ISRA and RL algorithms used for deconvolution of AO and HST images.

H. Lantéri, R. Soummer and C. Aime
UMR 6525 "Astrophysique" - Université de Nice - Sophia Antipolis
Faculté des Sciences - Parc Valrose - 06108 Nice Cedex 2

I – Introduction

The image restoration problem consists in the reconstruction of the best estimate of an object « *x* » from the knowledge of a blurred image « *y* » contaminated by noise. In the general case, the transformation suffered by « *x* » is described by a Fredholm integral equation of the first kind

$$\tilde{y}(r_1, r_2) = \iint h(r_1, r_2; \alpha, \beta)\ x(\alpha, \beta)\ d\alpha\ d\beta \qquad (1)$$

The function $h(r_1, r_2; \alpha, \beta)$ is the kernel of the integral equation, and $\tilde{y}$ is the blurred image without noise. In the following, we analyze two different aspects of the noise process. In the deconvolution problem, the kernel *h* is space invariant and is known as the Point Spread Function (PSF). Relation (1) then becomes:

$$\tilde{y}(r_1, r_2) = \iint h(r_1 - \alpha, r_2 - \beta)\ x(\alpha, \beta)\ d\alpha\ d\beta \qquad (2)$$

The deconvolution problem is then to obtain « *x* » knowing a noise contaminated version « *y* » of $\tilde{y}$. This is an ill-posed problem in the sense of Hadamar, and the difficulties specific to such problems have been extensively described in the literature [1 - 4]. A prior knowledge of properties of the admissible solutions is required to obtain stable and physically suitable results; for the image deconvolution problem, classical constraints are positivity and smoothness of the solution.

A smoothness constraint requires the choice of the smoothing operator and the strength of the constraint, that is a compromise between the consistency with the data and the amount of details wanted in the reconstructed image.

In this Communication, we use the matrix notation corresponding to the discrete problem; relation (2) is then written in the form of the linear system $\tilde{y} = Hx$, but we can only try to obtain *x* using y instead of $\tilde{y}$.

II –ISRA and RL Algorithms.

We present in this Section two algorithms giving solutions corresponding to the maximum of likelihood for additive Gaussian noise and for Poisson noise [5, 6]. The constraint for positive solutions in the optimization process allows us to obtain the Image Space Reconstruction Algorithm (ISRA) and the Richardson – Lucy (RL) algorithm. Moreover, we show how these algorithms can be analyzed as gradient algorithms.

Concerning the Astrophysical applications, ISRA may be more suitable for infrared detection and solar observations, while the RL algorithm is more convenient for visible wavelengths at low light level.

1- Gaussian noise: ISRA.

In this case, we consider that the data are contaminated by an additive, zero-mean, white Gaussian noise « *n* », with variance σ_i, and that the data « y_i » are independent random variables. The image formation model may be written as $y = Hx + n$. For a given pixel « *i* », the data « y_i »is the sum of a deterministic part $(Hx)_i$, which is the mean of the intensity in the considered pixel, and a random noise component 'n_i'. The likelihood, that is the conditional probability to have data '*y*' knowing the mean '*Hx*' is given by:

$$L(x) = P(y/x) \approx \prod_i \exp[-(y_i - (Hx)_i)^2 / 2\sigma_i^2] \qquad (3)$$

The corresponding log-likelihood is then:

$$Log[P(y/x)] \approx -\frac{1}{2}\sum_i (y_i - (Hx)_i)^2 / \sigma_i^2 \qquad (4)$$

and

$$J(x) = -Log[P(y/x)] = \frac{1}{2}\sum_i [y_i - (Hx)_i]^2 / \sigma_i^2 \qquad (5)$$

The solution '*x*' is obtained by minimization of the weighted Euclidean distance between the raw data and the restored mean. The weighting factor is the variance of the noise.

The minimization of *J(x)* can be done by any iterative descent method; the simplest is the constant stepsize gradient method, which can be written in the form :

$$x^{k+1} = x^k + \alpha\,[-\nabla J(x^k)] \qquad (6)$$

with

$$\nabla J(x^k) = H^T R H\, x^k - H^T R y \qquad (7)$$

where the descent stepsize α is such that $0 < \alpha < 2/\left\|H^T RH\right\|$ to ensure the convergence of the algorithm, and with $R = diag(1/\sigma_i^2)$.

The solution given by this algorithm when $k \to \infty$ is the minimum norm least squares (MNLS) solution, which may be formally expressed as $\nabla J(x^*) = 0$, that is $x^* = (H^T RH)^{-1} H^T Ry$.

However, the deconvolution problem is an ill-posed problem, and the solution given by the previous algorithm exhibit numerical instabilities. The noise affecting the data is amplified during the restoration process, and the solution may have negative parts without physical meaning.

Several methods have been proposed to ensure a positive solution, leading to ISRA. This algorithm was initially proposed by Daube-Witherspoon and Muehllehner [7], developed par De Pierro [8 - 10] and Titterington [11].

ISRA may be written as:

$$x_i^{k+1} = x_i^k . \frac{(H^T Ry)_i}{(H^T RHx^k)_i} \qquad \forall i \qquad (8)$$

It can straightforwardly be rewritten in the following additive form:

$$x^{k+1} = x^k + diag[\frac{x_i^k}{(H^T RHx^k)_i}] H^T R(y - Hx^k) , \qquad (9)$$

which is clearly a gradient form with stepsize:

$$\alpha_i = \frac{x_i^k}{(H^T RHx^k)_i} \qquad (10)$$

We emphasize that the stepsize depends on the considered pixel and on the value of the solution at the current step 'k'.

In our problem, the matrix H and R have positive entries, all the components of the initial estimate have positive values to satisfy the constraint. We have then $\alpha_i \geq 0 \;\; \forall i$, and the descent properties of the algorithm are verified; the convergence of this algorithm was shown in references [8] and [9].

2- Poisson noise: RL algorithm.

For a perfect photon detection, the intensity in the pixel 'i', is a random variable that follows a Poisson law of mean *(Hx)ᵢ* .The likelihood may be written as:

$$L(x) = P(y/x) = \prod_i \frac{[(Hx)_i]^{y_i}}{y_i!} \exp[-(Hx)_i] \tag{11}$$

Using Stirling's formula, the negative Log-likelihood becomes:

$$T(x) = -Log[L(x)] \approx \sum_i [(Hx)_i - y_i] + y_i Log \frac{y_i}{(Hx)_i} \tag{12}$$

Such function is known as the Csiszär [12] I-divergence between (Hx) and y. It is the quantity we seek to minimize, keeping the constraint that '*x*' must be positive. Dropping in *T(x)* the terms independent of '*x*', we must minimize under positivity constraint, the quantity:

$$D(x) = \sum_i (Hx)_i - y_i Log(Hx)_i \tag{13}$$

This problem is generally treated by means of the Expectation-Maximization (EM) technique [13] or using the Kuhn-Tucker optimality conditions for the constrained minimization problem [14]. In both cases, the same algorithm is obtained in the form:

$$x_i^{k+1} = \frac{1}{a_i} x_i^k [H^T \frac{y}{Hx^k}]_i \, , \tag{14}$$

where the ratio $\frac{y}{Hx^k}$ is a vector computed componentwise, and with $a_i = \sum_j H_{j,i}$.

This algorithm was proposed in astronomy by Richardson [15] and Lucy [16]. As in the case of ISRA, all successive estimates remain positive if the initial estimate is positive. Moreover, if a component of the solution becomes equal to zero, it remains equal to zero for all successive iterations. Using matrix operations, we can write the algorithm in the following additive form:

$$x^{k+1} = x^k + diag[\frac{x_i^k}{a_i}] \, H^T diag[\frac{1}{(Hx^k)_i}](y - Hx^k) \tag{15}$$

Taking into account the expression of the gradient:

$$\nabla D(x) = H^T diag[\frac{1}{(Hx^k)_i}](Hx^k - y) \, , \tag{16}$$

we recognize a gradient algorithm of the form:

$$x^{k+1} = x^k + diag[\frac{x_i^k}{a_i}] [-\nabla D(x^k)] \tag{17}$$

The stepsize $\alpha_i = \frac{x_i^k}{a_i}$ depends on the pixel considered, but remains always positive or zero; the descent property of the algorithm is kept. In our particular problem, a supplementary simplification is due to the fact that $a_i = 1 \quad \forall i$.

3. Notes.

In the two algorithms written in an additive form, we emphasize the analogous roles played by the weighting matrix $R = diag \frac{1}{\sigma_i^2}$ in relation (9) and by the matrix $diag \frac{1}{(Hx^k)_i}$ in relation (15). This analogy is fully justified since σ_i^2 is the variance of the intensity in a pixel for the Gaussian process, while $(Hx^k)_i$ is the variance in a pixel for a Poisson process (the variance is equal to the mean).

In the Gaussian case, *J (x)* is a quadratic form, the convexity of this objective function allows to obtain the solution by solving the linear system $\nabla J(x) = 0$. The Gradient or Conjugate Gradient methods in their classical form leading to linear algorithms may be used in this case and the stepsize is known explicitly. We must observe that the non-linearity of the algorithm ISRA is produced by the positivity constraint.

In the Poisson case, the functional D(x), although convex, is not quadratic. Thus, the non-linearity of the RL algorithm used in this case is due to the particular form of *D (x)* as well as to the positivity constraint.

Note that in the RL algorithm we have for all *k*: $\sum_i x_i^k = \sum_i x_i^0$, which is not the case for ISRA.

The two algorithms are not regularized. A positivity constraint, although necessary, does not produce smooth solutions. When the iteration number is too high, the classical phenomena of instability appear in the solutions. On the practical point of view, the iteration number must be limited to obtain an acceptable compromise between resolution and stability of the solutions. Alternatively, an explicit regularization may be done by introducing a supplementary constraint to attenuate the high frequency components in the solution.

On the computational point of view, we used the continuous function forms of the algorithms where the convolutions are easily obtained via FFTs. The algorithms are thus written as follows (where $\otimes$ is the convolution symbol):

$$(6) \Leftrightarrow x^{k+1}(r_1,r_2) = x^k(r_1,r_2) + \alpha * H(-r_1,-r_2) \otimes [y(r_1,r_2) - H(r_1,r_2) \otimes x^k(r_1,r_2)] \qquad (18)$$

$$(9) \Leftrightarrow x^{k+1}(r_1,r_2) = x^k(r_1,r_2) + x^k(r_1,r_2) * [\frac{H(-r_1,-r_2) \otimes y(r_1,r_2)}{H(-r_1,-r_2) \otimes H(r_1,r_2) \otimes x^k(r_1,r_2)} - 1] \qquad (19)$$

$$(15) \Leftrightarrow x^{k+1}(r_1,r_2) = x^k(r_1,r_2) + x^k(r_1,r_2) * [H(-r_1,-r_2) \otimes \left\{ \frac{y(r_1,r_2)}{H(r_1,r_2) \otimes x^k(r_1,r_2)} - 1 \right\}] \qquad (20)$$

III- Determination of the level of regularization by comparison with the Wiener filter.

1- Principle.

As shown in the previous section, the reconstruction of the object must be made using some kind of regularization, in order to prevent an unwanted amplification of the noise. This is generally done by limiting the number k of iterations and/or by adding explicitly a distance to a smooth solution, to the distance describing the consistency of the results with the data.

The determination of this regularization is probably the main problem encountered in the reduction of image blurring. It is generally performed in an empirical way.

We propose here a method to determine the strength of regularization on physical considerations. The idea is to use a comparison of the results of these algorithms (RL or ISRA) with the optimum Wiener filter that has the advantage of being defined from considerations on SNR.

The comparison may be performed in the Fourier plane by comparing the amplification of the spatial frequency power produced by RL or ISRA, to that produced by the Wiener optimum filter. Denoting *u* the two dimensional spatial frequency, the Fourier transforms of the observed image and of the object will be denoted Y(u) and X(u) respectively. Let $X^{(k)}(u)$, the Fourier transform of the reconstructed object at the k^{th} iteration for ISRA or RL, and X_w (u), the Fourier transform of the reconstructed object for the Wiener filter. For the sake of simplicity, we just consider the iteration number k, as the regularization parameter, but a full analysis would include the type of regularization function and the rate of regularization γ. Consequently, we propose to compare the ratio $Abs\left[X^{(k)}(u)/Y(u)\right]$ obtained in the iterative deconvolution technique, to $Abs\left[X_w(u)/Y(u)\right]$.

For the Wiener filter, this later quantity is given by:

$$\frac{X_w(u)}{Y(u)} = H^{-1}(u)\,\Phi(u)\,, \tag{21}$$

where $H^{-1}(u)$ is the raw inverse filter and $\Phi(u)$ is a zero phase filter defined by:

$$\Phi(u) = \frac{P_{Hx}(u)}{P_{Hx}(u) + P_N(u)} \quad , \tag{22}$$

and where $P_{Hx}(u)$ and $P_N(u)$ are the power spectra of the noiseless image and of the noise, respectively.

The resulting image is recovered by an inverse Fourier transform of the quantity $Y(u)\,H^{-1}(u)\,\Phi(u)$. The Wiener filter preserves the high SNR frequency components, while attenuating the low SNR frequency components.

Aside the difficulties of estimating or modeling acceptable values of $P_{Hx}(u)$ and $P_N(u)$, the term $\Phi(u)$ generally appear as a low-pass filter. The overall effect is that the amplitudes of $Y(u)$ are modified by a function $Abs\left[H^{-1}(u)\,\Phi(u)\right]$ that mainly amplifies the medium range of frequencies in the image; the effect on the phase is that of the inverse filter.

The drawback of the Wiener approach is that there is no constraint on the positivity of the recovered image x_w.

The algorithms RL and ISRA are not linear, therefore we cannot write a formal linear relation between the Fourier transform of the observed image *Y (u)* and of the reconstructed object $X^k(u)$. However, for each spatial frequency *u*, we can compute numerically the ratio $X^{(k)}(u)/Y(u)$ that we consider as an equivalent inverse filter.

Since the Wiener filter gives a good estimate of the SNR in the *u* plane, we expect to obtain satisfactory reconstruction x^k when the two estimated ratios best matches.

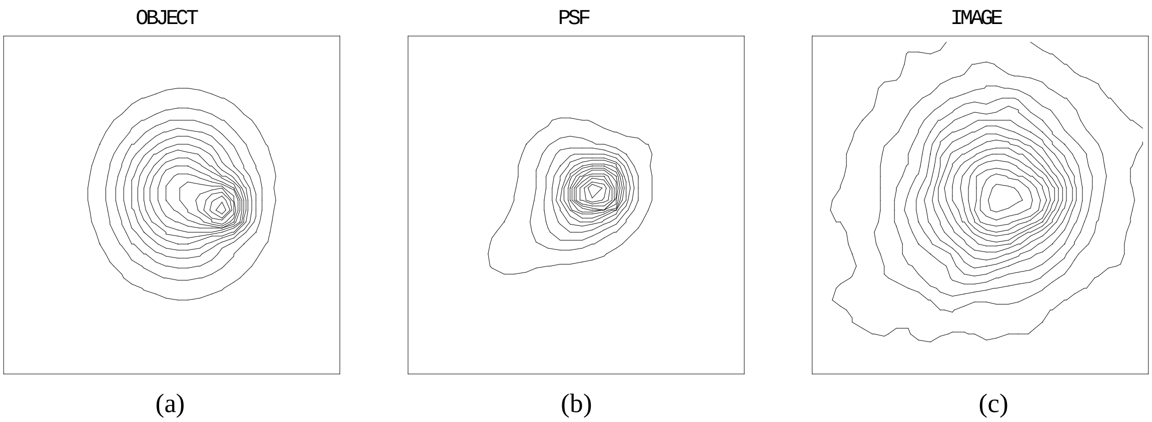


Figure 1. Simulated Objec:t (a), PSF: (b) and noisy image: (c). The simulation is made assuming 30000 photons in the image. The object was chosen so that after convolution with the experimental PSF of HST, the result resembles to the observation of the star Betelgeuse [17].

2- Numerical simulations

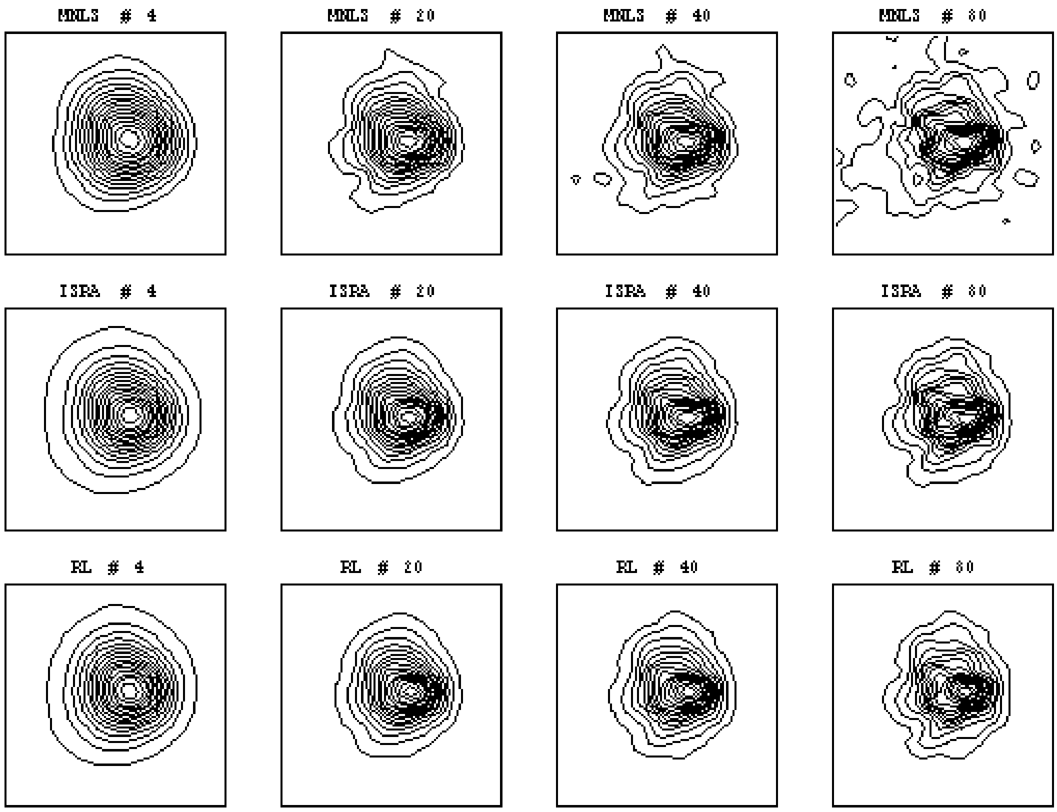
We have performed a numerical simulation that roughly represents the conditions of observation of Betelgeuse with the Faint Object Camera of the Hubble Space Telescope [17]. The object x, PSF and observed image y are represented in figure 1. The observed image is corrupted by photon noise.

We present in figure 2, results of deconvolutions using iterative methods for various iterations number k. As k increases, the image becomes sharper, but the noise increases too. This is more visible for the unconstrained minimization of the Euclidean distance (MNLS) because negative parts appear in the result. We get rid of this problem with RL and ISRA algorithms, but we still have the presence of noise that produces artifacts in the reconstructed image. This problem of amplification of the noise may be partly solved by using regularization procedures. The results are shown in figure 3.

It is interesting to represent the effects of these algorithms in the frequency plane. When k increases, the MNLS method reconstructs exactly the inverse filter. RL and ISRA are non-linear algorithms but as previously mentioned, we can still express an equivalent inverse filter.

In the low frequency range, both iterative algorithms gives the inverse filter; their behavior is different for the high frequencies where their transmission factor shows a speckled form, with peaks larger than the raw inverse filter (figure 4).

We have performed comparisons between the equivalent filter and the Wiener filter in figures 5 and 6. The non-linearity also produces energy outside the cut-off frequency of the original image and the phase differs from that of H^{-1} as shown in figure 7.



<u>Figure 2</u>. Result of the deconvolution of the noisy image of Figure 1(c) using the PSF shown in Figure 1(b), as a function of the iteration number k. From top to bottom the raw Euclidean distance, ISRA and RL. By comparison with the object (Figure 1 a), one can see that artifacts appear for k greater than 40.

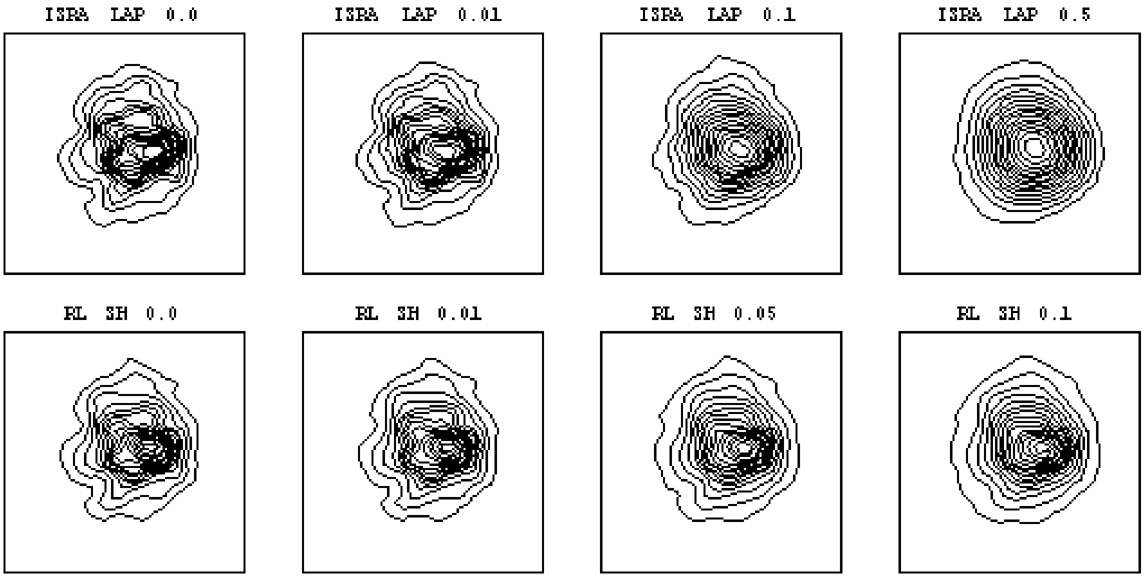


Figure 3. Result of the deconvolution for an iteration number k=100 with regularized algorithms. ISRA is regularized by a Laplacian operator with a regularization coefficient varying from left to right from 0 (no regularization) to 0.5. The RL algorithm is regularized by the Shannon entropy function as smoothness constraint with coefficients 0 to 0.1.

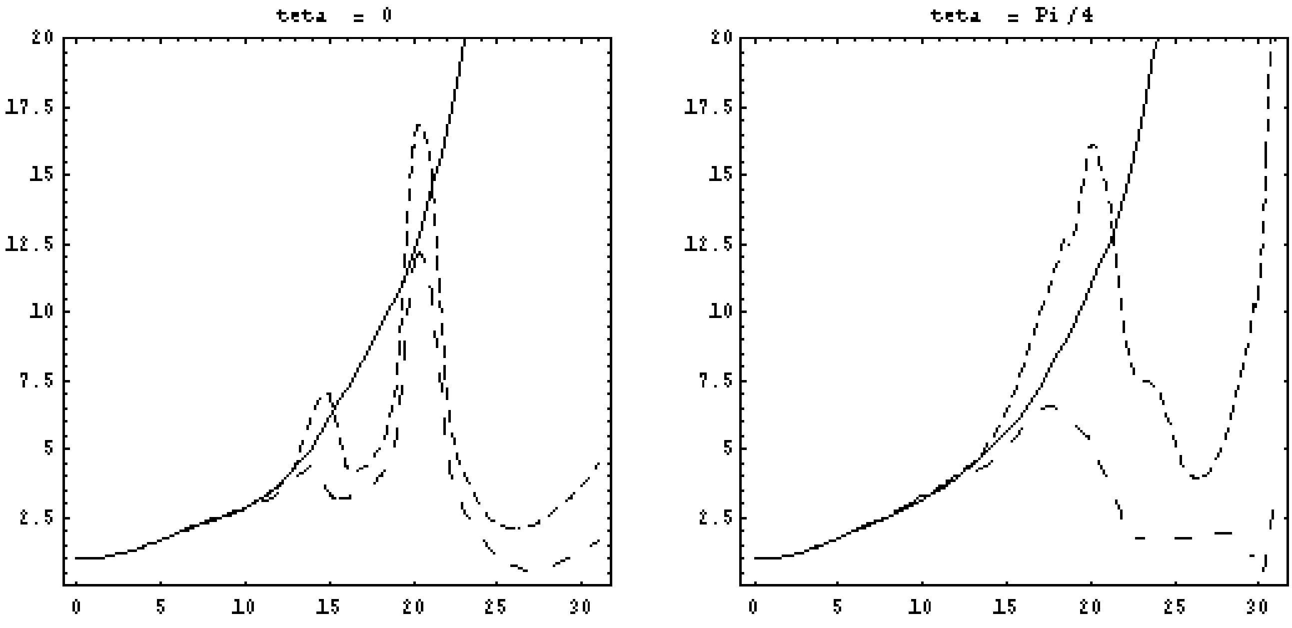


Figure 4. Examples of $Abs\left[X^{(k)}(u)/Y(u)\right]$ *for the Inverse Filter (continuous line), ISRA (long dashed line) and RL (short dashed line), for two angular direction of the Fourier plane. The amplitude of correction of ISRA and RL may be higher than that of the inverse filter. This is partly because these algorithms are non-linear and may create unexpected power at some spatial frequencies.*

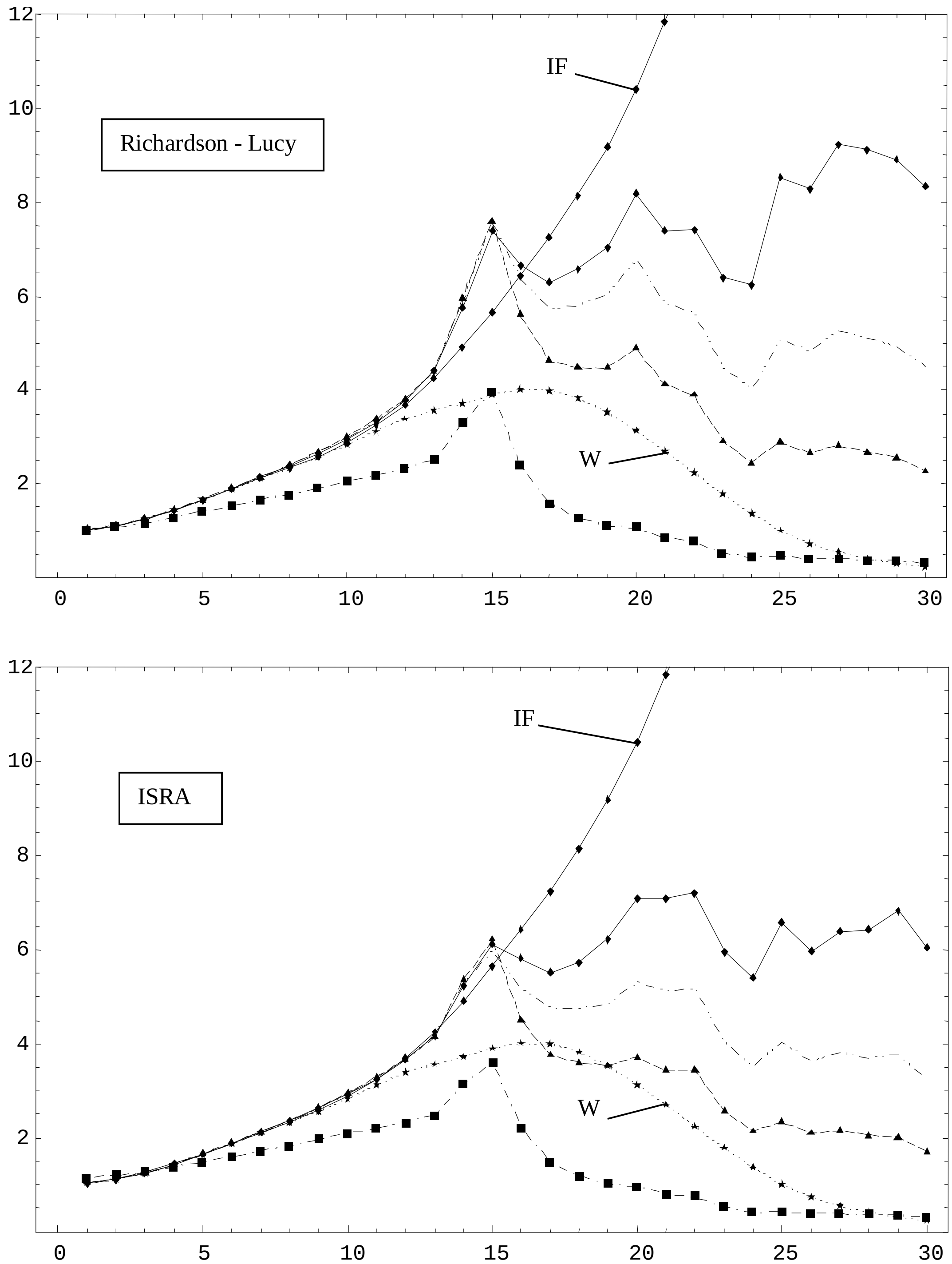


Figure 5. *Averaged values of* $Abs\left[X^{(k)}(u)/Y(u)\right]$ *for constant value of* $|u|$. *Top: RL, bottom, ISRA, for various values of the iteration number (lower to higher curves: k = 4, 20, 40 and 80) compared to the Inverse Filter (IF) and the Wiener Filter (W). For low frequencies, all curves (with sufficient value of k) fit the Inverse Filter, ISRA being closer than RL over a wider region. In the high frequencies, the best agreement will probably be obtained for k equal 20 or slightly lower. This agrees with the visual inspection of Figure 2.*

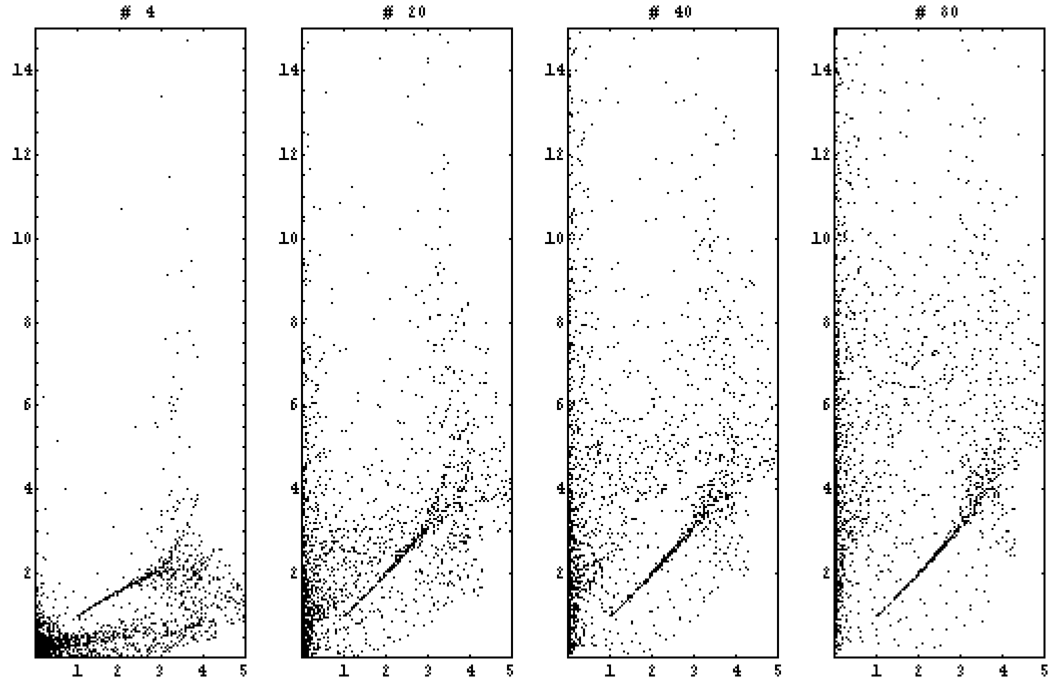


Figure 6. Alternate representation for the comparison between RL and Wiener. As indicated in the body of the Communication, we have represented $Abs\left[X^{(k)}(u)/Y(u)\right]$ *as function of* $Abs\left[X_w(u)/Y(u)\right]$*. Each spatial frequency u gives one point in this figure. The iteration number takes the values 4, 20, 40, 80, from left to right. The main diagonal corresponds mainly to the low frequencies. Large fluctuations in the module ratios are visible here while not observable in figure 5. We may consider that the best agreement is obtained for k ~ 20. Similar results are obtained with ISRA.*

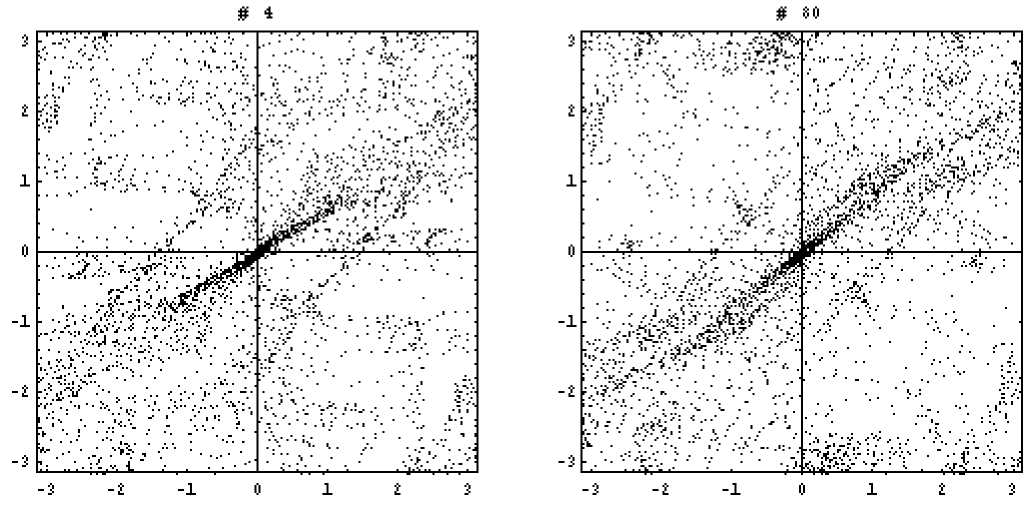


Figure 7. Diagram giving the modification of the phase in the reconstruction process. We have represented $Arg\left[X^{(k)}(u)\right]$ *as function of* $Arg\left[Y(u)\right]$*. As in the figure above, each spatial frequency u gives one point in the figure. A simple Wiener filter would give the first bisector in this representation.*

IV – Conclusion

We have analyzed the algorithm of Richardson-Lucy (RL) and the Image Space Reconstruction Algorithm (ISRA) for image deconvolution. These iterative algorithms keep the reconstructed image positive while maximizing the likelihood for a Poisson Process (RL) and a Gaussian process (ISRA).

We have shown that these algorithms can be written as gradient algorithms for which the stepsize depends on the considered pixel and on the value of the solution at the current step ‘k’. This might be used to improve the speed of convergence of these algorithms, but was not implemented here.

To prevent an amplification of the noise, these algorithms need to be regularized, either by limiting the iteration number or by introducing a penalty term (Tikhonov or entropy).

From numerical simulations, we have analyzed the effects of these algorithms in the Fourier plane. Their non-linearity clearly appears in amplitude and phase modifications of the reconstructed images, in particular by adding frequency components outside the diffraction limit.

In order to optimally adjust the parameters of regularization (iteration number, regularization coefficient) of the deconvolution, we have proposed to use a comparison of RL and ISRA with the Wiener filter. This comparison is made in the Fourier plane for amplification of the modulus of spatial frequencies components. It was found experimentally that the best-reconstructed images were obtained when the correction factors of RL and ISRA were comparable to those proposed by the Wiener filter.

This comparison was made for a given simulated image and a given level of noise. The parameters of the present simulation were defined to fit the observation of the star Betelgeuse by HST [17]. At that level of noise, no strong difference was found between the results of ISRA and RL although the simulated noise was of Poisson origin.

A more general analysis is now required to check the robustness of our approach, involving various objects and PSFs, with different kinds and levels of noises.

References

[1] - Bertero M., De Mol C. and Viano G.A. (1980) – « The stability of inverse problems » in Inverse scattering problems in optics – H.P.Baltes Ed. – Springer Verlag series Topics in current physics.

[2] - Bertero M. (1989) – Linear inverses and ill-posed problems – in Advances in Electronics and Electron Physics V.75, 1 – 121

[3] - Bertero M., Boccacci P. and Maggio F. (1995) – Regularization methods in image restoration: an application to HST images Int. J. of Imaging Systems and Technology , 6. 376-386.

[4] - Demoment G. (1986) – Deconvolution des signaux – Cours de l’Ecole Supèrieure d’Electricité

[5] - Lane R.G. (1996) - Methods for maximum – likelihood deconvolution. J.Opt.Soc.Am.A 13. N°10, 1992 – 1998.

[6] - Zacchéo T.S. and Gonsalves R.A. (1996) – Iterative maximum – likelihood estimators for positively constrained objects . J.Opt.Soc.Am.A, 13, N°2, 236 – 242.

[7] - Daube – Witherspoon M.E. and Muehllehner G. (1986) – An Iterative Image Space Reconstruction Algorithm suitable for volume ECT, I.E.E.E. Trans. Med. Imaging, MI - 5, N°2, 61 – 66.

[8] - De Pierro A.R. (1987) –On the convergence of the Iterative Image Space Reconstruction Algorithm for volume ECT. , I.E.E.E. Trans. Med. Imaging, MI- 6, N°2, 124 –125.

[9] - De Pierro A.R. (1990) – Non linear relaxation methods for solving symmetric linear complementarity problems. J. Optimization Theory and Appl., 64, N°1, 87 – 99.

[10] - De Pierro A.R. (1993) – On the relation between the ISRA and the EM algorithm for Positron Emission Tomography , I.E.E.E. Trans. Med. Imaging, MI-12, N°2, 328 – 333.

[11] - Titterington D.M. (1987) – On the Iterative Image Space Reconstruction Algorithm for ECT., I.E.E.E. Trans. Med. Imaging , MI- 6, N°1, 52 – 56.

[12] - Csiszär I. (1991) – Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, The annals of statistics, 19, N°4, 2032-2066

[13] - Shepp L.A. and Vardi Y. (1982) – Maximum likelihood reconstruction for emission tomography., I.E.E.E. Trans. Med. Imaging, MI - 6, N°2,113 – 122.

[14] - Byrne C.L. (1993) – Iterative image reconstruction algorithms based on cross - entropy minimization. I.E.E.E. Trans. Image Processing, 2, N°1, 96 – 103.

[15] - Richardson W.H. (1972) – Bayesian based iterative method of image restoration. J.Opt.Soc.Am. 62. N°1, 55 – 59

[16] - Lucy L.B. (1974) – An iterative technique for the rectification of observed distributions. Astron. J., 79, 745 – 754.

[17] - Gilliland R.L. and Dupree A.K. (1996) – First image of the surface of a star with the Hubble Space Telescope. Astrophys. J. 463, L29-L32.