

Least-squares methods with Poissonian noise: Analysis and comparison with the Richardson-Lucy algorithm

R. Vio^{1,3}, J. Bardsley², and W. Wamsteker³

¹ Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S. Liberale di Marcon, 30020 Venice, Italy
e-mail: robertovio@tin.it

² Department of Mathematical Sciences, The University of Montana, Missoula, MT 59812-0864, USA
e-mail: bardsleyj@mso.umt.edu

³ ESA-VILSPA, Apartado 50727, 28080 Madrid, Spain
e-mail: willem.wamsteker@esa.int

Received 12 September 2004 / Accepted 28 February 2005

Abstract. It is well-known that the noise associated with the collection of an astronomical image by a CCD camera is largely Poissonian. One would expect, therefore, that computational approaches that incorporate this a priori information will be more effective than those that do not. The Richardson-Lucy (RL) algorithm, for example, can be viewed as a maximum-likelihood (ML) method for image deblurring when the data noise is assumed to be Poissonian. Least-squares (LS) approaches, on the other hand, are based on the assumption that the noise is Gaussian with fixed variance across pixels, which is rarely accurate. Given this, it is surprising that in many cases results obtained using LS techniques are relatively insensitive to whether the noise is Poissonian or Gaussian. Furthermore, in the presence of Poisson noise, results obtained using LS techniques are often comparable with those obtained by the RL algorithm. We seek an explanation of these phenomena via an examination of the regularization properties of particular LS algorithms. In addition, a careful analysis of the RL algorithm yields an explanation as to why it is more effective than LS approaches for star-like objects, and why it provides similar reconstructions for extended objects. Finally a comparative convergence analysis of the two algorithms is carried out, with a section devoted to the convergence properties of the RL algorithm. Numerical results are presented throughout the paper. The subject treated in this paper is not purely academic. In comparison with many ML algorithms, the LS algorithms are much easier to use and to implement, are computationally more efficient, and are more flexible regarding the incorporation of constraints on the solution. Consequently, if little to no improvement is gained in the use of an ML approach over an LS algorithm, the latter will often be the preferred approach.

Key words. methods: data analysis – methods: statistical – techniques: image processing

1. Introduction

The restoration of images is a common problem in Astronomy. Astronomical images are blurred due to several factors: the refractive effects of the atmosphere, the diffractive effects of the finite aperture of the telescope, the statistical fluctuations inherent in the collection of images by a CCD camera, and instrumental defects. An example is represented by the spherical aberration of the primary mirror of the *Hubble Space Telescope* (White 1991) that limited the quality of the images before the detector system was refurbished.

The widespread interest in this subject has resulted in the development of a large number of algorithms with different degrees of sophistication (for a review, see Starck et al. 2002). For example, recent wavelet-based approaches have been shown to provide excellent results (e.g., see Neelamani et al. 2004). Unfortunately, these algorithms are very expensive to implement, prohibiting their use on large-scale

image restoration problems and on problems that require the restoration of a large number of images. Consequently, for many restoration problems, less sophisticated and computationally more efficient algorithms must be used. In this respect, the algorithms based on a linear Least-Squares (LS) methodology represent an interesting class. In this paper we discuss two LS approaches: direct and iterative. Direct methods, which we discuss in Sect. 3.1, are the most computationally efficient, while iterative techniques, which we discuss in Sect. 3.2, allow for the straightforward incorporation of constraints.

In spite of the advantages of the LS-based algorithms, astronomers typically use techniques based on a non-linear approach. Such algorithms are usually more difficult to implement, are less flexible and often have slow convergence rates. In particular, the original Richardson-Lucy (RL) algorithm (Richardson 1972; Lucy 1974; Shepp & Vardi 1982) and later modifications have attracted much attention. RL can be viewed as an Expectation-Maximization (EM) algorithm associated

with a Poisson statistical noise model. Linear LS methods, on the other hand, can be viewed as the maximum-likelihood (ML) approach when the noise contaminating the image(s) of interest is additive Gaussian with constant variance across pixels. For CCD camera noise, the statistical assumption inherent in the RL algorithm is much more accurate than that of the LS approach (see Sect. 2). Nonetheless, it is often the case that these two methods provide results of similar quality (see Sect. 4). This is disappointing since it means that in certain instances the RL algorithm is not able to exploit the a priori statistical information. This is particularly relevant when the incorporation of the a priori information results in algorithms that are more expensive and difficult to implement.

The aims of the present paper are as follows: I) to determine the performance of the LS algorithms when the noise is predominantly Poissonian; II) to determine when the LS and RL approaches will give very similar qualitative results. Such questions are not only of academic interest. We believe that due to certain distinct computational advantages, the LS algorithms should be preferentially. However, the LS approach is not always the best choice.

In the next section, we present the statistical model for CCD camera image formation as well as the approximate model that we will use here. After some preliminary considerations in Sect. 2, in Sect. 3 we will explore the convergence properties of the two LS approaches. We will also discuss the performance of these algorithms on different objects. In Sect. 4 we will explore in detail the convergence properties of the RL algorithm. Throughout the paper we present numerical results. We give our conclusion in Sect. 5.

2. Statistical considerations

Astronomical image data is typically collected with a CCD camera. The following statistical model (see Snyder et al. 1993, 1995) applies to image data from a CCD detector array:

$$b(i, j) = n_{\text{obj}}(i, j) + n_{\text{bck}}(i, j) + n_{\text{rdn}}(i, j). \quad (1)$$

Here, $b(i, j)$ is the data acquired by a readout of the pixels of the CCD detector array; $i, j = 1, 2, \dots, N$ (without loss of generality, square images are considered); $n_{\text{obj}}(i, j)$ is the number of object-dependent photoelectrons; $n_{\text{bck}}(i, j)$ is the number of background electrons; $n_{\text{rdn}}(i, j)$ is the readout noise. The random variables $n_{\text{obj}}(i_1, j_1)$, $n_{\text{bck}}(i_1, j_1)$ and $n_{\text{rdn}}(i_1, j_1)$ are assumed to be independent of one another and of $n_{\text{obj}}(i_2, j_2)$, $n_{\text{bck}}(i_2, j_2)$ and $n_{\text{rdn}}(i_2, j_2)$ for $i_1, j_1 \neq i_2, j_2$.

For clearness we use matrix-vector notation. We rewrite Eq. (1) as

$$\mathbf{b} = \mathbf{n}_{\text{obj}} + \mathbf{n}_{\text{bck}} + \mathbf{n}_{\text{rdn}}, \quad (2)$$

where the vectors have been obtained by column stacking the corresponding two-dimensional arrays. The random vector \mathbf{n}_{obj} has a Poisson distribution with Poisson parameter \mathbf{Ax} , where \mathbf{x} is the true image (or object) and \mathbf{A} is a matrix that corresponds to the point spread function (PSF). Depending on the assumed boundary conditions, its structure is typically block-circulant or block-Toeplitz (e.g., see Vio et al. 2003; Vogel 2002). \mathbf{n}_{bck}

and \mathbf{n}_{rdn} are random vectors containing independent entries that have a Poissonian distribution with mean β and a Gaussian distribution with mean 0 and fixed variance σ_{rdn}^2 , respectively.

We will use the following notation to denote model (2):

$$\mathbf{b} = \mathcal{P}[\mathbf{Ax}] + \mathcal{P}[\beta \cdot \mathbf{1}] + \mathcal{N}_{(0, \sigma_{\text{rdn}}^2 \cdot \mathbf{1})}, \quad (3)$$

where $\mathcal{P}[\boldsymbol{\mu}]$ denotes a Poissonian random vector with independent entries and mean given by $\boldsymbol{\mu}$, $\mathcal{N}_{(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)}$ represents a Gaussian random vector with independent entries and mean and variance given by the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, respectively. For iid entries, $\boldsymbol{\mu} = \mu \cdot \mathbf{1}$ and $\boldsymbol{\sigma}^2 = \sigma^2 \cdot \mathbf{1}$. If σ_{rdn}^2 is not too small, we have $\mathcal{P}[\sigma_{\text{rdn}}^2] \approx \mathcal{N}_{(\sigma_{\text{rdn}}^2, \sigma_{\text{rdn}}^2)}$ (see Feller 1971), and hence, using the independence properties of the random variables in Eq. (1) we obtain the following approximation of Eq. (3):

$$\mathbf{b} = \mathcal{P}[\mathbf{Ax} + (\beta + \sigma_{\text{rdn}}^2) \cdot \mathbf{1}] - \sigma_{\text{rdn}}^2 \cdot \mathbf{1}. \quad (4)$$

In order to simplify the analysis, we assume the following simplified model

$$\mathbf{b} = \mathcal{P}[\mathbf{Ax}]. \quad (5)$$

The analysis is easily extended to the model given by Eq. (4). Furthermore, in regions of high intensity (\mathbf{Ax})_{*i*} is much larger than both β and σ_{rdn}^2 , in which case model (4) is well-approximated by model (5).

A further useful approximation is possible if the elements of \mathbf{b} are large. In this case model (5) can be well approximated by

$$\mathbf{b} = \mathbf{Ax} + \mathbf{z}, \quad (6)$$

where \mathbf{z} is a zero mean Gaussian random vector

$$\mathbf{z} \approx \boldsymbol{\gamma} \odot \mathcal{N}_{(0, \mathbf{1})}. \quad (7)$$

Here symbol “ \odot ” denotes Hadamard (element-wise) multiplication and $\boldsymbol{\gamma} = (\mathbf{Ax})^{1/2} = \{(\mathbf{Ax})_i^{1/2}\}$. In other words, through Eq. (6), the nonlinear noise model (5) can be approximated by a linear, additive, *nonstationary*, Gaussian noise model. Our own numerical experiments suggest that in order for approximation (6), (7) to be accurate, it is sufficient that $b_i > 40$. This condition is true in many situations of practical interest (recall that b_i is the number of photons detected by the *i*th pixel in the CCD camera).

From Eqs. (6) and (7) we see that \mathbf{z} has a flat spectrum, i.e., the expected modulus of its Discrete Fourier Transform (DFT) is constant. However, here, unlike a *stationary* Gaussian white noise process, the various Fourier frequencies are not independent of each other (e.g., see Fig. 1). The reason is that the point-wise multiplication in the spatial domain corresponds to convolution in the Fourier domain, and vice versa. Thus, from Eq. (7), we have

$$\widehat{\mathbf{z}}(i, j) = [\widehat{\boldsymbol{\gamma}} \otimes \widehat{\mathcal{N}}_{(0, \mathbf{1})}](i, j), \quad (8)$$

where the symbol “ $\widehat{}$ ” indicates DFT, “ \otimes ” denotes convolution, and (i, j) represents a two-dimensional, discrete frequency index. Since convolution is a linear operation, and $\{\widehat{\mathcal{N}}_{(0, \mathbf{1})}(i, j)\}$ are iid complex Gaussian random variables with zero mean and unit variance, $\{\widehat{\mathbf{z}}(i, j)\}$ are complex Gaussian random variables with zero mean and a constant variance equal to $\sum_{i, j} |\widehat{\boldsymbol{\gamma}}(i, j)|^2$. However, they are not independent of each other.

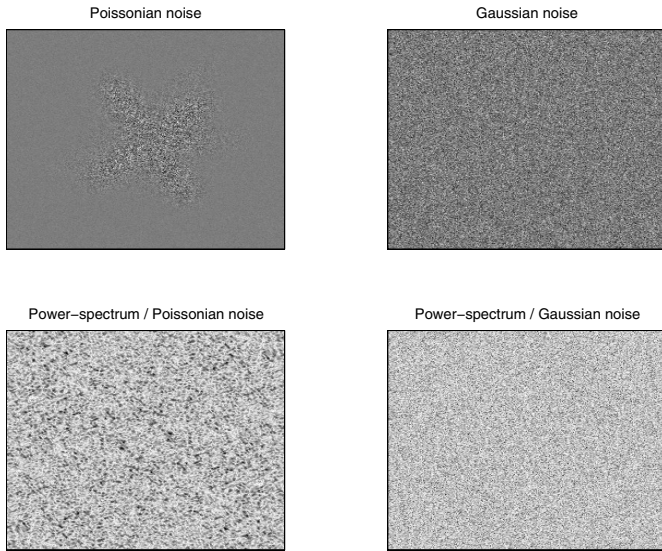


Fig. 1. Comparison of Poissonian vs. stationary Gaussian noise and corresponding power-spectra for the satellite image shown in Fig. 4. The variance of the two noises is the same.

3. Performance of the least-squares approach

The above comments regarding the noise process z provide some insight into the performance of the LS deblurring algorithms in the presence of Poissonian noise. In particular, due to the fact that the LS approach corresponds to the statistical assumption that the contaminating noise is stationary Gaussian, while z is well-approximated by a nonstationary Gaussian random variable, the possible worsening of the performance of LS algorithms has to be expected due to their inability to take into account the dependence between the Fourier frequencies that characterize the spectrum of z . The question then arises as to what happens if such a dependence is not taken into account. In order to understand this point, we remark that the LS estimate of the true image x in Eq. (6) given by

$$x_{LS} = \underset{x}{\operatorname{argmin}} (\|Ax - b\|_2^2); \tag{9}$$

$$= (A^T A)^{-1} A^T b, \tag{10}$$

is extremely unstable since the matrix A is highly ill-conditioned. The equality in Eq. (10) holds provided A has full rank, and the least-squares cost function in problem (9) is convex if A is positive semi-definite. Because of the instability of x_{LS} , a *regularized* or *biased* approximation is typically computed. Such an approximation is obtained by filtering out the high frequency components of x_{LS} (e.g., see Vogel 2002). This fact, together with the following two arguments, suggest that in many astronomical applications the consequences of the LS approach ignoring the dependence between the Fourier frequencies in z will in many instances be small:

1. The images b obtained from astronomical experiments have spectra in which only the lowest frequencies have intensities that are significantly different from zero. This is a consequence of the fact that the PSFs are nearly band

limited, i.e., they are very smooth functions. Furthermore, if a function is in C^k (i.e., it has k continuous derivatives) then its spectrum decreases at least as fast as $1/\nu^{k+1}$. Consequently, this constitutes the lower-limit with which the spectrum of the images can be expected to decrease;

2. The discrete Picard condition plus the Riemann-Lebesgue lemma (Hansen 1997) show that the only Fourier frequencies useful for the restoration of the image are (roughly) those where the contribution of the object is larger than that of the noise.

From such considerations, it is possible to conclude that in the construction of the deblurred image only a few frequencies (i.e., the lowest ones) are primarily used, whereas most of the remaining frequencies are of only marginal importance. For example, in the case of a star-like source (i.e., a non-bandlimited function) and Gaussian PSF with circular symmetry (a C^∞ function) and dispersion (in pixels) σ_p , the spectrum of the observed image is again a Gaussian function with circular symmetry and dispersion (in pixels) $\widehat{\sigma}_p$ given by:

$$\widehat{\sigma}_p \approx N/(2\pi\sigma_p). \tag{11}$$

In several of the numerical experiments presented below, we have used $N = 256$ and $\sigma_p = 12$. In this case, $\widehat{\sigma}_p \approx 3.5$. With the levels used in the simulations, it happens that in $\widehat{b}(i, j)$ the noise becomes dominant when (approximately) $i, j \geq \varrho$, where $10 \leq \varrho \leq 20$. Hence, the LS algorithms can be expected to be almost insensitive to the nature of the noise.

Although the set of frequencies most corrupted by noise are determined both by the noise level and by the spectrum of the PSF, it is the latter factor that has the most influence. To see this we note that, for the case of the star-like source and Gaussian PSF considered above, it is possible to show that the frequency index (i, j) where the spectrum of the signal and that of z have the same level is given by $r \approx N \sqrt{\ln(A_s/\sigma_z)}/(\pi\sigma_p \sqrt{2})$, where $r = (i^2 + j^2)^{1/2}$, A_s is the amplitude of the source, and σ_z is the level of the noise spectrum. In the two next sections, these arguments will be checked in the context of the direct and the iterative LS methods.

The approximation (7) motivates the use of a weighted least-squares (WLS) technique. In this approach, the ℓ_2 norm in Eq. (9) is replaced by an energy norm $\|\cdot\|_C$, where C is the diagonal covariance matrix of the (approximate) noise random variable z , and

$$\|v\|_C^2 = v^T C^{-1} v. \tag{12}$$

In the present context, the entries of C are given by $\{\gamma_i^2\}$, i.e., they depend on x . The WLS approach will provide more accurate reconstructions than LS in the case of nonstationary Gaussian noise, i.e. when C is not a constant matrix, but the direct approach that is discussed in the next section does not, in general, allow for a fast implementation in the case of WLS. An iterative approach can be used for WLS problems, but the cost of implementing these algorithms is the same as that of the RL algorithm. Thus, since RL arises from a more accurate statistical noise model, and is much more widely used in astronomical imaging, we will use RL in the comparisons below.

3.1. The direct approach

One of the most standard approaches for handling the ill-conditioning of problem (9) is known as Tikhonov regularization. In the Tikhonov approach, one obtains a stable estimate of the solution of the linear system of interest by imposing a norm penalization, or bias, to the solution itself. This means that we solve, instead, the penalized least-squares problem

$$\mathbf{x}_\lambda = \operatorname{argmin}_{\mathbf{x}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 \right), \quad (13)$$

or, equivalently, the linear system

$$(\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})\mathbf{x} = \mathbf{A}^T \mathbf{b}. \quad (14)$$

Here λ is a positive scalar known as the *regularization parameter*. The direct solutions of the system (14) can be obtained very efficiently using fast implementations of the DFT (Hunt 1977). Moreover, there are various reliable and well tested criteria that allow for the estimation of λ . A standard technique is known as the generalized cross-validation (GCV) method. With this approach, the optimal value of λ is estimated via the minimization of the GCV function

$$\operatorname{GCV}(\lambda) = \frac{\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{b}\|_2^2 / N^2}{[\operatorname{trace}(\mathbf{I} - \mathcal{A}(\lambda)) / N^2]^2}. \quad (15)$$

Here \mathcal{A} is the matrix that defines the estimator of $\mathbf{A}\mathbf{b}$, i.e., $\mathcal{A}\mathbf{b} = \mathbf{A}\mathbf{x}_\lambda$, and N^2 is the number of pixels in the image. For Tikhonov regularization

$$\mathcal{A}(\lambda) = \mathbf{A}(\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}. \quad (16)$$

It is useful to express model (13) and the GCV function (15) in the Fourier domain:

$$\widehat{x}_\lambda(i, j) = \frac{\widehat{A}^*(i, j)}{|\widehat{A}(i, j)|^2 + \lambda^2} \widehat{b}(i, j), \quad (17)$$

and

$$\operatorname{GCV}(\lambda) = N^2 \sum_{i,j=0}^{N-1} \left| \frac{\widehat{b}(i, j)}{|\widehat{A}(i, j)|^2 + \lambda^2} \right|^2 / \left| \sum_{i,j=0}^{N-1} \frac{1}{|\widehat{A}(i, j)|^2 + \lambda^2} \right|^2. \quad (18)$$

Here, one can compute both \widehat{x}_λ and the minimizer of $\operatorname{GCV}(\lambda)$ very efficiently.

Figures 2–5 compare the results obtainable with this method when the noise is stationary Gaussian and Poissonian, respectively. The image $b_p(i, j)$, contaminated by Poissonian noise, has been obtained by simulating a nonstationary Poissonian process with local mean given by the values of the pixels in the blurred images, i.e. using model (5). Four peak signal to noise (S/N) ratios have been considered¹: 20, 30, 40 and 60 dB. They correspond to situations of very high, intermediate, and very low noise levels. The PSF used in the simulations is a two-dimensional Gaussian function with circular

¹ Here $S/N = 20 \log(b_{\max}/b_{\max}^{1/2})$ dB, where b_{\max} is the maximum value in the image $\mathbf{A}\mathbf{x}$.

symmetry. The image $b_g(i, j)$, contaminated by Gaussian noise, has been obtained by the addition of a discrete stationary white noise process to the blurred images. Both the Gaussian and the Poissonian noises have been simulated through a classic *inverse distribution method* (e.g., see Johnson 1987) by transforming the same set of random uniform deviates. They have exactly the same variance. Here, the subject of interest is superimposed on a sky whose intensity, in the blurred image, is set to 1% of the maximum value of the image. This means that, contrary to the Gaussian case where the noise level is constant across the image, in the Poissonian case the noise is mostly concentrated in the pixels with highest values. In spite of this fact, these figures show that the results provided by Tikhonov coupled with GCV are quite similar regardless of whether the noise is Gaussian or Poissonian.

These results can be explained if one considers Eq. (17), where it is clear that the role of λ is to replace the Fourier coefficients $\widehat{A}(i, j)$ with small modulus, i.e., those coefficients that make the deblurring operation unstable. According to the two points mentioned above, the “optimal” value of λ should replace all the Fourier coefficients whose modulus is smaller than the expected level of the noise in the Fourier domain. Since in $b_p(i, j)$ and $b_g(i, j)$ the level of the noise is the same, such a replacement will be quite important and will have a similar effect for both images. This is shown by the results in Figs. 6–9. In particular, the *c*) panels show that GCV chooses λ so that the frequencies corresponding to the flat part of the spectra (i.e., those dominated by the noise) are filtered out. The consequence of this is that, for both Gaussian and the Poissonian noises, almost the same number of coefficients are filtered. Moreover, as is shown in the *d*) panels, the coefficients $|\widehat{b}_p(i, j)|$ and $|\widehat{b}_g(i, j)|$ corresponding to the coefficients $\widehat{A}(i, j)$ with the largest modulus, are very similar. From this, one can conclude that the deblurred images \mathbf{x}_λ can be expected to be very similar regardless of the nature of the noise.

The reason why the two GCV curves are almost identical (see the *b*) panels) is that, independently of the nature of the noise, in Eq. (18) the quantity $\widehat{b}(i, j)$ can be replaced by $\widehat{b}(i, j) = \widehat{A}(i, j)\widehat{x}_\lambda(i, j) + \widehat{z}_n(i, j)$, where $\widehat{z}_n(i, j)$ is given by Eq. (8) or by a stationary white noise process. Now, taking the expected value of the resulting $\operatorname{GCV}(\lambda)$, it is not difficult to show that

$$E[\operatorname{GCV}(\lambda)] = N^2 \sum_{i,j=0}^{N-1} \frac{|\widehat{A}(i, j)\widehat{x}_\lambda(i, j)|^2 + \sigma_{z_n}^2}{(|\widehat{A}(i, j)|^2 + \lambda^2)^2} / \left| \sum_{i,j=0}^{N-1} \frac{1}{|\widehat{A}(i, j)|^2 + \lambda^2} \right|^2. \quad (19)$$

Since the term $\sigma_{z_n}^2$ is constant, the $E[\operatorname{GCV}(\lambda)]$ function is independent of the specific nature of the noise. The same is not true for the variance. However, because of the arguments presented above, no instabilities are to be expected. This is supported by the fact that in our numerical experiments we have never experienced stability problems (see also Fig. 10).

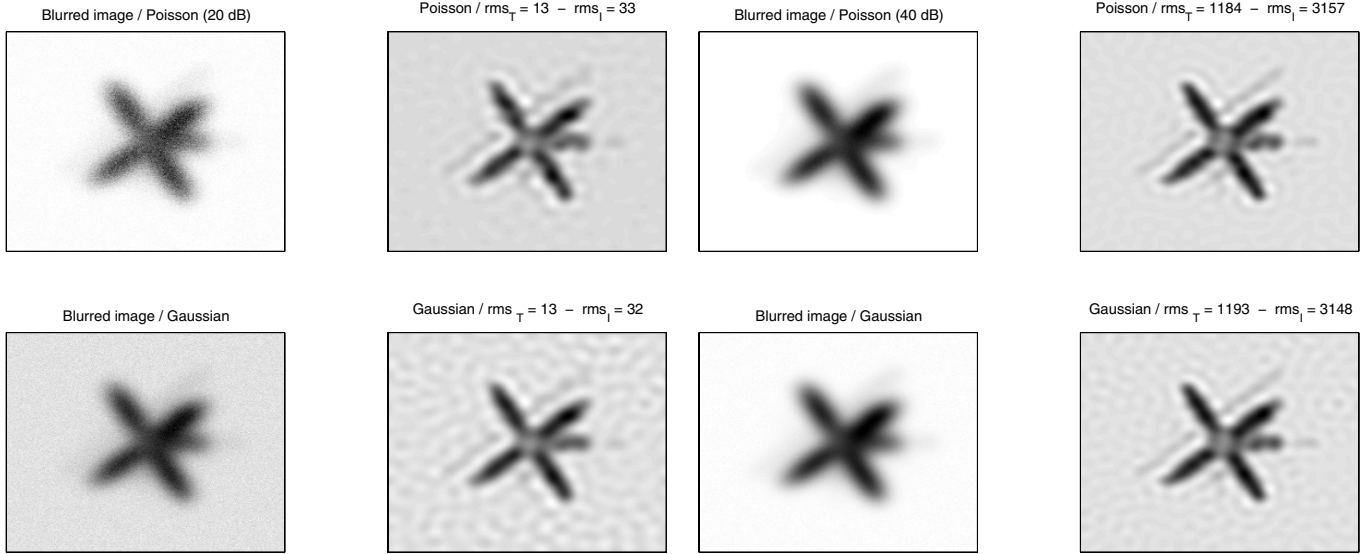


Fig. 2. Comparison of the results obtained by Tikhonov coupled with GCV in the case of Poissonian and Gaussian noises (see text). The images have sizes 256×256 pixels, the PSF is Gaussian with circular symmetry and dispersion set to 12 pixels, $S/N = 20$ dB. rms_T and rms_I are the rms of the true residuals calculated on the entire image and only on the pixels corresponding to the satellite, respectively.

Fig. 4. As in Fig. 2 but with $S/N = 40$ dB.

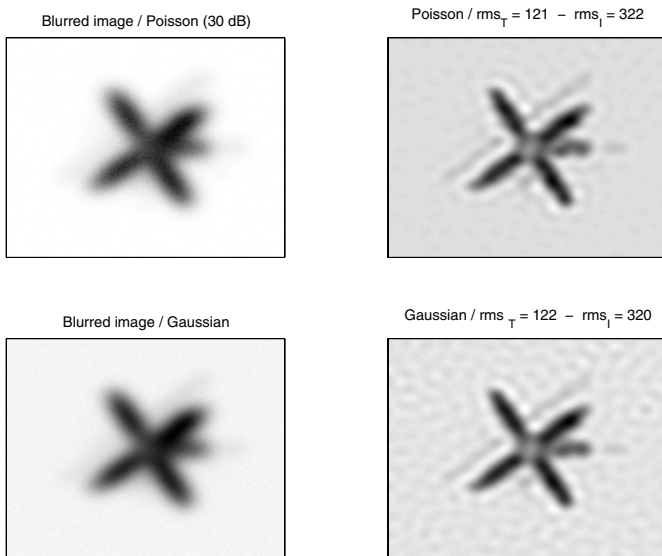


Fig. 3. As in Fig. 2 but with $S/N = 30$ dB.

It is not difficult to see that similar arguments hold also when in model (13) the norm penalization is substituted by a smoothness one, i.e., when

$$x_\lambda = \underset{x}{\operatorname{argmin}} \left(\|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2 \right). \quad (20)$$

Typically L is taken to be a discrete approximation of some derivative operator.

One can use an iterative approach together with Tikhonov regularization, i.e., one can use an iterative approach for solving problem (13). These approaches allow for the incorporation of nonnegativity constraints, while the direct approach does not. See Vogel (2002) for a thorough discussion.

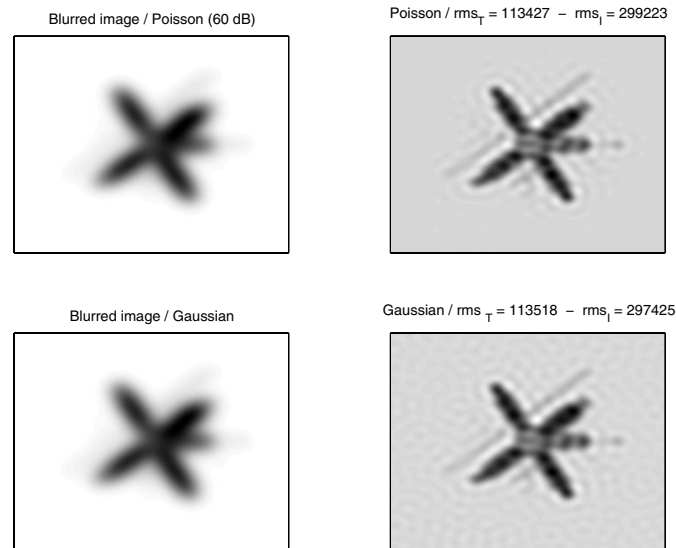


Fig. 5. As in Fig. 2 but with $S/N = 60$ dB.

3.2. The iterative approach

To deal with the ill-conditioning of problem (9) an iterative approach can also be used. Although computationally less efficient than the direct method discussed above, they are much more flexible in that they allow for the straightforward incorporation of constraints. These algorithms provide regularization via a semiconvergence property; that is, the iterates first reconstruct the low frequency components of the signal, i.e. those less contaminated by noise, and then the high frequency ones. In other words, the number of iterations plays the same role as the regularization parameter λ in the Tikhonov approach.

Semiconvergence has been rigorously proved only for a limited number of algorithms (see Vogel 2002; Lee & Nagy 2004). For others, some theoretical results are available but the primary evidence stems from many years of success in use in applied problems.

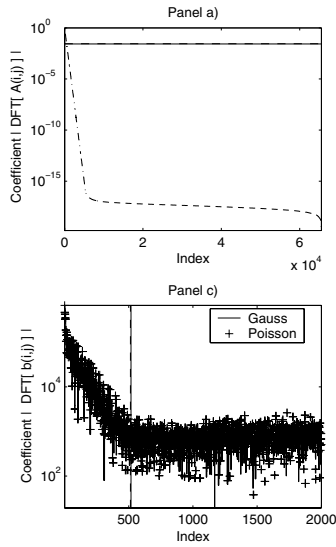


Fig. 6. Comparison of the results obtained by Tikhonov coupled with GCV in the case of Poissonian and Gaussian noises. This figures correspond to the experiment shown in Fig. 2. Panel **a)** the coefficients $|\widehat{b}_g(i, j)|$ and $|\widehat{b}_p(i, j)|$ in decreasing order. The two horizontal lines represent the values of λ for the two noises; **b)** corresponding GCV functions; **c)** coefficients $|\widehat{b}_p(i, j)|$ and $|\widehat{b}_g(i, j)|$ corresponding to the first 2000 coefficients of $|\widehat{A}(i, j)|$ shown in panel **a)**. The vertical lines show the indices of $|\widehat{b}(i, j)|$ closest to λ . **d)** $\Delta b(i, j) = |\widehat{b}_g(i, j) - \widehat{b}_p(i, j)| / |\widehat{b}_g(i, j)|$ vs. the corresponding first 2000 coefficients of $|\widehat{A}(i, j)|$ with the largest modulus. $S/N = 20$ dB.

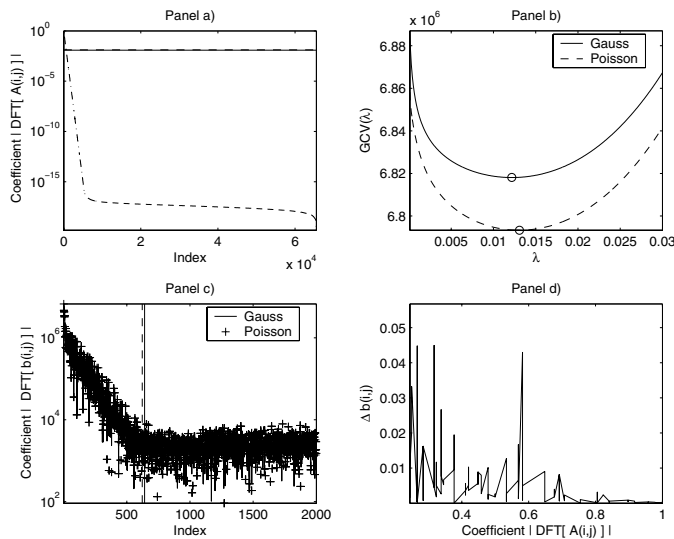


Fig. 7. As in Fig. 6 but with $S/N = 30$ dB.

The prototypical iterative regularization algorithm for least squares problems is the *Landweber method* (LW). This is a gradient descent algorithm for solving problem (9). In particular, it creates a sequence of iterates that, provided A has full column rank, converges to the solution of Eq. (10). Because of its slow convergence, it is not frequently used in practical applications. However, because of its simplicity, it is often utilized in the theoretical analysis of the potential performances of the LS approach.

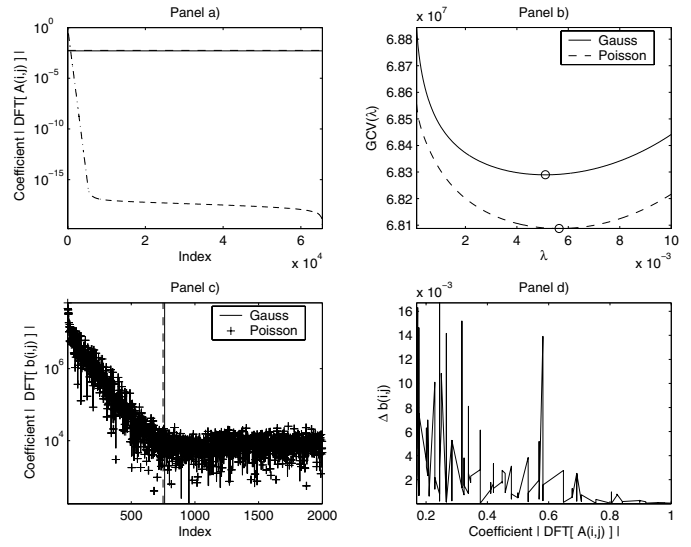


Fig. 8. As in Fig. 6 but with $S/N = 40$ dB.

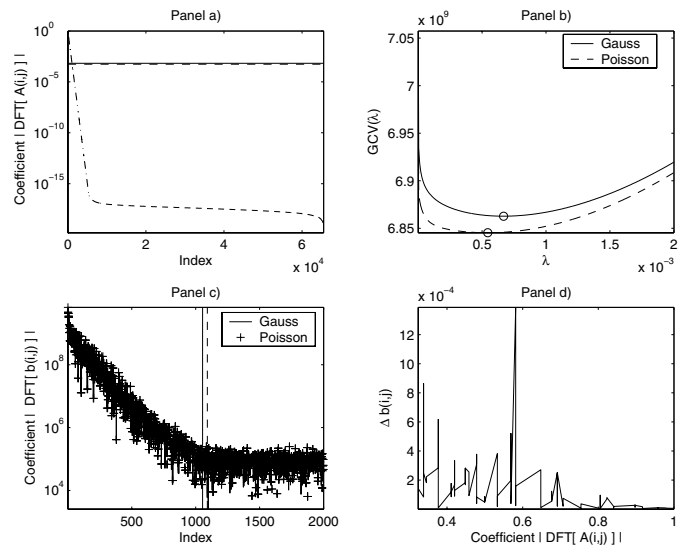


Fig. 9. As in Fig. 6 but with $S/N = 60$ dB.

If x_0 is the starting image (often $x_0 = \mathbf{0}$), then the iterations take the form

$$x_k = x_{k-1} + \omega A^T [b - Ax_{k-1}], \tag{21}$$

where, $k = 1, 2, \dots$, and ω is a real positive parameter satisfying $0 < \omega < 2 / \|A^T A\|$. The values of ω determine, in part, the convergence of the iteration. The semiconvergence property of this algorithm is typically proved using arguments based on the singular values decomposition of the matrix A (for a discussion of this, see Vogel 2002). However, it is, perhaps, more instructive to rewrite Eq. (21) in the Fourier domain, obtaining

$$\widehat{x}_k(i, j) = \frac{\widehat{b}(i, j)}{\widehat{A}(i, j)} \left[1 - (1 - \omega |\widehat{A}(i, j)|^2)^k \right], \tag{22}$$

with $0 < \omega < 2 / \max [|\widehat{A}(i, j)|^2]$. If, as usual, the PSF is assumed to have unit volume, then $\max [|\widehat{A}(i, j)|^2] = 1$ and $0 < \omega < 2$. From this equation, one can see that, for a given frequency index (i, j) , the closer the term $\omega |\widehat{A}(i, j)|^2$ is to one,

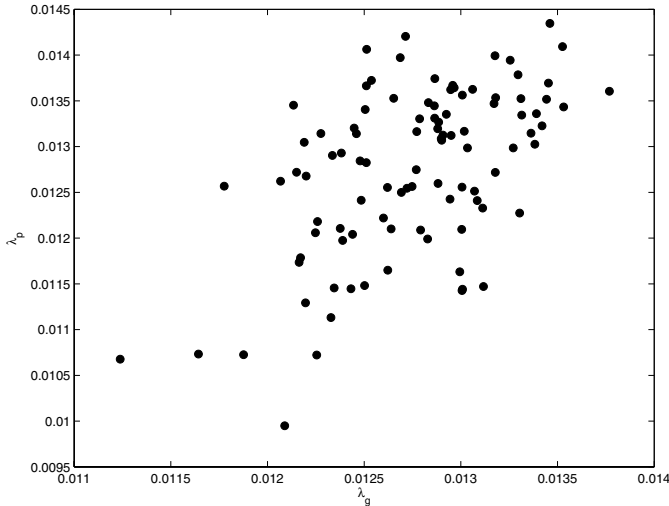


Fig. 10. Relationship between the estimates λ_g and λ_p , corresponding to the Gaussian and the Poissonian noise, respectively, obtained in 100 different realizations of the experiment shown in Fig. 3. The mean value and the standard deviation are 0.128 and 4.55×10^{-4} for λ_g , and 0.128 and 9.10×10^{-4} for λ_p .

the more rapid is the convergence to $\widehat{b}(i, j)/\widehat{A}(i, j)$, which corresponds to the unregularized solution. Since, as mentioned above, the largest values of the spectrum of $\widehat{A}(i, j)$ correspond to the lowest frequencies, it is evident from Eq. (22) that the lower frequencies are restored in early iterations, while progressively higher frequencies are restored as the iteration progresses.

There are gradient descent algorithms similar to Landweber for minimizing a least-squares function subject to nonnegativity constraints (e.g., see Vogel 2002).

3.2.1. Convergence properties

Equation (21) shows that the convergence of LW is driven by the rate with which the term

$$\widehat{\mathcal{K}}_k(i, j) = (1 - \omega|\widehat{A}(i, j)|^2)^k \quad (23)$$

goes to zero. In order to understand what this means in practical situations, it is useful to see what happens in the case of a noise-free image when the PSF is a two-dimensional Gaussian with circular symmetry and dispersion σ_p . Without loss of generality, we set $\omega = 1$. Then

$$\widehat{\mathcal{K}}_k(i, j) \approx [1 - \exp(-r^2/\sigma_p^2)]^k, \quad (24)$$

where $r^2 = i^2 + j^2$ and σ_p is the dispersion of the PSF in the frequency domain (see Eq. (11)). From this equation, even in the case of moderate values of k , the term within square brackets on the rhs of Eq. (22) can be well-approximated by the Boxcar function

$$\Pi_k(i, j) = \begin{cases} 1 & \text{if } 0 \leq |r| \leq r_{0.5,k}; \\ 0 & \text{otherwise} \end{cases}, \quad (25)$$

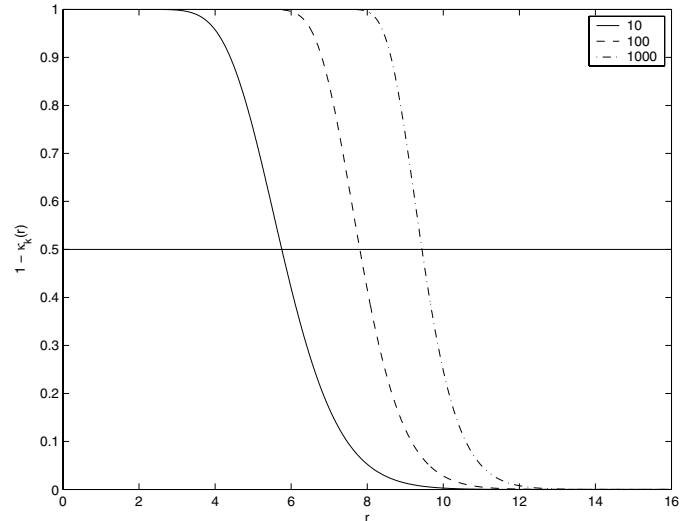


Fig. 11. Plot of $1 - \widehat{\mathcal{K}}_k(r)$ vs. r for different values of the iteration count k .

where $r_{0.5,k}$ is the value of r for which $\widehat{\mathcal{K}}_k(r) = 0.5$ (see also Fig. 11). Therefore, the iterate (22) can be approximated by

$$\widehat{x}_k(i, j) = \frac{\widehat{b}(i, j)}{\widehat{A}(i, j)} \Pi_k(i, j). \quad (26)$$

The requirement that $\widehat{\mathcal{K}}_k(i, j) \leq \epsilon$, with $0 < \epsilon < 1$ implies

$$k > \frac{\ln \epsilon}{\ln [1 - \exp(-r^2/\sigma_p^2)]}. \quad (27)$$

This result shows that the restoration of the highest frequencies (i.e., large r), requires a number of iterations that becomes rapidly huge. For the case of the star-like sources, where all the frequencies have to be restored, this means an extremely slow convergence. More specifically, from Eq. (27) one can see that for frequencies (i, j) such as $r \approx \sigma_p$, $k \propto r$, while for larger values of r , k increases exponentially. For example, some of the experiments presented in this paper are based on images with size 256×256 pixels and with a Gaussian PSF with $\sigma_p \approx 3.5$. In this case, if $\epsilon = 0.5$, in order to have $r_{0.5,k} = \sigma_p, 2\sigma_p, 3\sigma_p, 6\sigma_p$ (i.e., $r_{0.5,k} = 3.5, 7, 10.5, 21$), it is necessary that $k \approx 2, 4, 5600, 3 \times 10^{15}$, respectively.

The obvious conclusion is that LW is useful only for the restoration of objects for which the low frequencies are dominant, e.g. extended objects with smooth light distributions.

3.2.2. Numerical results

Since, regardless of the noise type, LW reconstructs the lower frequency components of the image first (i.e., the frequencies where the contribution of the noise is negligible), we expect the following for both $b_p(i, j)$ and $b_g(i, j)$:

1. The resulting deblurred images should be very similar;
2. In early iterations the convergence rate of the algorithms should be almost identical.

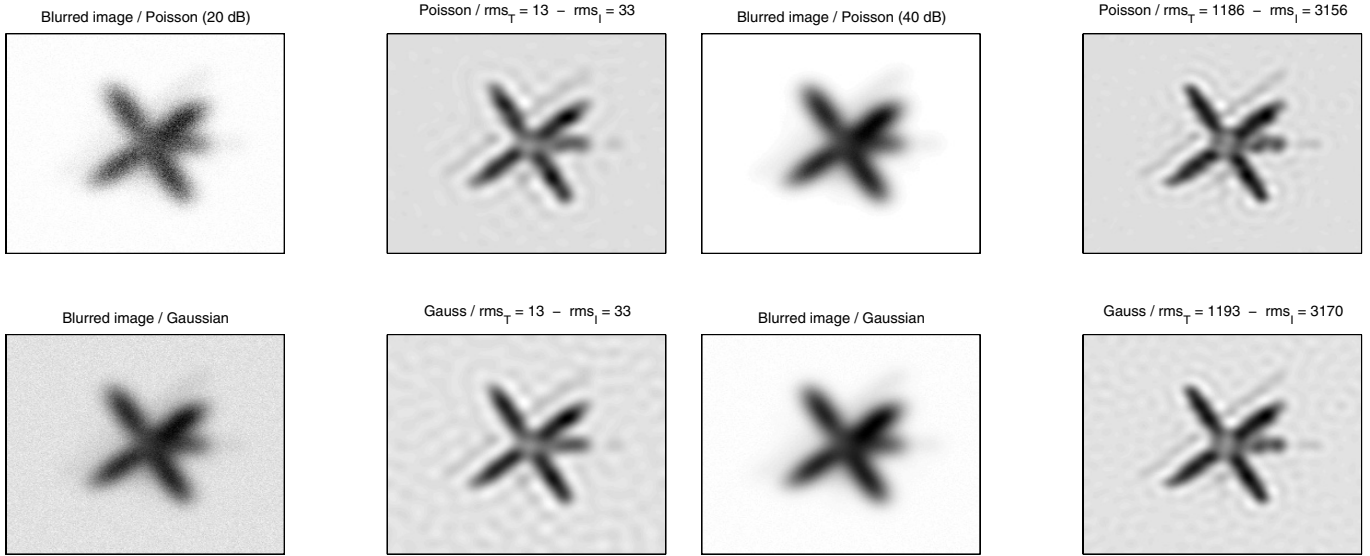


Fig. 12. LW deblurring of the images shown in Fig. 2 corresponding to the smallest value of the rms of the true residuals calculated on the entire image, say rms_T . rms_I is the rms calculated by using only the pixels containing the contribution of the satellite.

Fig. 14. LW deblurring of the images shown in Fig. 4. rms_T and rms_I as in Fig. 12.

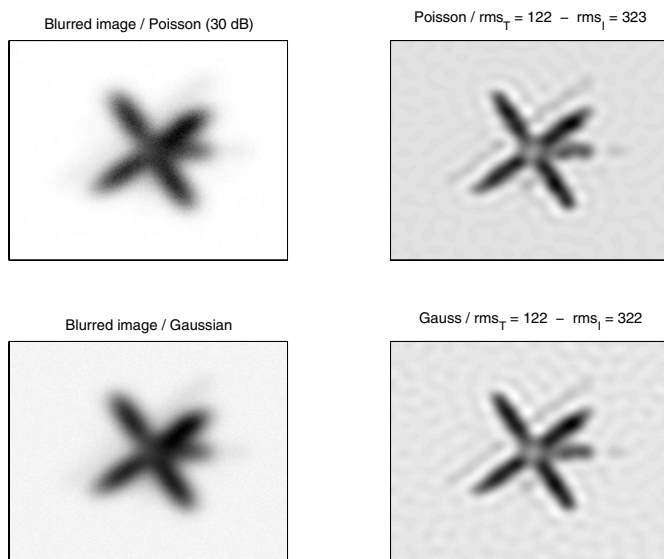


Fig. 13. LW deblurring of the images shown in Fig. 3. rms_T and rms_I as in Fig. 12.

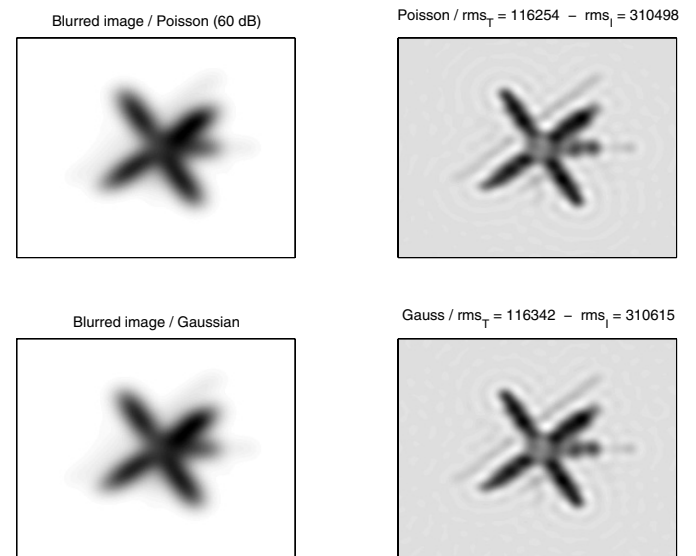


Fig. 15. LW deblurring of the images shown in Fig. 5. rms_T and rms_I as in Fig. 12.

These statements are supported in Figs. 12–15 and Figs. 16–19. In particular, from the last set of figures one can see that the convergence curves are almost identical until the minimum rms of the true residual is reached. After that, because the high frequencies (the ones that are more sensitive to the nature of the noise) begin to be included in the restoration, the curves diverge.

4. Richardson-Lucy algorithm

In the previous sections we have shown that the LS methods are relatively insensitive to the specific nature of the noise. However, this does not mean that they are optimal. In principle, methods that exploit the a priori knowledge of the statistical

characteristics of the noise should be able to provide superior results.

As has been stated above, the least-squares problem (9) corresponds to the assumption of additive Gaussian, stationary noise statistics. In this case, the solution is known as the maximum likelihood estimator (MLE) of the true image. Assuming, instead, noise model (5), one can show that the corresponding MLE is the minimizer of the negative Poisson log-likelihood function

$$J(x) = \mathbf{1}^T [Ax - b \odot \log(Ax)], \tag{28}$$

subject to the nonnegativity constraint $x \geq \mathbf{0}$. The nonnegativity constraint can be motivated physically by noting that the photon count at each pixel is positive. The function J in Eq. (28) is convex. In fact, assuming A is positive definite, J is

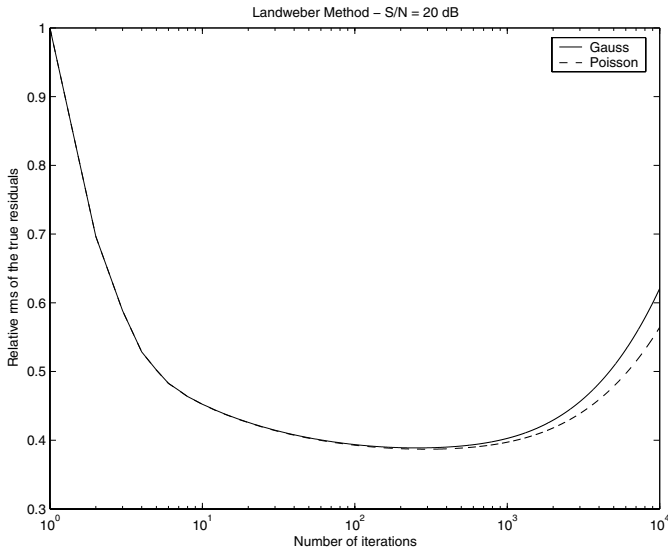


Fig. 16. Convergence rate of the LW algorithm applied to the problem of deblurring the images shown in Fig. 2.

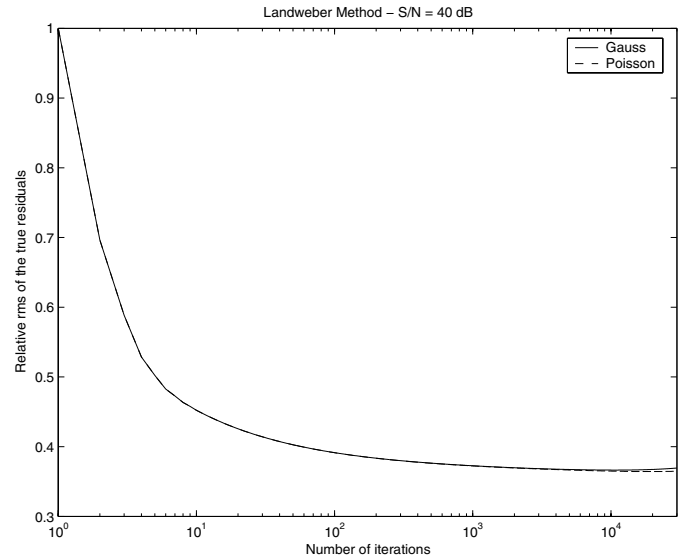


Fig. 18. Convergence rate of the LW algorithm applied to the problem of deblurring the images shown in Fig. 4.

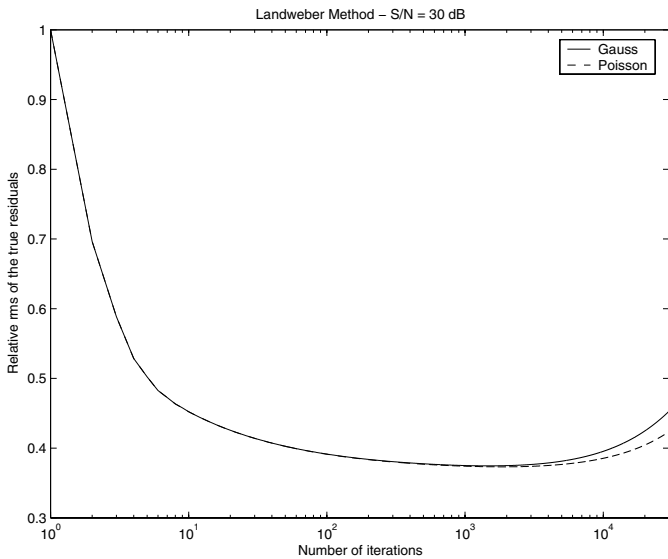


Fig. 17. Convergence rate of the LW algorithm applied to the problem of deblurring the images shown in Fig. 3.

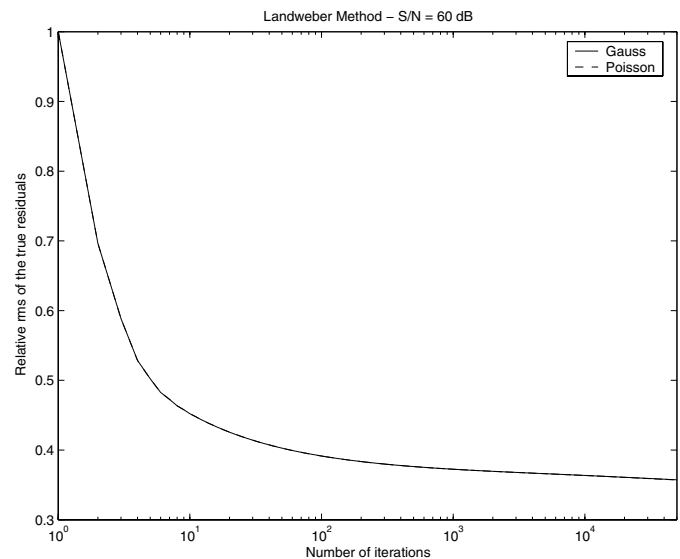


Fig. 19. Convergence rate of the LW algorithm applied to the problem of deblurring the images shown in Fig. 5.

strictly convex. In this case, the unique minimizer \mathbf{x} of \mathbf{J} subject to the nonnegativity constraint $\mathbf{x} \geq \mathbf{0}$ must satisfy the Kuhn-Tucker condition

$$\mathbf{x} \odot \nabla \mathbf{J}(\mathbf{x}) = \mathbf{0}, \tag{29}$$

where

$$\nabla \mathbf{J}(\mathbf{x}) = \mathbf{A}^T \left(\mathbf{1} - \frac{\mathbf{b}}{\mathbf{Ax}} \right). \tag{30}$$

If the standard assumption $\mathbf{A}^T \mathbf{1} = \mathbf{1}$ is made, Eq. (29) can then be written

$$\mathbf{x} = \mathbf{x} \odot \mathbf{A}^T \frac{\mathbf{b}}{\mathbf{Ax}}, \tag{31}$$

where the fraction of two vectors denote Hadamard (component-wise) division. From Eq. (31) we obtain the fixed point iteration

$$\mathbf{x}_{k+1} = \mathbf{x}_k \odot \mathbf{A}^T \frac{\mathbf{b}}{\mathbf{b}_k}, \tag{32}$$

where $\mathbf{b}_k = \mathbf{Ax}_k$. Iteration (32) is known as Richardson-Lucy (RL). It can be also obtained using EM formalism (e.g., see Vogel 2002; Shepp & Vardi 1982).

Since RL exploits the a priori knowledge regarding the statistics of photon counts, it should be expected to yield more accurate reconstructions than an approach that does not use this information. In reality, as shown by Figs. 20–27, the situation is not so clear. These figures provide the convergence rates and the performances of RL and LW methods for objects with a size that is increasing with respect to the size of the PSF

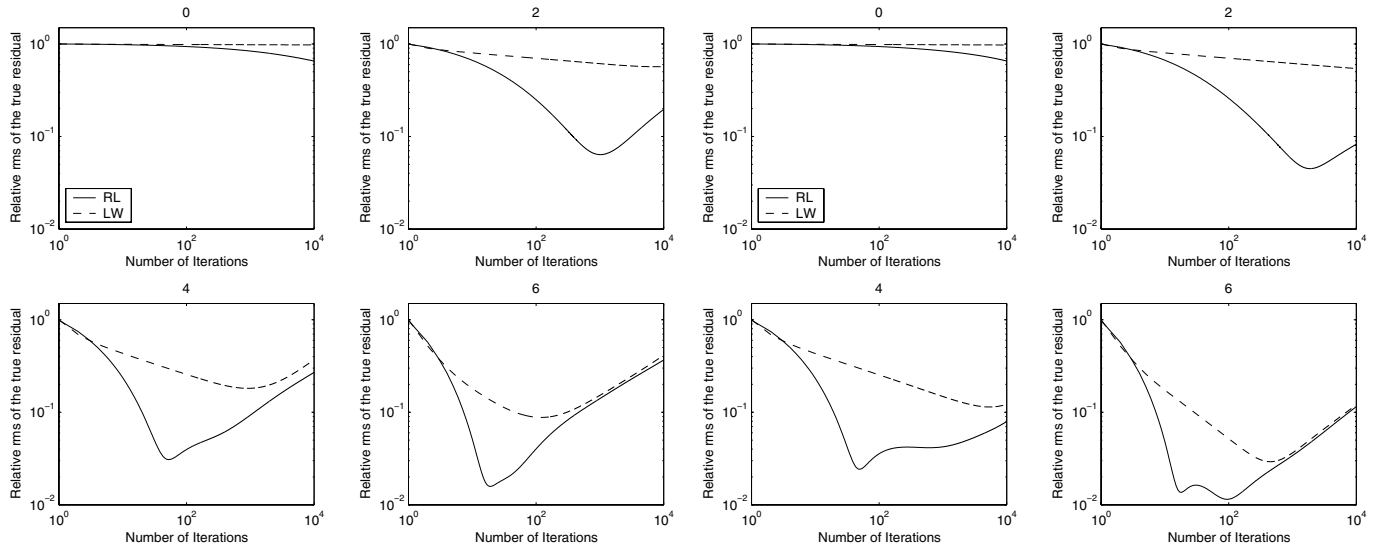


Fig. 20. $\|x_k - x\|/\|x\|$ vs. the number of iterations. The object of interest, located in the center of the image, is a two-dimensional Gaussian with circular symmetry and dispersion (in pixels) given in the top of each panel. It is superimposed on a background whose level is 1% of the peak value of the blurred image. The PSF is a two-dimensional Gaussian with a dispersion of 6 pixels. The size of the image are 256×256 pixels. The noise is Poissonian with peak $S/N = 30$ dB. Two deblurring algorithms are used: Richardson-Lucy (RL) and LW.

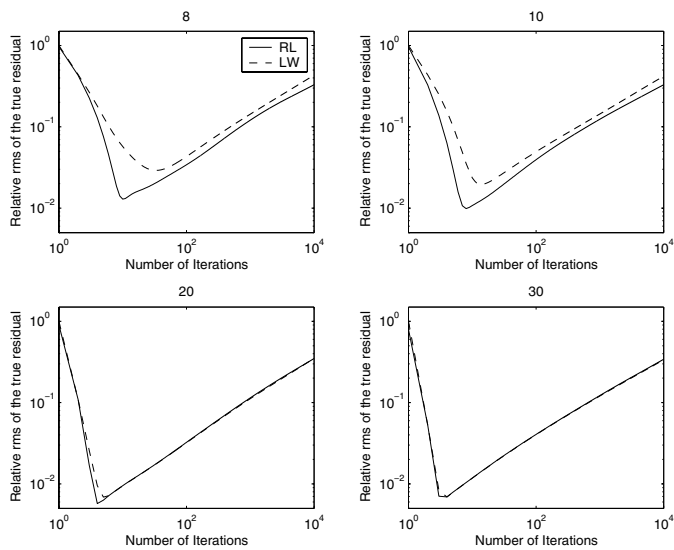


Fig. 21. As in Fig. 20.

(a two-dimensional Gaussian with circular symmetry). Two different types of objects are considered: a two-dimensional Gaussian and a rectangular function. Since the first target object presents an almost band-limited spectrum, whereas for the second target object the high-frequency Fourier components are important, their restorations represent very different problems. For both experiments, a background with an intensity of 1% of the maximum value in the blurred image has been added. Finally, two different levels of noise have been considered corresponding to a peak S/N of 30 and 40 dB, respectively. The first case provides a background with an expected number of counts

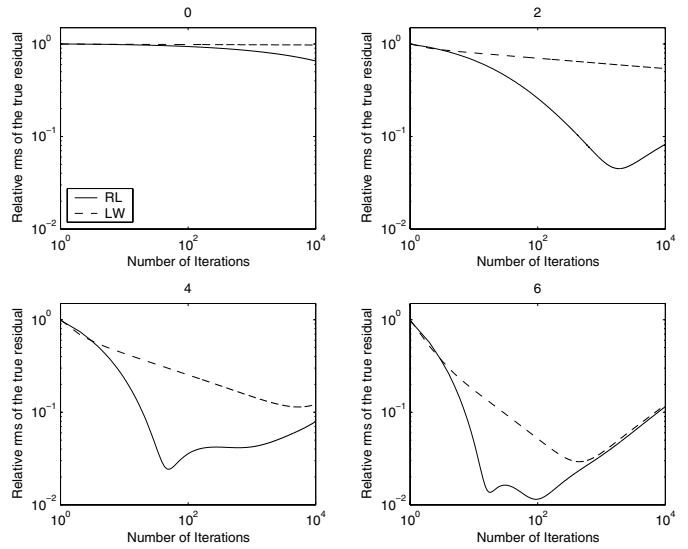


Fig. 22. As in Fig. 20 but with $S/N = 40$ dB.

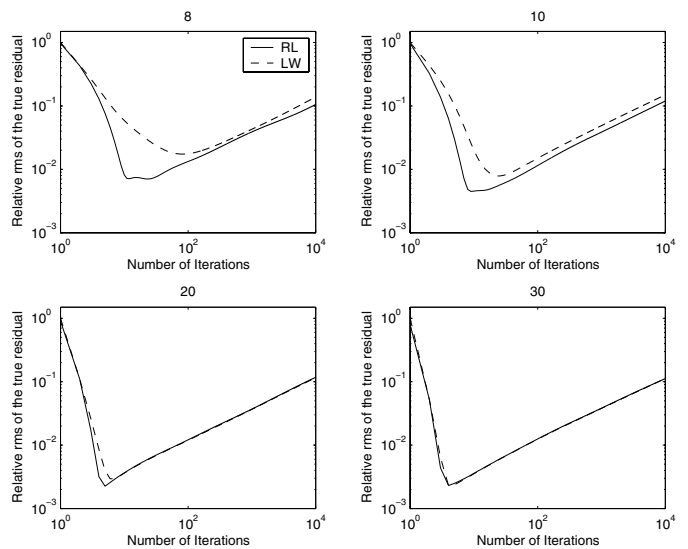


Fig. 23. As in Fig. 20 but with $S/N = 40$ dB.

approximately equal to 30, i.e., a level for which the Gaussian approximation of the Poissonian distribution is not very good.

From Figs. 20–27 it appears that the performance of RL for objects narrower than the PSF is in general superior to LW for the band-limited target. The same is not true for the other objects. Interestingly, though, for extended objects, i.e. smooth objects with high intensity profiles over large regions, the performance of RL is roughly equivalent to that of LS (to properly compare the convergence rate, it is necessary to take into account that, for each iteration, RL requires the computation of twice the number of two-dimensional DFT than is required by LW). This is especially true for the images characterized by the best S/N . Motivated by these numerical results, we seek answers to the following questions: (i) why does RL perform better than LS on star-like objects; and (ii) why do the RL and LS approaches yield roughly the same results on extended objects?

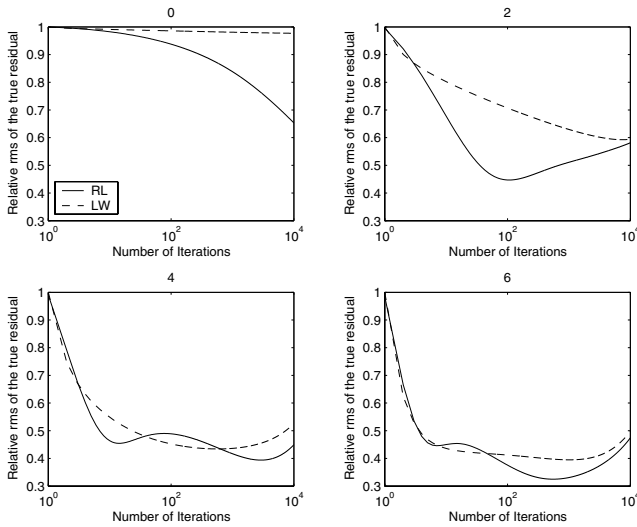


Fig. 24. $\|x_k - x\|/\|x\|$ vs. the number of iterations. The object of interest, located in the center of the image, is of a two-dimensional rectangular function, with side length (in pixels) given in the top of each panel. It is superimposed on a background whose level is 1% of the peak value of the blurred image. The PSF is a two-dimensional Gaussian with a dispersion of 6 pixels. The size of the image are 128×128 pixels. The noise is Poissonian with peak $S/N = 30$ dB. Two deblurring algorithms are used: Richardson-Lucy (RL) and LW.

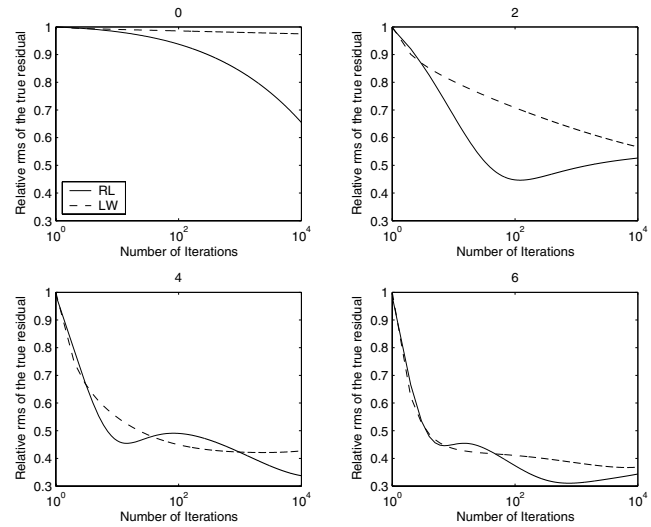


Fig. 26. As in Fig. 24 but with $S/N = 40$ dB.

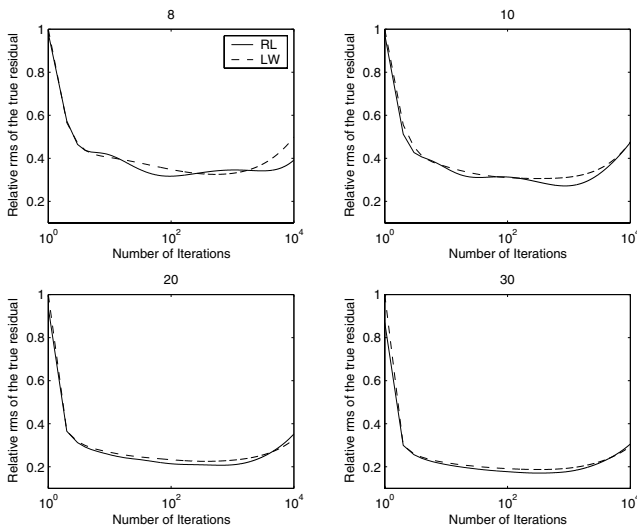


Fig. 25. As in Fig. 24.

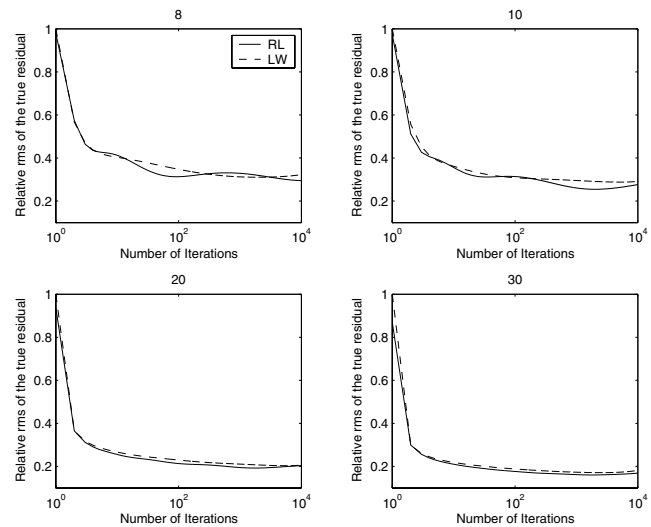


Fig. 27. As in Fig. 24 but with $S/N = 40$ dB.

4.1. RL vs. LS: Preliminary comments

In practice, when either the RL or LS approaches are used in solving image reconstruction problems, the exact computation of the maximum likelihood estimate (MLE) is not sought. For example, as was stated above, the LW iteration implements regularization via the iteration count. In fact, the objective when using LW is to stop the iteration late enough so that an accurate reconstruction is obtained, but before the reconstruction is corrupted by the noise in the high frequency components of the image. Notice, for example, that in Figs. 20–27 the relative error begins to increase at a certain point in both the RL and LW iterations.

As was stated above, one can show that the LW iterations provide regularized solutions of $A^T A x = A^T b$ via the singular value decomposition (SVD) of the matrix A . Unfortunately, such an analysis of RL is impossible due to the nonlinearity in the RL algorithm. In particular, note the Hadamard multiplication and division in algorithm (32). Instead, we first note that if A is an invertible matrix, RL iterations converge to the MLE corresponding to the statistical model (5) (see Wu 1983). The MLE is also the minimizer of the function (28) subject to $x \geq 0$. When the RL algorithm is used on image deblurring problems it exhibits a similar convergence behavior to that of the LW iteration. Specifically, for ill-conditioned problems, the RL iterates $\{x_k\}$ provide more accurate reconstructions in early iterations (semiconvergence property), while in later iterations blow-up occurs (Lucy 1974; Carasso 1999). To explain why this occurs, we note that at the positive components x_j of the MLE vector x , we must have, from Eqs. (29) and (30), that

$$\left[\mathbf{1} - \frac{\mathbf{b}}{A\mathbf{x}} \right]_j = 0, \tag{33}$$

or $[Ax]_j = [b]_j$. If $x_j > 0$ for all j and A is invertible, we have $x = A^{-1}b$, which is what LW iterations will converge to if A is invertible. Thus, it is not surprising that since blow-up occurs as $x_k \rightarrow x_*$ in the LW iteration when A is a poorly conditioned matrix, we will see the same results when we use RL. One consequence of this fact is that during reconstruction, RL uses only a few frequencies and therefore cannot fully exploit prior statistical information regarding the noise (this would require the use of the entire spectrum).

4.2. RL vs. LS: A sensitivity analysis

To obtain a deeper understanding of the RL and LS approaches and to answer the two questions posed above, we introduce the quantities

$$\Delta_{LW}(x_k) = -A^T r_k; \tag{34}$$

$$\Delta_{RL}(x_k) = -x_k \odot A^T \frac{r_k}{Ax_k}, \tag{35}$$

which provide the correction to the solution x_k at the k th iteration for LW and RL, respectively. Here,

$$r_k = Ax_k - b, \tag{36}$$

and, without loss of generality, we have set $\omega = 1$ in the LW iteration. We note that in order to obtain Eq. (35) we needed the identity $A^T \mathbf{1} = \mathbf{1}$ to hold. From these equations it is evident that at each iteration LW corrects the solution x_k with a quantity proportional to r_k , while the correction provided by RL is proportional to x_k itself. Thus it is not surprising that RL outperforms LW when applied to reconstructing objects composed of star-like sources on a flat back ground, since in the early stages of both iterations the entries of x_k are large and increasing in regions corresponding to the positions of the objects in the image, while the values of r_k are correspondingly small and decreasing.

However, as has been shown by the simulations presented above, RL does not outperform LW when applied to reconstructing objects with smooth light distribution and whose spatial extension is broader than the PSF. In order to understand this phenomenon, consider the negative Jacobian matrices of the quantities (34) and (35):

$$J_{LW}(x_k) = A^T A \tag{37}$$

and

$$J_{RL}(x_k) = \text{diag} \left[A^T \frac{r_k}{Ax_k} \right] + \text{diag}[x_k] A^T \text{diag} \left[\frac{b}{Ax_k \odot Ax_k} \right] A. \tag{38}$$

These matrices provide the sensitivities of the LW and RL algorithms to changes in the components of the iterate x_k . Equations (37) and (38) allow us to make several observations. We begin by considering Eq. (37). Since in general astronomical applications A is the discretization of a PSF with an almost limited spatial support, the LW sensitivity matrix $A^T A$ will also have spatial support that is almost limited. From this

observation, we can conclude that for a given pixel the corresponding component of the vector Δ_{LW} will be most sensitive to changes in the value of the pixel itself as well as in the values of “nearby” pixels. Here, the term “nearby” is defined by the characteristics of the PSF. More specifically, as the spread of the PSF increases, so does the collection of “nearby” pixels.

Perhaps an even more important observation is that the sensitivity of the LW iteration to perturbations in x_k is independent of both x_k and b . Consequently, the algorithm has no means of distinguishing between low and high intensity regions within the object, and hence perturbations of the same magnitude are allowed for components of x_k corresponding to regions of both low and high light intensity. This explains why in areas of low light intensity (where the margin of error is very small) LW, and the least-squares approach in general, does poorly.

The sensitivity matrix (38) for the RL iteration is more difficult to analyze. However, some simplification is possible when one considers the problem of the restoration of a flat background or of regions of an image in which the intensity distribution varies smoothly (e.g., the interior of the rectangular function considered in the simulations). In this case it is possible to define a region Ω where the image can be considered constant or almost constant. Because of the semiconvergence property of RL, in such regions the components of the vector r_k converge rapidly to zero (this has been verified via numerical simulations). Thus the first term in Eq. (38) converges to zero. The same is not true for the second term. Thus, it is reasonable to expect that it will provide an accurate approximation of the sensitivity of RL within Ω .

Provided that the spread of the PSF is small relative to the size of Ω , early in RL iterations the vector x_k is approximately constant and close to b , i.e. those pixels values are reconstructed rapidly, within Ω . Hence, the vector $b/(Ax_k \odot Ax_k) \approx 1/Ax_k$ is also approximately constant within Ω . In addition, if we define D_Ω to be the diagonal matrix with components

$$[D_\Omega]_{jj} = \begin{cases} 1, & j \in \Omega \\ 0, & j \notin \Omega \end{cases}, \tag{39}$$

then

$$D_\Omega Ax_k \approx D_\Omega A D_\Omega x_k \tag{40}$$

will be accurate within the interior of Ω . To obtain Eq. (40) we used the fact that x_k is approximately constant on Ω and that the spread of A is small compared to the size of Ω . Finally, the second term of Eq. (38) can be approximated within Ω as follows:

$$D_\Omega \text{diag}[x_k] A^T \text{diag}[b/(Ax_k \odot Ax_k)] A \approx \text{diag}[x_k] D_\Omega A^T D_\Omega \text{diag}[b/(Ax_k \odot Ax_k)] A \tag{41}$$

$$\approx \text{diag}[D_\Omega(x_k/Ax_k)] A^T A \tag{42}$$

$$\approx D_\Omega A^T A. \tag{43}$$

Approximation (41) follows from Eq. (40). Approximation (42) follows from the fact that, as stated above, early in RL iterations $b/(Ax_k \odot Ax_k) \approx 1/Ax_k$ is approximately constant. Thus we see that not only does the second term in Eq. (38) not converge to zero, it is well-approximated within Ω by the LW sensitivity (37). Recalling

that the first term in Eq. (38) converges rapidly to zero in RL iterations, it is therefore not surprising that RL and LW provide similar results in the interior of the rectangular object mentioned above. We can extend this discussion to extended objects in general by noting that such objects can be decomposed into a union of regions in which the light intensity is approximately constant. Hence, RL and LW should provide similar results for extended objects in general.

4.3. RL vs. LS: Convergence properties

As shown above, LW presents an acceptable convergence rate only in case of restoration of extended objects. Unfortunately, understanding the convergence properties of the RL algorithm (21) is difficult since it is not possible to carry out an analysis similar to that done in Sect. 3.2. For this reason, we consider, again, a noise-free signal \mathbf{b} and Gaussian PSF with circular symmetry and variance σ_p^2 . In addition, we suppose that the object of interest is a circular Gaussian source with variance σ_b^2 . The amplitude of the source is not considered since RL conserves the total number of counts. Due to the connection between the RL and LW iterations discussed in the previous section, an understanding of RL convergence may provide further understanding of the convergence of the LW iteration. For simplicity, in what follows we work in the continuous, and results will be later discretized.

If a Gaussian function with variance σ^2 is denoted by $G[\sigma^2]$, then

$$\frac{G[\sigma_1^2]}{G[\sigma_2^2]} = G\left[\frac{\sigma_1^2\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right]; \tag{44}$$

$$G[\sigma_1^2] \odot G[\sigma_2^2] = G\left[\frac{\sigma_1^2\sigma_2^2}{\sigma_2^2 + \sigma_1^2}\right]; \tag{45}$$

$$G[\sigma_1^2] \otimes G[\sigma_2^2] = G[\sigma_1^2 + \sigma_2^2]. \tag{46}$$

Here, the symbol “ \otimes ” indicates convolution. From these equations it is evident that the result of any of the above operations produces a new Gaussian function. Only the first operation requires a condition be satisfied, i.e., $\sigma_2^2 > \sigma_1^2$. This will always be satisfied during the RL iteration (see Eq. (49) below).

If we define σ^2 to be the variance of the Gaussians on the right hand side of Eqs. (44)–(46), then for Eq. (44) we have $\sigma_1^2 < \sigma^2$; in Eq. (45) we have $\sigma^2 < \sigma_1^2 < \sigma_2^2$; and in Eq. (46) we have $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Consequently, only the operation (45) results in a Gaussian function with a variance that is smaller than both σ_1 and σ_2 .

If the true object \mathbf{x} is a Gaussian with variance σ_0^2 , then using Eq. (46) and the fact that $\mathbf{Ax} = \mathbf{b}$, it is

$$\sigma_b^2 = \sigma_0^2 + \sigma_p^2. \tag{47}$$

As an initial guess in the RL algorithm, we take $\mathbf{x}_0 = \mathbf{1}$.

Now, let us suppose that $\mathbf{x}_k = G[\sigma_k^2]$. Using Eqs. (44)–(46) one can obtain

$$G[\sigma_{k+1}^2] = G\left[\frac{\sigma_k^2(\sigma_k^2\sigma_b^2 + \sigma_p^4 + \sigma_k^2\sigma_p^2)}{(\sigma_k^2 + \sigma_p^2)^2}\right]. \tag{48}$$

It is a straightforward exercise to show that

$$R(\sigma_k^2) = \frac{\sigma_{k+1}^2}{\sigma_k^2} = \frac{\sigma_k^2\sigma_b^2 + \sigma_p^4 + \sigma_k^2\sigma_p^2}{(\sigma_k^2 + \sigma_p^2)^2} < 1 \tag{49}$$

provided

$$\sigma_k^2 > \sigma_b^2 - \sigma_p^2. \tag{50}$$

To prove that Eq. (50) holds for all k , we use induction. Since $\mathbf{x}_0 = \mathbf{1}$, $\mathbf{A}^T\mathbf{1} = \mathbf{1}$, and $\mathbf{A}^T = \mathbf{A}$, we have that $\mathbf{x}_1 = \mathbf{A}^T\mathbf{b} = G[\sigma_p^2 + \sigma_b^2]$. Then $\sigma_1^2 = \sigma_p^2 + \sigma_b^2$, and hence, Eq. (50) is satisfied for $k = 1$. Now, we show that if Eq. (50) holds for k , it must hold also for $k + 1$. By replacing σ_{k+1}^2 in Eq. (50) by the argument of the Gaussian function on the right hand side of Eq. (48), one can obtain an equivalent inequality involving σ_k^2 given by

$$q(\sigma_k^2) > 0, \tag{51}$$

where $q(\sigma^2)$ defined by

$$\begin{aligned} q(\sigma^2) &= \sigma^2(\sigma^2\sigma_b^2 + \sigma_p^4 + \sigma^2\sigma_p^2) + (\sigma_p^2 - \sigma_b^2)(\sigma^2 + \sigma_p^2)^2 \\ &= 2\sigma_p^2(\sigma^2)^2 + (3\sigma_p^4 - 2\sigma_p^2\sigma_b^2)\sigma^2 + (\sigma_p^6 - \sigma_b^2\sigma_p^4). \end{aligned}$$

Notice that q is a quadratic function. We can therefore find its zeros via the quadratic formula. These are given by

$$\sigma^2 = -\frac{\sigma_p^2}{2}, \sigma_b^2 - \sigma_p^2. \tag{52}$$

Since $\sigma_p^2 > 0$, we know that the graph of q is an upward opening parabola. Furthermore, by Eq. (47) we have $\sigma_b^2 - \sigma_p^2 = \sigma_0^2 > 0$, and hence, we know that if $\sigma^2 > \sigma_b^2 - \sigma_p^2$, then $q(\sigma^2) > 0$. Thus Eq. (51) follows from the inductive hypothesis, and our proof is complete.

In light of these findings, it is possible to consider some convergence properties of the RL algorithm. We begin by showing that the sequence $\{\sigma_k^2\}$ converges to $\sigma_0^2 = \sigma_b^2 - \sigma_p^2$. First, note that Eqs. (49) and (50) imply that $\{\sigma_k^2\}$ is a decreasing sequence that is bounded below by $\sigma_b^2 - \sigma_p^2$. Hence, $\{\sigma_k^2\}$ converges to some $\sigma^* \geq \sigma_b^2 - \sigma_p^2$. From inequalities (49) and (50), we have that $R(\sigma^2) = 1$. Furthermore, the arguments used in the proof of Eq. (50) imply that if $\sigma^2 > \sigma_b^2 - \sigma_p^2$ then $R(\sigma^2) < 1$. Thus it must be that $\sigma^* = \sigma_b^2 - \sigma_p^2$.

With regard to the convergence rate of the RL algorithm, Eq. (49) shows that, almost independently of the characteristics of the object, in the first iteration, when $\sigma_b \gg \sigma_p$, we have

$$R(\sigma_0^2) = \frac{\sigma_0^2(\sigma_b^2 + \sigma_p^2)}{\sigma_0^4} \approx 0. \tag{53}$$

In fact, if $\mathbf{x}_0 = \mathbf{1}$ (i.e., $\sigma_0^2 = \infty$), it is not difficult to see that $\mathbf{x}_1 = \mathbf{Ab}$, i.e., the result of the first iteration is given by $G(\sigma_p^2 + \sigma_b^2)$. At this point, there are two possible situations:

1. For extended objects, we have $\sigma_b^2 \approx \sigma_1^2 \gg \sigma_p^2$. In this case, $R(\sigma_1^2) \approx 1$. In general, this means that we can expect rapid progress in early iterations; after that the convergence rate slows down remarkably. This behavior is similar to that of the LW algorithm;

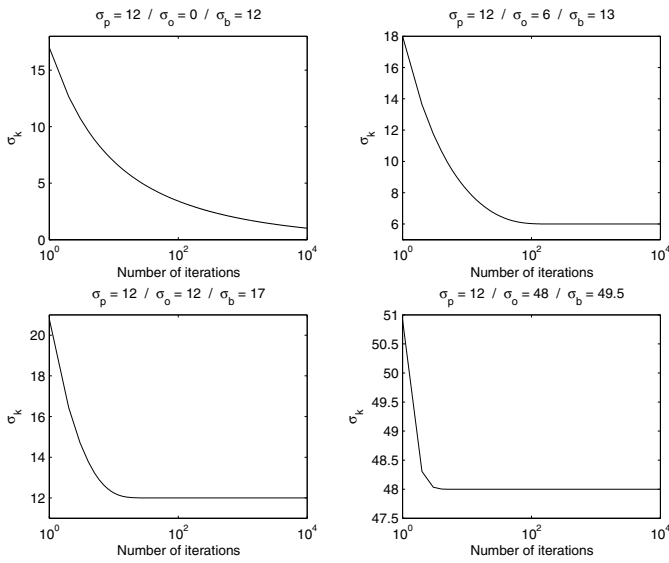


Fig. 28. σ_k vs. the number of iterations for RL in the case of a Gaussian source with various values of σ_o and a Gaussian PSF with dispersion $\sigma_p = 6$.

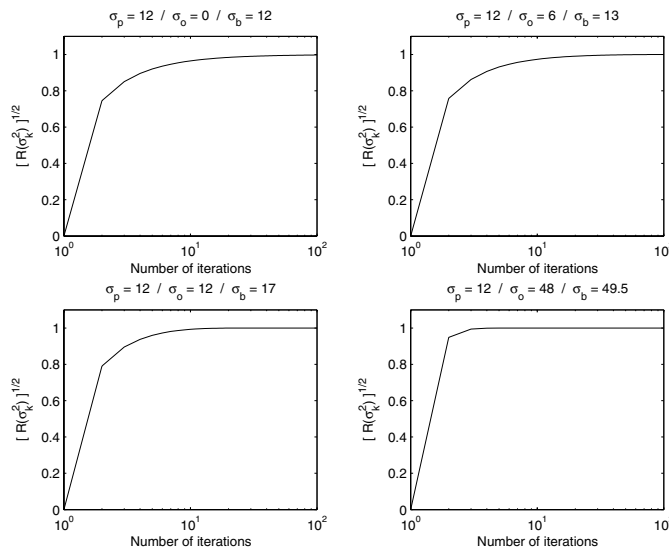


Fig. 29. $R^{1/2}(\sigma_k^2)$ vs. the number of iterations for the cases shown in Fig. 28.

2. For star-like objects, we have $\sigma_b^2 \approx \sigma_p^2$. Now, if we set $\sigma_k = \alpha \sigma_p$, then

$$R(\sigma_k^2) = \frac{1 + 2\alpha^2}{(1 + \alpha^2)^2}. \tag{54}$$

For example, when $\alpha = 1, 0.5, 1/3, 1/6$, then $R(\sigma_k^2) = 0.750, 0.960, 0.990, 0.999$, respectively. In other words, although the convergence rate of RL slows down as the iteration progresses, this effect is not as pronounced as it is for the LW algorithm.

These statements are confirmed by Figs. 28, 29.

A comparison between the RL solution for star-like sources at the k th iterate

$$\widehat{x}_k(i, j) \propto \exp(-r^2/2\widehat{\sigma}_k^2) \tag{55}$$

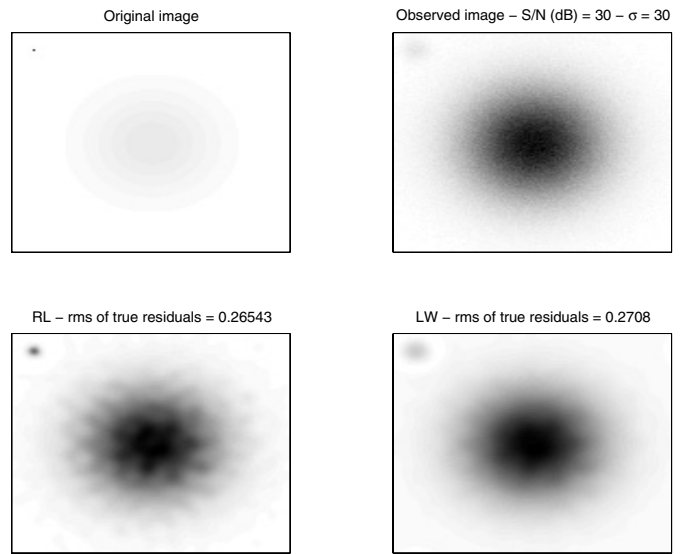


Fig. 30. Deblurring results obtained with RL and LW for a star-like object superimposed on an extended source given by a Gaussian with circular symmetry and dispersion set to 30 pixels. An additional background is present with a level set to 1% the maximum of the noise-free blurred image. The PSF is a Gaussian with circular symmetry and dispersion set to 6 pixels. The size of the image are 128×128 pixels. The noise is Poissonian with peak $S/N = 30$ dB. The image presented corresponds to the result with the smallest rms of the true residuals.

with the corresponding LW solution

$$\widehat{x}_k(i, j) = \Pi_k(i, j), \tag{56}$$

provides some additional insight into the convergence properties of these algorithms. From Eq. (55) it is evident that although the high frequencies are filtered in the RL algorithm, the filter is less stringent for high frequencies than is the Landweber filter. The consequence is that, in general, at a given k , RL has available a broader range of frequencies to restore the object. On the one hand, this can improve the convergence rate of RL compared to LW; on the other hand this could create problems when one or more star-like objects are superimposed on an extended object. A few RL iterations are sufficient to restore the extended object. The same is not true for the star-like objects. Therefore, more iterations are necessary. However, because of the amplification of the noise, this means the degradation of the results in the parts of the image not occupied by the star-like objects.

This effect is clearly visible in the experiment shown in Fig. 30.

5. Conclusions

In this paper we provide explanations for why, in spite of the incorporation of a priori information regarding the noise statistics of image formation, the RL deblurring algorithm often does not provide results that are superior to those obtained by techniques based on an LS approach. In particular, we have identified a possible explanation in the regularization approaches of the specific algorithms. The adoption of a penalty term in

the Tikhonov approach, or the need to stop the iterations before blow-up occurs in the iterative approaches, e.g. both LW and RL, do not permit the full exploitation of the information contained in the highest Fourier frequencies, i.e., those where the specific nature of the noise has the largest influence. This has two consequences: I) the performance of the LS algorithms is almost insensitive to whether the noise is Gaussian or Poissonian; II) the RL algorithm does not fully benefit from the fact that it incorporates the specific statistical model of the noise. In other words, the regularization of the solution implies a levelling out of the possible performances. In this respect, much more than a detailed knowledge of the nature of the noise is needed. Specifically, some rough a priori information regarding the solution, e.g. is it an extended or star-like object, is needed before one can know whether or not RL will provide superior results. Our numerical experiments support these conclusions. In particular, the fact that reconstructions obtained via the RL algorithm are often comparable to those of LW, i.e., an unsophisticated and very slow algorithm, indicates that resorting to advanced and often complex techniques is not always justified.

We stress that such conclusions are not only of academic interest. With respect to the ML algorithms, in general the LS algorithms are much easier to implement, are more flexible concerning the incorporation of constraints, are more amenable to a theoretical analysis of their characteristics and are computationally less costly. Consequently, unless the use of a different approach is justified, they should be considered the standard approach.

References

- Carasso, A. S. 1999, *SIAM J. Numer. Anal.*, 36, 1659
- Feller, W. 1971, *An Introduction to Probability Theory and its Applications* (Philadelphia: Wiley)
- Hansen, P. C. 1997, *Rank-Deficient and Discrete Ill-Posed Problems* (Philadelphia: SIAM)
- Hunt, B. R. 1977, *IEEE Trans. Comm.*, C-26, 219
- Johnson, M. E. 1987, *Multivariate Statistical Simulation* (New York: John Wiley & Sons)
- Lee, K. P., & Nagy, J. G., *BIT* 2004, in press
- Lucy, M. 1974, *AJ*, 79, 745
- Neelamani, R., Hyeokho, C., & Baraniuk, R. 2004, *IEEE Trans. Signal Proc.*, 52, 418
- Richardson, W. 1972, *J. Opt. Soc. Am.*, 62, 55
- Shepp, L., & Vardi, Y. 1984, *IEEE Trans. Med. Imag.*, 2, 113
- Snyder, D. L., Hammoud, A. M., & White, R. L. 1993, *JOSA A*, 10, 1014
- Snyder, D. L., Helstrom, C. W., Lanterman Faisal, M., & White, R. L. 1995, *JOSA A*, 12, 272
- Starck, J. L., Pantin, E., & Murthag, F. 2002, *PASP*, 114, 1051
- Vio, R., Nagy, J., Tenorio, L., et al. 2003a, *A&A*, 401, 389
- Vogel, C. R. 2002, *Computational Methods for Inverse Problems* (Philadelphia: SIAM)
- White, R. L. 1991, in *Proc. 1991 Conf. Inf. Sciences and Systems*, 665, The Johns Hopkins University, Baltimore
- Wu, C. F. J. 1983, *The Anals of Statistics*, 11, 95