# FAST RELATIVE NEWTON ALGORITHM FOR BLIND DECONVOLUTION OF IMAGES

*Alexander M. Bronstein, Michael Zibulevsky, Yehoshua Y. Zeevi*

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

*Michael M.Bronstein*

Department of Computer Science
Technion - Israel Institute of Technology
Haifa 32000, Israel

**ABSTRACT**

We present an efficient Newton-like algorithm for quasi-maximum likelihood (QML) blind deconvolution of images. This algorithm exploits the sparse structure of the Hessian. An optimal distribution-shaping approach by means of sparsification allows one to use simple and convenient sparsity prior for processing of a wide range of natural images. Simulation results demonstrate the efficiency of the proposed method.

## 1. INTRODUCTION

Two-dimensional *blind deconvolution* (BD) is a special case of a more general problem of *image restoration*. The goal of BD is to reconstruct the original scene from an observation degraded by the action of a linear shift invariant (LSI) system, when no or very little *a priori* information about the scene and the degradation process is available, hence the term "blind". BD is critical in many fields, including astronomy, remote sensing, biological and medical imaging and microscopy.

According to the convolution model, the observed sensor image $X$ is created from the *source image* $S$ passing through an LSI system characterized by the point spread function $W$,

$$X = W * S. \quad (1)$$

We assume that the action of $W$ is invertible (at least approximately), i.e. there exists some other kernel $H$ such that $W * H \approx \delta$. This assumption holds well especially in the case of blurring kernels resulting from scattering (such kernels are usually Lorenzian-shaped and their inverse can be approximated by small FIR kernels). The aim of BD is to find such a *deconvolution* (*restoration*) kernel $H$ that produces an estimate $\tilde{S}$ of $S$ up to integer shift and scaling factor:

$$\hat{S}_{mn} = (H * X)_{mn} \approx c \cdot S_{m-\Delta_M, n-\Delta_N}. \quad (2)$$

Unlike approaches estimating the image and the blurring kernel, we estimate the restoration kernel only, which results in a lower dimensionality of the problem [1]. Here we present a QML BD algorithm, which generalizes the fast relative Newton algorithm previously proposed for blind source separation [2]. We also propose optimal distribution-shaping approach (sparsification), which allows to use simple and convenient sparsity prior for a wide class of images. For technical details see [3, 4].

## 2. QML BLIND DECONVOLUTION

Denote by $Y = H * X$ the source estimate and let us assume that $S$ is zero-mean i.i.d. In the zero-noise case, the normalized minus-log-likelihood function of the observed signal $X$, given the restoration kernel $H$, is

$$\ell(X; H) = -\frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log |\mathcal{F}H(\xi, \eta)| \; d\xi d\eta + \frac{1}{M_X N_X} \sum_{m,n} \varphi(Y_{mn}), \quad (3)$$

where $\varphi(s) = -\log p_s(s)$, $p_s(s)$ stands for the source probability density function (PDF), $M_X \times N_X$ is the observation sample size, and $\mathcal{F}H(\xi, \eta)$ denotes the Fourier transform of $H_{mn}$. We will henceforth assume that $H$ is a FIR, supported on $[-M, ..., M] \times [-N, ..., N]$, and denote $K_M = 2M+1$, $K_N = 2N + 1$. Cost functions similar to (3) were also obtained in the 1D case using negative joint entropy and information maximization considerations [5].

### 2.1. The choice of $\varphi(s)$

Source images arising in most applications have usually multi-modal non-log-concave distributions. These are difficult to model and are not suitable for optimization. However,

This research has been supported by the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission, and by the Technion Ollendorff Center for Vision and Image Sciences.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

consistent estimator of $S$ can be obtained by minimizing $\ell(X; H)$ even when $\varphi(s)$ is not exactly equal to $-\log p_S(\cdot)$. Such *quasi-ML estimation* has been shown to be practical in instantaneous blind source separation [6, 2, 7] and blind deconvolution of time signals [3]. For example, when the source is super-Gaussian (sparse), a smooth approximation of the absolute value function is a good choice for $\varphi(s)$ [8, 9]. Although natural images are usually far from being sparse, they can be transformed into a space of a sparse representation. We will therefore focus our attention on modelling super-Gaussian distributions using a family of convex smooth functions

$$\varphi_\lambda(s) = |s| - \lambda \log\left(1 + \frac{|s|}{\lambda}\right) \tag{4}$$

with $\lambda$ being a positive smoothing parameter; $\varphi_\lambda(s) \to |s|$ as $\lambda \to 0^+$.

### 2.2. Approximation of the log-likelihood function

In practice, it is difficult to evaluate the first term of $\ell(X; H)$ containing the integral. However, it can be approximated with any desired accuracy by [3]

$$\frac{1}{M_F N_F} \sum_{k=0}^{M_F} \sum_{l=0}^{N_F} \log |F_{kl}|, \tag{5}$$

where $F_{kl} = \mathcal{F}H\left(\frac{2\pi k}{M_F}, \frac{2\pi l}{N_F}\right)$ are the 2D DFT coefficients of $H_{mn}$ [3], zero-padded over $M_F \times N_F$. The approximation error vanishes as $M_F, N_F$ grow to infinity. $M_F$ and $N_F$ should be chosen as integer powers of 2 to allow using FFT.

### 2.3. Gradient and Hessian of $\ell(X; H)$

The optimization algorithm discussed in Section 3 requires knowledge of the gradient and the Hessian of $\ell(X; H)$. The gradient of $\ell(X; H)$ w.r.t $H_{ij}$ is given by (for derivation see [4]):

$$\frac{\partial \ell}{\partial H_{ij}} = -Q_{-i,-j} + \frac{1}{M_X N_X} \sum_{m,n} \varphi'(Y_{mn})\, X_{m-i,n-j}, \tag{6}$$

where $Q_{mn}$ is the inverse DFT of $F_{kl}^{-1}$. The Hessian of $\ell(X; H)$ is:

$$\frac{\partial^2 \ell}{\partial H_{ij} \partial H_{kl}} = \frac{1}{M_X N_X} \sum_{m,n} \varphi''(Y_{mn})\, x_{m-i,n-j} x_{m-k,n-l} + R_{-(i+j),-(k+l)}, \tag{7}$$

where $R_{mn}$ is the inverse DFT of $F_{kl}^{-2}$. Both the gradient and the Hessian can be evaluated efficiently using FFT.

## 3. THE RELATIVE NEWTON METHOD

A fast relative optimization algorithm for blind source separation, based on the Newton method was introduced in [2]. In [3] it was used for BD of 1D signals. Here we use the relative optimization framework for BD of images. The main idea of relative optimization is to iteratively produce an estimate of the source signal and use it as the observed signal at the subsequent iteration:

***Relative optimization algorithm***

1. Start with initial estimates of the restoration kernel $H^{(0)}$ and the source $X^{(0)} = X$.
2. For $k = 0, 1, 2, \ldots$, until convergence
3. Start with $X^{(k+1)} = \delta$.
4. Using an unconstrained optimization method, find $H^{(k+1)}$ such that $\ell(X^{(k)}; H^{(k+1)}) < \ell(X^{(k)}; \delta)$.
5. Update source estimate: $X^{(k+1)} = H^{(k+1)} * X^{(k)}$.

6) End

The restoration kernel estimate at $k$-th iteration is $\hat{H} = H^{(0)} * \ldots * H^{(k)}$, and the source estimate is $\hat{S} = X^{(k)}$. This method allows to construct large restoration kernels growing at each iteration, using a set of relatively low-order factors. It can be seen easily that the relative optimization algorithm has uniform performance, i.e. its step at iteration $k$ depends only on $W * H^{(0)} * \ldots * H^{(k-1)}$.

Step 4 can be carried out using any unconstrained optimization algorithm. Particulary, it was found that a single Newton step can be used, yielding very fast convergence. In the standard Newton method [10], the descent direction $d$ at each iteration is given by solution of the linear system

$$\nabla^2 \ell \cdot d = -\nabla \ell. \tag{8}$$

In order to guarantee descent direction, positive definiteness of the Hessian is usually forced by using modified Cholesky factorization, which requires about $\mathcal{O}\left(\frac{1}{6} K_M^3 K_N^3 + K_M^2 K_N^2\right)$ operations [10]. Having the direction $d$, the new iterate $H^{(k+1)}$ is given by

$$H^{(k+1)} = H^{(k)} + \alpha d,$$

where $\alpha$ is the step size determined by either exact line search or by backtracking line search, which guarantees monotonic decrease of the objective function at every iteration.

Practical use of the relative Newton step is limited to small values of $M, N$ and $M_X, N_X$ due to the complexity of Hessian construction, and solution of the Newton system. However, this complexity can be significantly reduced if special Hessian structure is exploited. Near the solution point, $X^{(}k) \approx cS$, hence $\ell(X;\delta)$ evaluated at each relative Newton iteration becomes approximately $\ell(cS;\delta)$. For a zero-mean i.i.d. source and sufficiently large sample size (in practice, $M_X N_X > 10^2$), the following approximation holds:

$$\frac{\partial^2 \ell}{\partial H_{ij}\partial H_{kl}} \approx \delta_{i+j,k+l} + c^2 \cdot \mathbb{E}\left\{\varphi''(cS_{00})S_{ij}S_{kl}\right\}$$
$$= \begin{cases} \mathbb{E}\left\{\varphi''(cS)(cS)^2\right\}+1 & : \quad (i,j)=(k,l)=\mathbf{0}, \\ \mathbb{E}\varphi''(cS)\cdot \mathbb{E}(cS)^2 & : \quad (i,j)=(k,l)\neq \mathbf{0}, \\ 1 & : \quad (i,j)=-(k,l)\neq \mathbf{0}, \\ 0 & : \quad otherwise. \end{cases}$$

Using this approximation, only the main diagonal of the Hessian matrix has to be evaluated at each iteration, and the solution of the Newton system (8) separates into the set of $2\times 2$ systems of the form

$$\begin{pmatrix} \nabla \ell_{-i,-j} \\ \nabla \ell_{ij} \end{pmatrix} = -\begin{pmatrix} \nabla^2 \ell_{-i,-j,-i,-j} & 1 \\ 1 & \nabla^2 \ell_{ijij} \end{pmatrix} \begin{pmatrix} d_{-i,-j} \\ d_{ij} \end{pmatrix}$$

for $(i,j)\neq \mathbf{0}$, and an additional equation

$$\nabla \ell_{00} \quad = \quad -\nabla^2 \ell_{0000}\, d_{00}.$$

We will henceforth refer to this approximate Newton step as to the *fast relative Newton method*, since its complexity is of the same order as that of the gradient-based methods.

## 4. SPARSIFICATION

The sparsity prior used in the QML function (3) is valid for sparse sources and not valid for natural images in their native space. On the other hand, it is especially convenient for the underlying optimization problem due to its convexity. Moreover, deconvolution of sparse sources is especially accurate. While it is difficult to model actual distributions of natural images, it is much easier to transform an image in such a way that it fits the sparsity prior. This idea was previously exploited successfully in blind source separation [8, 9, 11].

Let us assume that there exist a *sparsifying transformation* $\mathcal{T}_S$, which yields a sparse representation of the source $S$, such that our algorithm is likely to produce in such a case a good source estimate of the restoration kernel $H$. The problem is that in the BD setting, $S$ is not available, and we can apply $\mathcal{T}_S$ to the observation $X$ only. Hence, it is necessary that the sparsifying transformation commute with the convolution operation, i.e.

$$(\mathcal{T}_S S) * W = \mathcal{T}_S (S * W) = \mathcal{T}_S X, \qquad (9)$$

such that applying $\mathcal{T}_S$ to $X$ is equivalent to applying it to $S$. It is obvious that $\mathcal{T}_S$ must be a shift-invariant (SI) transformation. For simplicity, we limit our attention to linear shift-invariant (LSI) transformations, i.e. $\mathcal{T}$ that can be represented by convolution with a *sparsifying kernel* $\mathcal{T}S = T * S$. An example of a source image sparsified by using the corner-detecting kernel is presented in Figure 1.

The sparsifying kernel can be constructed based on *a priori* information about the source image, such as image structure, whether it has sharp edges, corners, etc. Alternatively, the sparsifying kernel can be obtained from *training*, minimizing some sparsity criterion over a representative set of source images $\left\{S^{(k)}\right\}$. The latter can be achieved by solving, for example, the following problem [4]

$$\min_T \sum_{m,n,k} \left| \left( T * S^{(k)} \right)_{mn} \right| \qquad : \qquad \|T\| = 1.$$

## 5. SIMULATION RESULTS

We demonstrate the performance of the fast relative Newton algorithm in a simulation. A $100 \times 100$ aerial image (Fig. 1, left) was blurred by a Lorenzian-shaped kernel, approximating a typical point spread function of scattering medium (Fig. 2, left). In this case, image restoration can be well accomplished by the proposed method, since the blurring kernel can be inverted by a relatively short FIR restoration filter. Here we present the noiseless case; for examples of deconvolution in the presence of noise see [4].

Blind deconvolution was performed with a $3 \times 3$ FIR kernel. A $2 \times 2$ corner detector was used as the sparsifying kernel. The smoothing parameter was set to $\lambda = 10^{-2}$. The iterative optimization algorithm was terminated when $\|\nabla \ell\|$ fell below $10^{-10}$. In a typical case of application of our algorithm, convergence is obtained in 10-20 iterations, requiring about 0.1 sec per iteration on a PC workstation. Restoration results are depicted in Figure 2 (right). Restoration quality of SIR = 17.04 dB and $\text{SIR}_\infty$ = 23.55 dB is achieved. SIR refers to the interference energy, whereas $\text{SIR}_\infty$ to the maximum interference. Additional examples are presented in [4].

Convergence of the fast relative Newton method has been compared with that obtained using the standard Newton method, with restoration kernels of different sizes, used directly for minimization of $\ell(X;H)$. An i.i.d. sparse image was used as the source. The proposed method demonstrates a significantly faster convergence (Fig. 3).

## 6. CONCLUSIONS

The proposed Newton-like QML BD algorithm, based on the sparsity prior and the special Hessian structure, is computationally efficient. It is applicable to a wide class of images, which can be represented sparsely by application of a shift-invariant transformation. In cases where some prior knowledge regarding the structure of the image and/or

Source image

Sparsified source

**Fig. 1**. The source image – an aerial photo (left), and its sparsified version, obtained by using a corner-detecting kernel (right).

Observed image

Restoration result

**Fig. 2**. An image, created by blurring the source shown in Fig. 1 by convolving it with a kernel simulating scattering medium (left). Restoration result obtained by using the fast relative Newton method applied to the sparsified observed image (right).

the physics of the imaging conditions is available, a better choice of the proper sparsifying transformation can be made. Otherwise, such a transformation can be found by training. Good performance was achieved on simulated data in moderate noise conditions. Possible applications are microscopy, optical tomography, *in vivo* optical imaging, etc.

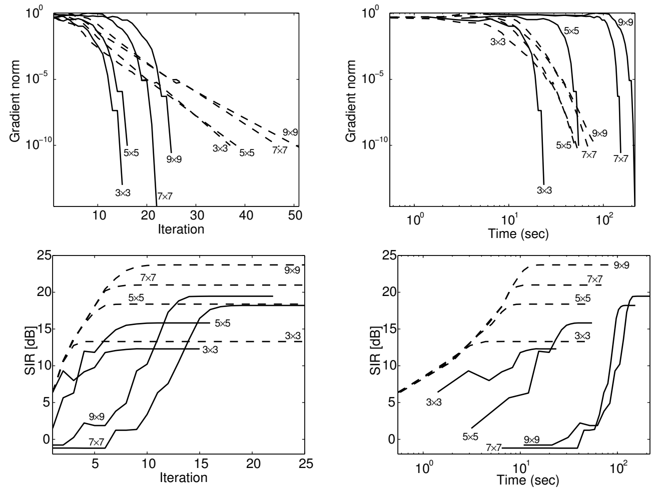


**Fig. 3**. Convergence of the Newton method (solid) and the fast relative Newton method (dashed) for different restoration kernel sizes (indicated on the plots).

## 7. REFERENCES


[1] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–35, 1978.

[2] M. Zibulevsky, "Sparse source separation with relative Newton method," in *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, April 2003, pp. 897–902.

[3] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky, "Relative optimization for blind deconvolution," *IEEE Sig. Proc.*, 2004, to appear.

[4] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Blind deconvolution of images using optimal sparse representations," *IEEE Image Proc.*, 2004, to appear.

[5] S.-I. Amari, A. Cichocki, and H. H. Yang, "Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach," in *Proc. SYSID*, July 1997, pp. 1057–1062.

[6] D. Pham and P. Garrat, "Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Sig. Proc.*, vol. 45, pp. 1712–1725, 1997.

[7] P. Kisilev, M. Zibulevsky, and Y.Y. Zeevi, "Multiscale framework for blind separation of linearly mixed signals," *JMLR*, vol. 4, pp. 1339–1363, 2003.

[8] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition," *Neural Computation*, vol. 13, no. 4, 2001.

[9] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, "Blind source separation via multinode sparse representation," in *Proc. NIPS*, 2002.

[10] D. P. Bertsekas, *Nonlinear Programming (2nd edition)*, Athena Scientific, 1999.

[11] A. M. Bronstein, M.M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Separation of reflections via sparse ICA," in *Proc. IEEE ICIP*, 2003.