below the conventional resolution limit if there is sufficient evidence in the data for such features. The ‘smoothing’ effect of the maximum-entropy prior will ensure that if no such super-resolution is possible then no spurious sub-resel structure will be put in. The size of the image in terms of the number of pixels N needs to be set large enough to include the maximum possible extent of the image based on prior information.

Given these initial parameters, the BSMEM program uses iterative gradient-descent algorithms to find the constrained maximum-entropy solution, starting from the model image M . The progress of this gradient descent in terms of the values of χ^2 and the entropy can be monitored during the running of the

program and intermediate images can be plotted. A situation often arises that the algorithm stagnates because it cannot find an image which fits the data adequately. This is most likely because such an image does not exist, because the noise on the data has been underestimated. In this case it may help to edit the data to put in more realistic error estimates, or to remove data points which appear (for justifiable reasons) to be outliers.

9.7.3 Image reconstruction from coherently averaged data

Data which have been coherently averaged can straightforwardly be converted into bispectrum and power-spectrum data for use in image reconstruction programs which expect these as input. However, the noise model assumed for the data will be wrong since, for example, the bispectrum amplitudes and the power-spectrum amplitudes will be strongly correlated, and the closure phases will be correlated as explained in Section 9.6.4.

There are many image-reconstruction programs that can be used directly with coherently averaged visibility data. These are mostly written with data from radio interferometers in mind, but the forward model for coherently averaged data in an optical interferometer is the same as the forward model for a phase-unstable radio interferometer and so the noise model used is appropriate. Image reconstruction from data corrupted by antenna-dependent phase errors is known in radio interferometry as *self-calibration* (Pearson and Readhead, 1984), referring to the fact that external calibration for these phase errors is not needed.

The program VLBMEM (Sivia, 1987) illustrates the general technique, but other image-reconstruction programs exist that are based on similar principle. The VLBMEM program uses an iterative ‘hybrid mapping’ scheme to solve for two sets of model parameters: the image pixels $\{i_p\}$ and the antenna-dependent phase errors $\{\epsilon_i\}$. There are two steps in each iteration, one attempting to solve for each set. In the first step, the model image from the previous iteration (or a point-source starting image) is used together with the visibility data to solve for the phase errors $\{\epsilon_i\}$. This is achieved through a least-squares fit between the noiseless visibilities predicted from the model image $\{V(\mathbf{u}_{ij})e^{i[\epsilon_i - \epsilon_j]}\}$ and the measured data $\{V_{ij}\}$, with the set of $\{\epsilon_i\}$ as free parameters. In the second step, the $\{\epsilon_i\}$ are held fixed at the values output from the previous step and the image pixels $\{i_p\}$ are adjusted to fit the data using a maximum-entropy method. These two steps are repeated until convergence is achieved at a self-consistent solution for the image and the corrupting antenna-based phases.

Other self-calibration programs exist for dealing with phase-corrupted radio interferometric data. They typically use the same two-step process but differ in the way to solve for the image pixels in the second stage. A popular method for this ‘deconvolution’ step is the CLEAN algorithm (Hogbom, 1974).

9.8 Image quality

The images reconstructed from interferometric data can be of high quality when the (u, v) sampling is adequate and the SNR of the visibility data is high. Figure 9.4 shows a series of infrared (2- μm wavelength) images of the dust shell around the carbon star IRC+10216. These images were reconstructed from visibility-modulus and closure-phase data from aperture-masking observations made at the Keck telescope (Tuthill *et al.*, 2000a).

The images show complex structures and most of these structures are repeated in all the images, despite being taken at different times, with different aperture masks, different filters and different rotations of the masks. Epoch-to-epoch differences seen in these images can in many cases be attributed to dust motion. Thus, the basic reliability of these image reconstructions is not in doubt, despite the fact that several possible sources of ambiguity, including incomplete (u, v) coverage and the use of closure phases rather than visibility phases, are present.

Nevertheless, like all experimental data, these images are not perfect representations of what the object looks like. First, the image resolution is limited by the baseline coverage. The maximum baseline used in these reconstructions is about 10 m, (limited by the size of the Keck telescope) and so the nominal resolution is $\lambda/B_{\text{max}} \approx 40$ milliarcseconds. The images show structures on scales perhaps slightly smaller than this, indicating some level of super-resolution in the image reconstruction.

Secondly, the structure at lowest contour levels, at the level of 1% of the peak brightness, is perhaps less reliable, showing epoch-to-epoch variations which are unlikely to be physical changes in the source but rather are likely to be noise in the image. Thus the ‘dynamic range’ of these images, which is defined as the range of the peak brightness in the image to the brightness of the weakest believable features, is perhaps between 50:1 and 100:1.

The dynamic range is the most commonly used metric of the ‘fidelity’ of an interferometric image, that is to say how well the image approximates to the true object brightness distribution. It is a useful metric for understanding how suitable the reconstructed images are for answering scientific questions such as the detectability of faint companions.

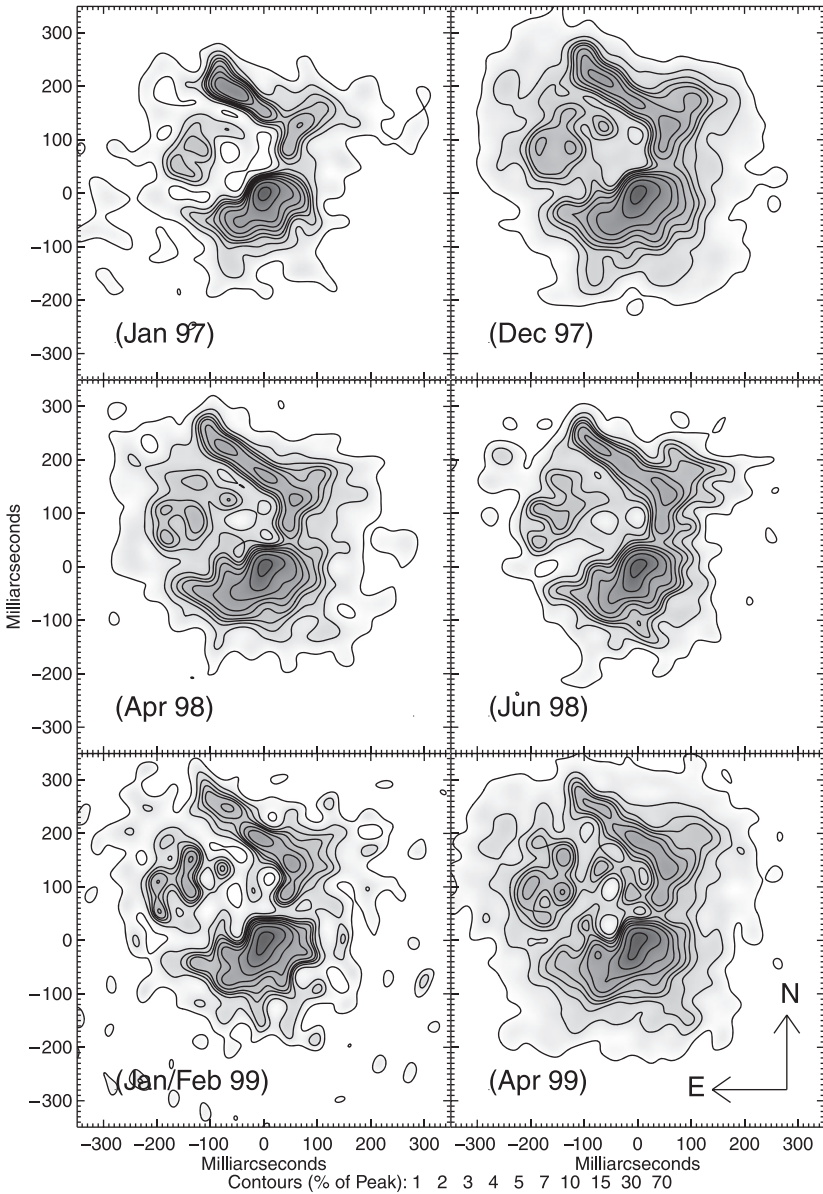


Figure 9.4 Images at several epochs of the evolved star IRC+10216, showing the complex structure of the dust shells around the star. From Tuthill *et al.* (2000a).

The dynamic range has the advantage of being readily applicable to experimental data where the true object brightness distribution is not available for comparison, but where there are regions of blank sky within the imaged region which are unlikely to contain any flux. Any features seen in this region can be taken to be the noise level and the ratio of the peak flux in the image to the flux in these features will give a measure of the dynamic range. However, some caution must be taken in interpreting the dynamic range derived in this way as the effect of regularisation such as that produced by maximum-entropy methods can suppress the level of noise peaks.

A rule of thumb for the dynamic range that can be expected from an interferometric image can be obtained from a simple model of the imaging process. In this model, the data $\{F_{ij}\}$ are taken at N_{uv} sample points in the (u, v) plane using a phase-stable interferometer. The data are corrupted by zero-mean additive noise $\{n_{ij}\}$ so that

$$F_{ij} = F(\mathbf{u}_{ij}) + n_{ij}. \quad (9.24)$$

The noisy synthesised image, in other words the inverse Fourier transform of the noisy data, will be given by

$$I_{\text{noisy}}(\boldsymbol{\sigma}) = \hat{I}(\boldsymbol{\sigma}) + \sum_{i,j} \left(n_{ij} e^{2\pi i \mathbf{u}_{ij} \cdot \boldsymbol{\sigma}} + n_{ij}^* e^{-2\pi i \mathbf{u}_{ij} \cdot \boldsymbol{\sigma}} \right), \quad (9.25)$$

where $\hat{I}(\boldsymbol{\sigma})$ is the noiseless synthesised image given in Equation (2.21). The two noise terms in Equation (9.25) come from the fact that every coherent flux sample at location \mathbf{u}_{ij} also gives a measurement of the value at $-\mathbf{u}_{ij}$ due to the Hermitian symmetry of the Fourier transform of a real function.

If the coherent flux noise distribution has circular symmetry in the complex plane and the noise values are uncorrelated, then the noise at any location in the synthesised image is twice the real part of a ‘random walk’ in the complex plane. The standard deviation of the noise at a given point in the image will

$$\sigma_{\text{image}} = \sqrt{2 \sum_{i,j} \sigma_{ij}^2}, \quad (9.26)$$

where the factor of $\sqrt{2}$ arises from taking twice the real part of a noise distribution, which is circularly symmetric.

If the object is a point source so that the coherent flux modulus $|F(\mathbf{u})| = F$ everywhere in the (u, v) plane, then the peak flux in the synthesised image will be given by

$$I_{\text{max}} \approx 2N_{uv}F \quad (9.27)$$

while if the noise is constant so that $\sigma_{ij} = \sigma$,

$$\sigma_{\text{image}} = \sqrt{2N_{uv}}\sigma. \quad (9.28)$$

Thus, the dynamic range of the image will be given by

$$\text{dynamic range} \sim \frac{2N_{uv}F}{\sqrt{2N_{uv}}\sigma} = \sqrt{2N_{uv}}\text{SNR}_F, \quad (9.29)$$

where

$$\text{SNR}_F = \frac{F}{\sigma}. \quad (9.30)$$

Thus, the dynamic range is proportional to the SNR of the coherent flux data and the square root of the number of data points.

This rule of thumb has been found to give a good indication of the dynamic ranges of images reconstructed from closure-phase and visibility-modulus data, providing that the (u, v) sampling is adequate. For the images in Figure 9.4, the number of (u, v) samples going into the reconstruction is of order 100, so it can be inferred that the average SNR of the coherent flux samples is probably of order 5. Given that this object is very bright the SNR inferred is likely to be representative of the typical random calibration error level rather than being due to detection noise.

If the (u, v) sampling is not adequate for the field-of-view and/or the noise is highly correlated, for example because of systematic calibration errors, then the dynamic range computed from Equation (9.29) may not give a good estimate of the reliability of the image. Figure 9.5 shows a set of ten image reconstructions from the same interferometric dataset, taken on the dust-enshrouded supergiant VY CMa (Monnier *et al.*, 2014). The data have approximately 2% errors on the visibility modulus and 1° errors on the closure phases, corresponding to an SNR of order 50. About 300 different (u, v) samples were taken. Using Equation (9.29) would yield a dynamic-range estimate of order $\sqrt{2} \times 300 \times 50 \approx 170 : 1$, implying a high-quality image.

In reality it can be seen that on this object different reconstruction packages (or even the same reconstruction package used by different teams) tend to agree on the strongest small-scale features but disagree to quite a large extent on the larger-scale features. This is likely due to a lack of short-baseline data: the object was previously imaged using aperture masking and this showed structure out to scales of order 200 milliarcseconds, but the data used in these reconstructions have a shortest baseline such that $|\mathbf{u}| > 5 \times 10^6$ (the (u, v) coverage is shown in Figure 9.2), and this corresponds to a spatial scale of order 40 milliarcseconds. As a result, the large-scale flux is poorly constrained and so the image reconstruction packages have a difficult time determining it.

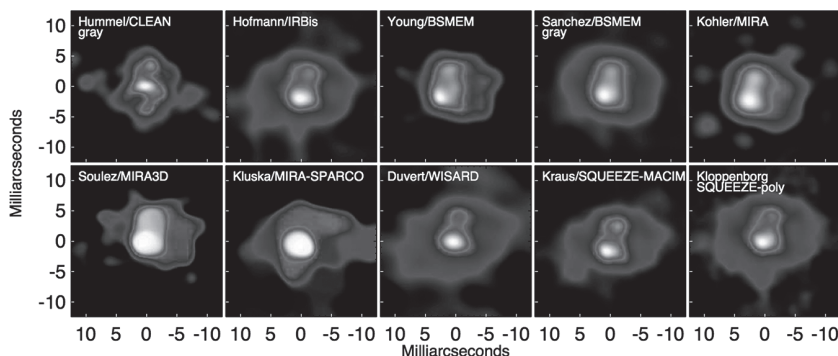


Figure 9.5 Image reconstructions of the dust around the supergiant VY Cma from the 2014 image ‘beauty contest’ (Monnier *et al.*, 2014).

Thus, even with apparently high-quality data, the reliability of the image reconstruction can be seriously compromised if the (u, v) sampling is not appropriate for the object. This must be borne in mind whenever interpreting any interferometric image reconstruction.

9.9 Practical case study: imaging and model-fitting on Betelgeuse

When analysing observational data, imaging and model-fitting are not mutually exclusive activities but are complementary. Crude model-fitting can be used to guide image reconstruction while the results from image reconstruction can be used to decide on which models to fit. This can be illustrated by using a practical example of imaging of Betelgeuse using aperture-masking data. The data presented are from 1990, but illustrate the scientific progress that can be made using relatively simple images when they have an angular resolution which is much higher than has been previously possible.

Betelgeuse is the nearest supergiant star, and as a result has one of the largest apparent angular diameters of any star in the sky. It was the first star to have its angular diameter measured using interferometry (Michelson and Pease, 1921). Supergiant stars are quite unlike main-sequence stars such as the Sun in that not only are they thousands of times larger, but the size of the convection cells at the surface of the star is comparable to the size of the star itself. Thus, the stellar surface might be expected to be pocked, not with the diminutive dark sunspots we see on the Sun, which are caused by localised cooling due to the magnetic field, but with giant bright ‘starspots’ corresponding to convective

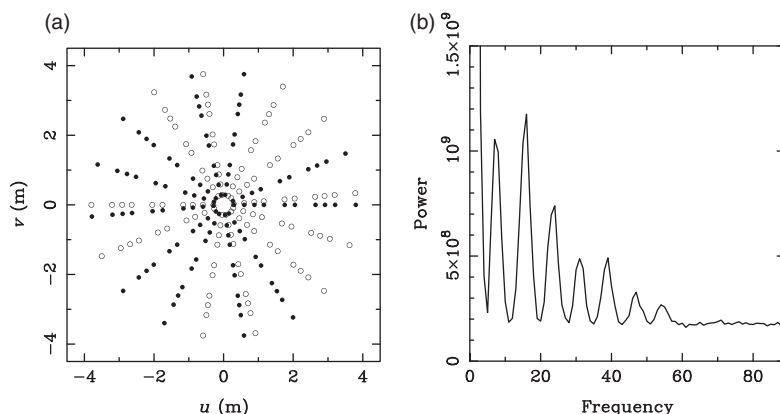


Figure 9.6 (a) The Fourier plane coverage of the Betelgeuse 710-nm wavelength observations. (b) The average power spectrum of 500 interferograms taken on Betelgeuse at 546-nm wavelength. These data were taken using a five-hole mask, but fringes are only visible on the shortest seven out of ten baselines because the source is resolved on longer baselines.

upwelling of hotter material from below the surface of the star. Despite this intrinsic interest, no convincing image of the features on the surface of this star had been made at the time of observation.

The observations were made using aperture masking on the 4.2-m William Herschel Telescope. The longest baseline in the aperture mask was 3.8 m and so at a wavelength of 700 nm the highest spatial frequency measured corresponded to a sinusoid on the sky with a peak-to-peak spacing of 38 mas. Betelgeuse has an angular diameter of 40–50 mas so the stellar surface is only just resolved.

The mask was arranged to be non-redundant so no two pairs of holes had the same spacing: for a five-hole mask fringes appeared at ten different frequencies. To achieve two-dimensional Fourier plane coverage the whole arrangement was rotated around the telescope axis to a number of different position angles with respect to the sky. With about ten different position angles and ten baselines per position angle, Fourier data at about 100 (u, v) points were obtained as shown in Figure 9.6. These are not all independent visibility points, but the resulting image has less than 10×10 independent resolution elements in it, so the coverage is adequate.

The observations at each wavelength yielded of order 100 calibrated visibility amplitudes and 100 closure phases. Figure 9.7 shows a subset of these data for illustration purposes. A number of important constraints on the resulting image can be inferred purely from examination of these data (a kind of ‘eyeball model fitting’):

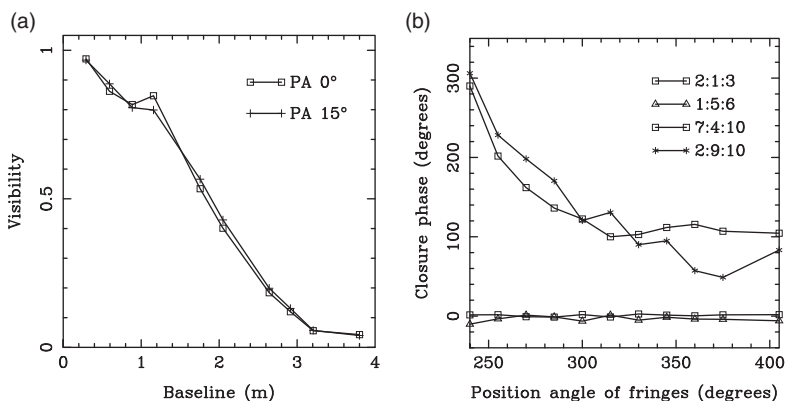


Figure 9.7 A subset of the calibrated visibility (a) and closure-phase (b) measurements on Betelgeuse at a wavelength of 710 nm. The visibility data are plotted for two similar position angles of the mask. The closure phase data are plotted for two sets of triples containing only short baselines and two sets of triples containing long baselines. The closure phases are labelled with the indices (1–10) of the baselines making up the triple.

- From the decrease in calibrated visibility with baseline it is clear that Betelgeuse is indeed resolved on the longest baselines and this is roughly consistent with a uniform disc diameter of around 50 milliarcseconds.
- The closure phases measured on baseline triangles involving the longest baselines show large departures from 0° and 180° , whereas the closure phases on the triangles containing only short baselines are close to zero. This means that there is significant asymmetric structure on scales comparable with the size of the stellar disc – the asymmetry cannot be due to a companion far away from the star.
- The closure phase varies over essentially 0° to 360° , indicating that the asymmetric structure is comparable to or larger than the symmetric structure at this resolution. The fraction of the flux that is unresolved at this resolution is given approximately by the visibilities on these baselines: these visibilities are of order 10%, hence the flux in the asymmetric structure is of the order of 10% of the total flux.

An additional inference which can be drawn is that the closure phase can be measured with very few systematics: the uncalibrated closure phases measured on the shortest baselines where Betelgeuse is effectively unresolved are within a degree of zero. This kind of accuracy in a conventional phase measurement would require an internal pathlength stability of better than 2 nm and yet no special precautions were taken in this experiment.

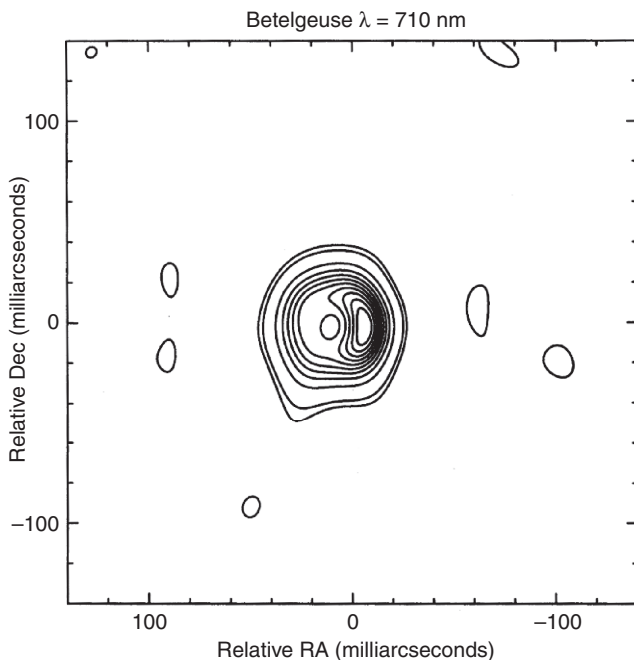


Figure 9.8 Contour map of the reconstructed image of Betelgeuse at 710 nm. The contours are at 1, 2, 10, 20, 30, ..., 90% of the peak intensity. From Buscher *et al.* (1990).

Image reconstruction on these data was performed with the VLBMEM program described in Section 9.7.3. This program was designed for radio interferometry data, and so expects coherently averaged data, but only incoherently averaged power-spectrum and bispectrum data were available.

The procedure used to adapt the incoherently-averaged data to the program was one which is a common way of using optical interferometry data with radio-interferometry software. This was to create a set of ‘fake’ coherently averaged data that were consistent with the bispectrum and power-spectrum data. This used a least-squares fit of a set of visibility phases to the closure phases. Note that the VLBMEM program assumes that the phases are corrupted by unknown antenna phases, so it was not necessary to solve for the antenna phases at this stage. The visibility modulus was derived from the square root of the calibrated power-spectrum estimates.

This procedure is less preferable than using a program which is designed to accept bispectrum and power-spectrum data directly, since the noise model used is not as accurate. Nevertheless, it produced image reconstructions of Betelgeuse (Figure 9.8) which agree with the constraints inferred from the

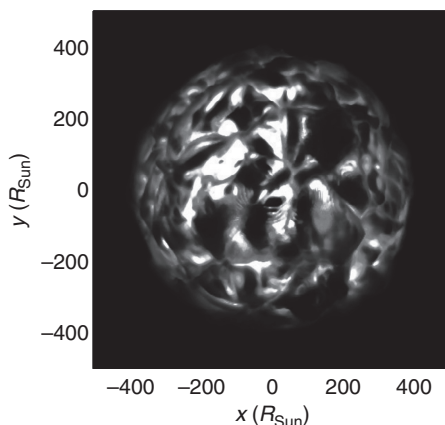


Figure 9.9 Image of a red supergiant star simulated using three-dimensional radiative hydrodynamic modelling. Note the size of the star – the axes are labelled in units of the Sun's radius. From Chiavassa *et al.* (2011).

data. The reconstructed image shows a stellar disc of about 50 milliarcseconds diameter and a region of increased intensity on one side of the disc.

Having made a model-independent image reconstruction, one can then fit simple models to the data to get quantitative information. These model-fits show that the 'hotspot' seen on the disc does indeed account for about 10% of the stellar flux. Further aperture-masking observations (Wilson *et al.*, 1997) showed that such hotspots appear and disappear on timescales of order several weeks and observations in the infrared (Young *et al.*, 2000) showed that these spots have significantly lower contrast at longer wavelengths.

The combination of image reconstruction and the model-fitting can then be used to constrain astrophysical models. Radiative hydrodynamic modelling of supergiants has shown that turbulent upwelling processes in stars like Betelgeuse can produce bright hotspots with the characteristics seen in these interferometric observations, and produces simulated images such as that shown in Figure 9.9. These images can be compared to the hotspots seen in interferometric images in different wavebands to constrain the models for the star.

Further understanding of these remarkable stars will require detailed images of these turbulent structures, preferably with sufficient time resolution to follow their evolution. Interferometry provides the only way to make 'movies' of structures like this, and hopefully future interferometers will provide us with pictures as beautiful and intriguing as seen in these simulations.