Pierre Léna • Daniel Rouan
François Lebrun • François Mignard
Didier Pelat

In collaboration with Laurent Mugnier

Observational Astrophysics

Translated by S. Lyle

Third Edition

Springer

9.6 From Data to Object: the Inverse Problem

In this section,[12] we discuss the procedures for obtaining the *best* possible estimate of the observed object, using all the information available about the instrument used for the observation and about sources of noise intrinsic to the instrument or to the received signal, but also the information we possess concerning the source itself, which is generally not totally unknown. We use the term *data* to refer to the complete set of measured values resulting from the observation process. By taking into account all this information in a suitable way, not only is the object better reconstructed by data processing, but error bars affecting this reconstruction can be estimated. Figure 9.13 illustrates these procedures schematically.

The developments and tools described in the following can be applied to a very broad range of data types encountered in astrophysical observation: images with any format and at all wavelengths, spectra, polarisation rates as a function of frequency, and so on. However, in order to make the following more concrete and to illustrate it with examples, we have chosen to discuss only wavefront sensing and *image processing*, and to develop examples taken from the areas of adaptive optics, which began in the 1990s (see Sect. 6.3), and interferometry (see Sect. 6.4). Indeed, these two areas have considerably stimulated such processing techniques. Readers will then be able to adapt the ideas discussed here to other cases they may encounter in radioastronomy, high energy observation, etc.

[12] This section was entirely contributed by Laurent Mugnier.

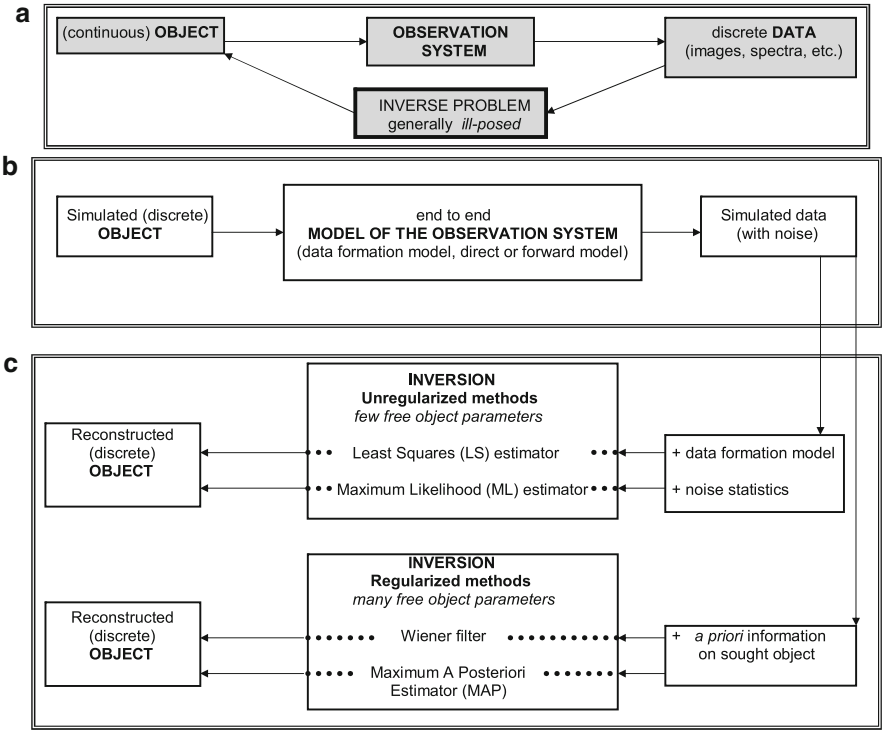

Fig. 9.13 Schematic view of the inverse problem and of the various solutions discussed in this section

9.6.1 Posing the Problem

Forward Problem

Consider the following classic problem in physics. An initial object such as an electromagnetic wave propagates through a system that will affect it in some way, e.g., atmospheric turbulence, an instrument involving aberrations and diffraction, etc. We can deduce the properties of the output wave if we know all the properties of the system, such as would be included in a *forward* or *direct model*, or *data formation model* for this system. This is the basic principle of *end-to-end models*, discussed in Sect. 9.2.

This can be illustrated by wavefront sensing as carried out in adaptive optics, i.e., the measurement of aberrations intrinsic to the telescope or due to the passage through turbulent atmospheric layers, using a (Hartmann–Shack) wavefront sensor. The relevant physical quantity here is the phase $\varphi(\boldsymbol{r}, t)$ on the exit pupil of the telescope, which contains all the information regarding the aberrations suffered by the wave. The measurement data are the average slopes of the phase in two perpendicular directions, viz., $\partial\phi(\boldsymbol{r}, t)/\partial x$ and $\partial\phi(\boldsymbol{r}, t)/\partial y$, on each Hartmann–Shack subaperture, which can be recorded together in a vector $\boldsymbol{i}$. The calculation of the slopes $\boldsymbol{i}$ given the phase φ is a classic *forward problem* in physics. It requires a choice of data formation model.

An end-to-end forward model includes models for data formation right up to detection, and even up to data storage when data are transmitted with compression. It thus takes into account photon noise, detector readout noise, digitiser quantisation noise, and data compression noise, if any (see Sect. 9.2).

Inverse Problem

The problem here is work back up from the data to knowledge of the object that originated the data. This is called the *inverse problem*. It involves inverting the data formation model. Our senses and our brains solve such inverse problems all the time, analysing the information received, e.g., by the retina, and deducing positions in space, 3D shapes, and so on. This is also the most common situation for the observer in astronomy, who makes deductions from observations and attempts to conclude as accurately as possible about some set of properties of the source.

In physics, and especially in astronomy, the processing of experimental data thus consists in solving an inverse problem, in practice, after a data reduction or preprocessing stage which aims to correct for instrumental defects in such a way that the data can be correctly described by the chosen forward model (see Sect. 9.4).

Estimating (see Sect. 9.5) or working back up to the phase φ from the slope vector $\boldsymbol{i}$ is the *inverse problem* of Hartmann–Shack wavefront sensing. It involves inverting the corresponding data formation model.

Let us consider another example, taken from another spectral region. The γ ray astronomy mission INTEGRAL (see Sect. 5.2.5) has three instruments capable of carrying out spectral analysis, namely, JEM-X, IBIS ISGRI+PICSIT, and SPI. Each of these can measure the spectrum of the observed source, here the black hole binary system Cygnus X-1. The data formation model is contained in a software called XSPEC, and an a priori model of the source containing a certain number of free parameters is injected into this. The simulated output of this software is then compared with measurements from each of the instruments, and the free parameters are fitted in the best possible way, in a sense to be made precise shortly, as shown in Fig. 9.14.

We shall see below that inversion can often take advantage of statistical knowledge regarding measurement error, generally modelled as *noise*. Naive inversion methods are often characterised by *instability*, in the sense that the inevitable measurement noise is amplified in an uncontrolled way during inversion, leading to an inacceptable solution. In this case, where the data alone are insufficient to obtain an acceptable solution, more sophisticated inversion methods, called *regularised inversion methods*, are brought to bear. They incorporate further constraints to impose a certain regularity on the solution, compatible with what is known *a priori* about it. This is a key point: in data processing, we introduce supplementary

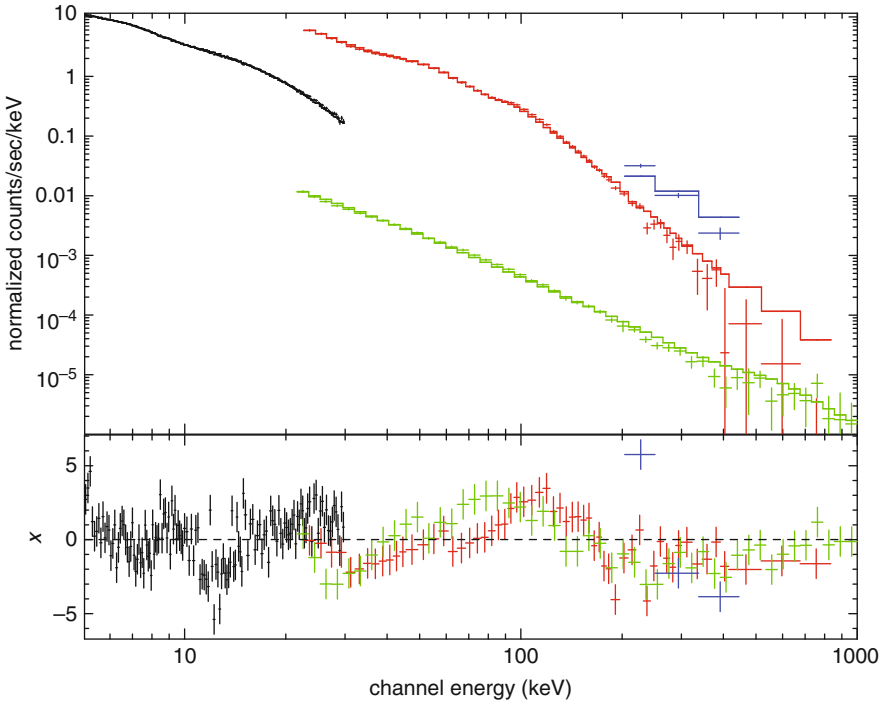

Fig. 9.14 Spectrum of the accreting binary system Cygnus X-1, viewed by the various instruments of the γ-ray observatory INTEGRAL. The model fitted to the data is a hybrid model, in which electrons in the corona have a non-thermal distribution. From Cadolle-Bel M. et al., Soc. Fr. Astron. Astroph., *Scientific Highlights* 2006, p. 191

information as dictated by our prior knowledge of the object. For example, a brightness can only take positive values, an object such as a stellar disk is bounded spatially by a contour, a temporal fluctuation cannot contain frequencies above a certain threshold, and so on.

It is extremely productive to conceive of data processing explicitly as the inversion of a forward problem. To begin with, this forces us to model the whole data formation process, so that each step can be taken into account in the inversion. It also allows us to analyse existing methods, e.g., a typical deconvolution, or filtering, as discussed in Sect. 9.1.3, and to clearly identify the underlying assumptions, or the defects. It is then possible to conceive of methods that take advantage of knowledge concerning the data formation process as well as knowledge of the source or the relevant physical quantity, available *a priori*, i.e., before the measurements are made.

In the following, we discuss the basic notions and tools required to tackle the resolution of an inverse problem, a subject that has been deeply transformed by

the advent of powerful computational tools. For the purposes of this textbook,[13] the tools discussed are illustrated on several relatively simple, i.e., linear, inverse problems encountered in astronomy: wavefront reconstruction from measurements by a Hartmann–Shack sensor, adaptive optics corrected image restoration and multispectral image reconstruction in nulling interferometry (see Sect. 6.6). We shall not discuss extensions to non-linear problems, which do exist, however.

9.6.2 Well-Posed Problems

Let the relevant physical quantity be o, referred to as the *object* in what follows. It could be the phase of the wave in the case of a wavefront sensor, a star or galaxy in the case of an image in the focal plane of a telescope, the spectrum of a quasar in the case of a spectrograph, etc. Let i be the *measured data*.[14] This could be the slopes measured by the Hartmann–Shack sensor, an image, a spectrum, etc. We shall consider for the moment that o and i are continuous quantities, i.e., functions of space, time, or both, belonging to Hilbert spaces denoted by X and Y.

The forward model, deduced from physics and from the known properties of the instrument, can be used to calculate a model of the data in the case of observation of a known object. This is what is done in a *data simulation* operation:

$$i = H(o). \tag{9.70}$$

We shall restrict attention here to *linear* forward models, whence

$$i = Ho, \tag{9.71}$$

where H is a continuous linear operator mapping X into Y. It was in this context that Hadamard[15] introduced the concept of *well-posed problems*.

When the forward model is linear and translation invariant, e.g., for imaging by a telescope within the isoplanatic patch (see Sect. 6.2), H is a convolution operator and there is a function h called

[13] A good reference for any reader interested in inverse problems is Titterington, D.M. General structure of regularization procedures in image reconstruction, Astron. Astrophys. **144**, 381–387, 1985. See also the extremely complete didactic account by Idier, Bayesian approach to inverse problems, ISTE/John Wiley, London, 2008, which has partly inspired the present introduction.

[14] The notation in this section is as follows: the object o and image i are in italics when they are continuous functions, and in bold type when discretised. (In Chap. 6, the capital letters O and I were used.) Matrices and operators are given as italic capitals like M.

[15] Jacques Hadamard (1865–1963) was a French mathematician who contributed to number theory and the study of well-posed problems.

the point spread function (*PSF*) such that

$$i = Ho = h \star o. \tag{9.72}$$

We seek to invert (9.71), i.e., to find o for a given i. We say that the problem is *well posed* in the sense of Hadamard if the solution o satisfies the usual conditions for existence and uniqueness, but also the less well known condition of *stability*, i.e., the solution depends continuously on the data i. In other words, a small change in the data — in practice, another realisation of the random noise — only brings about a small change in the solution. These three conditions, known as Hadamard's conditions, are expressed mathematically in the following way:

- **Existence.** There exists $o \in X$ such that $i = Ho$, i.e., $i \in \mathrm{Im}(H)$, the image or range of H.
- **Uniqueness.** The kernel of H contains only zero, i.e., $\mathrm{Ker}(H) = \{0\}$.
- **Stability.** The inverse of H on $\mathrm{Im}(H)$ is continuous.

Note that for the stability condition the inverse of H on $\mathrm{Im}(H)$, denoted by H^{-1}, is well defined, because we assume that $\mathrm{Ker}(H) = \{0\}$. Note also that the stability condition is equivalent to requiring the set $\mathrm{Im}(H)$ to be closed, i.e., $\mathrm{Im}(H) = \overline{\mathrm{Im}(H)}$.

For many inverse problems, even the first two of these conditions are not satisfied, let alone the third. Indeed, on the one hand $\mathrm{Im}(H)$ is the set of possible images when there is no noise, a smaller set than the space of noisy data to which i belongs. For example, in imaging, $\mathrm{Im}(H)$ is a vector space that contains no frequency greater than the optical cutoff frequency of the telescope (D/λ), whereas the noise will contain such frequencies. In general, the existence of a solution is therefore not guaranteed. On the other hand, the kernel of H contains all objects *unseen* by the instrument. So for a Hartmann–Shack sensor, these are the unseen spatial modes, e.g., described in terms of Zernike polynomials, such as the piston mode (Z_1) or the so-called waffle mode. For an imager, these are the spatial frequencies of the object above the optical cutoff frequency of the telescope. The kernel is therefore generally not just $\{0\}$ and uniqueness is not granted.

The mathematician Nashed introduced the idea of a well-posed problem in the sense of least squares, which provides a way to ensure existence (in practice) and uniqueness of the solution and then to show that inversion in the least squares sense does not lead to a good solution of the inverse problem owing to its instability, also called *non-robustness* to noise.[16] It thus remains ill-posed.

We say that $\hat{o}_{\mathrm{LS}}$ is a least-squares solution to the problem (9.71) if

$$\|i - H\hat{o}_{\mathrm{LS}}\|^2 = \inf_o \|i - Ho\|^2, \tag{9.73}$$

where $\|\ \|$ is the Euclidean or L_2 norm. Nashed then showed the following:

[16]The term *robustness to noise*, often used when discussing inversion methods, means that the given method does not amplify the noise power in an exaggerated way.

- **Existence.** A least-squares solution exists if and only if $i \in \mathrm{Im}(H) + \mathrm{Im}(H)^{\perp}$. This condition is always satisfied if $\mathrm{Im}(H)$ is closed.
- **Uniqueness.** If several solutions exist, we choose the unique solution with minimal norm, i.e., we project the solution on to the orthogonal complement $\mathrm{Ker}(H)^{\perp}$ of the kernel. We denote this by $H^{\dagger}i$, and call it the *generalised inverse*. The operator $H^{\dagger}$ associates the least-squares solution of $i = Ho$ of minimal norm, i.e., the only solution in $\mathrm{Ker}(H)^{\perp}$, to all $i \in \mathrm{Im}(H) + \mathrm{Im}(H)^{\perp}$.

Moreover, it can be shown that $H^{\dagger}$ is continuous if and only if $\mathrm{Im}(H)$ is closed.

We say that the problem (9.71) is well-posed in the least-squares sense if there exists a unique least-squares solution (with minimal norm) and it is stable, i.e., it depends continuously on the data i. We then see that the problem (9.71) is indeed well posed in the sense of least squares if and only if $H^{\dagger}$ is continuous, i.e., if and only if $\mathrm{Im}(H)$ is closed. And for many operators, e.g., when H is the convolution with a square-integrable response h, this condition is not satisfied.

We can understand the meaning of such a solution intuitively by characterising the least-squares solutions. Let H be a continuous linear operator from a Hilbert space X to the Hilbert space Y. Then the following three propositions are equivalent:

1. $\|i - H\hat{o}_{\mathrm{LS}}\|^2 = \inf_{o \in X} \|i - Ho\|^2$.
2. $H^*H\hat{o}_{\mathrm{LS}} = H^*i$, where H^* is the adjoint of H (normal equation).
3. $H\hat{o}_{\mathrm{LS}} = Pi$, where P is the (orthogonal) projection operator of Y on $\overline{\mathrm{Im}(H)}$.

In particular, characterisation (3) tells us that the least-squares solution *exactly* solves the original equation (9.71) when the data i are projected onto the (closure of the) set of all possible data in the absence of noise, i.e., $\overline{\mathrm{Im}(H)}$.

In finite dimensions, i.e., for all practical (discretised) problems, a vector subspace is always closed. As a consequence, we can be sure of both the existence and the continuity of the generalised inverse. However, the ill-posed nature of the continuous problem does not disappear by discretisation. It simply looks different: the mathematical instability of the continuous problem, reflected by the non-continuity of the generalised inverse in infinite dimensions, resurfaces as a numerical instability of the discretised problem. The discretised inverse problem in finite dimensions is *ill-conditioned*, as we shall explain shortly. The conditioning of a discretised inverse problem characterises the robustness to noise during inversion. It is related to the dynamic range of the eigenvalues of H^*H (a matrix in finite dimensions), and worsens as the dynamic range increases.

9.6.3 Conventional Inversion Methods

In the following, we shall assume that the data, which have been digitised (see Sect. 9.1.3), are discrete, finite in number, and gathered together into a vector $\boldsymbol{i}$. In imaging, for an image of size $N \times N$, $\boldsymbol{i}$ is a vector of dimension N^2 which

concatenates the rows or the columns of the image (in a conceptual rather than computational sense).

The first step in solving the inverse problem is to discretise also the sought object $\boldsymbol{o}$ by expanding it on to a finite basis, that is, a basis of pixels or cardinal sine functions for an image, or a basis of Zernike polynomials for a phase representing aberrations. The model relating $\boldsymbol{o}$ and $\boldsymbol{i}$ is thus an approximation to the continuous forward model of (9.70) or (9.71). By explicitly incorporating the measurement errors in the form of additive noise $\boldsymbol{n}$ (a vector whose components are random variables), this can be written

$$\boldsymbol{i} = H(\boldsymbol{o}) + \boldsymbol{n}, \tag{9.74}$$

in the general case, and

$$\boldsymbol{i} = H\boldsymbol{o} + \boldsymbol{n}, \tag{9.75}$$

in the linear case, where H is a matrix. In the special case where H represents a discrete convolution, the forward model can be written

$$\boldsymbol{i} = \boldsymbol{h} \star \boldsymbol{o} + \boldsymbol{n}, \tag{9.76}$$

where $\boldsymbol{h}$ is the PSF of the system and $\star$ denotes the discrete convolution.

Note that, in the case of photon noise, the noise is not additive in the sense that it depends on non-noisy data $H\boldsymbol{o}$. Equation (9.74) then abuses notation somewhat.[17]

Least-Squares Method

The most widely used method for estimating the parameters $\boldsymbol{o}$ from the data $\boldsymbol{i}$ is the least-squares method. The idea is to look for $\hat{\boldsymbol{o}}_{\mathrm{LS}}$ which minimises the mean squared deviation between the data $\boldsymbol{i}$ and the data model $H(\boldsymbol{o})$:

$$\hat{\boldsymbol{o}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{o}} \|\boldsymbol{i} - H(\boldsymbol{o})\|^2, \tag{9.77}$$

where arg min is the argument of the minimum and $\|\ \|$ is the Euclidean norm. This method was first published by Legendre[18] in 1805, and was very likely discovered by Gauss[19] a few years earlier but without publishing. Legendre used the

[17] Some authors write $\boldsymbol{i} = H\boldsymbol{o} \diamond \boldsymbol{n}$ for the noisy forward model to indicate a noise contamination operation which may depend, at each data value, on the value of the non-noisy data.

[18] Adrien-Marie Legendre (1752–1833) was a French mathematician who made important contributions to statistics, algebra, and analysis.

[19] Carl Gauss (1777–1855), sometimes called *Princeps mathematicorum* or the prince of mathematicians, was a German astronomer and physicist who made deep contributions to a very broad range of areas.

least-squares method to estimate the ellipticity of the Earth from arc measurements, with a view to defining the metre.

When the measurement model is linear and given by (9.75), the solution is analytic and can be obtained by setting the gradient of the criterion (9.77) equal to zero:

$$H^{\mathrm{T}} H \hat{\boldsymbol{o}}_{\mathrm{LS}} = H^{\mathrm{T}} \boldsymbol{i} \,. \tag{9.78}$$

If $H^{\mathrm{T}} H$ can be inverted, i.e., if the rank of the matrix H is equal to the dimension of the unknown vector $\boldsymbol{o}$, then the solution is unique and given by

$$\hat{\boldsymbol{o}}_{\mathrm{LS}} = \left(H^{\mathrm{T}} H\right)^{-1} H^{\mathrm{T}} \boldsymbol{i} \,. \tag{9.79}$$

Otherwise, as in infinite dimension (see Sect. 9.6.2), there are infinitely many solutions, but only one of them has minimal norm (or 'energy'). This is the *generalised inverse*, written $H^{\dagger} \boldsymbol{i}$.

Relation Between Least Squares and Inverse Filter

When the image formation process can be modelled by a convolution, the translation invariance of the imaging leads to a particular structure of the matrix H. This structure is approximately the structure of a circulant matrix (for a 1D convolution), or block circulant with circulant blocks (for a 2D convolution). In this approximation, which amounts to making the PSF $\boldsymbol{h}$ periodic, the matrix H is diagonalised by discrete Fourier transform (DFT), which can be calculated by an FFT algorithm, and its eigenvalues are the values of the transfer function $\tilde{\boldsymbol{h}}$ (defined as the DFT of $\boldsymbol{h}$). The minimal norm least-squares solution of the last section can then be written in the discrete Fourier domain:[20]

$$\tilde{\hat{\boldsymbol{o}}}_{\mathrm{LS}}(\boldsymbol{u}) = \frac{\tilde{\boldsymbol{h}}^*(\boldsymbol{u}) \tilde{\boldsymbol{i}}\,(\boldsymbol{u})}{|\tilde{\boldsymbol{h}}(\boldsymbol{u})|^2} = \frac{\tilde{\boldsymbol{i}}}{\tilde{\boldsymbol{h}}}(\boldsymbol{u}) \qquad \forall\ \boldsymbol{u} \ \text{ such that } \ \tilde{\boldsymbol{h}}(\boldsymbol{u}) \neq 0, \quad \text{and } 0 \text{ if } \tilde{\boldsymbol{h}}(\boldsymbol{u}) = 0, \tag{9.80}$$

where the tilde represents the discrete Fourier transform. In the case of a convolutive data model, the least-squares solution is thus identical to the *inverse filter*, up to the above-mentioned approximation.

Maximum Likelihood Method

In the least-squares method, the choice of a quadratic measure of the deviation between the data $\boldsymbol{i}$ and the data model $H(\boldsymbol{o})$ is only justified by the fact that the

[20] The transfer function $\tilde{\boldsymbol{h}}^*$ in the discrete Fourier domain corresponds to the matrix H^{T} and the matrix inverses give simple inverses.

solution can then be obtained by analytical calculation. Furthermore, this method makes no use whatever of any knowledge one may possess about the statistical properties of the noise. But this information about the noise can be used to interpret the least-squares method, and more importantly to extend it.

We model the measurement error $\boldsymbol{n}$ as noise with probability distribution $p_{\mathrm{n}}(\boldsymbol{n})$. According to (9.74), the distribution of the data $\boldsymbol{i}$ conditional to the object, i.e., for a given object $\boldsymbol{o}$ (hence supposed known), is then[21]

$$p(\boldsymbol{i}\,|\boldsymbol{o}) = p_{\mathrm{n}}\big(\boldsymbol{i} - H(\boldsymbol{o})\big). \tag{9.81}$$

Equation (9.81) can be used to draw realisations of noisy data knowing the object, i.e., to *simulate* data. On the other hand, in an inverse problem, one has only one realisation of the data, namely those actually measured, and the aim is to estimate the object. The maximum likelihood (ML) method consists precisely in reversing the point of view on $p(\boldsymbol{i}\,|\boldsymbol{o})$ by treating $\boldsymbol{o}$ as variable, with $\boldsymbol{i}$ fixed equal to the data, and seeking the object $\boldsymbol{o}$ that maximises $p(\boldsymbol{i}\,|\boldsymbol{o})$. The quantity $p(\boldsymbol{i}\,|\boldsymbol{o})$ viewed as a function of $\boldsymbol{o}$ is then called the *likelihood* of the data, and the object $\hat{\boldsymbol{o}}_{\mathrm{ML}}$ which maximises it is the one which makes the actually observed data the most likely:[22]

$$\hat{\boldsymbol{o}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{o}} p(\boldsymbol{i}\,|\boldsymbol{o}). \tag{9.82}$$

The most widely used model for noise is without doubt the centered (i.e. zero mean) Gaussian model, characterised by its covariance matrix C_{n}:

$$p(\boldsymbol{i}\,|\boldsymbol{o}) \propto \exp\left\{-\frac{1}{2}\big[\boldsymbol{i} - H(\boldsymbol{o})\big]^{\mathrm{T}} C_{\mathrm{n}}^{-1}\big[\boldsymbol{i} - H(\boldsymbol{o})\big]\right\}. \tag{9.83}$$

The noise is called *white noise* (see Appendix B) if its covariance matrix is diagonal. If this matrix is also proportional to the identity matrix, then the noise is called *stationary* or *homogeneous* (*white noise*). The readout noise of a CCD detector (see Sect. 7.4.6) is often modelled by such a stationary centered Gaussian noise. Photon noise is white, but it has a Poisson distribution (see Appendix B), which can be shown, for high fluxes, to tend towards a non-stationary Gaussian distribution with variance equal to the signal detected on each pixel.

Maximising the likelihood is obviously equivalent to minimising a criterion defined to be the negative of the logarithm of the likelihood and called the *neg-log-likelihood*:

$$J_i(\boldsymbol{o}) = -\ln p(\boldsymbol{i}\,|\boldsymbol{o}). \tag{9.84}$$

In the case of Gaussian noise, the neg-log-likelihood is

$$J_i(\boldsymbol{o}) = \frac{1}{2}\big[\boldsymbol{i} - H(\boldsymbol{o})\big]^{\mathrm{T}} C_{\mathrm{n}}^{-1}\big[\boldsymbol{i} - H(\boldsymbol{o})\big]. \tag{9.85}$$

---

[21] The equation reads: the probability of $\boldsymbol{i}$ given $\boldsymbol{o}$ is the noise probability distribution at $\boldsymbol{i} - H(\boldsymbol{o})$.

[22] The notation arg max refers to the *argument of the maximum*, i.e., the value of the variable which maximises the ensuing quantity, in this case $p(\boldsymbol{i}\,|\boldsymbol{o})$.

If the noise is also white, we have

$$J_i(\boldsymbol{o}) = \frac{1}{2}\sum_k \frac{|\boldsymbol{i}(k) - H(\boldsymbol{o})(k)|^2}{\boldsymbol{\sigma}_{\mathrm{n}}^2(k)}, \tag{9.86}$$

where $\boldsymbol{\sigma}_{\mathrm{n}}^2(k)$ are the elements on the diagonal of the matrix C_{n}. $J_i(\boldsymbol{o})$ is a *weighted least squares criterion*. If the noise is also stationary with variance σ_{n}^2, then $J_i(\boldsymbol{o}) = \|\boldsymbol{i} - H(\boldsymbol{o})\|^2/2\sigma_{\mathrm{n}}^2$ is precisely the *ordinary*, as opposed to weighted, least squares criterion.

The least squares method can thus be interpreted as a maximum likelihood method in the case of stationary white Gaussian noise. Conversely, if the noise distribution is known but different, the maximum likelihood method can take this knowledge of the noise into account and then generalises the least squares method.

Example: Wavefront Reconstruction by the Maximum Likelihood Method

Consider a Hartmann–Shack wavefront sensor that measures aberrations due to atmospheric turbulence on a ground-based telescope. The phase at the exit pupil is expanded on a basis of Zernike polynomials Z_k, the degree of which is necessarily limited in practice to some maximal value $k_{\max}$:

$$\varphi(x, y) = \sum_{k=1}^{k_{\max}} o_k Z_k(x, y). \tag{9.87}$$

We thus seek $\boldsymbol{o}$, the set of coefficients o_k of this expansion, by measuring the average slopes and inserting them into a measurement vector $\boldsymbol{i}$. The data model is precisely (9.75), viz., $\boldsymbol{i} = H\boldsymbol{o} + \boldsymbol{n}$, where H is basically the differentiation operator and is known as the interaction matrix.

In the simulation presented below, the sensor contains 20×20 subapertures of which only 276 receive light, owing to a central obstruction of 33% by the secondary mirror. This gives 552 slope measurements. The true turbulent phase φ is a linear combination of $k_{\max} = 861$ Zernike polynomials, drawn from a Kolmogorov distribution. The matrix H thus has dimensions 552×861. The noise on the slopes is stationary white Gaussian noise with variance equal to the variance of the non-noisy slopes, implying a signal-to-noise ratio of 1 on each measurement. Under such conditions, the maximum likelihood estimate of the phase is identical to the least squares solution (see last subsection). The matrix $H^{\mathrm{T}}H$ in (9.79) has dimensions $k_{\max} \times k_{\max}$ and cannot be inverted because we only have 552 measurements. The generalised inverse solution cannot be used because it is completely dominated by noise. A remedy often adopted here is to reduce the dimension $k_{\max}^{\mathrm{rec}}$ of the space of unknowns $\boldsymbol{o}$ during reconstruction of the wavefront. This remedy is one of the known *regularisation* methods, referred to as *regularisation by dimension control*. An example of reconstruction for different values of $k_{\max}^{\mathrm{rec}}$ is shown in Fig. 9.15.

For $k_{\max}^{\mathrm{rec}} = 210$, a value well below the number of measurements, the reconstructed phase is already unacceptable. The particular shape is due to the fact that the telescope has a central obstruction, whence there are no data at the centre of the pupil.

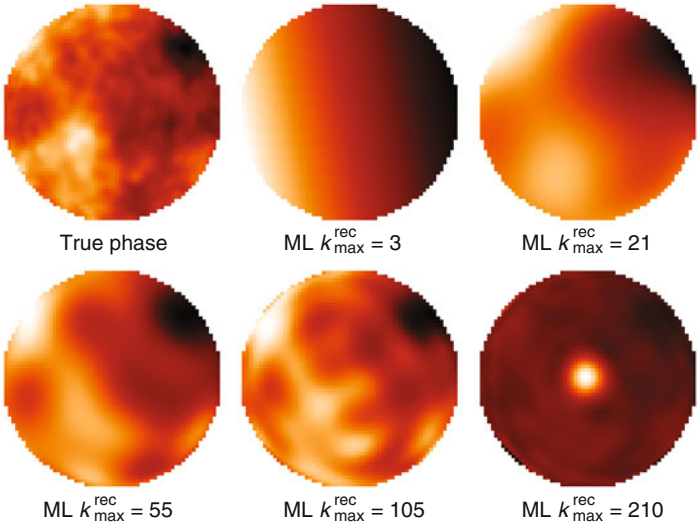

Fig. 9.15 Wavefront sensing on a pupil. *Top left*: Simulated phase (called *true phase*). *Other figures*: Phases reconstructed by maximum likelihood (ML) by varying the number of reconstructed modes

This example clearly illustrates the case where, although the matrix $H^{\mathrm{T}}H$ can be inverted, it is ill-conditioned, i.e., it has eigenvalues very close to 0, leading to uncontrolled noise amplification.

Truncating the solution space to lower values of $k_{\max}^{\mathrm{rec}}$ produces more reasonable solutions, but on the one hand the optimal choice of $k_{\max}^{\mathrm{rec}}$ depends on the levels of turbulence and noise and this adjustment is difficult to achieve in practice, while on the other hand this truncation introduces a modelling error, since it amounts to neglecting all the components of the turbulent phase beyond $k_{\max}^{\mathrm{rec}}$. We shall see below how the introduction of a priori knowledge regarding the spatial regularity of the phase (Kolmogorov spectrum) can lead to a better solution without this *ad hoc* reduction of the space of unknowns.

Interpreting the Failure of Non-Regularised Methods

The failure of the maximum likelihood method illustrated in the above example may seem surprising given the good statistical properties of this method, which is an estimator that converges towards the true value of the parameters when the number of data tends to infinity, which is asymptotically efficient, etc. However, these good properties are only asymptotic, i.e., they only concern situations where there is good *statistical contrast*, defined simply as the ratio of the number of measurements to the number of unknowns.

In practice, and in particular in the previous example, the problem is often to estimate a number of parameters of the same order of magnitude as the number of measurements, and sometimes greater, in which case these asymptotic properties are of little help. In this commonly encountered situation of unfavourable statistical contrast, the inversion is unstable, i.e., highly sensitive to noise, which can often be interpreted as arising due to the ill-posed nature of the underlying infinite dimensional problem.

Image restoration is another typical case of this situation. Obviously, for an image containing $N \times N$ pixels, we try to reconstruct an object of the same size, and if we increase the image size, we also increase the number of unknown parameters relating to the object, without improving the statistical contrast, which remains of order unity. We have seen that the least-squares solution was in this case given by the inverse filter, and we know that this is highly unstable with regard to noise. This instability is easily understood by reinserting the measurement equation (9.76), in the Fourier domain, into the solution (9.80):

$$\tilde{\hat{o}}_{\mathrm{LS}} = \frac{\tilde{i}}{\tilde{h}} = \tilde{o} + \frac{\tilde{m}}{\tilde{h}}. \tag{9.88}$$

According to this, it is clear that the noise is highly amplified for all frequencies for which the value of the transfer function $\tilde{h}$ is close to zero. One way to reduce noise amplification would be to modify (9.80) and divide $\tilde{i}$ by $\tilde{h}$ only for frequencies u where the transfer function $\tilde{h}$ is not too small. This is also a form of regularisation by controlling the dimension of the solution, very similar to the choice of $k_{\mathrm{max}}^{\mathrm{rec}}$ when reconstructing a wavefront, and it suffers from the same ad hoc character and the same difficulty in its tuning.

To sum up, simple inversion methods like least squares or maximum likelihood only give satisfactory results if there is good statistical contrast. For example, maximum likelihood can be successfully applied to problems like image registration, where we seek a 2D vector in order to register two images comprising a large number of pixels. More generally, it applies to the search for the few variables of a parsimonious parametric model from a large enough data set, e.g., estimating a star's diameter from visibilities in optical interferometry, estimating the Earth's ellipticity from arc measurements, etc.

In many problems where the statistical contrast is not favourable, the problem is ill-conditioned and regularisation, i.e., the addition during inversion of a priori knowledge and constraints on the solution, becomes highly profitable, as we shall now show.

9.6.4 Inversion Methods with Regularisation

Regularisation of an inverse problem corresponds to the idea that the data alone cannot lead to an acceptable solution, so that a priori information about the regularity of the considered observed object must necessarily be introduced. Here the term *regularity* implies that, for physical reasons intrinsic to the object, it must

have certain properties, or obey certain rules regarding sign, size, or frequency, for example. The solution then results from a compromise between the requirements of the object's postulated regularity and of the data fidelity.

Indeed, several very different solutions, some very poor and some rather good, may be compatible with the data. For instance, in the previous example of wavefront reconstruction, the true and reconstructed phases for $k_{\max}^{\text{rec}} = 210$ (Fig. 9.15) give very similar likelihood values, whence they are both faithful to the data. In addition, the *smoother*, more regular solution obtained for $k_{\max}^{\text{rec}} = 55$ is less well fitted, i.e., less faithful to the data than the one obtained for $k_{\max}^{\text{rec}} = 210$, since it was obtained by optimising in fewer degrees of freedom, and yet it is much closer to the true phase.

Bayesian Estimation and the Maximum a Posteriori (MAP) Estimator

Bayesian estimation, presented briefly here, provides a natural way of combining the information brought by measurements and information available a priori. Let us assume that we have expressed our a priori knowledge of the observed object in the form of a probability distribution $p(\boldsymbol{o})$, called the a priori distribution. The idea is not to assume that the observed object *really* is the realisation of a random phenomenon with distribution $p(\boldsymbol{o})$. This distribution is simply taken to be representative of our a priori knowledge, i.e., it takes small values for reconstructed objects that would be barely compatible with the latter and larger values for highly compatible objects.

The Bayes rule[23] provides a way of expressing the probability $p(\boldsymbol{o} \mid \boldsymbol{i})$ for the object $\boldsymbol{o}$ that is conditional upon the measurement data $\boldsymbol{i}$ as a function of the *prior probability distribution* $p(\boldsymbol{o})$ of the object and the probability $p(\boldsymbol{i} \mid \boldsymbol{o})$ of the measurements conditional upon the object. The latter contains our knowledge of the data formation model, including the noise model. This probability $p(\boldsymbol{o} \mid \boldsymbol{i})$ is called the *posterior distribution*, since it is the probability of the object given the results of the measurements. The Bayes rule is (see Sect. 9.5)

$$p(\boldsymbol{o} \mid \boldsymbol{i}) = \frac{p(\boldsymbol{i} \mid \boldsymbol{o}) \times p(\boldsymbol{o})}{p(\boldsymbol{i})} \propto p(\boldsymbol{i} \mid \boldsymbol{o}) \times p(\boldsymbol{o}). \tag{9.89}$$

Using the last expression, how should we choose the reconstructed object $\hat{\boldsymbol{o}}$ that would best estimate the true object? One commonly made choice is the maximum a posteriori (MAP) estimator. The idea is to define as solution the object that maximises the posterior distribution $p(\boldsymbol{o}|\boldsymbol{i})$, so that

[23] Thomas Bayes (1702–1761) was a British mathematician whose main work concerning inverse probabilities was only published after his death.

$$\boldsymbol{o}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{o}} p(\boldsymbol{o} \mid \boldsymbol{i}) = \arg\max_{\boldsymbol{o}} p(\boldsymbol{i} \mid \boldsymbol{o}) \times p(\boldsymbol{o}). \tag{9.90}$$

This is the most likely object given the data and our prior knowledge. The posterior distribution, viewed as a function the object, is called the *posterior likelihood*. Through (9.89), it takes into account the measurement data $\boldsymbol{i}$, the data formation model, and any prior knowledge of the object. In particular, it should be noted that, when the prior distribution $p(\boldsymbol{o})$ is constant, i.e., when we have no information about the object, the MAP estimator reduces to the maximum likelihood method.

It can be shown that choosing the MAP estimator amounts to minimising the mean risk (see Sect. 9.5) for a particular cost function called the *all-or-nothing cost function*. However, this goes beyond the scope of this introduction to inverse problems. Other choices of cost function, such as a choice of $\hat{\boldsymbol{o}}$ minimising the mean squared error with respect to the true object $\boldsymbol{o}$ under the posterior probability distribution, can also be envisaged, but they generally lead to longer computation times for obtaining the solution.[24]

Equivalence with Minimisation of a Regularised Criterion

Maximising the posterior likelihood is equivalent to minimising a criterion $J_{\mathrm{MAP}}(\boldsymbol{o})$ defined as minus its logarithm. According to (9.90), this criterion can be written as the sum of two terms, viz.,

$$J_{\mathrm{MAP}}(\boldsymbol{o}) = -\ln p(\boldsymbol{i} \mid \boldsymbol{o}) - \ln p(\boldsymbol{o}) = J_{\mathrm{i}}(\boldsymbol{o}) + J_{\mathrm{o}}(\boldsymbol{o}), \tag{9.91}$$

where J_{i} is the data fidelity criterion deduced from the likelihood [see (9.84)], which is often a least-squares criterion, while $J_{\mathrm{o}}(\boldsymbol{o}) \triangleq -\ln p(\boldsymbol{o})$ is a regularisation or penalisation criterion (for the likelihood) which reflects faithfulness to prior knowledge.

The expression for the MAP solution as the object that minimises the criterion (9.91) shows clearly that it achieves the compromise between faithfulness to the data and faithfulness to prior knowledge asserted at the beginning of this section.

When $\boldsymbol{o}$ is not a realisation of a random phenomenon with some probability distribution $p(\boldsymbol{o})$, e.g., in the case of image restoration, $J_{\mathrm{o}}(\boldsymbol{o})$ generally includes a multiplicative factor called the regularisation coefficient or hyperparameter.[25] Its value controls the exact position of the compromise. Unsupervised, i.e., automatic,

[24] A good reference for the theory of estimation is Lehmann, E. *Theory of Point Estimation*, John Wiley, 1983.

[25] The term *hyperparameter* refers to all the parameters of the data model (e.g., the noise variance) or of the a priori model (e.g., atmospheric turbulence and hence the parameter r_0) which are kept fixed during the inversion.

fitting methods exist for this coefficient, but they go beyond the scope of this introductory account.[26]

The Linear and Gaussian Case. Relation with the Wiener Filter

Here we consider the case where the data model is linear [see (9.75)], the noise is assumed to be Gaussian, and the prior probability distribution adopted for the object is also Gaussian,[27] with mean $\overline{\boldsymbol{o}}$ and covariance matrix C_{o} :

$$p(\boldsymbol{o}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{o}-\overline{\boldsymbol{o}})^{\mathrm{T}} C_{\mathrm{o}}^{-1}(\boldsymbol{o}-\overline{\boldsymbol{o}})\right]. \tag{9.92}$$

Using (9.85), we see that the criterion to minimise in order to obtain the MAP solution is

$$J_{\mathrm{MAP}}(\boldsymbol{o}) = \frac{1}{2}(\boldsymbol{i}-H\boldsymbol{o})^{\mathrm{T}} C_{\mathrm{n}}^{-1}(\boldsymbol{i}-H\boldsymbol{o}) + \frac{1}{2}(\boldsymbol{o}-\overline{\boldsymbol{o}})^{\mathrm{T}} C_{\mathrm{o}}^{-1}(\boldsymbol{o}-\overline{\boldsymbol{o}}). \tag{9.93}$$

This criterion is quadratic in $\boldsymbol{o}$ and thus has analytic minimum, obtained by setting the gradient equal to zero:

$$\hat{\boldsymbol{o}}_{\mathrm{MAP}} = \left(H^{\mathrm{T}} C_{\mathrm{n}}^{-1} H + C_{\mathrm{o}}^{-1}\right)^{-1}\left(H^{\mathrm{T}} C_{\mathrm{n}}^{-1}\boldsymbol{i} + C_{\mathrm{o}}^{-1}\overline{\boldsymbol{o}}\right). \tag{9.94}$$

Several remarks can throw light upon this otherwise slightly dull looking result. To begin with, the maximum likelihood solution is obtained by taking $C_{\mathrm{o}}^{-1} = 0$ in this equation. Incidentally, this shows that this solution corresponds in the Bayesian framework to assuming an infinite energy for $\boldsymbol{o}$. Then if we also take C_{n} proportional to the identity matrix, we see that we recover precisely the least-squares solution of (9.79).

Finally, the case of deconvolution is particularly enlightening. The noise is assumed stationary with power spectral density (PSD) S_{n}, while the prior probability distribution for the object is also assumed stationary with PSD S_{o}. For all the relevant quantities, we make the same periodicity approximation as on p. 583 when examining the relation between least squares and inverse filter. All the matrices in (9.94) are then diagonalised in the same discrete Fourier basis, and the solution can be written in this basis, with ordinary rather than matrix multiplications and inverses:

---

[26] See, for example, Idier, J. op. cit, Chaps. 2 and 3.

[27] There is no assumption here that the *shape* of the object is Gaussian, but only that its *probability distribution* is Gaussian. This happens, for example, when $\boldsymbol{o}$ is a turbulent phase obeying Kolmogorov statistics (Sect. 2.6). The central limit theorem (see Appendix B) often leads to Gaussian random objects when the perturbation of the object is due to many independent causes.

$$\tilde{\hat{o}}_{\mathrm{MAP}}(\boldsymbol{u}) = \frac{\tilde{\boldsymbol{h}}^*(\boldsymbol{u})}{|\tilde{\boldsymbol{h}}|^2(\boldsymbol{u}) + S_\mathrm{n}/S_\mathrm{o}(\boldsymbol{u})}\tilde{\boldsymbol{i}}(\boldsymbol{u}) + \frac{S_\mathrm{n}/S_\mathrm{o}(\boldsymbol{u})}{|\tilde{\boldsymbol{h}}|^2(\boldsymbol{u}) + S_\mathrm{n}/S_\mathrm{o}(\boldsymbol{u})}\tilde{\overline{\boldsymbol{o}}}(\boldsymbol{u}). \tag{9.95}$$

In this expression, $S_\mathrm{n}/S_\mathrm{o}(\boldsymbol{u})$ is the reciprocal of a signal-to-noise ratio at the spatial frequency $\boldsymbol{u}$ and $\tilde{\overline{\boldsymbol{o}}}$ is the Fourier transform of the a priori object, generally taken to be zero or equal to an object of constant value.

This expression is just the *Wiener filter* (see Sect. 9.1.3) for the case where the a priori object is not zero. For frequencies where the signal-to-noise ratio is high, this solution tends toward the inverse filter, and for frequencies where this ratio is low, the solution tends toward the a priori object. It can even be seen that, at each spatial frequency $\boldsymbol{u}$, the solution is on a segment that connects the maximum likelihood solution (inverse filter) to the a priori solution, with the position on the segment given by the signal-to-noise ratio:

$$\tilde{\boldsymbol{o}}_{\mathrm{MAP}}(\boldsymbol{u}) = \alpha\frac{\tilde{\boldsymbol{i}}(\boldsymbol{u})}{\tilde{\boldsymbol{h}}} + (1-\alpha)\tilde{\overline{\boldsymbol{o}}}(\boldsymbol{u}), \tag{9.96}$$

where

$$\alpha = \frac{|\tilde{\boldsymbol{h}}|^2(\boldsymbol{u})}{|\tilde{\boldsymbol{h}}|^2(\boldsymbol{u}) + S_\mathrm{n}/S_\mathrm{o}(\boldsymbol{u})}. \tag{9.97}$$

This expression clearly illustrates the fact that regularisation achieves a compromise between bias and variance that enables us to reduce the estimation error. The mean squared error on the estimated object is the square root of the sum of the squared bias and the variance of the estimator. The solution of the inverse filter (obtained for $S_\mathrm{n}/S_\mathrm{o} = 0$) has zero bias but amplifies the noise in an uncontrolled way, i.e., has unbounded variance. Compared with the maximum likelihood solution, the solution (9.95) is biased toward the a priori solution $\overline{\boldsymbol{o}}$. By accepting this bias, we can significantly reduce the variance and globally reduce the mean squared error on the estimated object.

Application to Wavefront Reconstruction

We return to the example of wavefront reconstruction from data provided by a Hartmann–Shack sensor, as discussed in Sect. 9.6.3. The noise is still assumed to be Gaussian and white with covariance matrix $C_\mathrm{n} = \sigma^2 I$. We assume that the phase obeys Kolmogorov statistics and is therefore Gaussian, with known covariance matrix C_o, and depends only on the Fried parameter r_0 which quantifies the strength of turbulence. The true phase has spatial variance $\sigma_\varphi^2 = 3.0\ \mathrm{rad}^2$. The most likely phase given the measurements and taking into account this a priori information is the MAP solution given by (9.94). This solution is shown in Fig. 9.16. The estimation error corresponds to a spatial variance $\sigma_\mathrm{err}^2 = 0.7\ \mathrm{rad}^2$, which is lower than the best solutions obtained previously by truncating the representation of the phase to a small number of Zernike polynomials ($\sigma_\mathrm{err}^2 = 0.8\ \mathrm{rad}^2$, obtained for $k_\mathrm{max}^\mathrm{rec} = 55$).

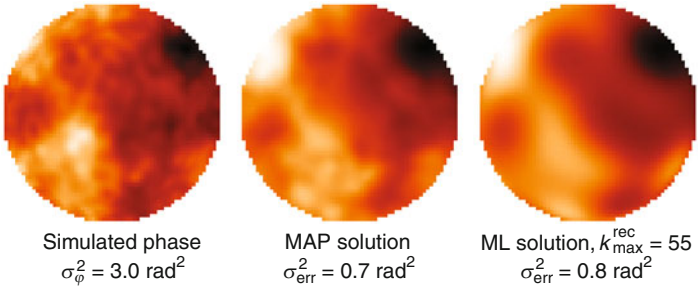

Fig. 9.16 Comparing reconstructed phases on a pupil. *Left*: Simulated turbulent phase, to be estimated. *Centre*: Phase estimated by MAP. *Right*: Best phase estimated by maximum likelihood on a truncated basis

The MAP solution takes advantage of a priori knowledge of the spatial statistics of the turbulence. To use this in adaptive optics, where the sample rate is generally well above $1/\tau_0$, it is judicious to opt for a natural extension of this estimator that also exploits a priori knowledge of the temporal statistics of the turbulence. This extension is the optimal estimator of Kalman filtering.

9.6.5 Application to Adaptive Optics Imaging

Here we apply these tools to solve a specific inverse problem, namely, adaptive optics imaging. We shall illustrate with several examples, either by simulation or by the processing of genuine astronomical observations.

Ingredients of the Deconvolution

Long exposure images corrected by adaptive optics (AO) must be deconvolved, because the correction is only partial. Considering that the PSF $\boldsymbol{h}$ is known, the object $\hat{\boldsymbol{o}}_{\text{MAP}}$ estimated by MAP is given by (9.90), i.e., it minimises the criterion (9.91). Let us see how to define the two components of this criterion.

To be able to measure objects with high dynamic range, which are common in astronomy, the data fidelity term J_{i} must incorporate a fine model of the noise, taking into account both photon noise and electronic noise. This can be done by treating the photon noise as a non-stationary Gaussian noise, and it leads to a weighted least-squares criterion J_{i} [see (9.86)] rather than an ordinary least-squares term.

For objects with sharp edges, such as artificial satellites, asteroids, or planets, a Gaussian prior (like the one leading to the Wiener filter on p. 590), or equivalently, a quadratic regularisation criterion, tends to smooth the edges and introduce spurious oscillations or *ringing* in their vicinity. One interpretation of this effect is that, when minimising the criterion $J_{\text{MAP}}(\boldsymbol{o})$, a quadratic

regularisation attributes to a step a cost proportional to the square of its value, e.g., at the edge of an object, where there is a large difference in value between adjacent pixels. One solution is then to use an *edge-preserving criterion*, such as the so-called quadratic–linear or L_2L_1 criteria. These are quadratic for small discontinuities and linear for large ones. The quadratic part ensures good noise smoothing and the linear part cancels edge penalisation.

In addition, for many different reasons, we are often led to treat the PSF $\boldsymbol{h}$ as imperfectly known. Carrying out a *classic* deconvolution, i.e., assuming that the point spread function is known but using an incorrect point spread function, can lead to disastrous results. Conversely, a so-called *blind* deconvolution, where the same criterion (9.91) is minimised but simultaneously seeking $\boldsymbol{o}$ and $\boldsymbol{h}$, is highly unstable, rather like unregularised methods. A *myopic deconvolution* consists in jointly estimating both $\boldsymbol{o}$ and $\boldsymbol{h}$ in a Bayesian framework with a natural regularisation for the point spread function and without having to fit an additional hyperparameter. The joint MAP estimator is given by

$$(\hat{\boldsymbol{o}}, \hat{\boldsymbol{h}}) = \arg\max_{\boldsymbol{o},\boldsymbol{h}} p(\boldsymbol{o}, \boldsymbol{h}|\boldsymbol{i}) = \arg\max_{\boldsymbol{o},\boldsymbol{h}} p(\boldsymbol{i}|\boldsymbol{o}, \boldsymbol{h}) \times p(\boldsymbol{o}) \times p(\boldsymbol{h})$$

$$= \arg\min_{\boldsymbol{o},\boldsymbol{h}} \Big[J_i(\boldsymbol{o}, \boldsymbol{h}) + J_o(\boldsymbol{o}) + J_h(\boldsymbol{h}) \Big],$$

where J_h is a regularisation criterion for $\boldsymbol{h}$, which introduces constraints on the possible variability of the PSF.

The next section presents experimental results obtained by the MISTRAL restoration method,[28] which combines the three ingredients discussed above: fine noise modelling, non-quadratic regularisation, and myopic deconvolution.

Image Restoration from Experimental Astronomical Data

Figure 9.17a shows a long exposure image of Ganymede, a natural satellite of Jupiter, corrected by adaptive optics. This image was made on 28/09/1997 on ONERA's adaptive optics testbed installed on the 1.52 m telescope at the Observatoire de Haute-Provence in France. The imaging wavelength is $\lambda = 0.85$ μm and the exposure time is 100 s. The estimated total flux is 8×10^7 photons and the estimated ratio D/r_0 is 23. The total field of view is 7.9 arcsec, of which only half is shown here. The mean and variability of the point spread function were estimated from recordings of 50 images of a bright star located nearby. Figures 9.17b and c show restorations obtained using the Richardson–Lucy algorithm (ML for Poisson

[28] This stands for *Myopic Iterative STep-preserving Restoration ALgorithm*. This method is described in Mugnier, L.M. et al., MISTRAL: A myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images, J. Opt. Soc. Am. A **21**, 1841–1854, 2004.

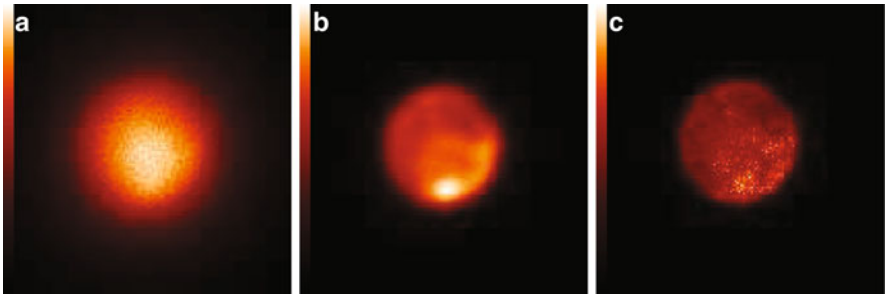

Fig. 9.17 (**a**) Corrected image of Ganymede obtained using the ONERA adaptive optics testbed on 28 September 1997. (**b**) Restoration using the Richardson–Lucy algorithm, stopped after 200 iterations. (**c**) Likewise, but stopped after 3 000 iterations. From Mugnier, L. et al., Chap. 10 of Idier, J. op. cit.

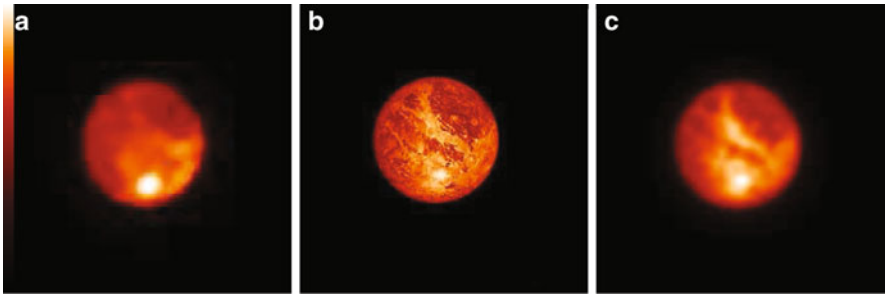

Fig. 9.18 (**a**) Deconvolution of the Ganymede image in Fig. 9.17 by MISTRAL. (**b**) Comparison with a wide band synthetic image obtained using the NASA/JPL database. (**c**) The same synthetic image, but convolved by the perfect point spread function of a 1.52 m telescope. From Mugnier, L.M. et al., MISTRAL: A myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images, J. Opt. Soc. Am. A **21**, 1841–1854, 2004

noise), stopped after 200 and 3 000 iterations, respectively.[29] In the first case, similar to restoration with quadratic regularisation, the restored image is somewhat blurred and displays ringing, while in the second, very similar to the result of inverse filtering, noise dominates the restoration.

Figure 9.18a shows a myopic deconvolution implementing an edge-preserving prior, while Fig. 9.18b is a wide band synthetic image obtained by a NASA/JPL space probe[30] during a Ganymede flyby. Comparing the two, we find that many

[29]The idea of stopping a non-regularised algorithm before convergence is still a widespread method of regularisation, but extremely ad hoc.

[30]See space.jpl.nasa.gov/ for more details.

features of the moon are correctly restored. A fairer comparison consists in jointly examining the myopic deconvolution carried out by MISTRAL with the image of Fig. 9.18b convolved by the perfect PSF of a 1.52 m telescope, as shown in Fig. 9.18c.

9.6.6 Application to Nulling Interferometry

We now discuss a second example of the inversion problem, this time relating to the detection of extrasolar planets by means of a nulling interferometer (see Sect. 6.6). With the Darwin instrument, or NASA's *Terrestrial Planet Finder Interferometer*, both under study during the 2000s, data will be very different from an image in the conventional sense of the term, and their exploitation will require implementation of a specific reconstruction process. They will consist, at each measurement time t, of an intensity in each spectral channel λ. This intensity can be modelled as the integral over a certain angular region of the instantaneous transmission map of the interferometer, denoted by $R_{t,\lambda}(\boldsymbol{\theta})$, which depends on the time t owing to rotation of the interferometer relative to the plane of the sky, multiplied by the intensity distribution $o_\lambda(\boldsymbol{\theta})$ of the observed object. The data model is thus linear, but notably non-convolutive, thus very different from the one used in imaging. The transmission map is a simple sinusoidal function in the case of a Bracewell interferometer, but becomes more complex when more than two telescopes interfere simultaneously.

By judiciously combining the data, and with asymmetrical transmission maps, the contribution to the measured signal of the components of the observed object with even spatial distribution can be eliminated. These components are stellar leakage, exozodiacal light, and a fortiori zodiacal light and thermal emission from the instrument itself (which have constant level in the field of view). It is then possible to seek out only planets during image reconstruction, which corresponds to the following object model:

$$\boldsymbol{o}_\lambda(\boldsymbol{\theta}) = \sum_{k=1}^{N_{\text{src}}} F_{k,\lambda}\,\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k), \tag{9.98}$$

where N_{src} is the number of planets, assumed known here, and $F_{k,\lambda}$ is the spectrum of the k th planet in a spectral interval $[\lambda_{\min}, \lambda_{\max}]$ fixed by the instrument. This parametric model can substantially constrain the inversion in so as to counterbalance the fact that the data are distinctly poorer than an image.

With this model of the object, the data formation model is

$$i_{t,\lambda} = \sum_{k=1}^{N_{\text{src}}} R_{t,\lambda}(\boldsymbol{\theta}_k) F_{k,\lambda} + n_{t,\lambda}, \tag{9.99}$$

where $n_{t,\lambda}$ is the noise, assumed to be white Gaussian, whose variance $\sigma^2_{t,\lambda}$ can be estimated from the data and is assumed known here. The inverse problem to

be solved is to estimate the positions $\boldsymbol{\theta}_k$ and spectra $F_{k,\lambda}$ of the planets, these being grouped together into two vectors $(\boldsymbol{\theta}, \boldsymbol{F})$. The ML solution is the one that minimises the following weighted least-squares criterion, given the assumptions made about the noise:

$$J_{\mathrm{i}}(\boldsymbol{\theta}, \boldsymbol{F}) = \sum_{t,\lambda} \frac{1}{\sigma_{t,\lambda}^2} \left[i_{t,\lambda} - \sum_{k=1}^{N_{\mathrm{src}}} R_{t,\lambda}(\boldsymbol{\theta}_k) F_{k,\lambda} \right]^2 . \tag{9.100}$$

As we shall see from the reconstruction results, the inversion remains difficult under the high noise conditions considered here. The object model (9.98), separable into spatial and spectral variables, already contains all spatial prior information concerning the object. It is nevertheless possible to constrain the inversion even more by including the further knowledge that the spectra we seek are positive quantities (at all wavelengths), and furthermore that they are relatively *smooth functions* of the wavelength. The latter fact is taken into account by incorporating a spectral regularisation into the criterion to be minimised, which measures the roughness of the spectrum:

$$J_{\mathrm{o}}(\boldsymbol{F}) = \sum_{k=1}^{N_{\mathrm{src}}} \mu_k \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left(\frac{\partial^m F_{k,\lambda}}{\partial \lambda^m} \right)^2 , \tag{9.101}$$

where the m th derivative of the spectrum ($m = 1$ or 2 in practice) is calculated by finite differences and where the μ_k are hyperparameters used to adjust the weight allocated to the regularisation. The MAP solution is the one minimising the composite criterion $J_{\mathrm{MAP}}(\boldsymbol{\theta}, \boldsymbol{F}) = J_{\mathrm{i}}(\boldsymbol{\theta}, \boldsymbol{F}) + J_{\mathrm{o}}(\boldsymbol{F})$. It is a rather delicate matter to implement this minimisation because there are many local minima. We use the fact that, for each assumed position $\boldsymbol{\theta}$ of the planets, the MAP estimate of the spectra, $\hat{\boldsymbol{F}}(\boldsymbol{\theta})$, can be obtained simply because J_{MAP} is quadratic in the spectra $\boldsymbol{F}$. If the latter are replaced by $\hat{\boldsymbol{F}}(\boldsymbol{\theta})$ in J_{MAP}, we obtain a partially optimised function for minimisation, which now only depends explicitly on the positions:

$$J^{\dagger}_{\mathrm{MAP}}(\boldsymbol{\theta}) = J_{\mathrm{MAP}}\big[\boldsymbol{\theta}, \hat{\boldsymbol{F}}(\boldsymbol{\theta})\big]. \tag{9.102}$$

This criterion is minimised by a sequential search for the planets, as in the CLEAN algorithm. Figure 9.19 shows the maps of $J^{\dagger}_{\mathrm{MAP}}$ obtained for a single planet as a function of the prior information used. It is clear that the constraints of positivity and smoothness imposed on the spectra significantly improve estimates of the position of the planet, by discrediting (Fig. 9.19 right) positions compatible with the data (Fig. 9.19 left and centre) but corresponding to highly *chaotic* spectra.

Figure 9.20 shows the estimated spectrum of an Earth-like planet. As expected, spectral regularisation avoids noise amplification and has a beneficial effect on the estimate.

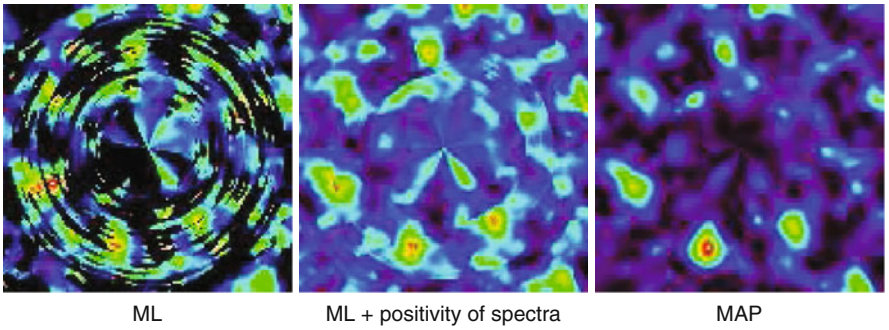

Fig. 9.19 Likelihood maps for the position of a planet. *Left*: Likelihood alone. *Centre*: Likelihood under the constraint that spectra are positive. *Right*: MAP, i.e., likelihood penalised by a spectral regularisation criterion. The true position of the planet is at the bottom, slightly to the left, and clearly visible on the right-hand image. From Mugnier, L., Thiébaut, E., Belu, A., in *Astronomy with High Contrast Imaging III*, EDP Sciences, Les Ulis, 2006, and also Thiébaut, E., Mugnier, L., Maximum a posteriori planet detection with a nulling interferometer, in Proceedings IAU Conf. 200, Nice, 2006

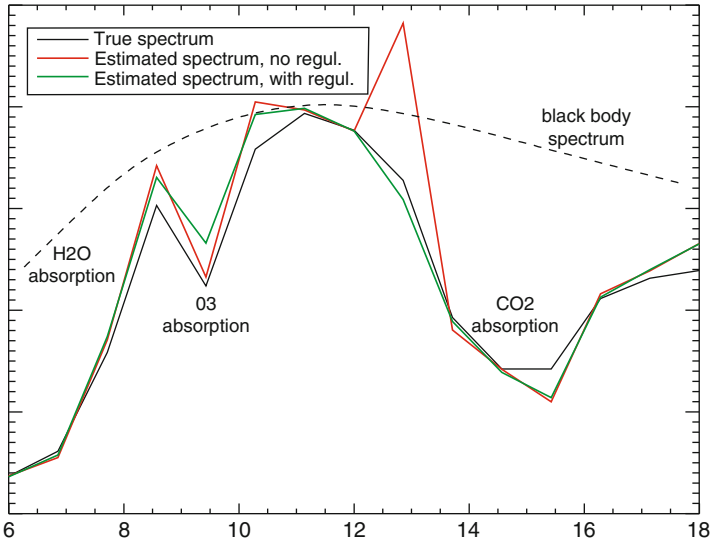

Fig. 9.20 Reconstructed spectrum at the estimated position of the planet. *Red*: Without regularisation. *Green*: With regularisation. *Black*: True spectrum of the Earth. Same reference as for Fig. 9.19