Optimization issues in blind deconvolution algorithms

Eric Thiébaut

CRAL / Observatoire de Lyon, Saint Genis Laval, France

ABSTRACT

Modern blind deconvolution algorithms combine agreement with the data and regularization constraints into a single criteria (a so-called penalizing function) that must be minimized in a restricted parameter space (at least to insure positivity). Numerically speaking, blind deconvolution is a constrained optimization problem which must be solved by iterative algorithms owning to the very large number of parameters that must be estimated. Additional strong difficulties arise because blind deconvolution is intrinsically ambiguous and highly non-quadratic. This prevent the problem to be quickly solved. Various optimizations are proposed to considerably speed up blind deconvolution. These improvements allow the application of blind deconvolution to very large images that are now routinely provided by telescope facilities. First, it is possible to explicitly cancel the *normalization ambiguity* and therefore improve the condition number of the problem. Second, positivity can be enforced by gradient projection techniques without the need of a non-linear re-parameterization. Finally, superior convergence rates can be obtained by using a small sub-space of ad-hoc search directions derived from the effective behavior of the penalizing function.

Keywords: inverse problem, constrained optimization, blind deconvolution.

1. INTRODUCTION

1.1. Blind Deconvolution

Within this isoplanatic patch of the instrument (+ transmitting media), an observed image $g(\mathbf{x})$ follow the image formation equation:

$$g(\mathbf{x}) = (f \star h)(\mathbf{x}) + n(\mathbf{x}) = \int f(\mathbf{y})\, h(\mathbf{x}-\mathbf{y})\, \mathrm{d}\mathbf{y} + n(\mathbf{x}) \tag{1}$$

where $f(\mathbf{x})$ is the brightness distribution of the source, $h(\mathbf{x})$ is the point spread function (PSF) and $n(\mathbf{x})$ is the noise contribution. The discrete version of Eq. (1) is suitable for sampled data:

$$\mathbf{g} = \mathbf{m} + \mathbf{n} \tag{2}$$

where the model $\mathbf{m}$ of the data $\mathbf{g}$ is:

$$\mathbf{m} \equiv \mathbf{f} \star \mathbf{h} = \mathbf{F}^{-1} \cdot (\hat{\mathbf{f}} \times \hat{\mathbf{h}}) \tag{3}$$

where $\star$ and $\times$ denote discrete convolution and elementwise multiplication respectively, $\hat{\mathbf{f}} \equiv \mathbf{F} \cdot \mathbf{f}$ is the discrete Fourier transform of $\mathbf{f}$, and the inverse Fourier transform operator is $\mathbf{F}^{-1} = \frac{1}{N_{\mathrm{pixel}}}\mathbf{F}^{\mathrm{H}}$ (in 1D: $F_{u,x} = \mathrm{e}^{-\mathrm{i}2\pi ux/N}$).

Under certain restrictions (constraints), it is possible to recover some approximation of both $\mathbf{f}$ and $\mathbf{h}$ from the data $\mathbf{g}$ alone. Such a process is an inverse problem known as *blind deconvolution* (see Fig. 1).

The obvious advantage of blind deconvolution with respect to deconvolution is that it requires no knowlegde of the PSF. This is very important in situations where calibration of the effective PSF is not possible, or very difficult, or time consuming, as it is the case, e.g., in high angular resolution astronomy or medical imaging. The drawbacks of the blind deconvolution is that it is far more difficult and, in practice, it takes much more time to be solved than conventional deconvolution.

Further author information:
E. Thiébaut: E-mail: thiebaut@obs.univ-lyon1.fr, Telephone: 33-(0)4 78 86 85 48, Address: CRAL / Observatoire de Lyon, 9 avenue Charles André, F-69561 Saint Genis Laval Cedex, France

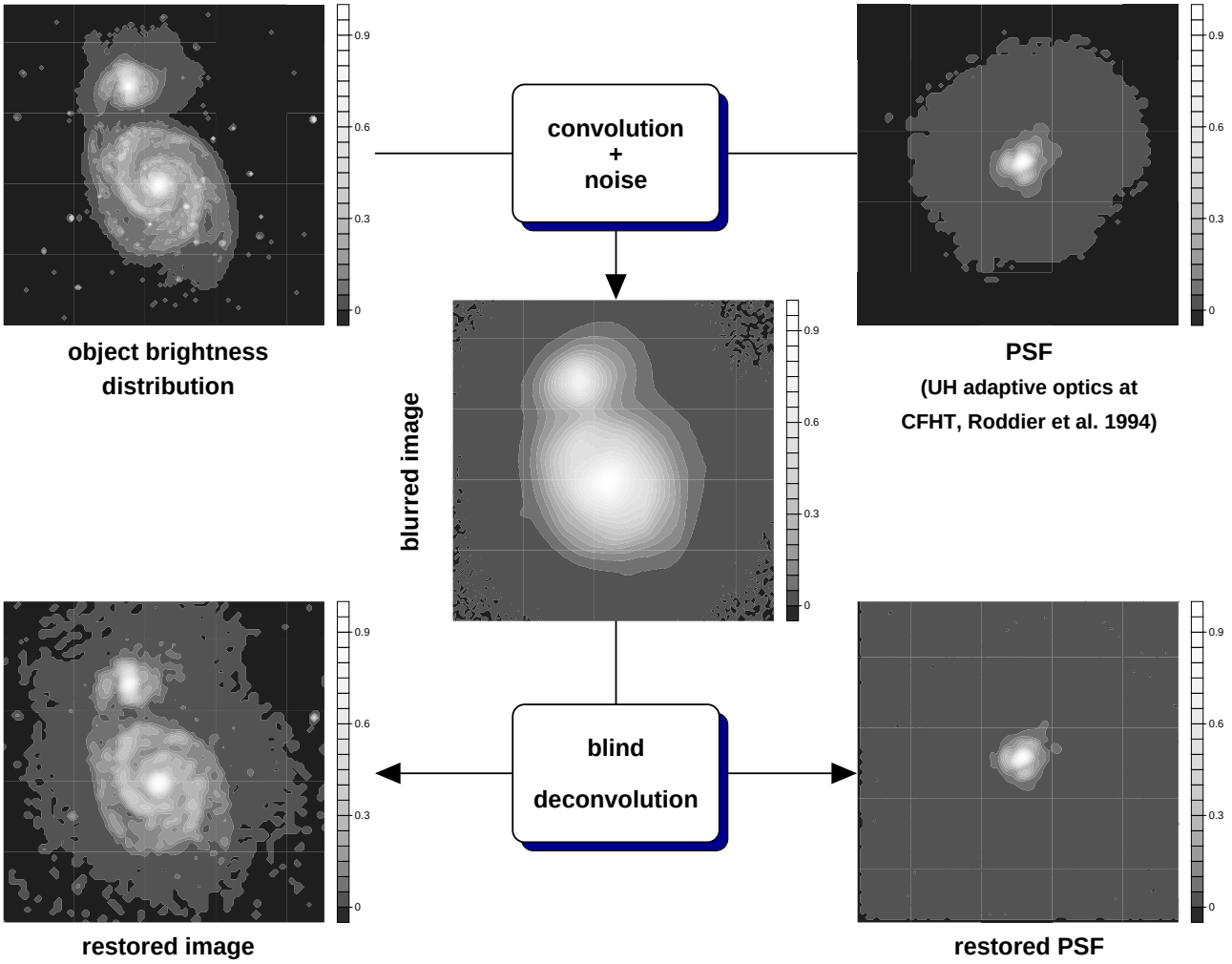

Figure 1. Top and center images: simulation of an astronomical image obtained with a partially compensated adaptive optics system. Bottom images: result of the blind deconvolution algorithm applied to the simulated image.

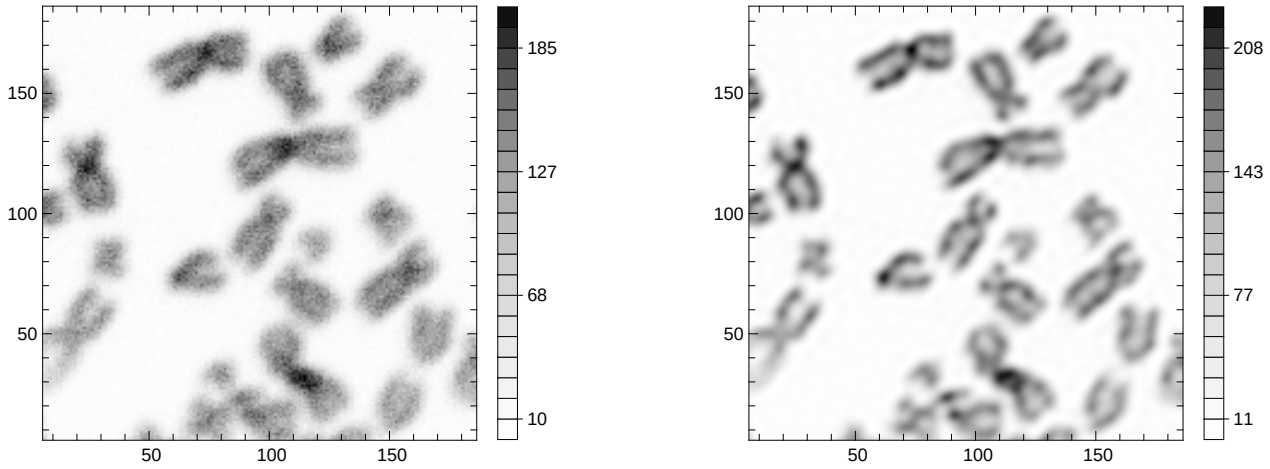

Figure 2. Microscopic image of chromosomes used in the paper to compare the performances ot the various optimization methods. left: original image. Right: blindly deconvolved image.

1.2. Blind Deconvolution Stated as a Constrained Optimization Problem

The blind deconvolution process is by nature ill-posed (there are more unknowns than measurements) and ill-conditioned (a small variation of the data, e.g. due to noise, could produce arbitrarily large changes in the solution). Additional constraints are therefore required to regularize the inverse problem and find an acceptable and unique solution given the data. Some intrinsic ambiguities of the blind deconvolution can be easily removed thanks to additional constraints or regularization. This is the case for, e.g., the *component ambiguities* which holds because $\mathbf{f}$ and $\mathbf{h}$ can be exchanged. Other ambiguities may be more difficult to solve. For instance, one of the component, $\mathbf{f}$ and/or $\mathbf{h}$, may be itself the convolution of several probability distribution functions[2] (PDF's). Finally, $(\alpha\mathbf{f}, \mathbf{h}/\alpha)$ gives the same model image as $(\mathbf{f}, \mathbf{h})$ for any $\alpha > 0$, such a *normalization ambiguity* may be responsible of numerical problems.

As it is frequently the case when solving ill-conditioned inverse problems, the solution of the blind deconvolution process can be obtained from a *constrained optimization*:

$$(\mathbf{f}, \mathbf{h})_{\text{bdec}} = \arg \min_{\substack{\mathbf{f}\in\Omega_{\text{img}} \\ \mathbf{h}\in\Omega_{\text{psf}}}} \phi(\mathbf{f}, \mathbf{h}|\mathbf{g}) \tag{4}$$

where the domains Ω_{img} and Ω_{psf} represent strict constraints such as positivity ($f_x \geq 0$ and $h_x \geq 0$, $\forall x$), or support. The penalizing function to optimize has the general form:

$$\phi(\mathbf{f}, \mathbf{h}|\mathbf{g}) = \varphi(\mathbf{f}, \mathbf{h}|\mathbf{g}) + \xi_{\text{img}}(\mathbf{f}) + \xi_{\text{psf}}(\mathbf{h}) \tag{5}$$

where the so-called *likelihood term* $\varphi(\mathbf{f}, \mathbf{h}|\mathbf{g})$ enforces agreement of the model with the data, whereas the regularization terms $\xi_{\text{img}}(\mathbf{f})$ and $\xi_{\text{psf}}(\mathbf{h})$ enforce subjective constraints for the object brightness distribution and the PSF respectively. For gaussian noise, the likelihood penalty is:

$$\varphi(\mathbf{f}, \mathbf{h}|\mathbf{g}) = (\mathbf{g} - \mathbf{m})^{\mathrm{T}} \cdot \mathbf{C}_{\text{noise}}^{-1} \cdot (\mathbf{g} - \mathbf{m}) \tag{6}$$

where $\mathbf{C}_{\text{noise}}$ is the covariance matrix of the noise. Many different kinds of regularization have been used in image reconstruction. Maximum entropy methods (MEM) make use of:

$$\xi_{\text{MEM}}(\mathbf{f}) = \sum_k \left[p_k - f_k + f_k \log(f_k/p_k)\right]$$

where $\mathbf{p}$ is some prior distribution for $\mathbf{f}$. Gaussian prior can be enforced by means of a quadratic regularization:

$$\xi_{\text{quad}}(\mathbf{f}) = (\mathbf{f}-\mathbf{p})^{\mathrm{T}} \cdot \mathbf{W} \cdot (\mathbf{f}-\mathbf{p})$$

where $\mathbf{p}$ is a prior distribution for $\mathbf{f}$ and $\mathbf{W}$ is a positive symmetric matrice (likely related to some a priori covariance). Tikhonov regularization is a special case of the previous one:

$$\xi_{\text{Tikhonov}}(\mathbf{f}) = \| \mathbf{f} \| = \mathbf{f}^{\mathrm{T}} \cdot \mathbf{f}.$$

Since the highest spatial frequencies are likely to be more ill-conditioned, proper regularization can be obtained by enforcing the smoothness of, e.g., $\mathbf{f}$ with:

$$\xi_{\text{smoothness}}(\mathbf{f}) = \| \mathbf{D} \cdot \mathbf{f} \|$$

where $\mathbf{D}$ is some finite difference operator. The smoothness of a PDF can also be enforced by a regularization of its power-spectrum:

$$\xi_{\text{spectral}}(\mathbf{f}) = \sum_u w_u |\hat{f}_u|^2 \tag{7}$$

where the weighting $w_u \in \mathbb{R}_+$ is a non decreasing function of the length of the spatial frequency. For instance:

$$w_u = |\mathbf{u}|^\beta \quad \text{with} \quad \beta > 0. \tag{8}$$

Obviously, regularisation should only depends on the *shape* of $\mathbf{f}$ and $\mathbf{h}$, not on the particular value of $\sum_x f_x$ nor that of $\sum_x h_x$. This would also prevent the normalization ambiguity to have an incidence on the values of the regularization terms. To that end, it is sufficient to define the regularization terms as functions of the normalized PDF's: $\mathbf{f}/\sum_x f_x$ and $\mathbf{h}/\sum_x h_x$. This prescription was used for all the tests reported in this paper using the regularization as defined by Eqs. (7) and (8) and with $\beta = 2$:

$$\xi_{\text{img}}(\mathbf{f}) = \xi_{\text{spectral}}\left(\frac{\mathbf{f}}{\sum_x f_x}\right) \tag{9}$$

and similarly for $\mathbf{h}$. Also note that all the tests were carried out with the image of chromosomes shown in Fig. 2.

2. CONSTRAINED OPTIMIZATION

Assuming that regularization insures the uniqueness of the solution, the blind deconvolution problem can be solved by some efficient constrained optimization algorithm. Owning to the very large number of parameters (up to several 10^6) in the problem, only iterative methods can be used. Since the penalty function is non-quadratic (at least due to the likelihood term), non-linear algorithms should behave better. Finally, the parameters are bounded (constrained to be positive) which introduces some complication in the optimization process.

2.1. Conjugate Gradients Optimization

Conjugate gradients is widely used in multi-dimensional optimization and only requires the computation of the penalty function ϕ and its gradient $\nabla\phi$ with respect to the parameters. Conjugate gradients method seeks for the new estimate as:

$$(\mathbf{f},\mathbf{h})^{(n+1)} = (\mathbf{f},\mathbf{h})^{(n)} + \alpha^{(n)}\,\mathbf{s}^{(n)} \tag{10}$$

where the search direction is:

$$\mathbf{s}^{(n)} = \beta^{(n)}\mathbf{s}^{(n-1)} - (\nabla_{\mathbf{f}}\phi, \nabla_{\mathbf{h}}\phi)^{(n)} \tag{11}$$

The optimal value of $\beta^{(n)}$ can be derived from the previous and current gradients (Fletcher-Reeves or Polak-Ribiere methods; Polak-Ribiere with positive β is more suitable for a non-quadratic penalty[6] as it is the case for the blind deconvolution). The value of $\alpha^{(n)}$ is obtained by a *line search algorithm* (1D optimization). Precise search along $\mathbf{s}^{(n)}$ requires many number of evaluations of ϕ because it is non-quadratic (about 20 evaluations to find $\alpha^{(n)}$ with a relative precision of 10^{-8}). Thanks to the *inexact line search* algorithm of Moré & Thuente,[7]

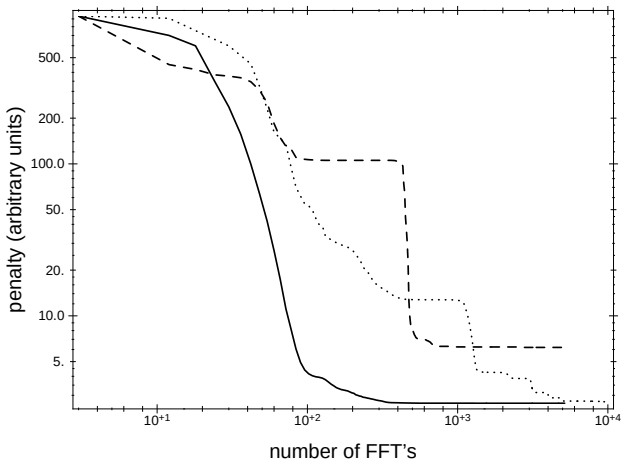

Figure 3. Convergence rate of blind deconvolution with respect to the normalization ambiguity. In all the cases, the shape of the starting solutions were the same (i.e. only the normalization change) and the optimization method was VMLM-B with 5 directions memorized. **Dashed curve:** start with $\sum_x f_x = \sum_x g_x$ and $\sum_x h_x = 1$. **Dotted curve:** start with $\sum_x f_x = \sum_x h_x = (\sum_x g_x)^{1/2}$. **Solid curve:** start with $\sum_x f_x = \sum_x h_x = 1$ and use optimal rescaling of the model at every iteration, see Eq. (13).

a value $\alpha^{(n)}$ satisfying a sufficient decrease condition can be found in one or two evaluations of the penalty function at each iteration.

Because the same step size is used for $\mathbf{f}$ and $\mathbf{h}$, the efficiency of the conjugate gradient optimisation (in other words, the *condition number* of the problem) depends on the ratio $\sum_x f_x / \sum_x h_x$ (see Fig. 3). This is also true for other methods such as variable metric optimization.

2.2. Normalization Issues

In order to reduce the problems related to the *normalization ambiguity*, I propose to apply the following prescriptions:

1. For optimization methods such as conjugate gradient or variable metric that consider all the parameters globally, it is possible to improve the condition number of the problem by starting with $\mathbf{f}$ and $\mathbf{h}$ with appropriate scaling. In practice, starting with $\mathbf{f}$ and $\mathbf{h}$ such that $\sum_x f_x = \sum_x h_x$ may considerably reduce the number of required iterations comparing to, e.g. $\sum_x f_x = \sum_x g_x$ and $\sum_x h_x = 1$.

2. It is possible to rewrite the likelihood term so as it no longer depends on the actual normalization of $\mathbf{f}$ and $\mathbf{h}$. For instance, for gaussian noise:

$$\varphi(\mathbf{f},\mathbf{h}|\mathbf{g}) = \psi(\gamma,\mathbf{f},\mathbf{h}|\mathbf{g})|_{\gamma=\tilde{\gamma}} = (\mathbf{g}-\tilde{\gamma}\,\mathbf{m})^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{noise}}^{-1} \cdot (\mathbf{g}-\tilde{\gamma}\,\mathbf{m}) \tag{12}$$

with $\mathbf{m} \equiv \mathbf{f} \star \mathbf{h}$ and $\tilde{\gamma}$ such that $\partial\psi/\partial\gamma|_{\gamma=\tilde{\gamma}} = 0$, i.e.:

$$\tilde{\gamma} = \frac{\mathbf{m}^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{noise}}^{-1} \cdot \mathbf{g}}{\mathbf{m}^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{noise}}^{-1} \cdot \mathbf{m}}. \tag{13}$$

This optimal scaling is very easy to implement since it does not change very much the gradient of φ, e.g.:

$$\begin{aligned} \frac{\partial\varphi(\mathbf{f},\mathbf{h}|\mathbf{g})}{\partial f_x} &= \left.\frac{\partial\psi(\gamma,\mathbf{f},\mathbf{h}|\mathbf{g})}{\partial f_x}\right|_{\gamma=\tilde{\gamma}} + \frac{\partial\tilde{\gamma}}{\partial f_x} \underbrace{\left.\frac{\partial\psi(\gamma,\mathbf{f},\mathbf{h}|\mathbf{g})}{\partial\gamma}\right|_{\gamma=\tilde{\gamma}}}_{\equiv 0} \\ &= -2\,\tilde{\gamma}\,\frac{\partial\mathbf{m}}{\partial f_x}^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{noise}}^{-1} \cdot (\mathbf{g}-\tilde{\gamma}\,\mathbf{m}). \end{aligned}$$

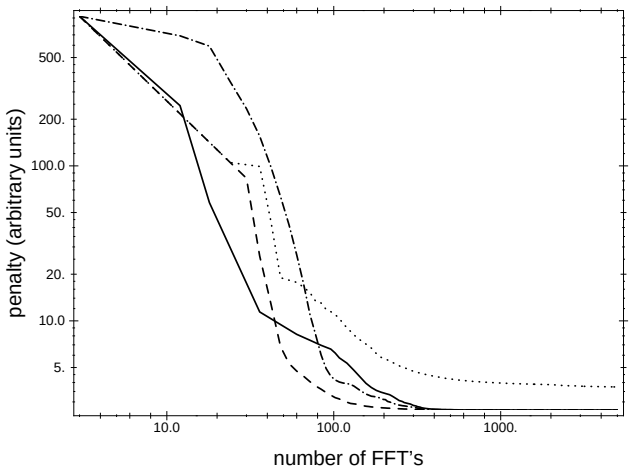

Figure 4. Convergence rate of blind deconvolution for different means to enforce positivity. **Dotted curve:** non-linear conjugate gradient with reparameterization $f_x = \exp(p_x)$ (and similarly for $\mathbf{h}$). **Dashed curve:** non-linear conjugate gradient with reparameterization $f_x = p_x^2$ (and similarly for $\mathbf{h}$). **Solid curve:** non-linear conjugate gradient with gradient projection. **Dash-dotted curve:** VMLM-B with 5 memorized directions.

Figure 3 clearly demonstrates that these prescriptions considerably improve the rate of convergence of the blind deconvolution.

2.3. Positivity Constraints

Many different ways to implement positivity constraints have been used. Lane[1] introduced additional penalty terms:

$$\ldots + \lambda \sum_{x, f_x<0} f_x^2 + \mu \sum_{x, h_x<0} h_x^2 \,;$$

but this only enforces loose positivity and requires more hyperparameters to tune (λ and μ). Another possibility is to introduce a change of parameters,[2] e.g.:

$$f_x = p_x^2 \quad \text{or} \quad f_x = \exp(p_x)$$

where p_x are the new parameters; but this makes the penalty function even more non-quadratic.

Recent optimization algorithms make use of the so-called gradient projections and/or active set of parameters to account for bound constraints. If l_x and u_x are the lower and upper bounds (possibly $\pm\infty$) respectively for the parameter f_x (for a simple positivity constraint: $l_x \equiv 0$ and $u_x \equiv \infty$), then the constraints can be enforced at each iteration as follow:

$$f_x^{(n+1)} = \mathrm{mid}(u_x, f_x^{(n)} + s_x^{(n)}, l_x) = \begin{cases} u_x & \text{if } f_x^{(n)} + s_x^{(n)} < u_x \\ l_x & \text{if } f_x^{(n)} + s_x^{(n)} > l_x \\ f_x^{(n)} + s_x^{(n)} & \text{otherwise} \end{cases} \tag{14}$$

where $\mathbf{f}^{(n)}$ and $\mathbf{s}^{(n)}$ are the parameter value and step at n-th iteration. If this scheme is applied, then it is easy to realise that the steepest descent effectively becomes:

$$d_x = \begin{cases} 0 & \text{if } f_x = l_x \text{ and } \partial\phi/\partial f_x > 0 \text{ or, if } f_x = u_x \text{ and } \partial\phi/\partial f_x < 0 \\ -\partial\phi/\partial f_x & \text{otherwise} \end{cases} \tag{15}$$

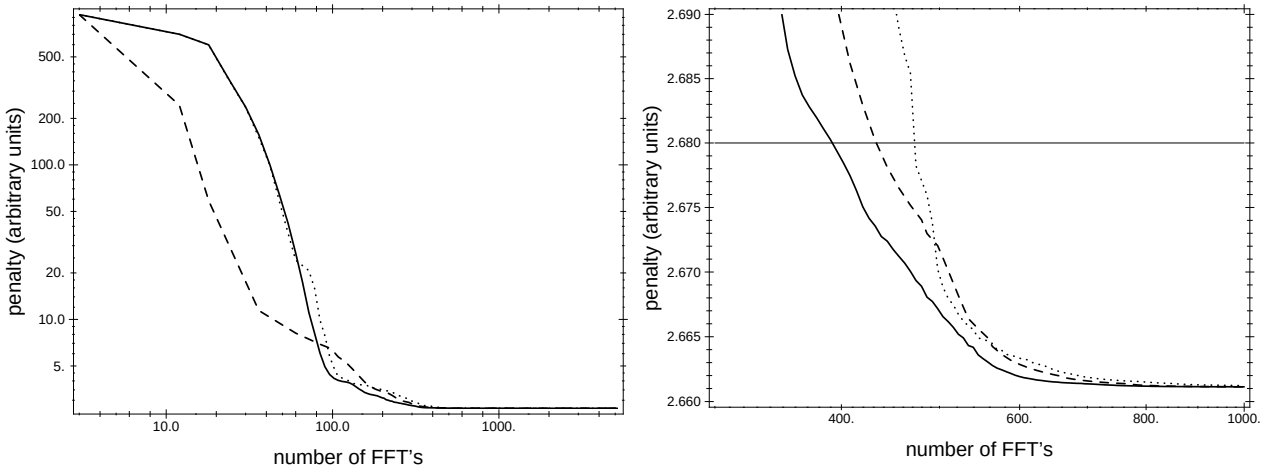

Figure 5. Comparison of the conjugate gradient and VMLM-B optimization algorithms for blind deconvolution. **Dashed curve:** non-linear conjugate gradient with bound constraints. **Dotted curve:** VMLM-B with 1 memorized direction. **Solid curve:** VMLM-B with 5 memorized directions. The rightmost figure is a magnified part of the leftmost one with a thin horizontal line to indicate the penalty level below which there is no noticeable improvement in the restored image.

which is the so-called projected (anti-)gradient that accounts for the bound constraints. Similarly the set of *active parameters* is defined as the parameters that would not violate the bound constraints if the steepest descent direction is followed:

$$l_x < f_x < u_x \quad \text{or} \quad f_x = l_x \quad \text{and} \quad \frac{\partial \phi}{\partial f_x} < 0 \quad \text{or} \quad f_x = u_x \quad \text{and} \quad \frac{\partial \phi}{\partial f_x} > 0 \tag{16}$$

It is possible to modify the conjugate gradient method to account for the bounds by choosing the new search direction as:

$$\mathbf{s}^{(n)} = \beta^{(n)} \mathbf{s}^{(n-1)} + \mathbf{d}^{(n)} \tag{17}$$

where $\mathbf{d}^{(n)}$ is the projected (anti-)gradient defined in Eq. (15) and where $\beta^{(n)}$ is provided by Polak-Ribiere formula (with positive β).

Figure 4 compares the performances of the blind deconvolution with different methods to enforce positivity. Surprinsingly, the monotonic change of parameters $f_x = \exp(p_x)$ shows very poor convergence whereas $f_x = p_x^2$ behaves the best. Nevertheless, I observed that using the projected gradient is generally superior for other kind of constrained optimization problems (e.g., conventional deconvolution). This is probably because using $f_x = p_x^2$ introduces local minima due to the non-monotonic behavior the reparameterization.

2.4. Variable Metric Method

As conjugate gradients methods, variable metric methods only require the computation of the penalty function to be minimized and its gradient. Nevertheless, their convergence rate is generally superior to that of conjugate gradients. There exist limited memory versions of the variable metric method that are suitable to solve large optimization problem (e.g., VMLM[8]) and that memorize only a small number of previous parameters and gradient changes. It is possible to modify the VMLM algorithm in order to account for bound constraints:

1. Obtain a new step thanks to the VMLM recursion but restricted to the set of current active parameters as determined by Eq. (16).

2. Use the inexact line search algorithm to find a suitable step size, the bound constraints being enforced as indicated by Eq. (14) for each try.

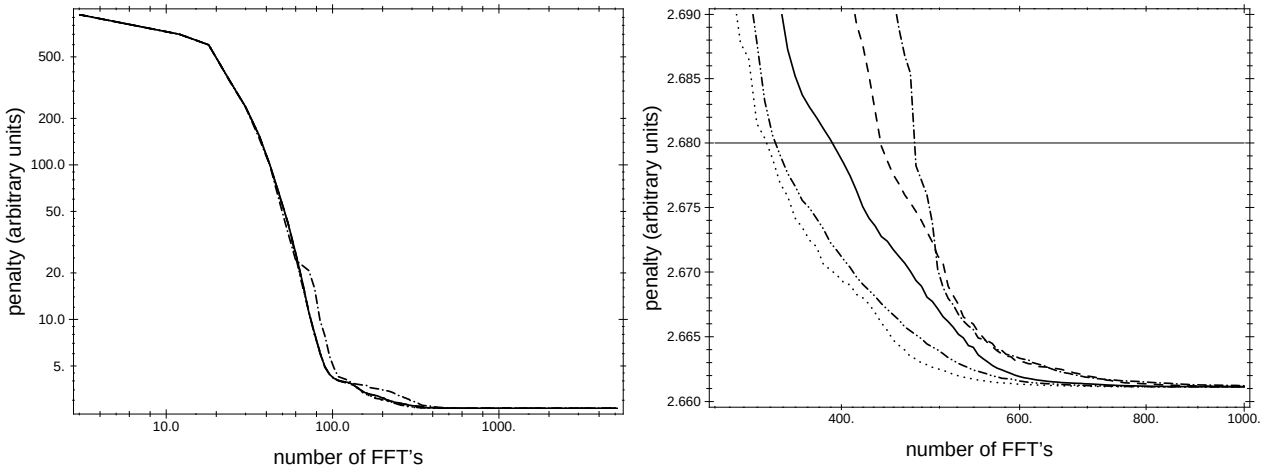

Figure 6. Convergence rate of blind deconvolution by the VMLM-B optimization algorithm. **Dash-dotted curve:** 1 memorized direction. **Solid curve:** 5 memorized directions. **Dashed curve:** 5 memorized directions and recursion restarted every 50 iterations. **Dotted curve:** 10 memorized directions. **Dash-dot-dotted curve:** 15 memorized directions. The rightmost figure is a magnified part of the leftmost one with a thin horizontal line to indicate the penalty level below which there is no noticeable improvement in the restored image.

3. Remember the effective step taken and proceed with next iteration.

I call the resulting algorithm VMLM-B (B for *bounded*).

Figure 5 shows that VMLM-B has a better convergence rate compared to constrained conjugate gradients: it takes 392 FTT's to find the solution with VMLM-B (with 5 memorized directions) instead of 433 with conjugate gradients. The number of directions memorized by VMLM-B can be tuned to increase its efficiency. For the test problem and as shown by Fig. 6, 10 memorized directions appears to be the optimum and requires only 337 FTT's to achieve the solution.

2.5. Optimization in a Local Subspace

Following Skilling & Bryan[3] and Pichon & Thiébaut,[4] it should be more efficient to use several search directions rather than a single global direction. This prescription can be further improved in the case of blind deconvolution: because $\mathbf{f}$ and $\mathbf{h}$ may behave differently it is certainly better to consider distinct sets of search directions for each component (in particular this should cancel the dependency of the condition number with the actual normalization of $\mathbf{f}$ and $\mathbf{h}$). In this case, the new estimates read:

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} + \sum_{i=1}^{m} \alpha_i^{(n)} \mathbf{s}_i^{(n)} \tag{18}$$

$$\mathbf{h}^{(n+1)} = \mathbf{h}^{(n)} + \sum_{i=m+1}^{2m} \alpha_i^{(n)} \mathbf{s}_i^{(n)} \tag{19}$$

where $\{\mathbf{s}_1^{(n)}, \ldots, \mathbf{s}_m^{(n)}\}$ and $\{\mathbf{s}_{m+1}^{(n)}, \ldots, \mathbf{s}_{2m}^{(n)}\}$ are the sets of search directions for $\mathbf{f}$ and $\mathbf{h}$ respectively. At least the projected steepest descent and the previous effective step can be used as potential search directions. Following Lucy[5] and Pichon &Thiébaut,[4]

$$\mathbf{s} = \mathbf{f} \times \nabla_{\mathbf{f}} \phi \qquad \text{(elementwise multiplication)}$$

may also be a good search direction for $\mathbf{f}$ and similarly for $\mathbf{h}$.

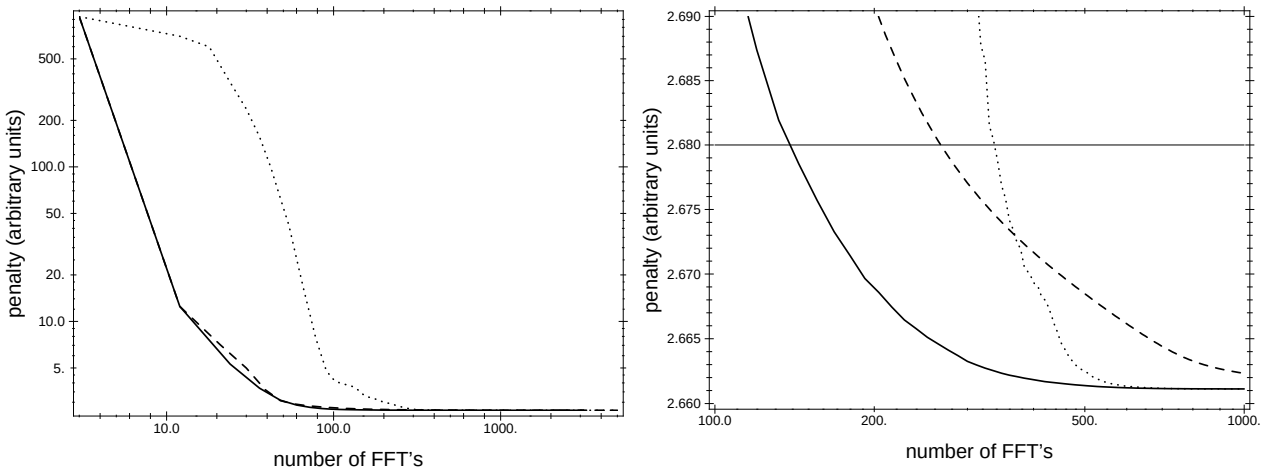

Figure 7. Convergence rate of blind deconvolution by VMLM-B and the local sub-space method. **Dotted curve:** VMLM-B with 10 memorized directions. **Dashed curve:** local sub-space with 2 search directions (steepest descent for each component). **Solid curve:** local sub-space with 4 search directions (steepest descent plus previous effective step for each component). The rightmost figure is a magnified part of the leftmost one with a thin horizontal line to indicate the penalty level below which there is no noticeable improvement in the restored image.

The best step lengths along the search directions can be obtained from a local quadratic approximation of the variation of ϕ with respect to $\boldsymbol{\alpha}$:

$$\phi\left(\mathbf{f}^{(n)}+\sum_{i=1}^{m}\alpha_i\,\mathbf{s}_i^{(n)},\mathbf{h}^{(n)}+\sum_{i=m+1}^{2m}\alpha_i\,\mathbf{s}_i^{(n)}|\mathbf{g}\right)-\phi\left(\mathbf{f}^{(n)},\mathbf{h}^{(n)}|\mathbf{g}\right)\simeq-\boldsymbol{\alpha}^{\mathrm{T}}\cdot\mathbf{b}^{(n)}+\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\cdot\mathbf{A}^{(n)}\cdot\boldsymbol{\alpha}$$

which is optimized for:

$$\boldsymbol{\alpha}^{(n)}=[\mathbf{A}^{(n)}]^{-1}\cdot\mathbf{b}^{(n)}$$

Since the local quadratic approximation is only valid in a small region, the length of $\boldsymbol{\alpha}$ must be controlled. This can be done in the same way as in Levenberg-Marquardt method by using a Lagrange parameter $\lambda>0$ and taking the step size to be:

$$\boldsymbol{\alpha}(\lambda)^{(n)}=[\mathbf{A}(\lambda)^{(n)}]^{-1}\cdot\mathbf{b}^{(n)}\quad\text{with}\quad A(\lambda)_{i,j}^{(n)}=(1+\delta_{i,j}\lambda)\,A_{i,j}^{(n)}.$$

λ must be augmented to reduce the step length whenever the quadratic approximation appears to be only valid in a narrower region.

As shown by Fig. 7, local subspace method outperforms the VMLM-B method: 297 FFT's (against 337 for VMLM-B with optimal number of directions) are required for the local subspace method with only 2 search directions: the two projected anti-gradient for $\mathbf{f}$ and $\mathbf{h}$. If 2 more search directions are used, i.e. the previous step along $\mathbf{f}$ and $\mathbf{h}$, then the local subspace method takes only 139 FFT's to find the solution.

3. CONCLUSION

This paper demonstrates that the speed of blind deconvolution algorithms can be greatly improved thanks to an optimal rescaling of the model (the gain may be **huge**, see Fig. 3). This improves the condition number of the inverse problem which could otherwise be almost arbitrarily altered by the normalization of the sought image and PSF.

This paper also explains how to modify existing optimization algorithms (variable metric and conjugate gradients) in order to account for bound constraints. This leads to algorithms faster than conventional conjugate gradient. Even faster convergence (estimated by the number of required FFT's to achieve a good solution) can be obtained by varying the image and the PSF along distinct directions that may scale differently.

ACKNOWLEDGMENTS

The image of the chromosomes was given by Jean-Claude Bernengo from the Centre Commun de Quantimétrie (Université Claude Bernard, Lyon, France) and was processed by Cyril Mavré using the software described in this paper.

The PSF used for the astronomical simulation was kindly given by François Roddier and is a real PSF of the partially compensated adaptive optics system of the CFH telescope (ancestor of PUEO).

All the processing was done with Yorick, a free interactive data processing language written by David Munro (`ftp://ftp-icf.llnl.gov/pub/Yorick/doc/index.html`).

REFERENCES

1. R. Lane, *Blind deconvolution of speckle images*, J. Opt. Soc. Am. A **9**, pp. 1508–1514, 1992.
2. E. Thiébaut and J.-M. Conan, *Strict a priori constraints for maximum likelihood blind deconvolution*, J. Opt. Soc. Am. A **12**, pp. 485–492, 1995.
3. J. Skilling and R. K. Bryan, *Maximum entropy image reconstruction: general algorithm*, MNRAS **211**, pp. 111–124, 1984.
4. C. Pichon and E. Thiébaut, *Recovering distribution functions from observed flat and round galactic disks*, MNRAS **301**, p. 419, 1998.
5. L .B. Lucy, *Optimum Strategies for Inverse Problems in Stastitical Astronomy*, A&A **289**, p. 983, 1994.
6. J. C. Gilbert and J. Nocedal, *Global Convergence Properties of Conjugate Gradient Methods*, SIAM Journal on Optimization **2**, pp. 21–42, 1992.
7. J. J. Moré and D. J. Thuente, *Line search algorithms with guaranteed sufficient decrease*, ACM Transactions on Mathematical Software **20**, pp. 286–307, 1994.
8. J. Nocedal, *Updating Quasi-Newton Matrices with Limited Storage*, Mathematics of Computation **35**, pp. 773–782, 1980.