E. WOLF, PROGRESS IN OPTICS VVV

X

SUPER-RESOLUTION BY DATA INVERSION

BY

Mario Bertero(*) and Christine De Mol(+)

(*) *Dipartimento di Fisica, Università di Genova*
Via Dodecaneso 33, I–16146 Genova, Italy
(+) *Département de Mathématique,*
Université Libre de Bruxelles, Campus Plaine CP 217
Boulevard du Triomphe, B–1050 Bruxelles, Belgium

CONTENTS*

* Typeset in Plain TEX with the macros provided by Elsevier.

§ 1. Introduction

Super-resolution is a widely-used keyword, which appears in the literature in several contexts and with quite different significations. Unfortunately, this has contributed to create some confusion about the corresponding concept. In the present tutorial paper, we are trying to clarify the situation by giving a precise meaning to the word 'super-resolution' in some specific and well-defined case-examples.

The conventional concept of resolving power of an optical instrument dates back to the end of last century with the well-known classical work by Abbe and Rayleigh. In §2, we briefly recall how the classical Rayleigh resolution limit is introduced and defined in terms of the overlap between the images of two point sources. Next we show that, in the framework of modern Fourier optics, the resolving power is characterized instead by specifying the spatial-frequency band associated with the instrument. Indeed, due to diffraction effects, the set of frequencies transmitted by an optical system is confined to some finite region in Fourier space, called the band of the system. A link between the two viewpoints is provided by the sampling theorems for bandlimited functions, which we also recall in §2.

In the last decades, the development of micro-informatics and of computer-assisted instruments has in fact deeply modified the classical concept of resolving power. Indeed, in many modern optical devices, numerical algorithms can be implemented to process the recorded data, in order to get a better representation of the probed object. Such 'reconstructions' of the object may eventually allow to resolve finer details than those visible in the unprocessed image. The corresponding enhancement in resolution is often called 'super-resolution', although we will adopt in the following a slightly more restrictive definition of this word. We suppose to know a good model for describing the imaging process in a given instrument. In other words, we assume that we can write down an explicit mathematical relationship between any object under examination and its image produced by the instrument. The determination of the image of a given object according to such model is called the 'direct' imaging problem. Conversely, the recovery or 'restoration' of the object from its image is called the 'inverse' imaging problem. As usual in Fourier optics, we assume moreover that the imaging model is linear, so that the corresponding direct and inverse problems are linear as well. This means in particular that we do not consider here any partial-coherence effects or nonlinear inverse problems such as the well-known 'phase problem', which consists in retrieving the phase of an object with known modulus.

For the sake of simplicity, we split the inverse imaging problem into two parts and we analyze separately each of the two sub-problems. The first one, which we call 'deblurring', consists in recovering the Fourier spectrum of the object inside the band, i.e. for those frequency components which are transmitted by the instrument. Deblurring allows to correct e.g. for the effect of aberrations, which are responsible for phase distortion and attenuation inside the band. *The second step, for which we reserve the denomination of 'super-resolution', is the attempt at recovering the object spectrum outside the band of the instrument or, in other words, at restoring the object beyond the diffraction limit.* As we will see later, this step involves an extrapolation of the object spectrum outside the band. Such an extrapolation appears to be feasible under some assumptions, such as the analyticity of the object spectrum, which holds e.g. for an object vanishing outside some finite region in space. Notice that both steps require to process – more precisely to 'invert' – the image data in order to get estimates of the object spectrum inside or outside the band.

The aim of the present study is to analyze the resolution enhancement that can be achieved by such data inversion processes and to quantify the corresponding gain in resolving power. The natural framework for investigating this question is provided by the so-called 'linear inversion theory', which is a collection of methods, referred to as 'regularization' methods, allowing to define stable solutions of linear inverse imaging problems. In §3, we recall some important features of such inversion methods, to the extent that they relate to the assessment of resolution limits. The main difficulty encountered in solving inverse problems is their sensitivity to noise in the data, which can be the source of major instabilities in the solutions. The role of regularization is to prevent such instabilities to occur. However, the price to pay for stability is that we can only recover the solution with a limited resolution. In this framework, the resolution limits arise as a practical limitation imposed by the necessity to control the noise amplification inherent in all inversion procedures. Accordingly, the resolving power appears no longer as a purely intrinsic characteristic of the instrument itself, but rather as a combined property of the hardware (the optical components) and of the implemented inversion software. The concept of 'overall impulse response' introduced in §3.1 allows to take both aspects into account and to characterize the ultimate resolution capabilities of the instrument. In §3.2, we introduce an optimal filtering method, known as the Wiener filter, for the solution of convolution equations. We apply it to the deblurring problem and we show that the object spectrum can only be recovered on some 'effective' band, which depends on the level of noise in the recorded data through a parameter called the signal-to-noise ratio. This effective band is in general smaller than the band of the instrument, so that no super-resolution can be achieved in such a case. To recover the object spectrum outside the effective band, one has to take into account a priori information about the localization of the object. This is done in §3.3 where we analyze some well-known inversion methods in terms of the singular system of the imaging operator. Indeed, because of the object localization, this operator is no longer a pure convolution operator but instead a compact integral operator. The corresponding overall impulse response allows to assess the achievable resolution limits. Accordingly, we show that the resolving power depends on the signal-to-noise ratio and, to a lesser extent, on the choice of the data inversion algorithm. We also define the useful concept of 'number of degrees of freedom', which generalizes the Shannon number used in information theory.

In the following sections, we consider in more detail three particular problems which can be viewed as paradigms for super-resolution problems. In §4, we analyze the problem of extrapolating the object spectrum outside the band – or the effective band – under the assumption that the object vanishes outside some finite known domain. We show how to assess the amount of achievable super-resolution and we are led to the conclusion that a significant resolution enhancement with respect to the Rayleigh limit is obtained only in the case of a very low space-bandwidth product, or equivalently of a very small number of degrees of freedom. In §5, we consider a situation where this condition is fulfilled in practice, namely the case of scanning microscopy. We focus on confocal microscopy and we show how the use of data inversion techniques allows to enhance the resolving power of such microscopes. Finally, in §6, we consider the problem of inverse diffraction from plane to plane, which consists in back-propagating towards the source plane a field propagating in free space. When the data plane is situated in the far-field region, the problem essentially reduces to that of out-of-band extrapolation considered in §4. Hence, no significant super-resolution is achievable, except when the space-bandwidth product is very low, which means that the source should be small compared to the wavelength of the illuminating field. Such conclusion, however, no longer holds when near-field data are available. We show that in such a case the effective bandwidth is significantly increased due to the presence of evanescent waves and we assess the improvement as a function of the signal-to-noise ratio and of the distance between the source and data planes. Near-field imaging techniques provide in fact an alternative way for beating the far-field diffraction limit, as demonstrated nowadays e.g. by the resolving capabilities of near-field scanning microscopes.

§ 2. Resolution limits and bandwidth

The resolving power of an optical device can be characterized in various ways. We show that, instead of the classical Rayleigh criterion recalled in §2.1, it is preferable to analyze and specify the band of the spatial frequencies transmitted by the system and to define the resolution limit on the basis of sampling theorems for bandlimited functions (see §2.2). Then we argue that because of noise the concept of band has to be replaced by the notion of 'effective band', and we define 'deblurring' as object restoration inside the effective band and 'super-resolution' as out-of-band extrapolation of the object spectrum (see §2.3).

2.1. THE RAYLEIGH RESOLUTION LIMIT

The resolving power of an imaging system is a measure of its ability to separate the images of two neighbouring points. According to geometrical optics, in the absence of aberrations, the image of a point source is a perfectly sharp point, and hence the resolving power is unlimited. However, because of diffraction effects, the image is never just a point, but instead a small light patch called the *diffraction pattern*. When two point sources get closer, their diffraction patterns start progressively overlapping, until it is no longer possible to discriminate between one or two point sources. The limit distance between the two sources down to which the discrimination can be done may depend in practice on many factors (including e.g. the sensitivity of the human eye), which can hardly be quantified. It appears nevertheless useful to have some simple and objective criteria to compare the performances of optical systems. Different criteria have been proposed to this purpose, the most famous being the Rayleigh criterion (see the original paper by Rayleigh [1879] or the book by Born and Wolf [1980]). According to the original version of this criterion, the diffraction patterns are considered as just resolved if the central maximum of the first coincides with the first minimum of the other. For the so-called Airy pattern, which is the Fraunhofer (far-field) intensity diffraction pattern at a circular aperture, the distance between the principal maximum and the first zero or dark ring is given by $R = 1.22\ (\lambda/2\alpha)$, where λ is the wavelength of the light and α is equal to the radius of the diffracting aperture divided by the distance between the aperture and image planes. The distance R is known as the *Rayleigh resolution limit* and $1/R$ as the *resolving power*. The above formula provides as well the resolution limit of a microscope if α denotes its numerical aperture (Abbe [1873]), and also the angular resolution limit of a telescope if α represents the radius of the objective aperture (see Born and Wolf [1980]).

2.2. BAND OF AN OPTICAL SYSTEM AND SAMPLING THEOREMS

Since the introduction of Fourier methods in optics, it has become more usual to characterize the resolution of an optical system in terms of its *bandwidth*. Diffraction effects are indeed responsible for the existence of a *cut-off frequency*, due to the fact that not all the spatial frequencies of the object are transmitted by the *pupil* of the instrument. The optical system is then said to be *bandlimited* or *diffraction-limited*. More precisely, in the framework of Fourier optics (see e.g. Born and Wolf [1980], Goodman [1968] or Papoulis [1968]), an optical system is viewed as a linear system, i.e. as a black box fully characterized by its impulse response. Hence the equation describing the imaging process is as follows

$$g(\mathbf{r}) = \int \mathrm{d}\mathbf{r}'\ S(\mathbf{r},\mathbf{r}')\ f(\mathbf{r}')\ . \tag{2.1}$$

The functions $f(\mathbf{r})$ and $g(\mathbf{r})$ are respectively the object (input of the system) and the image (output of the system); they represent scalar light amplitudes in the case of coherent imaging and light intensities in the case of incoherent imaging. In full generality, the spatial vector coordinate $\mathbf{r} = (x, y, z)$ is three-dimensional (3D), since it includes the depth coordinate z along the optical axis. In many instances, however, a two-dimensional (2D) description can be used, in terms of the vector $\boldsymbol{\rho} = (x, y)$ formed by the transverse cartesian coordinates in the object and image planes. The impulse response $S(\mathbf{r}, \mathbf{r}')$, which is assumed to be known, represents the image at point $\mathbf{r}$ of a point source situated at point $\mathbf{r}'$. For space-invariant ('isoplanatic') systems, this impulse response depends only on the difference of the variables $\mathbf{r}-\mathbf{r}'$ and the image $g(\mathbf{r})$ is then given by the following convolution integral

$$g(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S(\mathbf{r} - \mathbf{r}') \; f(\mathbf{r}') \; . \tag{2.2}$$

Notice that from now on, unless explicitly specified, the limits of integration are supposed to be infinite. The function $S(\mathbf{r})$ is usually called in optics the *point spread function* (PSF) and its Fourier transform $\widehat{S}(\mathbf{k})$ is called the *transfer function* (TF) of the optical system. We use the following definition of the Fourier transform

$$\widehat{S}(\mathbf{k}) = \int \mathrm{d}\mathbf{r} \; \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \; S(\mathbf{r}) \; . \tag{2.3}$$

In the case of an optical system consisting of well-corrected lenses ('ideal' system) and in the case of coherent illumination, $\widehat{S}(\mathbf{k})$ is the *pupil function* (except for a scaling of coordinates), i.e. the function equal to one inside the pupil and to zero outside. In mathematical terms, $\widehat{S}(\mathbf{k})$ is called the characteristic function of the pupil. In any case, and also for aberrated systems and incoherent illumination, because of diffraction, $\widehat{S}(\mathbf{k})$ vanishes outside some bounded set $\mathcal{B}$ called the band of $S(\mathbf{r})$ or the *band of the optical system.* In such a case the function $S(\mathbf{r})$ is said to be *bandlimited.* Then all images $g(\mathbf{r})$ are also bandlimited with the same band as $S(\mathbf{r})$, as follows from the Fourier convolution theorem:

$$\widehat{g}(\mathbf{k}) = \widehat{S}(\mathbf{k}) \; \widehat{f}(\mathbf{k}) \; . \tag{2.4}$$

Therefore, the sharpest details in the object, which correspond to high-frequency components lying outside the band, are smoothed off and no longer noticeable in the image, which appears as a blurred version of the object. Rayleigh's criterion can then be reinterpreted in the framework of communication and information theory (see Gabor [1961] and Toraldo di Francia [1969]), and related to the sampling theory for bandlimited functions. In particular, in one-dimensional (1D) situations, the Rayleigh resolution distance coincides with the Nyquist sampling distance. To clarify this point, we first consider a 1D ideal coherent imaging system, having as TF the characteristic function of the interval $[-K, +K]$ and hence as PSF $S(x) = \sin(Kx)/(\pi x)$. In such a case, according to Rayleigh's criterion, the resolution distance is $R = \pi/K$. On the other hand, we see from the 1D version of eq. (2.4) that the image has the same spatial *cut-off frequency* K as the PSF. One says then that the function $g(x)$ is bandlimited with *bandwidth* K (or in short, K-bandlimited). A basic result in information theory – widely used in optics, as discussed by Gabor [1961], – is the following sampling theorem, named after Whittaker [1915] and Shannon [1949] (see Jerri [1977] for a review): any K-bandlimited function can be represented by the sampling expansion

$$g(x) = \sum_{n=-\infty}^{+\infty} g\left(n\frac{\pi}{K}\right)\ \mathrm{sinc}\left[\frac{K}{\pi}\left(x - n\frac{\pi}{K}\right)\right] \tag{2.5}$$

where

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}\quad . \tag{2.6}$$

The theorem implies that any K-bandlimited function $g(x)$ can be represented by a sequence of its samples without any loss in information, provided that the samples are taken at equidistant points spaced by the distance π/K, called the *Nyquist distance*. We see that in this simple case the Nyquist and Rayleigh distances coincide.

Let us observe that the set of all the K-bandlimited functions is a linear subspace of the space of all square-integrable functions on the real line. In the mathematical literature, such a subspace is usually called a Paley-Wiener space of functions. Consider now a function $f(x)$ which is not bandlimited and let $f_K(x) = (B_K f)(x)$ be its projection on the subspace of the K-bandlimited functions. In Fourier space, the bandlimiting operator B_K acts by simply multiplying the spectrum $\widehat{f}(k)$ by the characteristic function of the band $[-K, +K]$. Hence by the convolution theorem, we get

$$f_K(x) = (B_K f)(x) = \int_{-\infty}^{+\infty} \mathrm{d}y \, \frac{K}{\pi} \, \mathrm{sinc}\left[\frac{K}{\pi}(x-y)\right] f(y) \; . \tag{2.7}$$

For the ideal system considered above, the image $g(x)$ is precisely the projection $f_K(x)$ of the object. We will say that knowing $f_K(x)$ is equivalent to knowing $f(x)$ within the *limit of resolution* $\Delta_x = \pi/K$.

For 2D images, the Rayleigh resolution distance is defined only in the case of a circular band, while sampling expansions can be derived for any geometrical shape of the band. We define the band $\mathcal{B}$ of a 2D image $g(\boldsymbol{\rho}) = g(x, y)$ as the support of its Fourier transform, i.e. the set of all the spatial frequencies $\boldsymbol{\kappa} = \{k_x, k_y\}$ such that $\widehat{g}(\boldsymbol{\kappa}) \neq 0$. We say that the function $g(\boldsymbol{\rho})$ is $\mathcal{B}$-bandlimited if $\mathcal{B}$ is a bounded subset of the frequency plane.

When the band $\mathcal{B}$ is the rectangle $|k_x| \leq K_x, |k_y| \leq K_y$, the sampling expansion (2.5) is replaced by the following one

$$g(x,y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} g\left(\frac{n\pi}{K_x}, \frac{m\pi}{K_y}\right) \mathrm{sinc}\left[\frac{K_x}{\pi}\left(x - \frac{n\pi}{K_x}\right)\right] \mathrm{sinc}\left[\frac{K_y}{\pi}\left(y - \frac{m\pi}{K_y}\right)\right] \; . \tag{2.8}$$

As in dimension one, a function $f(\boldsymbol{\rho})$ which is not bandlimited can be projected on the subspace of the $\mathcal{B}$-bandlimited functions and its projection is denoted by $f_{\mathcal{B}}(\boldsymbol{\rho})$. When we know that an image is bandlimited to a rectangle, the expansion (2.8) suggests the definition of two limits of resolution, one in the x-variable, $\Delta_x = \pi/K_x$, and one in the y-variable, $\Delta_y = \pi/K_y$.

More generally, one can introduce a direction-dependent resolution limit: given a unit vector $\boldsymbol{\theta} = \{\cos\phi, \sin\phi\}$, if $K_{\boldsymbol{\theta}}$ is the bandwidth of the 1D function $f_{\mathcal{B},\boldsymbol{\theta}}(s) = f_{\mathcal{B}}(s\boldsymbol{\theta})$, then the limit of resolution in the direction $\boldsymbol{\theta}$ is $\Delta_{\boldsymbol{\theta}} = \pi/K_{\boldsymbol{\theta}}$. It is easy to see that $K_{\boldsymbol{\theta}}$ is related to the length of the interval which is obtained by projecting $\mathcal{B}$ orthogonally on the direction $\boldsymbol{\theta}$ (see Appendix A). This results holds for any shape of the band $\mathcal{B}$. When $\mathcal{B}$ is the rectangle considered hereabove, $K_{\boldsymbol{\theta}} = K_x \cos\phi +$

$K_y \sin\phi$. When $\mathcal{B}$ is a disc of radius K, then $K_{\boldsymbol{\theta}} = K$ and the resolution limit $\Delta_{\boldsymbol{\theta}} = \pi/K$ does not depend on $\boldsymbol{\theta}$. Let us remark that this limit does not coincide exactly with the resolution distance $R = 1.22\,(\pi/K)$ prescribed by the Rayleigh criterion and resulting from the position of the first zero of the Fourier transform of the characteristic function of the disc of radius K.

For functions of two variables, it can also be useful to introduce the concept of *resolution cell*, which in the case of a rectangular band is the fundamental cell of the rectangular lattice characterized by the vectors $\boldsymbol{\rho}_x = (\pi/K_x)\boldsymbol{\theta}_x$ and $\boldsymbol{\rho}_y = (\pi/K_y)\boldsymbol{\theta}_y$, i.e. the rectangle with side lengths π/K_x and π/K_y, $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ being respectively the unit vectors along the x-axis and the y-axis. Notice that the number of resolution cells per unit area is the number of independent sampling values, per unit area, of the image $f_{\mathcal{B}}(\boldsymbol{\rho})$. In a sense, it is the number of independent information elements, per unit area, contained in the image $f_{\mathcal{B}}(\boldsymbol{\rho})$. When $\mathcal{B}$ is not a rectangle, it is always possible to find a rectangle of minimum area containing $\mathcal{B}$. However, in such a case, the sampling expansion (2.8) is not in general the most efficient one. It has been proved by Petersen and Middleton [1962] that a function with a bounded band $\mathcal{B}$ can be reconstructed in several ways from its samples taken over a periodic lattice, and that the most efficient lattice – in the sense that it requires the minimum number of sampling points per unit area – is not in general rectangular. The fundamental cell of this optimum periodic lattice is the natural extension of the resolution cell defined above. For example, in the case of a circular band with radius K, Petersen and Middleton [1962] show that the unique optimum sampling lattice is the 120° rhombic lattice with a spacing between adjacent sampling points equal to $2\pi/\sqrt{3}K = 1.16\,\pi/K$.

The previous analysis can also be extended to deal with functions representing 3D images. This extension is important, for instance, in the case of confocal microscopy, a basic technique for obtaining 3D images of biological objects. Then, the definition of a direction-dependent resolution limit is necessary, since the resolving power of the instrument is different in terms of lateral resolution and of axial resolution (see §5).

From the previous considerations, it follows that the concept of resolution limit may be somewhat ambiguous in many instances. We believe that a more relevant and unambiguous concept is the band of the optical instrument. A further example to support this statement is the case of imaging by a partially obscured pupil. Indeed, the limit of resolution given by the sampling theorem, the Nyquist distance, coincides with that of the corresponding completely filled pupil. On the other hand, according to the Rayleigh criterion, the resolving power of the partially obscured pupil is greater than that of the completely filled one – see again Born and Wolf [1980]. We claim that neither the Rayleigh nor the Nyquist distance is fully satisfactory for characterizing the resolving power of such a system. Indeed, when analyzing the available bands it is clear that a partially obscured pupil provides less information about the object than the corresponding completely filled pupil.

2.3. DEBLURRING AND SUPER-RESOLUTION

To get a complete understanding of the imaging performances of a given optical system, let us observe that even the concept of band introduced in the previous subsection is not yet sufficient. In general, indeed, the image $g(\mathbf{r})$ is not simply a bandlimited approximation of $f(\mathbf{r})$ because $\widehat{S}(\mathbf{k})$ is not necessarily constant on the band (see eq. (2.4)). This happens e.g. in incoherent imaging and in the presence of aberrations. In order to get a bandlimited approximation of $f(\mathbf{r})$ from the image, one should divide $\widehat{g}(\mathbf{k})$ by $\widehat{S}(\mathbf{k})$ for all $\mathbf{k} \in \mathcal{B}$. Anticipating on the next section, let us already observe that this operation is not feasible for the following reasons:

(i) in general $\widehat{S}(\mathbf{k})$ tends to zero when $\mathbf{k}$ tends to points belonging to the boundary of the band $\mathcal{B}$;

(ii) if the image is detected by an instrument, then the recorded image $g(\mathbf{r})$ is affected by instrumental noise and therefore eq. (2.4) must be replaced by

$\widehat{g}(\mathbf{k}) = \widehat{S}(\mathbf{k})\widehat{f}(\mathbf{k}) + \widehat{e}(\mathbf{k})$, where $\widehat{e}(\mathbf{k})$ is the Fourier transform of the function $e(\mathbf{r})$ modelling the experimental errors or noise.

If we take into account these two points, we see that the main effect resulting from the division of $\widehat{g}(\mathbf{k})$ by $\widehat{S}(\mathbf{k})$ is expected to be the amplification of the noise in a neighbourhood of the boundary of $\mathcal{B}$. This is a well-known difficulty of 'inverse filtering' (see e.g. Frieden [1975], Andrews and Hunt [1977]).

Appropriate methods for the recovery of the object $f(\mathbf{r})$ from its image $g(\mathbf{r})$ through any linear optical system will be discussed in §3. Let us just observe now that, in general, it will only be possible to recover a bandlimited approximation of $f(\mathbf{r})$ on a band which is smaller than the band $\mathcal{B}$ of the instrument. In §3, we will define this 'effective band' $\mathcal{B}_{\text{eff}}$ as the set of frequencies for which the modulus of the TF $\widehat{S}(\mathbf{k})$ is larger than some threshold value depending on the data noise level. In the absence of noise or when $|\widehat{S}(\mathbf{k})| = 1$ on the band $\mathcal{B}$, the effective band $\mathcal{B}_{\text{eff}}$ coincides with the band. As already stated in §1, we call *deblurring* any restoration process taking place inside the band of the instrument and, more precisely, inside its effective band. Accordingly, the recovery of a bandlimited approximation of the object on the effective band is just a deblurring process. In general, it is the effective band which determines the 'effective' resolution capabilities, i.e. the true performances of the instrument in practical situations.

In some cases, however, we can gain access to spatial-frequency components of the object lying outside this effective band. We will call *super-resolution* the resulting enhancement in resolution. Super-resolution is achieved when the object spectrum can be somehow extrapolated outside the effective band, so that we can get a bandlimited approximation of the object on a larger band. This clearly requires further assumptions about the object, namely some extra information that can be exploited to restore a piece of its spectrum which is not transmitted by the optical system. The required *out-of-band extrapolation* can be performed, at least in principle, under the assumption that the object has a finite spatial extent (see §3). Indeed, in such a case, the object spectrum is an entire analytic function and hence is uniquely determined by the part of the spectrum transmitted by the optical system (see the papers by Wolter [1961] and Harris [1964]). This would allow to restore the object 'beyond the diffraction limit', i.e. with a better resolution than the Rayleigh or Nyquist distance. This idea for beating the Rayleigh limit is quite old and has been abundantly discussed in the literature (see e.g. the papers by Toraldo di Francia [1955], Wolter [1961], Harris [1964], McCutchen [1967], Rushforth and Harris [1968] and Toraldo di Francia [1969]). However, as shown by Viano [1976], this argument does not take into account the instability of analytic continuation in the presence of noise, which prevents in practice an easy recovery of the object spectrum beyond the cut-off frequency. The appropriate framework to address this question and to assess quantitatively the achievable resolution improvement appears to be the regularization theory for inverse problems, which we review in the next section. In such a framework it has been shown by Bertero, De Mol and Viano [1979, 1980] that, because of the ill-posedness of the inverse problem in the presence of noise, only very little super-resolution could be achieved in most practical situations. However, as will be shown in §4, such a conclusion does not hold for problems characterized by a very small *Shannon number* or equivalently *space-bandwidth product.*

When feasible, to enhance the resolving power of an instrument, an alternative to the above approach based on data processing is of course to modify the instrument itself. One way is to improve the data collection scheme to record data on a wider range and containing more information about the probed object. For example, one can think of measuring image data obtained with different wavelengths. Another possibility is to increase the aperture to make the band larger or to try to modify the PSF to make it narrower. The latter possibility is precisely the idea behind an early attempt at beating the Rayleigh limit, by means of so-called *super-resolving pupils* (Toraldo di Francia [1952]). A super-resolving pupil is a pupil which produces a diffraction pattern with a central peak narrower that the initial PSF of the instrument and surrounded by a sufficiently large dark ring. In principle, such a pupil can be realized by means of a suitable coating changing the TF without modifying the band of the instrument. Hence it is a deblurring technique, which could also be seen as a kind of 'apodization'. Let us recall however that, usually, the scope of apodization is to reduce – by changing the TF – the side-lobes of the PSF to avoid unpleasant artefacts in the image. It is well known that the usual price to pay for this is an increase of the width of the central peak of the PSF. In a super-resolving pupil, the shrinkage of this central peak is accompanied by an increase of the side-lobes, which may become quite huge. This major drawback is however irrelevant in some special circumstances: the side-lobes can just be ignored or chopped when the observed object is small compared to the resolution limit and isolated in an uninteresting background. Another serious limitation is then to keep enough light intensity in the central peak of the PSF. Moreover, the required coating of pupil should be realized with an extremely high accuracy, since even a small error would completely destroy its super-resolving properties. The design of a super-resolving pupil is a typical example of a *synthesis problem* and such problems present the same sensitivity to errors as inverse problems do (see §3). This explains why, as far as we know, no practical super-resolving pupil has ever been manufactured. Anyhow, synthesis problems are out of the scope of the present work where we focus on resolution enhancement *by means of data inversion*. We assume that the instrument is given and has a specified PSF, and that the wavelength at which it is operated as well as the type of the recorded data are also fixed. In other words, we work on the basis of a prescribed imaging equation like (2.1), with objects and images also defined on well-specified domains. Then the only way to get super-resolution consists in processing the data to recover the object beyond the diffraction limit. Let us nevertheless notice that super-resolving pupils can be emulated by a suitable post-processing of the image, as it has been shown by De Santis, Gori, Guattari and Palma [1986]. The method can be viewed as a particular linear estimator for object deblurring and hence will be discussed further in §3.2.

§ 3. Linear inversion methods and filtering

Because of noise, to perform both object restoration on the band of the instrument (deblurring) and outside the band (super-resolution), we need appropriate data inversion methods. In the present section, we summarize essential features of such methods and we introduce the concept of *overall impulse response* which is useful for assessing the achievable resolution (§3.1). For convolution equations, we use the classical Wiener filter to solve the deblurring problem and to define the *effective bandwidth* of an optical instrument (§3.2). To deal with super-resolution problems, we show how to make use of the required assumption about the localization (limited spatial extent) of the object and we introduce the *singular system* of the corresponding imaging operator (§3.3). This allows to define the notion of *number of degrees of freedom* (NDF), which is useful for analyzing super-resolution methods. Both the effective bandwidth and the NDF depend on the instrumental PSF and on a quantity called the *signal-to-noise ratio* (SNR) to be defined in §3.2.

3.1. THE OVERALL IMPULSE RESPONSE

The *inverse imaging problem* consists in estimating the original object $f(\mathbf{r})$ from a given recorded image $g(\mathbf{r})$ through an optical instrument with known impulse response. When tackling this problem, we have to take into account the fact that a measured image is inevitably contaminated by noise. Therefore we modify the imaging equation (2.1) as follows:

$$g(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S(\mathbf{r}, \mathbf{r}') \; f(\mathbf{r}') \; + \; e(\mathbf{r}) \tag{3.1}$$

where $e(\mathbf{r})$ is an unknown function modelling the experimental errors and noise in the data. This equation can be rewritten shortly in operator form as

$$g \; = \; Lf \; + \; e \; , \tag{3.2}$$

where L is the *imaging operator* which maps the object on its noise-free image defined by

$$(Lf)(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S(\mathbf{r}, \mathbf{r}') \; f(\mathbf{r}') \; . \tag{3.3}$$

We call it L to remind that it is a Linear operator describing the effect of the optical system, which is just a Lens in the simplest case. In the following we will also use the adjoint imaging operator L^*, which can be defined as

$$(L^* g)(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S^*(\mathbf{r}', \mathbf{r}) \; g(\mathbf{r}') \; , \tag{3.4}$$

denoting by S^* the complex conjugate of S.

Reconstructing the object $f(\mathbf{r})$ from its image is a typical *linear inverse problem* and we have to devise an appropriate method to perform the inversion. We will restrict ourselves to linear estimation and we will denote throughout by $\widetilde{f}(\mathbf{r})$ an estimate of the object. Then the most general linear estimate of the object is obtained as a linear superposition of the values of the recorded image. Hence we may write it in the following form:

$$\widetilde{f}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; M(\mathbf{r}, \mathbf{r}') \; g(\mathbf{r}') \; . \tag{3.5}$$

This linear solver or estimator is characterized by the function $M(\mathbf{r}, \mathbf{r}')$, which we call the *reconstruction kernel* or the *restoration function*. Replacing the imaging equation into (3.5), we get

$$\widetilde{f}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \int \mathrm{d}\mathbf{r}'' \; M(\mathbf{r}, \mathbf{r}') \; S(\mathbf{r}', \mathbf{r}'') \; f(\mathbf{r}'') \; + \int \mathrm{d}\mathbf{r}' \; M(\mathbf{r}, \mathbf{r}') \; e(\mathbf{r}') \; . \tag{3.6}$$

Reversing the order of integration and putting

$$T(\mathbf{r}, \mathbf{r}'') = \int \mathrm{d}\mathbf{r}' \; M(\mathbf{r}, \mathbf{r}') \; S(\mathbf{r}', \mathbf{r}'') \; , \tag{3.7}$$

we also obtain

$$\widetilde{f}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}'' \; T(\mathbf{r}, \mathbf{r}'') \; f(\mathbf{r}'') \; + \int \mathrm{d}\mathbf{r}' \; M(\mathbf{r}, \mathbf{r}') \; e(\mathbf{r}') \; . \tag{3.8}$$

In the absence of noise, perfect restoration – with unlimited resolution – would correspond to $T(\mathbf{r}, \mathbf{r}'') = \delta(\mathbf{r} - \mathbf{r}'')$, where $\delta(\mathbf{r})$ denotes the Dirac distribution. As already mentioned above, this cannot be achieved in practice because of the presence of noise. Indeed, the second term in eq. (3.8) – i.e. the noise term – has to be kept sufficiently small and, as will be shown later, the necessity of controlling the noise amplification implies a certain widening of the function $T(\mathbf{r}, \mathbf{r}'')$. In the absence of noise or with a negligible noise term, this function represents the restoration at point $\mathbf{r}$ of a point source situated at point $\mathbf{r}''$. Hence it plays the role of an impulse response for the object restoration process. Typically, as a function of the variable $\mathbf{r}$, $T(\mathbf{r}, \mathbf{r}'')$ presents a peak centered at $\mathbf{r}''$, with a few side-lobes. The width of this peak characterizes the achievable resolution in the restored object at point $\mathbf{r}''$. In fact, this function $T(\mathbf{r}, \mathbf{r}'')$ allows to take into account both the initial resolving power of the instrument, when no data processing is performed, and the subsequent improvement due to the use of a restoration algorithm. Indeed, as seen from eq. (3.7), it depends on the PSF $S(\mathbf{r}, \mathbf{r}')$ and on the reconstruction kernel $M(\mathbf{r}, \mathbf{r}')$. Therefore, we follow the terminology introduced by Gori and Guattari [1985] and call $T(\mathbf{r}, \mathbf{r}'')$ the *overall impulse response*, since it characterizes the entity 'instrument + implemented inversion algorithm'. When this response is space-invariant, we also call it the *overall PSF*.

Many different linear restoration algorithms have been proposed in the literature, corresponding to different choices of the restoration kernel $M(\mathbf{r}, \mathbf{r}')$. In the following, we present only some of the most common ones, which are sufficient for our purpose of assessing resolution limits. We want to emphasize the fact that there is no 'best all-purpose inversion algorithm'. The choice of a particular method is often a matter of convenience and is made according to the specificities of the particular inverse problem one has to deal with. Nevertheless, inversion algorithms rely on some common ground and present some basic similarities that can be understood from the examples we describe in the following.

3.2. OPTIMAL FILTERING FOR CONVOLUTION EQUATIONS

In the present subsection we assume that the imaging operator is a true convolution, namely that we have to deal with the equation

$$g(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S(\mathbf{r}-\mathbf{r}') \; f(\mathbf{r}') \; + \; e(\mathbf{r}) \; . \tag{3.9}$$

In Fourier space this equation becomes

$$\widehat{g}(\mathbf{k}) \; = \; \widehat{S}(\mathbf{k}) \; \widehat{f}(\mathbf{k}) \; + \; \widehat{e}(\mathbf{k}) \; . \tag{3.10}$$

Because of the space-invariance of the system, it is quite natural to consider only space-invariant restoration, i.e. to take an object estimate of the form

$$\widetilde{f}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; M(\mathbf{r}-\mathbf{r}') \; g(\mathbf{r}') \tag{3.11}$$

or in Fourier space

$$\widehat{\widetilde{f}}(\mathbf{k}) \; = \; \widehat{M}(\mathbf{k}) \; \widehat{g}(\mathbf{k}) \; . \tag{3.12}$$

The function $\widehat{M}(\mathbf{k})$ and the corresponding restoration method are usually called a *filter*. The restoration process is sometimes also referred to as 'image deconvolution' or 'image deblurring'.

In the absence of noise, we see from eq. (3.10) that perfect restoration of the object spectrum on the band of the system is achieved by means of the *inverse filter* $\widehat{M}(\mathbf{k}) = [\widehat{S}(\mathbf{k})]^{-1}$. No out-of-band restoration is thus obtained because this filter can be used only where $\widehat{S}(\mathbf{k})$ does not vanish. We also see that if the inverse filter is applied to the noisy data (3.10), the noise will be considerably amplified for those frequencies where $\widehat{S}(\mathbf{k})$ is very small and this will cause uncontrolled instabilities in the restored object. Therefore, the formal solution given by the inverse filter cannot be applied straightforwardly in practice and has to be replaced by a so-called *regularized algorithm*, for which stability with respect to noise is guaranteed. This is achieved for example by the well-known *Wiener filter* (see e.g. the book by

Bell [1962]), which provides the best linear estimate in the least-squares sense of the solution of eq. (3.9) (see also the papers by Strand and Westwater [1968], Franklin [1970], Turchin, Kozlov and Malkevich [1971] and Cesini, Guattari, Lucarini and Palma [1978]). We briefly recall how such filter is derived. One has to assume that the object f, the image g and the noise e are zero-mean stochastic processes. The power spectrum of the noise is assumed to be given by

$$\langle|\widehat{e}(\mathbf{k})|^2\rangle = \varepsilon^2 \ \widehat{\rho}_{\mathrm{ee}}(\mathbf{k}) \tag{3.13}$$

where $\langle \ \cdot \ \rangle$ denotes the expectation value or ensemble average. For the so-called 'white noise', the spectral density function $\widehat{\rho}_{\mathrm{ee}}(\mathbf{k})$ is equal to 1. We also assume the power spectrum of the object to be given by

$$\langle|\widehat{f}(\mathbf{k})|^2\rangle = E^2 \ \widehat{\rho}_{\mathrm{ff}}(\mathbf{k}) \tag{3.14}$$

and the object to be uncorrelated with the noise, i.e.

$$\langle\widehat{f}(\mathbf{k}) \ \widehat{e}^{\,*}(\mathbf{k})\rangle \ = \ 0 \tag{3.15}$$

(in other words, we assume the noise to be additive). Let us look now for the best linear estimate $\widetilde{f}$ of the object f in the least-squares sense, i.e. for the estimate which minimizes the quadratic or least-squares error $\langle\|\widetilde{f}-f\|^2\rangle$, where $\|f\|$ denotes the L^2-norm of f, namely

$$\|f\| \ = \ \left[\int \mathrm{d}\mathbf{r} \ |f(\mathbf{r})|^2\right]^{1/2} \ . \tag{3.16}$$

Through Parseval's relation, the norm $\|f\|$ is also equal to the L^2-norm $\|\widehat{f}\|$ of $\widehat{f}$ in Fourier space. Hence, equivalently, we have to find the function $\widehat{M}(\mathbf{k})$ minimizing the quantity

$$\langle\int \mathrm{d}\mathbf{k} \ |\widehat{M}(\mathbf{k}) \ \widehat{g}(\mathbf{k}) - \widehat{f}(\mathbf{k})|^2\rangle \ . \tag{3.17}$$

Using eq. (3.10) and the assumptions (3.13-15), we see that this quantity is also given by

$$\int \mathrm{d}\mathbf{k} \left[E^2 \ \widehat{\rho}_{\mathrm{ff}}(\mathbf{k}) \ |\widehat{M}(\mathbf{k}) \ \widehat{S}(\mathbf{k}) - 1|^2 \ + \ \varepsilon^2 \ \widehat{\rho}_{\mathrm{ee}}(\mathbf{k}) \ |\widehat{M}(\mathbf{k})|^2\right] \ . \tag{3.18}$$

It is easily seen that the minimum is provided by

$$\widehat{M}_{opt}(\mathbf{k}) = \frac{\widehat{S}^*(\mathbf{k})}{|\widehat{S}(\mathbf{k})|^2 \ + \frac{\varepsilon^2}{E^2} \frac{\widehat{\rho}_{\mathrm{ee}}(\mathbf{k})}{\widehat{\rho}_{\mathrm{ff}}(\mathbf{k})}} \quad , \tag{3.19}$$

which is the optimal Wiener filter.

For white-noise processes, i.e. when $\widehat{\rho}_{\mathrm{ee}}(\mathbf{k}) = \widehat{\rho}_{\mathrm{ff}}(\mathbf{k}) = 1$, we define the *signal-to-noise ratio* (SNR) to be the quantity E/ε. Let us point out that it does not coincide with the usual concept of SNR used in the field of image restoration, where it is defined as the ratio of the power of the blurred image to the power of the noise. Here the SNR refers to the ratio of magnitude of the object to the magnitude of the data noise. Let us also notice that when $\widehat{\rho}_{\mathrm{ff}}(\mathbf{k}) = 1$, i.e. in the case of a white-noise process, the term $E^2 \ \widehat{\rho}_{\mathrm{ff}}(\mathbf{k})$ integrated over all frequencies is divergent, corresponding to infinite energy. One should not worry about such divergences, which are easily eliminated e.g. by considering, instead of the unphysical model of purely white noise, 'quasi-white' or 'coloured' noise, with a flat spectrum extending beyond the support of $\widehat{S}(\mathbf{k})$.

For white- or coloured-noise processes such that $\widehat{\rho}_{\mathrm{ee}}(\mathbf{k})/\widehat{\rho}_{\mathrm{ff}}(\mathbf{k}) = 1$ and when introducing the parameter $\alpha = \varepsilon^2/E^2$, the Wiener filter estimate coincides in fact with Tikhonov's regularized solution, which corresponds to the restoration function

$$\widehat{M}_\alpha(\mathbf{k}) = \widehat{W}_\alpha(\mathbf{k}) \ \frac{1}{\widehat{S}(\mathbf{k})} \tag{3.20}$$

with

$$\widehat{W}_\alpha(\mathbf{k}) = \frac{|\widehat{S}(\mathbf{k})|^2}{|\widehat{S}(\mathbf{k})|^2 \ + \ \alpha} \tag{3.21}$$

(see e.g. Tikhonov and Arsenin [1977] or Groetsch [1984]). The function $\widehat{W}_\alpha(\mathbf{k})$ can be interpreted as a spectral window which has to be used to 'apodize' the inverse filter $[\widehat{S}(\mathbf{k})]^{-1}$ in order to prevent noise amplification. Notice that because of the presence of the positive parameter α at the denominator, the Tikhonov filter can be used for all frequencies, even where $\widehat{S}(\mathbf{k})$ vanishes. Indeed, from eq. (3.20) and (3.21), we see that $\widehat{M}_\alpha(\mathbf{k}) = 0$ when $\widehat{S}(\mathbf{k}) = 0$. In Tikhonov's method the parameter α is called the *regularization parameter* and the regularized estimate obtained through (3.12) and (3.20-21) is viewed as the solution of the variational problem of minimizing the functional

$$\Phi_\alpha\,[f] = \int \mathrm{d}\mathbf{k}\ |\widehat{S}(\mathbf{k})\ \widehat{f}(\mathbf{k}) - \widehat{g}(\mathbf{k})|^2 + \alpha \int \mathrm{d}\mathbf{k}\ |\widehat{f}(\mathbf{k})|^2\ , \tag{3.22}$$

which through Parseval's relation is also equal to

$$\Phi_\alpha\,[f] = \int \mathrm{d}\mathbf{r}\ |(Lf)(\mathbf{r}) - g(\mathbf{r})|^2 + \alpha \int \mathrm{d}\mathbf{r}\ |f(\mathbf{r})|^2\ . \tag{3.23}$$

The second term in these expressions can be interpreted as a penalization functional, the role of which is to stabilize the pure least-squares solution obtained when minimizing the first term alone. Hence Tikhonov's method is a constrained least-squares method. It is also a purely deterministic or functional method as opposed to stochastic methods like the Wiener filter. Various methods for choosing the regularization parameter α have been proposed in the literature (see e.g. Tikhonov and Arsenin [1977], Groetsch [1984], Bertero [1989] or Davies [1992]). As we have seen above, using a stochastic approach, α is naturally related to the power spectra of the object and of the noise. For white-noise processes, α is simply the inverse of the squared SNR.

The overall impulse response corresponding to (3.20) is clearly space-invariant and is given by the inverse Fourier transform of the filtering window (3.21), which is the *overall transfer function.* This overall TF cannot be equal to one in the presence of noise and vanishes where the transfer function $\widehat{S}(\mathbf{k})$ is zero. Hence we cannot recover the object spectrum outside the band. In other words, super-resolution – in the sense of out-of-band extrapolation as defined in §2.3 – cannot be achieved by means of the above method. Moreover, the presence of noise prevents perfect deblurring on the band and consequently affects the resolving power of the instrument. As discussed previously, in the absence of noise, the resolving power is determined by the band of the system, i.e. the support of $|\widehat{S}(\mathbf{k})|^2$. To assess the resolving power in the presence of noise, we can proceed as follows. Notice that the filtering window (3.21) is close to 1 for those frequencies for which $|\widehat{S}(\mathbf{k})|^2$ dominates the term α in the denominator. It becomes however close to zero when $|\widehat{S}(\mathbf{k})|^2$ is much smaller than α, so that the corresponding frequencies are damped and are not used in the restoration of the object. The resolution limits will be determined by an effective band which depends on the noise level and which can be most easily defined in the following way. Assume that the restoration filter (3.20) is such that $\widehat{W}_\alpha(\mathbf{k})$ is the characteristic function of some frequency band $\mathcal{B}_\alpha$:

$$\widehat{W}_\alpha(\mathbf{k}) = \begin{cases} 1\,, & \mathbf{k} \in \mathcal{B}_\alpha \\ 0\,, & \mathbf{k} \notin \mathcal{B}_\alpha\,. \end{cases} \tag{3.24}$$

Then we define the *effective band* $\mathcal{B}_{\text{eff}}$ as being the band $\mathcal{B}_\alpha$ which minimizes the least-squares error (3.18), allowing for all restoration filters of the form (3.20) with $\widehat{W}_\alpha(\mathbf{k})$ given by (3.24). From the expression of eq. (3.18) in this particular case, it is easily seen that the effective band is the set of frequencies such that

$$|\widehat{S}(\mathbf{k})|^2 > \frac{\varepsilon^2}{E^2}\,\frac{\widehat{\rho}_{\text{ee}}(\mathbf{k})}{\widehat{\rho}_{\text{ff}}(\mathbf{k})} \tag{3.25}$$

or, in the case of white-noise processes, such that

$$|\widehat{S}(\mathbf{k})| > \frac{\varepsilon}{E}\,. \tag{3.26}$$

When the effective band is a simply connected region, we can also define a direction-dependent effective cut-off $K_{\boldsymbol{\theta},\text{eff}}$ associated to the effective band $\mathcal{B}_{\text{eff}}$, as done in §2 for the band $\mathcal{B}$. This effective cut-off will determine the effective resolving power in the direction $\boldsymbol{\theta}$ in the presence of a given amount of noise.

Other criteria than optimality with respect to noise can be used in the design of linear object estimators and one can try to realize an overall PSF or TF having desirable properties. For example, if the restoration filter is chosen in such a way that the spectral window (3.21) is replaced by a classical apodizing window like the Hamming or Hanning functions, the overall PSF will have reduced side-lobes, at the price of an increased width of the central peak. An opposite requirement is to make this central peak narrower than that of the PSF $S(\mathbf{r})$ and in fact as narrow as possible, in order to mimic a super-resolving pupil. In an interesting paper, De Santis, Gori, Guattari and Palma [1986] have shown that this can be achieved by emulating numerically a super-resolving pupil. The price to pay here is the presence of large side-lobes and the sensitivity to noise. As already mentioned in §2.3, the effect of the side-lobes can be neglected in the case where the object has a finite spatial extent and a size of the order of the Rayleigh resolution distance. On the other hand, the amount of achievable resolution is limited by the necessity to control the instability with respect to noise. Although no out-of-band extrapolation is performed in such a way, the method appears to work under an assumption which is similar to that we will make in §4: the object should have small spatial extent. Then the emulated super-resolving pupil restores a bandlimited object, which obviously has an infinite spatial extent but nevertheless approximates the original space-limited object over its extent. This suggests that the two points of view may somehow be related, and we think that this question deserves further investigation to be fully understood.

3.3. FILTERED SINGULAR-SYSTEM EXPANSIONS FOR COMPACT OPERATORS

In the case of a pure convolution, no access to out-of-band frequencies – and hence no super-resolution according to our definition – is provided by any of the Fourier restoration filters described in the previous subsection. To be able to extrapolate outside the band or the effective band, we clearly need something more. For example, a further constraint on the object, expressing some a priori knowledge about its properties, can often do the job. A classical a priori constraint is to assume that the object has a finite spatial extent, i.e. vanishes outside some finite known region, which we will call the *domain of the object*. Then, by a theorem due to Paley and Wiener [1934], its Fourier transform $\widehat{f}(\mathbf{k})$ is an entire analytic function. Thanks to the uniqueness of analytic continuation, such a function is (at least in principle) entirely determined by its values on any finite domain, as for example the band of the system. Such a constraint expressing a priori knowlegde about the domain of the object can be taken into account by rewriting the imaging equation (3.9) as follows

$$g(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; S(\mathbf{r}-\mathbf{r}') \; P(\mathbf{r}') \; f(\mathbf{r}') \; + \; e(\mathbf{r}) \; , \tag{3.27}$$

where $P(\mathbf{r})$ is the characteristic function equal to 1 inside and to 0 outside the domain of the object. Notice that the localization of the object can also be expressed through more general and smooth 'profile functions' $P(\mathbf{r})$ such as e.g. a gaussian function. In such a case eq. (3.27) can also be used to describe imaging processes where the object is illuminated in a nonuniform way, e.g. by a laser spot with gaussian profile. For such a gaussian profile, the solution of the imaging equation with $e(\mathbf{r}) = 0$ is still unique, but uniqueness holds no longer for certain profiles such as bandlimited functions. The imaging problem associated to eq. (3.27) with different profiles has been analyzed in more detail by Bertero, De Mol, Pike and Walker [1984].

For profiles $P(\mathbf{r}')$ vanishing outside some finite domain or decreasing sufficiently fast at infinity, the imaging operator L defined by eq. (3.27) and (3.2) acquires the mathematical property of being compact in an appropriate function space, the most usual choice being a L^2-space of square-integrable objects. The spectral properties of compact operators are quite similar to those of finite-dimensional matrices, namely they have a discrete spectrum. Compact self-adjoint operators can be expressed in diagonal form by means of their eigensystem, but they have in general an infinite number of eigenvalues. Imaging operators, however, are not necessarily self-adjoint (e.g. because the image and object domains are different or because the optical system is affected by aberrations and hence the PSF is not real). Instead of the eigensystem, one uses then the *singular system* of the imaging operator L. To our knowledge, singular systems were first introduced in optics by Gori, Paolucci and Ronchi [1975], but without calling them in that way and without reference to a more general mathematical framework (for this see e.g. the book by Groetsch [1984] or the review paper by Bertero [1989]). The method was used by Gori, Paolucci and Ronchi [1975] and later by De Santis and Palma [1976] for investigating the number of degrees of freedom of an optical image in the presence of aberrations.

Let us recall that the singular system of L is the set of the triples $\{\sigma_n; u_n, v_n\}, n = 0, 1, ...,$ which solve the following 'shifted eigenvalue problem' (L^* denotes the adjoint operator):

$$Lv_n = \sigma_n u_n \ ; \qquad\qquad L^* u_n = \sigma_n v_n \ . \tag{3.28}$$

The singular values σ_n are real and positive by definition, and ordered as follows: $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \cdots$. Except in degenerate cases where they are in finite number, the sequence $\{\sigma_n\}$ of the singular values tends to zero when n tends to infinity. The singular functions or vectors $\{u_n\}$, eigenvectors of LL^*, constitute an orthonormal basis in the space of all possible noise-free images, whereas the singular functions or vectors $\{v_n\}$, eigenvectors of L^*L, form an orthonormal basis in the set of all objects if and only if the equation $Lf = 0$ has only the trivial solution $f = 0$. Otherwise they span the subspace orthogonal to the set of invisible (or transparent) objects. An invisible object is an object producing a zero-image, or equivalently, belonging to the so-called null-space of the imaging operator. When the null-space of L is not reduced to zero, the solution of the imaging equation with zero noise is not unique. Indeed, any object can be written as the sum of its visible and of its invisible parts. The invisible part cannot be retrieved from the data. Hence we are free to assign this component arbitrarily; the most usual choice is to set it equal to zero.

In view of the previous properties, data and solutions can be expanded on the singular functions of L, and restoration kernels as well. The following representations of the operators L and L^* hold also true:

$$Lf = \sum_{n=0}^{\infty} \sigma_n \ (f, v_n) \ u_n \tag{3.29}$$

$$L^* g = \sum_{n=0}^{\infty} \sigma_n \ (g, u_n) \ v_n \tag{3.30}$$

where (f, v_n) and (g, u_n) denote the usual scalar products in L^2. With the help of these formulas, one can easily find the object estimate which minimizes the least-squares error or, in the case of Tikhonov's regularization, the functional (3.23). The reconstruction kernel yielding the minimum is given by

$$M(\mathbf{r}, \mathbf{r}') = \sum_{n=0}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha} \; v_n(\mathbf{r}) \; u_n(\mathbf{r}') \tag{3.31}$$

and the corresponding overall impulse response by

$$T(\mathbf{r}, \mathbf{r}'') = \sum_{n=0}^{\infty} \frac{\sigma_n^2}{\sigma_n^2 + \alpha} \; v_n(\mathbf{r}) \; v_n(\mathbf{r}'') \; . \tag{3.32}$$

The filtering factor it contains, namely

$$W_{n,\alpha} = \frac{\sigma_n^2}{\sigma_n^2 + \alpha} \; , \tag{3.33}$$

is analogous to the spectral window (3.21). The effect of this filter can be discussed along the same lines as for the window (3.21). In particular, one realizes that only the singular values such that $\sigma_n > \sqrt{\alpha}$ yield a significant contribution to the reconstruction kernel (3.31). For this reason, a widely used kernel amounts to sharp truncation to the N first terms of the singular-system expansion, i.e.

$$M(\mathbf{r}, \mathbf{r}') = \sum_{n=0}^{N-1} \frac{1}{\sigma_n} \; v_n(\mathbf{r}) \; u_n(\mathbf{r}') \; , \tag{3.34}$$

corresponding to the following overall impulse response

$$T(\mathbf{r}, \mathbf{r}'') = \sum_{n=0}^{N-1} \; v_n(\mathbf{r}) \; v_n(\mathbf{r}'') \; . \tag{3.35}$$

Clearly, the truncation allows to avoid the unacceptable amplification of the noise term in (3.8) arising from the fact that singular values accumulate to zero. If one keeps only the terms corresponding to singular values larger than the square root of α, then it can be shown that, among all truncated solutions obtained through (3.34), this stopping criterion yields the minimum of the functional (3.23). Moreover for white-noise processes, the stopping criterion becomes

$$\sigma_n > \frac{\varepsilon}{E} \tag{3.36}$$

in complete analogy with (3.26). Here ε denotes the standard deviation (i.e. the square-root of the variance) of each noise component (e, u_n) and E the standard deviation of the object components (f, v_n) (for more details, see Bertero, De Mol and Viano [1979]).

As discussed by De Santis and Palma [1976] and by Bertero and De Mol [1981b], the resulting number N of terms in the expansion of the restored object can be considered as its *number of degrees of freedom* (NDF). Indeed, since the expansion is orthogonal, N represents the number of independent 'pieces of information' about the object which can be reliably (i.e. stably) retrieved from the noisy data. The number of degrees of freedom, which depends on the operator L and on the value of ε/E, can be viewed as a generalization of the Shannon number, which will be defined in §4 and represents the information content of a bandlimited signal. Notice that instead of sharp truncation of the singular-system expansion, one could also filter the smallest singular values more smoothly, by introducing gently decreasing weighting factors W_n in the expansions (3.34) and (3.35). For example, one can use the classical window shapes of Hanning, Hamming or a triangular filter.

Again the ultimate resolution capabilities of the instrument can be assessed by looking at the overall impulse response $T(\mathbf{r}, \mathbf{r}'')$. When the imaging equation $Lf = g$ has a unique solution, then, in the absence of noise, N tends to $+\infty$ and the singular system provides a resolution of the identity, i.e. $T(\mathbf{r}, \mathbf{r}'') = \delta(\mathbf{r} - \mathbf{r}'')$, which yields a perfect restoration. In the presence of noise, however, the expansion has to be truncated or filtered to avoid instabilities, and accordingly, $T(\mathbf{r}, \mathbf{r}'')$ acquires a certain width depending on the signal-to-noise ratio. Notice that when the solution of the imaging equation is not unique, the overall impulse response has always a finite width, even in the absence of noise, because of the existence of invisible objects that can never be retrieved from the data.

The method of singular-system expansions described in the present subsection applies also to the very important case where the data are discrete. In practice, data are always recorded by a finite set of detectors. Hence, instead of a continuous image, one measures a finite-dimensional data vector $\mathbf{g}$. In such a case the imaging operator is automatically compact, even without any assumption on the domain of the object. All formulas given hereabove still hold provided that one replaces the integrals on the x-variable by discrete sums on the components of the data vector and of the singular vectors u_n. The singular system can then be quite easily computed numerically. For a given imaging operator L, it can be computed once for all and then stored. When this is done, the restoration kernel (3.34) provides very fast numerical algorithms for recovering the object. For a thorough discussion of linear inversion from discrete data, we refer the interested reader to the two review papers by Bertero, De Mol and Pike [1985 and 1988]. In particular, a comparison is made between the linear inverse problem with a continuous data function described by eq. (3.27) and the corresponding problem with discrete data, obtained when only sampled values of this data function are known. It can be shown that provided the sampling points are adequately placed and in sufficient number – roughly of the order of the NDF – then the singular system of the discrete-data problem approximates in some sense the singular system of the corresponding continuous-data problem. Therefore all the considerations about resolution we make in the present work remain essentially valid in the case of discrete data, which therefore does not require a separate discussion.

Before concluding this section, we still want to point out that we have only considered here *linear* restoration methods. Moreover, the only kind of a priori information we use about the object is some knowledge about its localization. As seen, such type of a priori knowlegde is easily taken into account in the framework of linear methods. Our choice relies on the fact that it is only using linear inversion methods that one is able to define a resolution limit (through the overall impulse response) which is essentially independent of the particular object under study. However, let us stress the fact that many other inversion methods have been proposed for inverse problems and image restoration. Let us mention iterative methods with and without constraints and statistical methods such as maximum likelihood, maximum entropy, simulated annealing, etc. Most of these methods are non linear and therefore, the achievable resolution is in general strongly dependent on the image or object to be restored. We refer the interested reader to the recent review paper by Biemond, Lagendijk and Mersereau [1990] for iterative methods, to the paper by Meinel [1986] for maximum likelihood and to the paper by Donoho, Johnstone, Hoch and Stern [1992] for maximum entropy. A comparison of several inversion methods in the case of object restoration can be found in the paper by Bertero, Boccacci and Maggio [1995].

§ 4. Out-of-band extrapolation

As discussed in §3.2, by processing the image provided by an optical instrument, it is possible to estimate the Fourier spectrum of the object on the effective band $\mathcal{B}_{\text{eff}}$ and hence to recover a bandlimited approximation of the object. The problem of estimating $\widehat{f}(\mathbf{k})$ outside the effective band is then a particular case of the general problem of *out-of-band extrapolation*, which can be formulated as follows: given the noisy values of $\widehat{f}(\mathbf{k})$ on a bounded domain $\mathcal{B}$ of the frequency space, estimate $\widehat{f}(\mathbf{k})$ outside $\mathcal{B}$ from its known noisy values on $\mathcal{B}$. In the applications we consider, $\mathcal{B}$ represents either the band of the optical system or the effective band.

We first discuss this problem in the simple case of functions of a single variable x, denoting by k the conjugate variable in Fourier space. Let us assume to know $\widehat{f}(k)$ on the interval $\mathcal{B} = [-K, +K]$, with an error $\widehat{e}(k)$ which depends on k. Hence the data are given by

$$\widehat{g}(k) = \begin{cases} \widehat{f}(k) + \widehat{e}(k)\ , & |k| \leq K \\ 0\ , & |k| > K . \end{cases} \tag{4.1}$$

The inverse Fourier transform of $\widehat{g}(k)$ is given by

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}k\ \widehat{g}(k)\ \mathrm{e}^{\mathrm{i}kx} \tag{4.2}$$

or, using eq. (4.1), by

$$g(x) = \int_{-\infty}^{+\infty} \mathrm{d}x' \, \frac{\sin K(x-x')}{\pi(x-x')} \, f(x') \; + \; e(x) \; . \tag{4.3}$$

In other words, $g(x)$ is a noisy bandlimited approximation of $f(x)$. As already observed in §2.2, knowing $g(x)$ is equivalent to knowing $f(x)$ with a resolution π/K. Equation (4.3) is a special case of the imaging equation (3.9). The problem is to use this equation in order to estimate $\widehat{f}(k)$ on a broader band $[-K', +K']$ and therefore to know $f(x)$ with a better resolution π/K'. If we find a method for doing so, then we say that we have achieved *super-resolution* in the restoration of the object $f(x)$.

The extrapolation of $\widehat{f}(k)$ outside $[-K, +K]$ is not possible in general without any additional information about $f(x)$. Clearly, the solution of eq. (4.3) with $e(x) = 0$ is not unique, since the spectrum $\widehat{f}(k)$ is perfectly arbitrary outside the band. Uniqueness holds true, however, in the case of a finite-extent object, i.e. if $f(x)$ vanishes outside a finite interval, say $[-X, +X]$. Indeed, by the theorem of Paley and Wiener [1934] already mentioned in §3.3 (see also e.g. Papoulis [1968]), the Fourier transform of a finite-extent object is an entire analytic function and therefore the analytic continuation of $\widehat{f}(k)$ outside $[-K, +K]$ is unique, provided it exists. This point is carefully discussed by Wolter [1961], who also points out that small errors on the data can produce completely different analytic continuations. The same problem was discussed by Viano [1976] from the point of view of the regularization theory for ill-posed problems.

The main difficulties of the problem of analytic continuation can be summarized as follows. First, we do not know exactly $\widehat{f}(k)$ on $[-K, +K]$, but only $\widehat{g}(k)$ as given by eq. (4.1). Since the noise term $\widehat{e}(k)$ is not in general an analytic function, then also $\widehat{g}(k)$ is not a piece of an analytic function and there is no solution to the problem of analytic continuation of the function $\widehat{g}(k)$. Moreover, even if for a very peculiar $\widehat{e}(k)$, there would be a solution, then we could find many analytic functions reproducing the data on the band within a prescribed accuracy, but being completely different outside the band. Indeed, it is always possible to find an entire function $\widehat{h}(k)$ which is arbitrarily small on $[-K, +K]$ and arbitrarily large outside $[-K, +K]$. This shows the instability of analytic continuation in the presence of noise on the data.

To proceed further, let us write explicitly the corresponding imaging operator to be inverted, taking into account the a priori knowledge about the domain of the object

$$(Lf)(x) = \int_{-X}^{+X} \mathrm{d}x' \, \frac{\sin K(x - x')}{\pi(x - x')} \, f(x') \; , \quad -\infty < x < \infty \quad . \tag{4.4}$$

It is shown by Bertero and Pike [1982] that this integral operator is a compact operator from the space $L^2(-X, +X)$ of square-integrable functions on $[-X, +X]$ into the space $L^2(-\infty, +\infty)$ of square-integrable functions on the real line. Moreover, the inverse operator exists, and this is just a different way of stating the uniqueness of the out-of-band extrapolation. Since we are now in the case described in §3.3, we can use the singular system of the operator L. Its singular functions are in fact related to the *prolate spheroidal wave functions* (PSWF) introduced by Slepian and Pollack [1961]. To see this, we introduce the reduced variables $s = x/X$, $s' = x'/X$ and the parameter

$$c = KX \tag{4.5}$$

which we define as the *space-bandwidth product*. Then the imaging operator becomes

$$(Lf)(s) = \int_{-1}^{+1} \mathrm{d}s' \, \frac{\sin c(s - s')}{\pi(s - s')} \, f(s') \; , \quad -\infty < s < +\infty \quad , \tag{4.6}$$

whereas its adjoint is given by

$$(L^*g)(s) = \int_{-\infty}^{+\infty} \mathrm{d}s' \, \frac{\sin c(s - s')}{\pi(s - s')} \, g(s') \; , \quad -1 < s < +1 \quad . \tag{4.7}$$

It is now easy to compute the operator L^*L, which is precisely the operator

$$(L^*Lf)(s) = \int_{-1}^{+1} \mathrm{d}s' \, \frac{\sin c(s - s')}{\pi(s - s')} \, f(s') \; , \quad -1 < s < +1 \quad , \tag{4.8}$$

considered by Slepian and Pollack. Hence the singular system $\{\sigma_n; u_n, v_n\}$ of the operator L is given by

$$\sigma_n = \sqrt{\lambda_n} \quad ; \quad u_n(s) = \psi_n(c,s) \quad ; \quad v_n(s) = \frac{1}{\sqrt{\lambda_n}}\, \psi_n(c,s) \tag{4.9}$$

for $n = 0, 1, 2, ...$, where the $\psi_n(c,s)$ are the PSWF, i.e. the eigenfunctions of the operator (4.8), and the λ_n are the corresponding eigenvalues, which tend to zero when $n \to \infty$. Notice that the functions $u_n(s)$ are defined on the real line, whereas the functions $v_n(s)$ are defined on the interval $[-1, +1]$.

The solution of the out-of-band extrapolation problem can be estimated by means of the methods described in §3.3. For example, an approximate and stable solution is given by a truncated singular-system expansion, the number of terms being equal to the number of singular values greater than the inverse of the SNR – see eq. (3.34) and eq. (3.36). This number is also defined in §3.3 as the number of degrees of freedom. In order to estimate the NDF as a function of c and of the SNR, we need to know the dependence of the prolate eigenvalues λ_n on the space-bandwidth product c. Let us recall that, when c is sufficiently larger than one, the eigenvalues λ_n are approximately equal to 1 for $n < 2c/\pi$ and tend to zero very rapidly for $n > 2c/\pi$. The quantity

$$S = \frac{2c}{\pi} = \frac{2KX}{\pi} \ , \tag{4.10}$$

called the *Shannon number* by Toraldo di Francia [1969], has a very simple meaning: it is the number of sampling points, spaced by the Nyquist distance $R = \pi/K$, interior to the interval $[-X, +X]$. In communication theory, it is interpreted as the information content of the finite-extent object or 'signal' transmitted by a bandlimited system or 'channel' with cut-off frequency K. From the behaviour of the prolate eigenvalues, we can conclude that, when the space-bandwidth product is large, the NDF is approximately equal to the Shannon number S and that, accordingly, no significant out-of-band extrapolation can be achieved.

The previous conclusion, however, holds true only when the Shannon number is large, i.e. when the size $2X$ of the object is large compared to the sampling distance π/K. When S is not much larger than one, the sharp stepwise behaviour of the eigenvalues λ_n is no longer observed and the NDF can be significantly larger than S. To demonstrate this, let us assess the number of singular values $N(E/\varepsilon, c)$ satisfying the truncation condition

$$\sqrt{\lambda_n(c)} \geq \frac{\varepsilon}{E} \ . \tag{4.11}$$

This number can be easily determined from the published tables of the prolate eigenvalues (see, for instance, Frieden [1971]). In table 1, we report the NDF $N(E/\varepsilon, c)$ for various values of the SNR and for different values of c corresponding to small Shannon numbers. Let us make a few comments about the interpretation of table

Insert table 1 here.

1. To take an example, in the case $c = 5$, we have $S = 3.18$ and therefore the size of the object is approximately three times the sampling distance. Then, if $E/\varepsilon = 10^2$, we see that the NDF is equal to 7. Hence the number of parameters which can be stably estimated is approximately twice the number of sampling points associated with the initial bandwidth. This result implies that an improvement in resolution by a factor of 2 can be achieved, i.e. that it is possible to extrapolate $\widehat{f}(k)$ from $[-K, +K]$ into $[-K', +K']$ with $K' \simeq 2K$. More generally, we notice from table 1 that the NDF behaves roughly as follows:

$$N\left(\frac{E}{\varepsilon}, c\right) \cong S + A + B \log_{10}\left(\frac{E}{\varepsilon}\right) \ , \tag{4.12}$$

where A and B are some constants of the order of 1 (in fact, allowing for a slow variation of A and B with c would help to reproduce more accurately the behaviour of the NDF). This result implies that *the increase of the NDF with respect to the Shannon number depends only logarithmically on the signal-to-noise ratio.* A similar result holds true for the gain in bandwidth, if we define the bandwidth of the extrapolated image as follows

$$K' = \frac{\pi}{2X}\ N\left(\frac{E}{\varepsilon}, c\right) \cong K + \frac{\pi}{2X}\ A + \frac{\pi}{2X}\ B \log_{10}\left(\frac{E}{\varepsilon}\right) = \tag{4.13}$$
$$= K\left[1 + \frac{1}{S}\ A + \frac{1}{S}\ B\log_{10}\left(\frac{E}{\varepsilon}\right)\right] \ .$$

The corresponding new Nyquist or Rayleigh distance is given by $R' = \pi/K'$, namely

$$R' = \frac{\pi}{K'} = \frac{S}{S + A + B\log_{10}\left(\frac{E}{\varepsilon}\right)}\ R \ . \tag{4.14}$$

The main conclusions which can be derived from the previous analysis are the following:

(i) super-resolution, in the sense of out-of-band extrapolation, is feasible only when the known size of the object is not too large compared to the resolution limit of the imaging system;

(ii) the amount of achievable super-resolution depends on the space-bandwidth product $c = KX$ and on the signal-to-noise ratio, even if the latter dependence is rather weak (since it is only logarithmic).

To actually perform the out-of-band extrapolation of the object spectrum, one needs a numerical method. As discussed above, one possibility is to use a truncated singular-system expansion, but this method is not very efficient from the computational point of view. It presents also the drawback of producing restorations which may not be satisfactory at the edges of the object domain. Indeed, the PSWF become quite large at the endpoints ± 1, for large values of the index n, and therefore the truncated singular-system expansion cannot reproduce accurately a continuous object vanishing at those endpoints.

A simple iterative method which is easier to implement has been proposed by Gerchberg [1974]. In the case where the object domain is the interval $[-1, +1]$ and the band is the interval $[-c, +c]$, the procedure works as follows. We start from the noisy data function $\widehat{g}(k)$ given by eq. (4.1) with $K = c$, or equivalently from its Fourier transform $g(s)$. The initial object estimate $f_1(s)$ is obtained by truncating

$g(s)$ to the object domain $[-1,+1]$. Then the Fourier transform $\widehat{f_1}(k)$ of $f_1(s)$ is computed and the values of this function in the interval $[-c,+c]$ are replaced by the known values of $\widehat{g}(k)$. The inverse Fourier transform of the resulting composite function, truncated to the interval $[-1,+1]$, forms the new object estimate $f_2(s)$, and so on. If we introduce the function

$$\mathrm{rect}(x) = \begin{cases} 1\,, & |x| \le 1 \\ 0\,, & |x| > 1 \end{cases} \tag{4.15}$$

and if we recall that $\widehat{g}(k) = 0$ for $|k| > c$, this iterative procedure can be written as follows, with $j = 1, 2, ...$:

$$\begin{cases} f_1(s) & = \mathrm{rect}(s)\ g(s) \\ \widehat{g}_{j+1}(k) & = \widehat{g}(k) + [1 - \mathrm{rect}\,(k/c)]\,\widehat{f}_j(k) \\ f_{j+1}(s) & = \mathrm{rect}(s)\ g_{j+1}(s). \end{cases} \tag{4.16}$$

In the absence of noise, the convergence of this method to the solution of the equation $Lf = g$, with L given by eq. (4.6), has been proved by Papoulis [1975] and by De Santis and Gori [1975]. A generalization of the Gerchberg method to the case of an image produced by an optical instrument with an arbitrary non-negative TF has been proposed by Gori [1975]. This method applies, for instance, to both coherent and incoherent imaging through ideal systems. It was later recognized by Abbiss, De Mol and Dhadwal [1983], that the Gerchberg method – as well as its generalization due Gori – is a particular form of the method of Landweber [1951] for the solution of first-kind Fredholm integral equations, such as are for example the imaging equations of the form $Lf = g$ considered here. Indeed, we show in Appendix B that the iterative procedure (4.16) coincides with the following one

$$\begin{cases} f_0 & = 0 \\ f_{j+1} & = L^*g + [I - L^*L]f_j \end{cases} \tag{4.17}$$

where I denotes the identity operator and L^* and L^*L are, respectively, the operators (4.7) and (4.8). Equation (4.17) defines precisely the Landweber iterative method applied to the integral equation $Lf = g$. In Appendix B, we also derive from eq. (4.17) a result due to De Santis and Gori [1975] and Gori [1975]: the j-th estimate f_j can be obtained by applying a suitable filter to the solution of the integral equation. In fact, the number of iterations plays the role of a regularization parameter and therefore, for any given noisy function to be extrapolated, there exists an optimum number of iterations which depends on the function and on the noise. The application of this iterative method to the case of regularized solutions was investigated by Cesini, Guattari, Lucarini and Palma [1976] and by Abbiss, De Mol and Dhadwal [1983]. The repeated Fourier transforms used by the Gerchberg method are usually implemented numerically by means of the FFT algorithm, which makes the method work faster, even though, in some cases, the required number of iterations can be quite large.

The previous analysis of the out-of-band extrapolation problem can be extended in the two following directions:

(i) *The use of other constraints, in addition to that on the domain of the solution.* One of the most frequent is the positivity of the solution, which can be imposed at each step of the Gerchberg procedure by simply setting to zero all negative values. The constraint of positivity was considered by Lent and Tuy [1981], who also proved a convergence result. More general classes of constraints were considered by Youla and Webb [1982] (for a tutorial, see Youla [1987]).

(ii) *The extrapolation of functions of two and three variables.* The analysis of the 1D case based on singular-system expansions can be extended to 2D and 3D problems by means of the generalized spheroidal wave functions introduced by Slepian [1964]. Estimates of the amount of super-resolution have been derived by Bertero and Pike [1982] for functions of two variables, assuming that both the band of the image and the domain of the object are either squares or discs. To our knowledge, no estimates have been obtained for more general shapes and for functions of three variables. On the other hand, the Gerchberg method extends quite naturally to the case of two and three variables, and is still then equivalent to the Landweber iterative method. In the expansion (B.11), one has to substitute the PSWF by the generalized prolate spheroidal wave functions, making also the corresponding substitution for the eigenvalues.

An example of the application of the Gerchberg method to the problem of limited-angle tomography, with the additional constraint of positivity, is given in the paper by Lent and Tuy [1981] quoted above. In that case, the shape of the available band is an angular sector determined by the directions accessible to measurements. In three dimensions, out-of-band extrapolation can also be applied to inverse scattering problems at fixed energy in the Born approximation, for which measurements only provide access to the so-called Ewald sphere (Wolf and Habashy [1993], Habashy and Wolf [1994]).

Let us still observe that, for didactic purposes, we have in fact split the problem of super-resolving the image provided by an imaging system into two steps:

(a) the first step, mainly considered in §3.2, consists in the solution of an image deblurring problem; the result is an estimate of the Fourier transform of the object on the effective band of the imaging system;

(b) the second step consists in the extrapolation of the Fourier transform of the object out of the effective band of the imaging system.

This separation is useful for clarifying the role of the finite domain of the object in the second step. It is important to mention, however, that one can find methods mixing together the two steps. The simplest one is the constrained Landweber method applied directly to the imaging equation. This possibility is mentioned, for instance, by Schafer, Mersereau and Richards [1981], although in their paper, the method is written in the form originally due to van Cittert [1931] and the proof of convergence is not correct. In such a method, it is easy to introduce, at each iteration step, the constraint on the domain of the object as well as other constraints such as positivity. Another method, which probably requires further investigation, is the constrained conjugate-gradients method (see Lagendijk, Biemond and Boekee [1988]).

Statistical methods can be used as well for this problem, such as maximum entropy (see e.g. the paper by Frieden [1975]) and maximum likelihood (see e.g. the papers by Richardson [1972], Lucy [1974] and Shepp and Vardi [1982]). It has been shown by Donoho, Johnstone, Hoch and Stern [1992] that the method of maximum entropy has a super-resolving effect in the case of so-called *nearly-black objects*, i.e. of objects which are essentially zero in the vast majority of samples. This result seems to be in agreement with the analysis of super-resolution presented in this section.

§ 5. Confocal microscopy

The results of the analysis made in the previous section indicate that a considerable resolution enhancement is to be expected if the theory can be put into practice in some optical devices. As we have seen, to get a significant amount of super-resolution, the size of the object must be of the same order as or smaller than the resolution distance of the imaging system. This condition is rarely satisfied in usual imaging systems. One exception we can quote is the problem of resolving a double star, which cannot be resolved by the available telescope, in the case where this object is isolated in a dark background.

Another situation where the above requirement is satisfied can be found in scanning microscopy, as noticed by Bertero and Pike [1982] who suggested the possibility of using inversion techniques, based on singular-system expansions, to enhance the resolution of such microscopes. In a scanning device, indeed, for each scanning position, only a very small portion of the object is illuminated at a time, so that the resulting space-bandwidth product is indeed very low. Bertero and Pike [1982] treat the case of uniform illumination of the object through some small diaphragm. More general illumination profiles have been considered by Bertero, De Mol, Pike and Walker [1984], resulting in an imaging equation of the type (3.27) for each scanning position. Because of the great practical relevance of this technique, we will focus here on the case of *confocal scanning microscopy* (Sheppard and Choudury [1977], Brakenhoff, Blom and Barends [1979]; for a survey, see the book by Wilson and Sheppard [1984]).

In a confocal microscope, focused laser illumination allows to achieve low space-bandwidth products as well as high SNR values, resulting as we will see in an enhanced resolving power with respect to ordinary microscopes. Each portion of the specimen is illuminated by means of a sharp laser spot focused by a first lens, the *illumination lens*. The image is formed by means of a second lens, the *imaging* or *collector lens*, whose focal plane also coincides with the plane containing the specimen. This plane is called the *confocal plane*, while the common focus of the two lenses is called the *confocal point*. Finally, the central part of this image is detected by placing a pinhole in front of the detector (which is e.g. a photomultiplier). A scan through the specimen is then performed by translating it with respect to the confocal system. The situation described above corresponds to the so-called transmission mode of the confocal microscope. In the reflection mode, the same lens is used both as illumination and as imaging lens. Coherent imaging can be used, as well as incoherent one in the case of fluorescence microscopy. Fluorescence microscopy is widely used because it allows to achieve also 3D imaging, with a good resolution in the axial direction as well (Wijnaendts-van-Resandt, Marsman, Kaplan, Davoust, Stelzer and Stricker [1985], Brakenhoff, van der Voort, van Spronsen and Nanninga [1986]).

Let us first consider the case of 2D images and planar objects. In such a case, the imaging equations corresponding to the transmission and to the reflection mode have essentially the same structure. For each fixed scanning position, the imaging process can be modelled by an equation similar to (3.27), but with 2D transverse coordinates in the object and image planes, namely the equation

$$g(\boldsymbol{\sigma}) = \int \mathrm{d}\boldsymbol{\rho}' \; S_2(\boldsymbol{\sigma} - \boldsymbol{\rho}') \; S_1(\boldsymbol{\rho}') \; f(\boldsymbol{\rho}') \; , \tag{5.1}$$

which expresses that the object $f(\boldsymbol{\rho}')$ is first multiplied by the illumination profile, which is the PSF $S_1(\boldsymbol{\rho})$ of the illuminating lens, and then convolved with the PSF $S_2(\boldsymbol{\rho})$ of the imaging lens.

The images formed at the different scanning positions $\boldsymbol{\rho}$ are obtained by translating the object, i.e. they are given by

$$g(\boldsymbol{\sigma};\boldsymbol{\rho}) = \int \mathrm{d}\boldsymbol{\rho}' \; S_2(\boldsymbol{\sigma}-\boldsymbol{\rho}') \; S_1(\boldsymbol{\rho}') \; f(\boldsymbol{\rho}'+\boldsymbol{\rho}) \quad . \tag{5.2}$$

For each scanning position $\boldsymbol{\rho}$, only one value of the image is recorded by means of the detector, namely its central value on the optical axis (corresponding to $\boldsymbol{\sigma} = \mathbf{0}$) or, more exactly, a central value obtained by integration over the finite pinhole. Then, if $P(\boldsymbol{\sigma})$ denotes the characteristic function of the pinhole, the output of the detector is proportional to

$$G(\boldsymbol{\rho}) = \int \mathrm{d}\boldsymbol{\sigma} \; P(\boldsymbol{\sigma}) \; g(\boldsymbol{\sigma};\boldsymbol{\rho}) \; . \tag{5.3}$$

Replacing eq. (5.2) into eq. (5.3), by a change of the integration order and by a change of variable, one finds

$$G(\boldsymbol{\rho}) = \int d\boldsymbol{\rho}' \; H(\boldsymbol{\rho}-\boldsymbol{\rho}') \; f(\boldsymbol{\rho}') \; , \tag{5.4}$$

where

$$H(\boldsymbol{\rho}) = S_1(-\boldsymbol{\rho}) \int \mathrm{d}\boldsymbol{\sigma} \; P(\boldsymbol{\sigma}) \; S_2(\boldsymbol{\sigma}+\boldsymbol{\rho}) \; . \tag{5.5}$$

In a usual confocal microscope, no object restoration is performed and the recorded image $G(\boldsymbol{\rho})$ is taken as the estimate of the object. We see that it is simply the convolution of the object by the PSF $H(\boldsymbol{\rho})$. This PSF is also bandlimited, but its band is broader than the band of $S_1(\boldsymbol{\rho})$ or of $S_2(\boldsymbol{\rho})$. This leads to an improvement in resolution with respect to an ordinary microscope using uniform illumination of the object and only one imaging lens with PSF $S_2(\boldsymbol{\rho})$. For example, in an ideal 1D coherent-illumination case – when $S_1(x)$ and $S_2(x)$ are both sinc-functions with bandwidth K – and for a very small pinhole – i.e. $P(s) = \delta(s)$ –, $H(x)$ is a sinc^2-function with bandwidth $2K$. Since $\widehat{H}(k)$ has a triangular shape, the resulting bandwidth of the instrument is approximately twice the bandwidth of each single lens (the effective bandwidth in the presence of noise being slightly smaller since $\widehat{H}(k)$ goes to zero at the edges of the band). Unfortunately, one cannot use a very small pinhole, because one needs sensible values of the signal-to-noise ratio. The effect of the finite size of the pinhole is an additional reduction of the effective bandwidth of the instrument. It follows that, in practice, the effective gain in resolving power provided by confocal imaging is close to a factor 1.4 (Cox, Sheppard and Wilson [1982], Brakenhoff, Blom and Barends [1979]).

A way for further enhancing the resolving power of a confocal microscope has been suggested in a paper by Bertero and Pike [1982], and later investigated in a paper by Bertero, De Mol, Pike and Walker [1984] and in a series of subsequent papers (Bertero, Brianzi and Pike [1987], Bertero, De Mol and Pike [1987], Bertero, Boccacci, Defrise, De Mol and Pike [1989], Bertero, Boccacci, Davies and Pike [1991]). The original idea – a more recent modification will be discussed in a moment – was to replace the single on-axis detector of the conventional confocal microscope by an array of detectors in order to measure, at each scanning position, the complete diffraction image. In other words, one should record, for each fixed scanning position $\boldsymbol{\rho}$, the image $g(\boldsymbol{\sigma};\boldsymbol{\rho})$, given by eq. (5.2), and solve this equation in order to estimate $f(\boldsymbol{\rho})$. We notice that, up to a change of origin, the imaging equation is the same for each scanning position, so that one can use the same linear estimator at every point.

As already remarked, the integral operator appearing in eq. (5.1) has the same structure as the imaging operator of eq. (3.27), and in fact it is not difficult to prove that it is a compact operator whenever $S_1(\boldsymbol{\rho})$ and $S_2(\boldsymbol{\rho})$ are bandlimited PSF (with or without aberrations). In 1D ideal coherent imaging, when both $S_1(x)$ and $S_2(x)$ are sinc-functions, a nice feature happens, namely that the singular system of the imaging operator defined by (5.1) can be determined analytically, as it was shown by Gori and Guattari [1985]. The properties of the singular systems of the imaging operators corresponding to various other situations (coherent and incoherent illumination, one-dimensional and bi-dimensional systems) have been investigated numerically in the series of papers by Bertero, Pike and coworkers mentioned above. The investigation of the 1D coherent and incoherent cases was motivated essentially by the fact that, due to the difficulty of the general problem, they provide sufficiently simple inverse problems, whose solution facilitates the understanding of the structure of the singular system corresponding to more general cases.

The method we used in the papers mentioned above for estimating the object $f(\boldsymbol{\rho}' + \boldsymbol{\rho})$ in eq. (5.2), on the optical axis (i.e. for $\boldsymbol{\rho}' = \mathbf{0}$) and for each scanning position $\boldsymbol{\rho}$, is a truncated singular-system expansion. The result is expressed by the following estimator

$$\widetilde{f}(\boldsymbol{\rho}) = \int \mathrm{d}\boldsymbol{\sigma}\; M(\boldsymbol{\sigma})\; g(\boldsymbol{\sigma}; \boldsymbol{\rho}) \tag{5.6}$$

where (see eq. (3.34))

$$M(\boldsymbol{\sigma}) = \sum_{n=0}^{N-1} \frac{1}{\sigma_n}\; v_n(\mathbf{0})\; u_n(\boldsymbol{\sigma}) \quad . \tag{5.7}$$

The result of these investigations is that the effective band of this modified confocal microscope essentially coincides with the band of the PSF $H(\boldsymbol{\rho})$, as given by eq. (5.5) with $P(\boldsymbol{\sigma}) = \delta(\boldsymbol{\sigma})$, and is therefore broader than the effective band of $H(\boldsymbol{\rho})$. Indeed, in general, $\widehat{H}(\mathbf{k})$ tends to zero rather rapidly at the boundary of the band and this causes a substantial loss in useful frequencies. In other words, the method decribed above is a way for utilizing efficiently the full theoretical band of the confocal system. It amounts to extending the object spectrum from the effective band of $H(\boldsymbol{\rho})$ to its full band. Hence, the modified confocal microscope is super-resolving compared to the usual confocal microscope where no data inversion is performed.

Experimental confirmations of the theoretical results discussed above were obtained by Walker [1983], Young, Davies, Pike and Walker [1989] and Young, Davies, Pike, Walker and Bertero [1989], using both coherent and incoherent confocal microscopes, with low numerical aperture. The use of multiple detectors presents however the disadvantage that the detectors or detector elements must be calibrated.

In addition, for the coherent case, signals relating to the complex amplitude are required, thus necessitating the use of interferometry or of some other method of phase measurement. However, looking at eq. (5.6), we remark that the only operations required for implementing the inversion method are the following: i) multiplication of the image $g(\boldsymbol{\sigma}, \boldsymbol{\rho})$ by the 'mask' function $M(\boldsymbol{\sigma})$; ii) integration of the result over the image plane. These operations can in fact be implemented by means of optical processors, without needing an array of detectors. The use of such optical processors has been proposed by Walker, Pike, Davies, Young, Brakenhoff and Bertero [1993] both for the coherent and the incoherent case. Properties of the optical masks $M(\boldsymbol{\sigma})$ corresponding to various situations were investigated theoretically by Bertero, Boccacci, Davies, Malfanti, Pike and Walker [1992]. Preliminary experimental results were obtained by Grochmalicki, Pike, Walker, Bertero, Boccacci and Davies [1993], again in good agreement with the theory.

A more general linear estimator than (5.6) has been considered by Defrise and De Mol [1992]. It is given by

$$\widetilde{f}(\boldsymbol{\rho}) = \int \mathrm{d}\boldsymbol{\sigma} \int \mathrm{d}\boldsymbol{\rho}' \; M(\boldsymbol{\sigma}; \boldsymbol{\rho} - \boldsymbol{\rho}') \; g(\boldsymbol{\sigma}; \boldsymbol{\rho}') \tag{5.8}$$

and is assumed to be shift-invariant with respect to the scanning position, but not with respect to the detector position, since the integration on $\boldsymbol{\sigma}$ may be limited to a finite domain. With the help of such estimator, one can derive some interesting results for the ideal 1D coherent case where $S_1(x) = S_2(x) = (K/\pi)\,\mathrm{sinc}(Kx/\pi)$. For instance, in the absence of noise, one can devise different choices of the restoration function leading to the following overall impulse response

$$T(x, x') = \frac{\sin[2K(x - x')]}{\pi(x - x')} \tag{5.9}$$

or, equivalently, to the overall transfer function $\widehat{T}(k) = 1$ on the band $[-2K, +2K]$ and $\widehat{T}(k) = 0$ outside. This corresponds to a resolving power effectively enhanced by a factor 2 with respect to an ordinary microscope with PSF $S_1(x)$. A first possibility for such restoration relies on the 1D exact inversion formula for the imaging equation (5.1) derived by Bertero, De Mol and Pike [1987]:

$$M(s;x) = \frac{4\pi}{K}\ \cos(Ks)\ \delta(x) \quad . \tag{5.10}$$

where s is the 1D coordinate corresponding to $\boldsymbol{\sigma}$. In this case we obtain an estimator having the same structure as the estimator (5.6). However, besides (5.10), and because of the redundancy of the data (5.2), a whole family of reconstruction kernels can be constructed, all yielding (5.9) as overall impulse response (Defrise and De Mol [1992]), including the following one

$$M(s;x) = \frac{\pi}{K}\ \delta(x - \frac{s}{2}) \tag{5.11}$$

first proposed by Sheppard [1988].

Let us now consider the problem of producing 3D images of thick specimens. The usual case is that of a fluorescent object which is characterized by the distribution function of the fluorescent material, denoted by $f(\mathbf{r})$, where $\mathbf{r}$ is the 3D vector coordinate. We also use the notation $\mathbf{r} = (\boldsymbol{\rho}, z)$, where z denotes the depth coordinate along the optical axis and $\boldsymbol{\rho}$ is the transverse coordinate. In general, for 3D imaging, the confocal microscope is operated in the transmission mode and circularly polarized excitation light is used. Then, if we denote by $W(\mathbf{r}) = W(\boldsymbol{\rho}, z)$ the time-averaged electrical energy distribution in the focal region of the lens, the imaging equation (5.1) must be replaced by the following one (Bertero, Boccacci, Brakenhoff, Malfanti and van der Voort [1990]):

$$g(\boldsymbol{\sigma}) = \int \mathrm{d}\boldsymbol{\rho}'\ \mathrm{d}z'\ W(\boldsymbol{\sigma} - \boldsymbol{\rho}', z')\ W(\boldsymbol{\rho}', z')\ f(\boldsymbol{\rho}', z')\ . \tag{5.12}$$

We have neglected, for simplicity, the effect of the difference between the primary and the emission wavelengths. A more refined imaging model is discussed by van der Voort and Brakenhoff [1990]. The function $g(\boldsymbol{\sigma})$ is proportional to the intensity distribution in the image plane. For lenses with a high numerical aperture, the Debye approximation is not adequate and one must use the more accurate approximation of Richards and Wolf [1959] for the computation of $W(\boldsymbol{\rho}, z)$.

A 3D scanning is performed and the images at different scanning positions $\mathbf{r} = (\boldsymbol{\rho}, z)$ are given by

$$g(\boldsymbol{\sigma};\boldsymbol{\rho},z) = \int \mathrm{d}\boldsymbol{\rho}' \; \mathrm{d}z' \; W(\boldsymbol{\sigma}-\boldsymbol{\rho}',z') \; W(\boldsymbol{\rho}',z') \; f(\boldsymbol{\rho}'+\boldsymbol{\rho},z'+z). \tag{5.13}$$

Therefore, if one takes into account the effect of the pinhole as in the 2D case, one obtains the following expression for the recorded 3D image

$$G(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \; H(\mathbf{r}-\mathbf{r}') \; f(\mathbf{r}') \;, \tag{5.14}$$

where

$$H(\mathbf{r}) = W(-\mathbf{r}) \int \mathrm{d}\boldsymbol{\sigma} \; P(\boldsymbol{\sigma}) \; W(\boldsymbol{\sigma}+\boldsymbol{\rho},z) \;, \tag{5.15}$$

which is an extension of eq. (5.5). The function $H(\mathbf{r})$ is bandlimited but its band is not a sphere. If the optical system has a circular symmetry around the optical axis z, the band of $H(\mathbf{r})$ has a circular symmetry around the k_z-axis, but normally its extension in the direction of the k_z-axis is roughly one third of its extension in the transverse directions and, consequently, axial resolution is poorer than lateral resolution (see the paper by van der Voort and Brakenhoff [1990]).

Methods of image deconvolution similar to those discussed in §3.2 have been applied to some 3D images produced by a confocal microscope by Bertero, Boccacci, Brakenhoff, Malfanti and van der Voort [1990], demonstrating the possibility of obtaining an improvement of the image quality, especially in the axial direction. In the same paper, one can also find an estimation of the resolution improvement obtained by detecting, for each scanning position $\boldsymbol{\rho}$, the complete image $g(\boldsymbol{\sigma};\boldsymbol{\rho})$ and by solving the integral equation (5.13) to estimate the object $f(\mathbf{r})$. It was shown that this method leads to an improvement of the effective bandwidth in all directions.

§ 6. Inverse diffraction and near-field imaging

In this last section, we analyze another possibility for improving the resolution limits, which consists in recording – whenever possible – near-field data. The conclusions derived from the previous examples, which deal only with far-field data, hold no longer true in such a case. As we shall see, near-field imaging techniques allow to considerably enhance the resolving power compared to far-field imaging with the same wavelength. Strictly speaking, this is not true super-resolution in the sense of out-of-band extrapolation as defined above. Indeed, the near-field data contain information – conveyed by the so-called evanescent waves – about spatial frequencies of the object which are no longer present in the far-field region.

A good theoretical laboratory for investigating the resolution enhancement arising from the effect of evanescent waves is provided by the so-called *inverse diffraction problem*, which consists in back-propagating (towards the sources) a scalar field propagating in free space according to the Helmholtz equation. The simplest geometry to formulate this problem is the case of *inverse diffraction from plane to plane*, where the field propagates in the half-space $z \geq 0$ and where one has to recover the field on the boundary plane $z = 0$ from its values on the plane $z = a > 0$. This problem has been first considered by Sherman [1967] and Shewell and Wolf [1968] (see also the discussion in the book by Nieto-Vesperinas [1991]).

Let us consider a scalar monochromatic field propagating in the half-space $z \geq 0$, the complex field amplitude $u(\mathbf{r}) = u(x, y, z)$ being a solution of the Helmholtz equation

$$\Delta u \; + \; K^2 u \; = \; 0 \; , \tag{6.1}$$

satisfying the Sommerfeld radiation condition at infinity

$$\lim_{r \to \infty} [r(\frac{\partial u}{\partial r} - \mathrm{i}Ku)] = 0 \; , \quad |\theta| < \pi/2 \; , \tag{6.2}$$

where $r = |\mathbf{r}|$, θ is the polar angle and $K = 2\pi/\lambda$ is the wavenumber.

If the field amplitude is known on the source plane $z = 0$ and given by $u(x, y, 0) = f(x, y)$, then the field $u(x, y, a) = g_a(x, y)$ in the plane $z = a$, $a > 0$, is uniquely determined and given by

$$g_a \; = \; L_a f \; , \tag{6.3}$$

where the imaging operator L_a acts by convolving the field in the source plane with the forward propagator $S_a^+(\boldsymbol{\rho})$:

$$(L_a f)(\boldsymbol{\rho}) = \int \mathrm{d}\boldsymbol{\rho}' \; S_a^+(\boldsymbol{\rho} - \boldsymbol{\rho}') \; f(\boldsymbol{\rho}') \; . \tag{6.4}$$

We have denoted by $\boldsymbol{\rho} = (x, y)$ and $\boldsymbol{\rho}' = (x', y')$ the transverse vector coordinates in the image and source planes, respectively. The propagator $S_a^+(\boldsymbol{\rho})$ is given by

$$S_a^+(\boldsymbol{\rho}) = \left(\frac{1}{2\pi}\right)^2 \int \mathrm{d}\boldsymbol{\kappa} \; \mathrm{e}^{\mathrm{i}\boldsymbol{\kappa}\cdot\boldsymbol{\rho}} \; \mathrm{e}^{\mathrm{i}a \; m(\boldsymbol{\kappa})} \tag{6.5}$$

with $\boldsymbol{\kappa} = (k_x, k_y)$ and

$$m(\boldsymbol{\kappa}) = \begin{cases} \sqrt{K^2 - |\boldsymbol{\kappa}|^2} & \text{for } |\boldsymbol{\kappa}| \le K \\ \mathrm{i}\sqrt{|\boldsymbol{\kappa}|^2 - K^2} & \text{for } |\boldsymbol{\kappa}| > K \; . \end{cases} \tag{6.6}$$

The real part of $m(\boldsymbol{\kappa})$ corresponds to the so-called *homogeneous waves*, propagating without attenuation, whereas the imaginary part of $m(\boldsymbol{\kappa})$ corresponds to the *inhomogeneous* or *evanescent waves*, whose amplitude decreases exponentially with the distance a between the source and data planes.

In Fourier space the convolution (6.3) becomes simply

$$\widehat{g}_a(\boldsymbol{\kappa}) = \mathrm{e}^{\mathrm{i}a \; m(\boldsymbol{\kappa})} \; \widehat{f}(\boldsymbol{\kappa}) \; . \tag{6.7}$$

Hence we see that backward propagation is described by

$$\widehat{f}(\boldsymbol{\kappa}) = \mathrm{e}^{-\mathrm{i}a \; m(\boldsymbol{\kappa})} \; \widehat{g}_a(\boldsymbol{\kappa}) \; . \tag{6.8}$$

Because of the exponential damping, the restoration of the part of the object spectrum corresponding to evanescent waves is unstable. When the data plane goes to the far-field region, the evanescent waves are no longer present in the data. Only the spatial frequencies $|\boldsymbol{\kappa}| \le K$ propagate to the far-field and then the corresponding propagator $S_a^+(\boldsymbol{\rho})$ is just the Fourier transform of the disc of radius K, except for a phase factor. Hence, the inverse diffraction problem from far-field data is equivalent to inverse imaging through a circular pupil, in the presence of aberrations because of the phase factor $\mathrm{e}^{\mathrm{i}a \; m(\boldsymbol{\kappa})}$. The corresponding resolution limit is given by the Rayleigh distance $R = (1.22)\;(\lambda/2)$.

In the absence of further information about the unknown object, which is the field in the source plane, all we can recover from the far-field data g_a with $a >> \lambda$ is the following bandlimited approximation of $f(\boldsymbol{\rho})$:

$$f_{\mathcal{B}}(\boldsymbol{\rho}) = \left(\frac{1}{2\pi}\right)^2 \int_{|\boldsymbol{\kappa}|\leq K} \mathrm{d}\boldsymbol{\kappa}\ \mathrm{e}^{\mathrm{i}\boldsymbol{\kappa}\cdot\boldsymbol{\rho}}\ \mathrm{e}^{-\mathrm{i}a\ m(\boldsymbol{\kappa})}\ \widehat{g}_a(\boldsymbol{\kappa})\ . \tag{6.9}$$

This is equivalent to the inversion formula derived by Shewell and Wolf [1968]. On the band $|\boldsymbol{\kappa}| \leq K$, the inverse of the imaging operator is simply given by its adjoint L_a^*, which is the convolution operator whose kernel is the backward propagator $S_a^-(\boldsymbol{\rho}) = (S_a^+)^*(-\boldsymbol{\rho})$. In such a case no super-resolution is achieved in the inversion process.

For any value of a, the possibility of going beyond the Rayleigh limit through the use of regularization theory has been explored by Bertero and De Mol [1981a], using Fourier techniques and constraints on the spectrum of the object. The case of finite-extent objects, vanishing outside an a priori known domain, has been considered by Bertero, De Mol, Gori and Ronchi [1983]. As follows from eq. (6.9), in the case of far-field data, the problem of inverse diffraction is exactly equivalent to the out-of-band extrapolation discussed in §4, and the increase in resolution with respect to the bandlimited approximation given by eq. (6.9) can be assessed with the methods already described therein. Let us recall the main conclusion that the Rayleigh limit can be significantly improved only for objects of small spatial extent, corresponding to a small space-bandwidth product. Since the cut-off frequency is given here by $K = 2\pi/\lambda$, this means that the linear spatial dimensions of the object must be of the order of or smaller than the wavelength λ of the field. Such objects are sometimes referred to as *subwavelength sources.*

In the case one has access to near-field data, the use of the information conveyed by the evanescent waves before they get damped allows to increase the effective band available for the restoration. According to formula (3.26), the effective band is the set of spatial frequencies such that

$$|\widehat{S_a^+}(\boldsymbol{\kappa})| = \mathrm{e}^{-a\sqrt{|\boldsymbol{\kappa}|^2-K^2}} > \frac{\varepsilon}{E}\ , \tag{6.10}$$

where E/ε is the signal-to-noise ratio and $|\widehat{S_a^+}(\boldsymbol{\kappa})|$ is the modulus of the Fourier transform of the forward propagator. The corresponding effective cut-off frequency K_{eff} is defined by

$$a\sqrt{K_{\text{eff}}^2 - K^2} = \ln\left(\frac{E}{\varepsilon}\right) \tag{6.11}$$

or else

$$K_{\text{eff}} = K\left[1 + \frac{1}{(Ka)^2}\ln^2\left(\frac{E}{\varepsilon}\right)\right]^{1/2} \quad . \tag{6.12}$$

In table 2, we report some numerical values for the resolution improvement, i.e. for the ratio $R/R_{\text{eff}} = K_{\text{eff}}/K$, where R is the far-field resolution distance. Notice that also in this problem we have a logarithmic dependence on the SNR.

Insert table 2 here.

A further gain in resolution is to be expected from the knowledge of the domain of the object. As in the case of far-field data, a significant additional gain is obtained only for subwavelength sources. The resolution improvement can then be estimated through a numerical computation of the singular values of the imaging operator and of the corresponding number of degrees of freedom. Some numerical results are reported by Bertero, De Mol, Gori and Ronchi [1983] for the one-dimensional problem of a field amplitude invariant with respect to one of the lateral variables and with domain in a slit of width $2X$ in the boundary plane.

The conclusions of the previous analysis provide a theoretical model for understanding why and to what extent near-field imaging devices allow to supersede the classical far-field resolution limit of half a wavelength. Whereas the idea of a super-resolving scanning near-field microscope is already present in a theoretical paper by Synge [1928], the first experimental realization of such a device is due to Ash and Nicholls [1972] in the microwave range. Nowadays, near-field imaging techniques are rapidly developing and are implemented in various ways. Let us mention for example near-field acoustic holography (NAH – see Williams and Maynard [1980]) and scanning near-field optical microscopy (SNOM – see e.g. Betzig and Trautman [1992], Pohl and Courjon [1993], and the papers therein). However, in this latter case, it is not so simple to derive a good theoretical model for the imaging process and to write a closed-form imaging equation like eq. (6.3). As an example of recent attempts in that direction, let us quote a paper by Bozhevolnyi, Berntsen and Bozhevolnaya [1994].

The problem of inverse diffraction can be formulated in other geometries, corresponding to backward propagation from one closed surface to another. The corresponding uniqueness problem for the solution is extensively discussed by Hoenders [1978]. For simple surfaces such as spheres and cylinders, expansions in terms of Bessel or Hankel functions can be written for the solution. The corresponding imaging operator is then compact so that the regularization techniques decribed in §3.3 can be used (see the paper by Bertero, De Mol and Viano [1980] for a discussion of the case of inverse diffraction from cylinder to cylinder).

The results derived in the scalar case can also be generalized to the full vector case, described by the Maxwell equations (see the paper by Hoenders [1978]). Such a treatment is certainly requested for near-field imaging, where polarization effects cannot be neglected. Similar conclusions about resolution limits should hold true also in the vector case, in analogy with those derived above from the simpler model of scalar propagation.

Acknowledgements

The authors would like to thank Franco Gori and Michel Defrise for a critical reading of the manuscript. They are particularly grateful to Prof. Gori for his valuable suggestions and his precious help in clarifying some delicate points.

Christine De Mol is 'Maître de recherches' with the Belgian National Fund for Scientific Research.

Appendix A

Let $S(\mathbf{r})$ be the PSF of an imaging system. It is a function of two variables for a system providing 2D images and a function of three variables for a system providing 3D images (for example, a confocal microscope).

We define the *PSF in the direction* $\boldsymbol{\theta}$ as the function of one variable defined by

$$S_{\boldsymbol{\theta}}(s) = S(s\boldsymbol{\theta}) \ , \tag{A.1}$$

where $\boldsymbol{\theta}$ is a vector of unit length. The width of this function is related to the resolution achievable in the direction $\boldsymbol{\theta}$.

In this Appendix, we derive the relationship between the TF associated with $S_{\boldsymbol{\theta}}(s)$, i.e. $\widehat{S_{\boldsymbol{\theta}}}(k)$, and the TF associated with $S(\mathbf{r})$, i.e. $\widehat{S}(\mathbf{k})$. To this purpose we must recall some definitions related to the Radon transform, which is basic in computerized tomography – see e.g. the book by Natterer [1986].

We define the *projection of* $\widehat{S}(\mathbf{k})$ *in the direction* $\boldsymbol{\theta}$ as the function of one variable defined by

$$(R_{\boldsymbol{\theta}}\widehat{S})(k) = \int\limits_{\boldsymbol{\xi}\cdot\boldsymbol{\theta}=0} \mathrm{d}\boldsymbol{\xi} \ \widehat{S}(k\boldsymbol{\theta} + \boldsymbol{\xi}) \ . \tag{A.2}$$

In the case of a function of two variables, $(R_{\boldsymbol{\theta}}\widehat{S})(k)$ is the integral of $\widehat{S}(\mathbf{k})$ over the straight line perpendicular to $\boldsymbol{\theta}$ with signed distance k from the origin. In the case of a function of three variables, $(R_{\boldsymbol{\theta}}\widehat{S})(k)$ is the integral of $\widehat{S}(\mathbf{k})$ over the plane perpendicular to $\boldsymbol{\theta}$ with signed distance k from the origin.

A basic theorem of the theory of Radon transform is the so-called *projection theorem* or *Fourier slice theorem* - see Natterer [1986] - which provides the relationship we need between the transfer functions. For completeness we reproduce here the proof which is quite simple.

If we take the inverse Fourier transform of $R_{\boldsymbol{\theta}}\widehat{S}$, we get

$$\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \mathrm{d}k \ (R_{\boldsymbol{\theta}}\widehat{S})(k) \ \mathrm{e}^{\mathrm{i}sk} = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \mathrm{d}k \ \mathrm{e}^{\mathrm{i}sk} \int\limits_{\boldsymbol{\xi}\cdot\boldsymbol{\theta}=0} \mathrm{d}\boldsymbol{\xi} \ \widehat{S}(k\boldsymbol{\theta} + \boldsymbol{\xi}) \ . \tag{A.3}$$

Therefore if we introduce the variable $\mathbf{k} = k\boldsymbol{\theta} + \boldsymbol{\xi}$, we have

$$\mathrm{d}\mathbf{k} = \mathrm{d}k\ \mathrm{d}\boldsymbol{\xi}\ ; \qquad sk = s\boldsymbol{\theta} \cdot \mathbf{k}\ , \tag{A.4}$$

so that

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}k\ (R_{\boldsymbol{\theta}}\widehat{S})(k)\ \mathrm{e}^{\mathrm{i}sk} = \frac{1}{2\pi} \int \mathrm{d}\mathbf{k}\ \widehat{S}(\mathbf{k})\ \mathrm{e}^{\mathrm{i}s\boldsymbol{\theta}\cdot\mathbf{k}} = (2\pi)^{n-1} S(s\boldsymbol{\theta}) \tag{A.5}$$

where n is the number of variables. Therefore we conclude that *the inverse Fourier transform of the projection of* $\widehat{S}(\mathbf{k})$ *in the direction* $\boldsymbol{\theta}$ *is precisely the PSF in the direction* $\boldsymbol{\theta}$ *multiplied by* $(2\pi)^{n-1}$. We can also state the result in a direct form: *the Fourier transform of* $S_{\boldsymbol{\theta}}(s)$ *is the projection of* $\widehat{S}(\mathbf{k})$ *in the direction* $\boldsymbol{\theta}$ *multiplied by* $(2\pi)^{1-n}$.

In this way one can associate to each direction $\boldsymbol{\theta}$ a TF and a band, as well as a bandwidth or an effective bandwidth.

Appendix B

In this Appendix we prove that the Gerchberg method (4.16) coincides with the Landweber method (4.17). To this purpose, it is convenient to consider the functions f_j in eq. (4.17) as defined everywhere on $(-\infty, +\infty)$, by putting $f_j(s) = 0$ outside $[-1, +1]$. Then the operator L^* and L^*L, given respectively by eq. (4.7) and eq. (4.8), can be written as follows

$$(L^*g)(s) = \mathrm{rect}(s) \int_{-\infty}^{+\infty} \mathrm{d}s'\ \frac{\sin c(s-s')}{\pi(s-s')}\ g(s') \tag{B.1}$$

and

$$(L^*Lf_j)(s) = \mathrm{rect}(s) \int_{-\infty}^{+\infty} \mathrm{d}s'\ \frac{\sin c(s-s')}{\pi(s-s')}\ f_j(s') \tag{B.2}$$

where $\mathrm{rect}(s)$ is the function defined by eq. (4.15).

Now, if the Fourier transform of $g(s)$ is zero outside $[-c,+c]$, we have

$$g(s) = \int_{-\infty}^{+\infty} \mathrm{d}s' \ \frac{\sin c(s-s')}{\pi(s-s')} \ g(s') \tag{B.3}$$

and therefore, from eq. (B.1), we get for such a bandlimited function

$$(L^* g)(s) = \mathrm{rect}(s) \ g(s) \quad . \tag{B.4}$$

Noticing that from (4.17) we have $f_1 = L^* g$, we conclude from eq. (B.4) that the initial estimate of the procedure (4.17) coincides with the initial estimate of the procedure (4.16). Moreover, since $f_j(s) = \mathrm{rect}(s) \ f_j(s)$, from eq. (4.17), (B.4) and (B.2), it results that $f_{j+1}(s)$ can be written as follows

$$f_{j+1}(s) = \mathrm{rect}(s) \ g_{j+1}(s) \tag{B.5}$$

where

$$g_{j+1}(s) = g(s) + f_j(s) - \int_{-\infty}^{+\infty} \mathrm{d}s' \ \frac{\sin c(s-s')}{\pi(s-s')} \ f_j(s') \ . \tag{B.6}$$

By taking the Fourier transform of both sides of eq. (B.6) we obtain the second of the equations (4.16).

In order to obtain the expression of f_j in terms of the singular values and singular functions of the operator L, we use eq. (4.17) and the following representations of the operators L^* and $L^* L$, which can be derived from eq. (3.29-30):

$$L^* g = \sum_{n=0}^{\infty} \sigma_n \ (g, u_n)_\infty \ v_n \tag{B.7}$$

where $(g, u_n)_\infty$ is the scalar product in $L^2(-\infty, +\infty)$, and

$$L^* L f_j = \sum_{n=0}^{\infty} \sigma_n^2 \ (f_j, v_n)_1 \ v_n \tag{B.8}$$

where $(f_j, v_n)_1$ is the scalar product in $L^2(-1,+1)$. The singular values σ_n and singular functions u_n, v_n are given in eq. (4.9). Now, from eq. (4.17) we obtain, for any n,

$$(f_{j+1}, v_n)_1 = \sigma_n \ (g, u_n)_\infty + (1 - \sigma_n^2) \ (f_j, v_n)_1 \quad . \tag{B.9}$$

It is easy to show by induction that

$$(f_j, v_n)_1 = \sigma_n \left\{1 + (1-\sigma_n^2) + (1-\sigma_n^2)^2 + ... + (1-\sigma_n^2)^{j-1}\right\} (g, u_n)_\infty =$$

$$= \left[1 - \left(1-\sigma_n^2\right)^j\right] \frac{(g, u_n)_\infty}{\sigma_n} \ . \qquad (B.10)$$

Therefore one obtains for the j^{th} iterate the following filtered expansion on the PSWF

$$f_j = \sum_{n=0}^{\infty} W_{n,j} \ \frac{1}{\sigma_n} \ (g, u_n)_\infty v_n \ , \qquad (B.11)$$

the filter being given by

$$W_{n,j} = 1 - (1-\sigma_n^2)^j \ . \qquad (B.12)$$

Since σ_n^2 is always less than one (even if it can be very close to one), we have $0 < W_{n,j} < 1$. In fact, when σ_n^2 is close to one, then $W_{n,j} \simeq 1$. On the other hand, when σ_n^2 is much smaller than 1, then $W_{n,j} \simeq j\sigma_n^2$ and therefore $W_{n,j}$ is also very small if $j\sigma_n^2 \ll 1$. Comparing with eq. (3.33), we see that the inverse of the number of iterations j plays here the same role as the regularization parameter α in the Tikhonov method.

References

Abbe, E., 1873, Archiv. f. Microskopische Anat. **9**, 413.

Abbiss, J.B., C. De Mol and H.S. Dhadwal, 1983, Optica Acta **30**, 107.

Andrews, H.C., and B.R. Hunt, 1977, Digital Image Restoration (Prentice Hall, Englewood Cliffs, New York).

Ash, E.A., and G. Nicholls, 1972, Nature **237**, 510.

Bell, D.A., 1962, Information Theory and its Engineering Applications (Pitman, New York).

Bertero, M., 1989, Linear Inverse and Ill-Posed Problems, in: Advances in Electronics and Electron Physics, Vol. 75, P.W. Hawkes, ed., (Academic Press, New York) pp. 1-120.

Bertero, M., and C. De Mol, 1981a, IEEE Trans. Ant. and Prop. **AP-29**, 368.

Bertero, M., and C. De Mol, 1981b, Atti della Fondazione G. Ronchi **XXXVI**, 619.

Bertero, M., and E.R. Pike, 1982, Optica Acta **29**, 727 and 1599.

Bertero, M., C. De Mol and G.A. Viano, 1979, J. Math. Phys. **20**, 509.

Bertero, M., C. De Mol and G.A. Viano, 1980, The Stability of Inverse Problems, in: Inverse Scattering Problems in Optics, ed. H.P. Baltes (Springer-Verlag, Berlin) ch. V, pp. 161-214.

Bertero, M., C. De Mol, F. Gori and L. Ronchi, 1983, Optica Acta **30**, 1051.

Bertero, M., C. De Mol, E.R. Pike and J.G. Walker, 1984, Optica Acta **31**, 923.

Bertero, M., C. De Mol and E.R. Pike, 1985, Inverse Problems **1**, 301.

Bertero, M., P. Brianzi and E.R. Pike, 1987, Inverse Problems **3**, 195.

Bertero, M., C. De Mol and E.R. Pike, 1987, J. Opt. Soc. Am. **A 4**, 1748.

Bertero, M., C. De Mol and E.R. Pike, 1988, Inverse Problems, **4**, 573.

Bertero, M., P. Boccacci, M. Defrise, C. De Mol and E.R. Pike, 1989, Inverse Problems **5**, 441.

Bertero M., P. Boccacci, G.J. Brakenhoff, F. Malfanti and H.T.M. van der Voort, 1990, J. Microsc. **157**, 3.

Bertero, M., P. Boccacci, R.E. Davies and E.R. Pike, 1991, Inverse Problems **7**, 655.

Bertero, M., P. Boccacci, R.E. Davies, F. Malfanti, E.R. Pike and J.G. Walker, 1992, Inverse Problems **8**, 1.

Bertero, M., P. Boccacci and F. Maggio, 1995, Int. J. Im. Systems Tech. **6**, to appear.

Betzig, E., and J.K. Trautman, 1992, Science **257**, 189.

Biemond, J., R.L. Lagendijk and R.M. Mersereau, 1990, Proc. IEEE **78**, 856.

Born, M., and E. Wolf, 1980, Principles of Optics, 6th edition (Pergamon Press, Oxford).

Bozhevolnyi, S., S. Berntsen and E. Bozhevolnaya, 1994, J. Opt. Soc. Am. **A 11**, 609.

Brakenhoff, G.J., P. Blom and P. Barends, 1979, J. Microsc. **117**, 219.

Brakenhoff, G.J., H.T.M. van der Voort, E.A. van Spronsen and N. Nanninga, 1986, Annals N. Y. Acad. of Sc. **483**, 405.

Cesini, G., G. Guattari, G. Lucarini and C. Palma, 1978, Optica Acta **25**, 501.

Cox, I.J., C.J.R. Sheppard and T. Wilson, 1982, Optik **60**, 391.

Davies, A.R., 1992, Optimality in Regularization, in: Inverse Problems in Scattering and Imaging, eds. M. Bertero and E.R. Pike (Adam Hilger, Bristol) pp. 393-410.

Defrise, M., and C. De Mol, 1992, Inverse Problems **8**, 175.

De Santis, P., and F. Gori, 1975, Optica Acta **22**, 691.

De Santis, P., and C. Palma, 1976, Optica Acta **23**, 743.

De Santis, P., F. Gori, G. Guattari and C. Palma, 1986, Optics Comm. **60**, 13.

Donoho, D.L., I.M. Johnstone, J.C. Hoch and A.S. Stern, 1992, J. Roy. Statist. Soc. **B 54**, 41.

Franklin, J.N., 1970, J. Math. Anal. Applic. **31**, 682.

Frieden, B.R., 1971, Evaluation, design and extrapolation for optical signals, based on the use of prolate functions, in: Progress in Optics, Vol. IX, ed. E. Wolf (North Holland, Amsterdam) ch. VIII.

Frieden, B.R., 1975, Image Enhancement and Restoration, in: Picture Processing and Digital Filtering, ed. T.S. Huang (Springer-Verlag, Berlin) pp. 177-248.

Gabor, D., 1961, Light and Information, in: Progress in Optics, Vol. I, ed. E. Wolf (North-Holland, Amsterdam) ch. IV.

Gerchberg, R.W., 1974, Optica Acta **21**, 709.

Goodman, J.W., 1968, Introduction to Fourier Optics (McGraw-Hill, New York).

Gori, F., 1975, in: Digest of the International Optical Computing Conference, Washington, D.C., April 23-25, 1975 (IEEE Catalog No. 75 CH0941-5C), pp. 137-141.

Gori, F., and G. Guattari, 1985, Inverse Problems **1**, 67.

Gori, F., S. Paolucci and L. Ronchi, 1975, J. Opt. Soc. Am. **65**, 495.

Grochmalicki, J., E.R. Pike, J.G. Walker, M. Bertero, P. Boccacci and R.E. Davies, 1993, J. Opt. Soc. Am. **A 10**, 1074.

Groetsch, C.W., 1984, The theory of Tikhonov regularization for Fredholm equations of the first kind (Pitman, Boston).

Habashy, T., and E. Wolf, 1994, J. Mod. Optics **41**, 1679.

Harris, J.L., 1964, J. Opt. Soc. Am. **54**, 931.

Hoenders, B.J., 1978, The Uniqueness of Inverse Problems, in: Inverse Source Problems in Optics, ed. H.P. Baltes (Springer-Verlag, Berlin) pp. 41-82.

Jerri, A.J., 1977, Proc. IEEE **65**, 1565.

Lagendijk, R.L., J. Biemond and D.E. Boekee, 1988, IEEE Trans. Acoust., Speech and Signal Processing **ASSP-36**, 1874.

Landweber, L., 1951, Amer. J. Math. **73**, 615.

Lent, A., and H. Tuy, 1981, J. Math. Anal. Appl. **83**, 554.

Lucy, L., 1974, Astron. J. **79**, 745.

McCutchen, C.W., 1967, J. Opt. Soc. Am. **57**, 1190.

Meinel, E.S., 1986, J. Opt. Soc. Am. **A 3**, 787.

Natterer, F., 1986, The Mathematics of Computerized Tomography (Teubner, Stuttgart).

Nieto-Vesperinas, M., 1991, Scattering and Diffraction in Physical Optics (John Wiley, New-York).

Paley, R.E.A.C., and N. Wiener, 1934, Fourier Transforms in the Complex Domain (Amer. Math. Soc., Providence, RI).

Papoulis, A., 1968, Systems and Transforms with Applications in Optics (McGraw-Hill, New York).

Papoulis, A., 1975, IEEE Trans. Circuits and Systems **CAS-22**, 735.

Petersen, D.P., and D. Middleton, 1962, Inf. Control **5**, 279.

Pohl, D.W., and D. Courjon, eds., 1993, Near-Field Optics (Kluwer, Dordrecht).

Rayleigh, J.W.S., 1879, Phil. Mag. **8**, 261.

Richards, B., and E. Wolf, 1959, Proc. Roy. Soc. **A 253**, 358.

Richardson, W.H., 1972, J. Opt. Soc. Am. **62**, 55.

Rushforth, C.K., and R.W. Harris, 1968, J. Opt. Soc. Am. **58**, 539.

Schafer, R.W., R.M. Mersereau and M.A. Richards, 1981, Proc. IEEE **69**, 432.

Shannon, C.E., 1949, Proc. IRE **37**, 10.

Shepp, L.A., and Y. Vardi, 1982, IEEE Trans. Med. Imaging **MI-1**, 113.

Sheppard, C.J.R., 1988, Optik **80**, 53.

Sheppard, C.J.R., and A. Choudhury, 1977, Optica Acta **24**, 1051.

Sherman, G.C., 1967, J. Opt. Soc. Am. **57**, 1490.

Shewell, J.R., and E. Wolf, 1968, J. Opt. Soc. Am. **58**, 1596.

Slepian, D., 1964, Bell Syst. Tech. J. **43**, 3009.

Slepian, D., and H.O. Pollack, 1961, Bell Syst. Tech. J. **40**, 43.

Strand, O.N., and E.R. Westwater, 1968, SIAM J. Numer. Anal. **5**, 287.

Synge, E.H., 1928, Phil. Mag. **6**, 356.

Tikhonov, A.N., and V.Y. Arsenin, 1977, Solutions of Ill-posed Problems (Winston/Wiley, Washington).

Toraldo di Francia, G., 1952, Nuovo Cimento Suppl. **9**, 426.

Toraldo di Francia, G., 1955, J. Opt. Soc. Am. **45**, 497.

Toraldo di Francia, G., 1969, J. Opt. Soc. Am. **59**, 799.

Turchin, V.F., V.P. Kozlov and M.S. Malkevich, 1971, Soviet Phys. Uspekhi **13**, 681.

van Cittert, P.H., 1931, Z. Physik **69**, 298.

van der Voort, H.T.M., and G.J. Brakenhoff, 1990, J. Microsc. **158**, 43.

Viano, G.A., 1976, J. Math. Phys. **17**, 1160.

Walker, J.G., 1983, Optica Acta **30**, 1197.

Walker, J.G., E.R. Pike, R.E. Davies, M.R. Young, G.J. Brakenhoff and M. Bertero, 1993, J. Opt. Soc. Am. **A 10**, 59.

Whittaker, E.T., 1915, Proc. Roy. Soc. Edinburgh **A 35**, 181.

Wijnaendts-van-Resandt, R.W., H.J.B. Marsman, R. Kaplan, J. Davoust, E.H.K. Stelzer and R. Stricker, 1985, J. Microsc. **138**, 29.

Williams, E.G., and J.D. Maynard, 1980, Phys. Rev. Lett. **45**, 554.

Wilson, T., and C. Sheppard, 1984, Theory and Practice of Scanning Optical Microscopy (Academic Press, London).

Wolf, E., and T. Habashy, 1993, J. Mod. Optics **40**, 785.

Wolter, H., 1961, On Basic Analogies and Principal Differences Between Optical and Electronic Information, in: Progress in Optics, Vol. I, ed. E. Wolf (North Holland, Amsterdam) ch. V.

Youla, D.C., 1987, Mathematical Theory of Image Restoration by the Method of Convex Projections, in: Image Recovery, Theory and Applications, ed. H. Stark (Academic Press, New York) pp. 29-77.

Youla, D.C., and H. Webb, 1982, IEEE Trans. Med. Imaging **MI-1**, 81.

Young, M.R., R.E. Davies, E.R. Pike and J.G. Walker, 1989, Digest of Conference on Signal Recovery and Synthesis 3 (Optical Society of America, Washington, D.C.) paper WD4.

Young, M.R., R.E. Davies, E.R. Pike, J.G. Walker and M. Bertero, 1989, Europhysics Lett. **9**, 773.

TABLE 1

Table 1 - Number of degrees of freedom $N(E/\varepsilon, c)$ of the problem of out-of-band extrapolation, as a function of the signal-to-noise ratio E/ε and for various values of the Shannon number $S = 2c/\pi$.

c=1, S=0.64		c=2, S=1.27		c=5, S=3.18		c=10, S=6.37	
E/ε	$N(E/\varepsilon, c)$	E/ε	$N(E/\varepsilon, c)$	E/ε	$N(E/\varepsilon, c)$	E/ε	$N(E/\varepsilon, c)$
10	2	10	3	10	5	10	9
10^2	3	10^2	4	10^2	7	10^2	10
10^3	4	10^3	5	10^3	8	10^3	12
10^4	5	10^4	6	10^4	9	10^4	13
10^5	5	10^5	7	10^5	10	10^5	15
10^6	6	10^6	8	10^6	11	10^6	16

TABLE 2

Table 2 - Resolution ratio R/R_{eff} for inverse diffraction from near-field data, as a function of the signal-to-noise ratio E/ε and for various values of the distance a between the source and data planes.

$a = \lambda/20$		$a = \lambda/10$		$a = \lambda/2$		$a = \lambda$	
E/ε	R/R_{eff}	E/ε	R/R_{eff}	E/ε	R/R_{eff}	E/ε	R/R_{eff}
10	7.40	10	3.80	10	1.24	10	1.06
10^2	14.69	10^2	7.40	10^2	1.77	10^2	1.24
10^3	22.01	10^3	11.04	10^3	2.42	10^3	1.49
10^4	29.33	10^4	14.69	10^4	3.10	10^4	1.77
10^5	36.66	10^5	18.35	10^5	3.80	10^5	2.09
10^6	43.99	10^6	22.01	10^6	4.51	10^6	2.42