

Blind Deconvolution of Images Using Optimal Sparse Representations

Michael M. Bronstein, *Student Member, IEEE*, Alexander M. Bronstein, *Student Member, IEEE*, Michael Zibulevsky, and Yehoshua Y. Zeevi

Abstract—The relative Newton algorithm, previously proposed for quasi-maximum likelihood blind source separation and blind deconvolution of one-dimensional signals is generalized for blind deconvolution of images. Smooth approximation of the absolute value is used as the nonlinear term for sparse sources. In addition, we propose a method of sparsification, which allows blind deconvolution of arbitrary sources, and show how to find optimal sparsifying transformations by supervised learning.

Index Terms—Blind deconvolution, quasi-maximum likelihood, relative Newton optimization, sparse representations.

I. INTRODUCTION

TWO-DIMENSIONAL (2-D) *blind deconvolution* (BD) is a special case of a more general problem of *image restoration*. The goal of BD is to reconstruct the original image from an observation degraded by the action of a linear shift invariant (LSI) system, when no or very little *a priori* information about the image and the degradation process is available, hence the term “blind.” BD is critical in many fields, including astronomy [1], [2], remote sensing [3], biological and medical imaging [4], [5], microscopy [6], [7], etc. Typically, the image is degraded by imperfections of an optical system and distortions caused by light propagation in the medium, and can be presented in terms of convolution of the source image with some *blurring kernel* or *point spread function* (PSF); in such applications, the term *deblurring* is synonymous to deconvolution.

A. Problem Formulation

In the general setup of 2-D BD, the observed sensor image x is created from the *source image* s passing through a convolutive system defined by its impulse response a

$$x_{\mathbf{n}} = \sum_{\mathbf{m}} a_{\mathbf{m}} s_{\mathbf{n} - \mathbf{m}} + v_{\mathbf{n}}$$

and is possibly contaminated by additive *sensor noise* v . Sometimes, s can be contaminated by *source noise* u , usually multiplicative. For notation convenience, we use bold subscripts to

denote multi-indices consisting of two variables: $\mathbf{n} = (n_1, n_2)$. Summation is performed over both of them. We assume that the action of w is invertible, at least approximately. The aim of BD is to find such *deconvolution (restoration) kernel* w that produces an estimate \tilde{s} of s up to integer shift $\Delta = (\Delta_1, \Delta_2)$ and scaling factor c

$$\tilde{s}_{\mathbf{n}} = \sum_{\mathbf{m}} w_{\mathbf{m}} x_{\mathbf{n} - \mathbf{m}} \approx c s_{\mathbf{n} - \Delta}$$

or, equivalently, the *global system response* should be

$$g_{\mathbf{n}} = (a * w)_{\mathbf{n}} \approx c \delta_{\mathbf{n} - \Delta},$$

where $\delta_{\mathbf{n}}$ denotes the Krönecker delta (discrete impulse signal).

B. Previous Work

Among the various BD methods that have been previously proposed, we briefly outline only the basic approaches (for a comprehensive comparison see, e.g., [8]). Most of the BD approaches can be divided into *parametric* and *nonparametric*. In applications where the form of the PSF can be assumed in advance (e.g., motion blur or defocus), it is possible to use a parametric model of the PSF and instead of finding the PSF itself, one can try to estimate the parameters of its model; the advantages are, obviously, in having a smaller number of variables. However, in real applications, it is often difficult to derive a good model for the PSF.

BD approaches can be divided into those that estimate the blurring kernel, those estimating the source image and the blurring kernel simultaneously, and those estimating the restoration kernel. The first class includes the so-called *a priori* blur identification methods, which first estimate the blurring kernel and then employ a nonblind deconvolution algorithm to find the source estimate [9]–[11].

The second class includes methods based on statistical or deterministic priors of the source image, the blurring kernel, and the noise [12], [13]. Estimation of the source image is performed by maximizing some optimality criterion, which includes these priors. Since the variables in this problem are both the source image and the blurring kernel, the computational complexity is a major problem.

The third class of methods usually employs maximum likelihood (ML) estimators of the restoration kernel; such estimators usually incorporate priors on the image and the kernel and ignore the noise. Since, in this case, $\tilde{s} = w * x$, there is no need to estimate the source image; therefore, these approaches demand the solution of more modest optimization problems and, consequently, are more computationally efficient. However, the

Manuscript received January 3, 2004; revised June 14, 2004. This work was supported in part by the HASSIP Research Network Program HPRN-CT-2002-00285 and in part by the European Commission and by the Ollendorff Minerva Center. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vincent Caselles.

M. M. Bronstein and A. M. Bronstein are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: bronstein@iee.org; alexbron@iee.org).

M. Zibulevsky, and Y. Y. Zeevi are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: mzib@ee.technion.ac.il; zeevi@ee.technion.ac.il).

Digital Object Identifier 10.1109/TIP.2005.847322

exact source distribution, required for the ML approach, is often unknown.

A possible remedy is to use an approximate probability density function; such a modified ML approach is usually referred to as *quasi* ML, or QML. Such estimation techniques were successfully used in blind source separation (BSS) and BD [14]–[18]. QML BD has received significant interest after the introduction of the natural gradient algorithm, a robust and efficient numerical optimization method for solution of QML BD estimation problems [19].

In [20], the relative Newton QML framework for BD of one-dimensional (1-D) signals was introduced and was shown to have superior performance compared to the natural gradient algorithm, while preserving the same order of computational complexity per iteration. Here, we extend this method to the 2-D case. In addition, we present a novel learning approach of optimal sparse representation, which can be used for BD of source images with arbitrary distributions.

II. QML BLIND DECONVOLUTION

The convolution operation $w * x$ can be thought of as application of an infinite Toeplitz-block-Toeplitz operator \mathcal{W} , defined by the impulse response w . Denoting the source estimate by $y = w * x$ and assuming that s is i.i.d., the likelihood of the observation x given the restoration kernel w in zero-noise conditions is

$$p_{x|w}(x; w) = |\det \mathcal{W}| \cdot \prod_{\mathbf{n}} p_s(y_{\mathbf{n}}) \quad (1)$$

where $p_s(s)$ stands for the source probability density function (PDF). For convenience, instead of maximizing the likelihood $p_{x|w}(x; w)$, maximum likelihood estimators usually minimize $-\log p_{x|w}(x; w)$. Neglecting edge effects, the following normalized minus-log likelihood function is obtained [14], [19] (see [21] for details):

$$\ell(x; w) = -\frac{M_X N_X}{4\pi^2} \int_{\Pi} \log |\hat{w}(\boldsymbol{\xi})| d\boldsymbol{\xi} + \sum_{\mathbf{n}} \varphi(y_{\mathbf{n}}) \quad (2)$$

where $M_X N_X$ is the observation sample size, $\Pi = [-\pi, \pi] \times [-\pi, \pi]$, $\varphi(s) = -\log p_s(s)$, and

$$\hat{w}(\boldsymbol{\xi}) = \sum_{\mathbf{n}} w_{\mathbf{n}} e^{-i\mathbf{n}^T \boldsymbol{\xi}}$$

denotes the 2-D Fourier transform of $w_{\mathbf{n}}$. We will, henceforth, assume that $w_{\mathbf{n}}$ has finite impulse response (FIR), supported on $\mathbf{n} \in [-M, \dots, M] \times [-N, \dots, N]$. We will use $K_M = 2M + 1$ and $K_N = 2N + 1$ to denote the dimensions of the restoration kernel, and denote the first and the second terms of $\ell(x; w)$ as ℓ_1 and ℓ_2 , respectively.

In practice, the first term of $\ell(x; w)$, containing the integral, is difficult to evaluate. However, it can be approximated with any desired accuracy by [20]

$$\frac{1}{4\pi^2} \int_{\Pi} \log |\hat{w}(\boldsymbol{\xi})| d\boldsymbol{\xi} \approx \frac{1}{M_F N_F} \sum_{\mathbf{k}} \log |\hat{w}_{\mathbf{k}}|$$

where

$$\hat{w}_{\mathbf{k}} = \mathcal{F}\{w_{\mathbf{n}}\}_{\mathbf{k}} = \hat{w} \left(\frac{2\pi k_1}{M_F}, \frac{2\pi k_2}{N_F} \right) \quad (3)$$

are the 2-D DFT coefficients of $w_{\mathbf{n}}$, zero-padded to the support $\mathbf{k} \in [0, \dots, M_F - 1] \times [0, \dots, N_F - 1]$. \mathcal{F} denotes the 2-D DFT operator. The approximation error vanishes as M_F, N_F grow to infinity. Choosing M_F and N_F as integer powers of 2 allows to use 2-D fast Fourier transform. The gradient and the Hessian of $\ell(x; h)$ are given in the Appendix.

A. Equivariance

A remarkable property of the QML estimator of the restoration kernel is its *equivariance*, which implies that for any invertible kernel b , the estimator $\tilde{w}(x)$ of w given the observation x , obtained by minimization of the target function $\ell(x; w)$ obeys [21], [22]

$$\tilde{w}(b * x) = b^{-1} * \tilde{w}(x)$$

where b^{-1} denotes the inverse of b . In other words, the parameters to be estimated (in our case, coefficients $w_{\mathbf{n}}$ of the invertible restoration kernel) form a group with the convolution operation. It must be noted, however, that when the restoration filter is constrained to have a fixed finite support, equivariance holds only approximately.

B. Choice of $\varphi(s)$

A QML blind deconvolution algorithm consists of two essential parts: the choice of the nonlinear function $\varphi(s)$ and the numerical algorithm for minimization of $\ell(x; w)$. The choice of $\varphi(s)$ depends on the type of sources appearing in the problem. Natural images encountered in most applications are usually characterized by nonlog-concave, multimodal distributions, which are difficult to model and are not well-suited for optimization.¹

However, consistent estimator of s can be obtained by minimizing $\ell(x; w)$ even when $\varphi(s)$ is not exactly equal to $-\log p_s(s)$. Such QML estimation has been shown to be practical in instantaneous BSS [15], [18], [23], [24] and BD of 1-D signals [19], [20]. For example, when the source is sparse or sparsely representable, a smooth approximation of the absolute value function is a good choice for $\varphi(s)$ [20], [24], [25]. Although natural images are usually far from being sparse, they can be sparsely represented by a proper transformation [18], [26] (in Section IV, we will show how to transform general classes of natural images into sparse ones). We, therefore, focus our attention on modeling distributions of sparse signals using a family of convex smooth functions

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

where λ is a positive smoothing parameter [16]; $\varphi_{\lambda}(s) \rightarrow |s|$ as $\lambda \rightarrow 0^+$. For convenience, we, henceforth, omit λ from our

¹Non-log-concave distributions lead to nonconvex prior term in the minus log-likelihood function.

notation whenever possible, and refer to $\varphi_\lambda(s)$ without using the subscript.

Another important advantage of working with sparsely represented sources is the fact that the asymptotic restoration error is significantly smaller compared to the original ones, i.e., the deconvolution is more accurate. This issue is addressed in [21]. Moreover, sparse representation reduces the statistical dependence between the signal samples. Thus, natural images, which usually have high correlation between neighbor pixels, become less dependent and better fit the i.i.d. source assumption.

III. RELATIVE NEWTON ALGORITHM

In their fundamental work, Amari *et al.* [27] introduced an iterative algorithm for 1-D BD based on the natural gradient learning, which was originally used in the context of blind source separation (BSS) [22], [28], [29]. A more general relative optimization approach was introduced in [16], again, in the context of BSS. This method was extended in [20] to BD of time series. Here, we extend these results to BD of images.

A. Relative Optimization Algorithm

The main idea of relative optimization is to iteratively produce source estimate and use it as the current observation. This yields the following algorithm.

Relative Optimization Algorithm

- 1) Start with $w^{(0)}$, and with $x^{(0)} = x$.
- 2) **For** $k = 0, 1, 2, \dots$, until convergence:
 - a) Start with: $w = \delta$.
 - b) Using an unconstrained optimization method, find w^{k+1} such that $\ell(x^{(k)}; w^{(k+1)}) < \ell(x^{(k)}; \delta)$.
 - c) Update the source estimate $x^{(k+1)} = w^{(k+1)} * x^{(k)}$.
- 3) **End loop.**

The restoration kernel estimate at k -th iteration is

$$w = w^{(0)} * w^{(1)} * \dots * w^{(k)} \quad (5)$$

and the source estimate is $\tilde{s} = x^{(k)}$. The method allows to construct large restoration kernels using a set of low-order factors. The algorithm assumes infinite memory and produces a restoration kernel of order growing at each iteration. In real applications, it might be necessary to limit the support of the restoration kernel. This can be done by cropping w after each update. A remarkable property of the relative optimization algorithm is its *uniform performance*, i.e., its step at iteration k depends only on $g^{(k-1)} = a * w^{(0)} * \dots * w^{(k-1)}$.

Unfortunately, even when a convex $\varphi(s)$, such as the smoothed absolute value is used, the log-spectrum term ℓ_1 is nonconvex, and, consequently, the QML function is nonconvex (e.g., if w^* is a minimizer of $\ell(x; w)$, then $-w^*$ is also a minimizer). Yet, our observations show that the QML function is well behaving and when the relative update is used, the minimization algorithm does not converge to “wrong” local minima, especially when the restoration kernel stability is enforced. Similar behavior of QML estimators is known in the context of BSS as well [15], [16]. This remarkable fact deserves additional research.

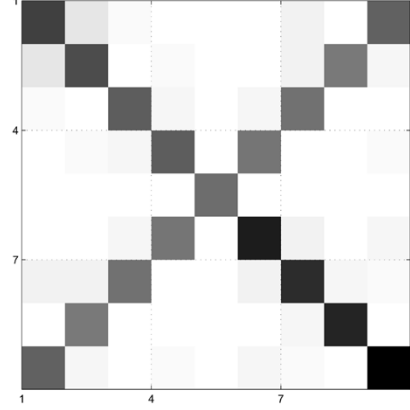


Fig. 1. Hessian structure for $w = \delta$ with $M = N = 1$ (3×3 kernel). White stands for near-zero elements of the matrix.

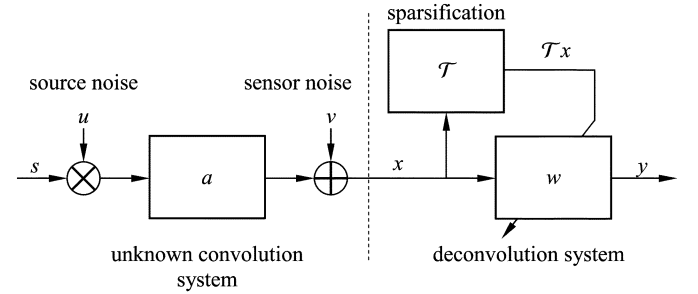


Fig. 2. Scheme of blind deconvolution using sparsification.

B. Newton Method

Newton method is often used for unconstrained optimization, since it provides a very fast (quadratic) rate of convergence. In this approach, the direction d at each iteration is given by solution of the linear system

$$\nabla^2 \ell \cdot d = -\nabla \ell. \quad (6)$$

Since the objective function is nonconvex, in order to guarantee descent direction, positive definiteness of the Hessian is forced by using modified Cholesky factorization [30]. Having the direction d , the new iterate $w^{(k+1)}$ is given by

$$w^{(k+1)} = w^{(k)} + \mu^{(k)} d^{(k)}$$

where $\mu^{(k)}$ is the step size determined, e.g., by backtracking line search restricted to the set where the inverse of the restoration kernel is stable, i.e., $\log |\det \mathcal{W}| > -\infty$ [21].

The Newton method is used in Step 4 of the relative optimization algorithm [16], [20]; such an optimization algorithm will be referred to as the *relative Newton* method. Apart from gradient and Hessian computation, required at each relative Newton iteration, additional $(1/3)K_M^3 K_N^3 + K_M^2 K_N^2$ operations are required for solution of the Newton system (6) using modified Cholesky factorization [30].

C. Fast Relative Newton Step

Practical use of the relative Newton step is limited to cases of small M, N, M_X, N_X , due to the complexity of Hessian construction and solution of the Newton system. This complexity

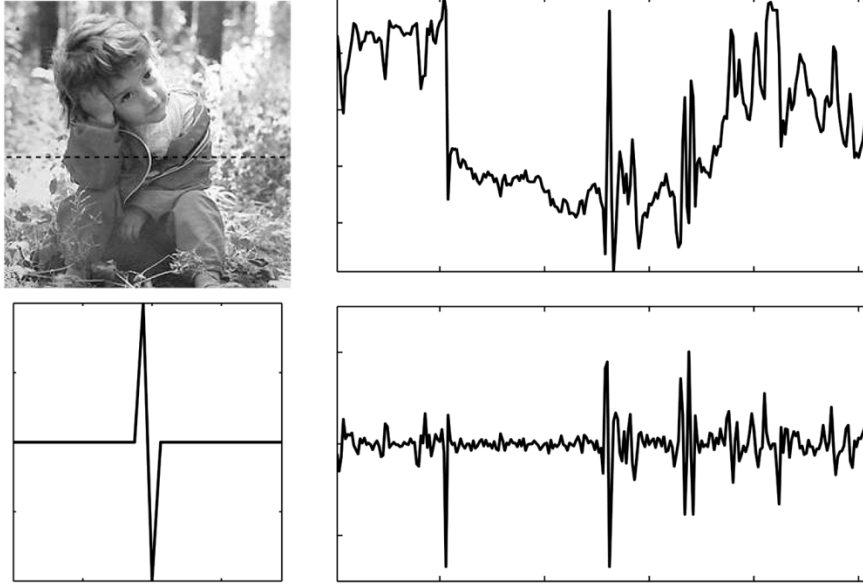


Fig. 3. One-dimensional example of optimal sparsification. First row, left: image; right: a 1-D signal (line 140 from the image). Second row, left: optimal sparsifying kernel; right: sparsified signal.

can be significantly reduced if special Hessian structure at the minimum is exploited.

Substituting $w = \delta$ and $x = cs$ to the Hessian (14) from the Appendix yields $\hat{w} \equiv 1$ and $y = cs$, from where

$$\frac{\partial^2 \ell}{\partial w_i \partial w_j} = \delta_{i+j} + \frac{1}{M_F N_F} \sum_{\mathbf{n}} \varphi''(cs_{\mathbf{n}}) cs_{\mathbf{n}-i} cs_{\mathbf{n}-j}.$$

It can be seen that the first term contributes a unit anti-diagonal to the Hessian. For sufficiently large sample size $M_X \times N_X$, the second term becomes approximately

$$\frac{\partial^2 \ell_2}{\partial w_i \partial w_j} \approx \mathbf{E} \varphi''(cs_{\mathbf{n}}) cs_{\mathbf{n}-i} cs_{\mathbf{n}-j}.$$

Let us assume that $s_{\mathbf{n}}$ is zero mean. Since s is i.i.d.

$$\frac{\partial^2 \ell_2}{\partial w_i \partial w_j} \approx \begin{cases} \alpha c^2, & i = j = 0 \\ \gamma \sigma'^2, & i = j \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\alpha = \mathbf{E} \varphi''(cs)(cs)^2$, $\gamma = \mathbf{E} \varphi''(cs)$, $\sigma^2 = \mathbf{E} s^2$, and $\sigma' = cs$. We conclude that $\nabla^2 \ell(cx; \delta)$ has an approximate diagonal-anti-diagonal form (see Fig. 1). When $\gamma \sigma'^2 \gg 1$, $\nabla^2 \ell(cx; \delta)$ is approximately diagonal. When $\gamma^2 \sigma'^4 < 1$, the Hessian at the solution point is not positive-definite, which means that the QML estimator is *asymptotically unstable*. This issue is addressed in depth in [21].

The use of the diagonal-anti-diagonal approximation of the Hessian, allows to reduce Newton system solution to regularized solution of a set of 2×2 systems of the form

$$\begin{pmatrix} (\nabla^2 \ell)_{kk} & 1 \\ 1 & (\nabla^2 \ell)_{K-k, K-k} \end{pmatrix} \begin{pmatrix} d_k \\ d_{K-k} \end{pmatrix} = \begin{pmatrix} -(\nabla \ell)_k \\ -(\nabla \ell)_{K-k} \end{pmatrix} \quad (7)$$

and an additional 1×1 system

$$(\nabla^2 \ell)_{K/2} \cdot d_{K/2} = -(\nabla \ell)_{K/2}$$

where $K = K_M + K_N + 1$. Regularization is performed by forcing positive definiteness of each of the 2×2 submatrices in (7) by inverting the sign of negative eigenvalues and forcing small eigenvalues to be larger than some positive threshold. When $\gamma \sigma'^2 \gg 1$, the main diagonal is dominant and, hence, the diagonal approximation can be used. Consequently, the Newton system (6) can be solved as a set of $K_M K_N$ independent linear equations.

The use of the diagonal-anti-diagonal approximation in a relative Newton iteration requires about $M_X N_X + 4M_X N_X \log_2 M_X N_X$ operations for approximate Hessian construction, which is of the same order as gradient computation. Additional $\mathcal{O}(K_M K_N)$ operations are required for approximate Hessian inversion in the case of diagonal approximation, and slightly more in the case of the diagonal-anti-diagonal approximation. This is compared to $\mathcal{O}(M_X N_X + K_M K_N [4M_X N_X \log_2 M_X N_X + M_X N_X])$ operations for exact Hessian evaluation and additional $(1/3)(K_M K_N)^3 + (K_M K_N)^2$ computations for exact Newton system solution required for the full relative Newton step.

IV. OPTIMAL SPARSE REPRESENTATIONS OF IMAGES

The QML framework presented in Section II is valid for sparse sources; this type of a prior source distribution is especially convenient for the underlying optimization problem due to its convexity and results in very accurate deconvolution. However, natural images arising in the majority of BD applications can by no means be considered to be sparse in their native representation (usually, they are sub-Gaussian).

Nevertheless, it is possible to transform an image in such a way that it fits some universal prior. In this study, we limit our attention to the sparsity prior and, thus, discuss sparsifying transformations, though the idea is general and is suitable for other priors as well. The idea of *sparse representations* has been successfully exploited in signal and image processing. In [31],

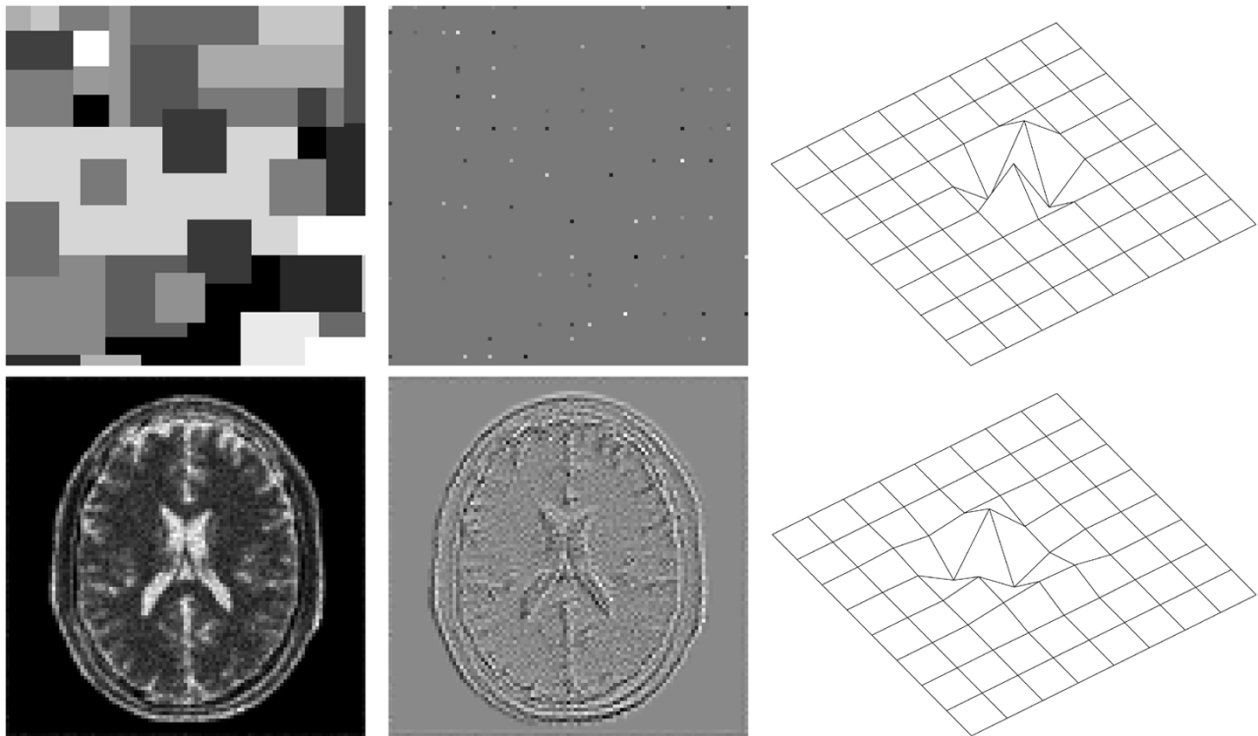


Fig. 4. Optimal sparsification of images. First column: Original image. Second column: Sparsified image. Third column: Optimal sparsifying kernel. First row: Synthetic block image. Second row: Brain image with Poisson noise.

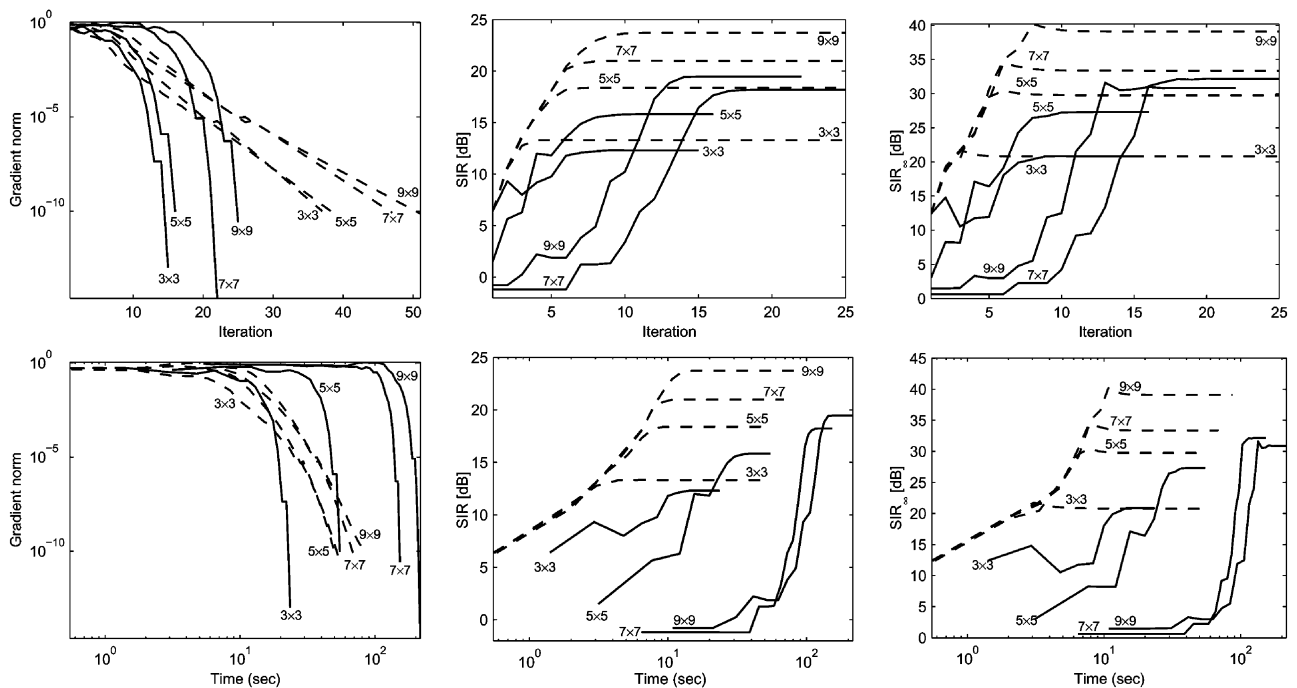


Fig. 5. Convergence of the (solid) Newton and of the fast relative (dashed) Newton methods, for various sizes of the restoration kernel versus (first row) the number of iterations and (second row) the CPU time.

Field and Olshausen proposed learning optimal bases for efficient sparse representation of signals and images. Sparsification was also proved efficient in BSS [18], [23], [26]. In [26], it was shown in that even such simple sparsifying transformation as a discrete derivative can be useful in BSS. Here, we present a criterion for finding optimal sparsifying transformations for BD of images.

Let us assume that there exists a *sparsifying transformation* $\mathcal{T}_{[s]}$, which makes the source s sparse (wherever possible, the subscript s in $\mathcal{T}_{[s]}$ will be omitted for brevity). In this case, our algorithm is likely to produce a good estimate of the restoration kernel w since the source properties are in accordance with the sparsity prior. The problem is, however, that in the BD setting, s is not available, and \mathcal{T} can be applied only to the observation x .

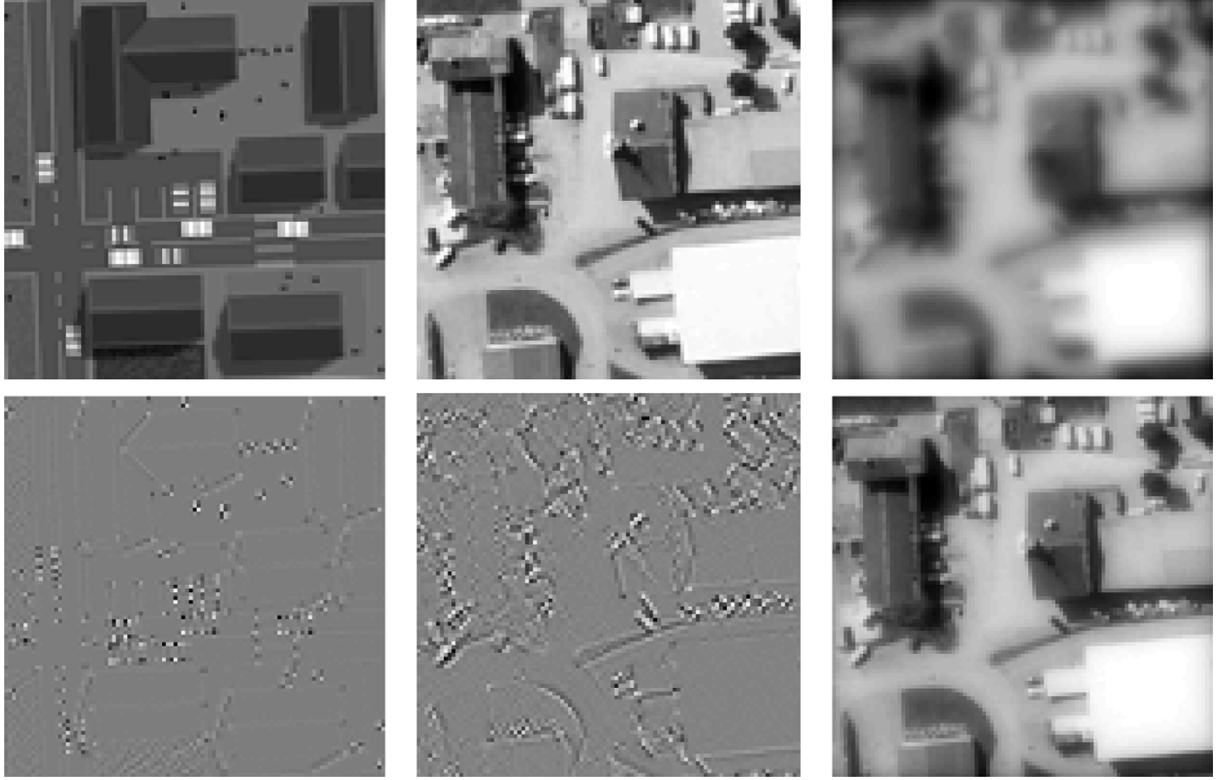


Fig. 6. First row, left to right: training synthetic image; source aerial image s ; blurred image $s * w$. Second row, left to right: sparsified training image; sparsified source; restored image.

Hence, it is necessary that the sparsifying transformation commute with the convolution operation, i.e.

$$(\mathcal{T}s) * a = \mathcal{T}(s * a) = \mathcal{T}x \quad (8)$$

such that applying \mathcal{T} to x is equivalent to applying it to s . Obviously, \mathcal{T} must be a shift-invariant (SI) transformation.²

In this study, we limit our attention to linear shift-invariant (LSI) transformations, i.e., \mathcal{T} that can be represented by convolution with a *sparsifying kernel*

$$\mathcal{T}s = t * s. \quad (9)$$

Thus, we obtain a general BD algorithm, which is not limited to sparse sources. We first sparsify the observation data x by convolving it with t (which has to be found in a way described below), and then apply the sparse BD algorithm on the result $\tilde{x} * t$. The obtained restoration kernel w is then applied to x to produce the source estimate (see Fig. 2).

A. Learning the Sparsifying Kernel

For a general source s , which does not fit the prior, $\tilde{w}(x) \neq ca^{-1}$; yet, we would like to find such a sparsifying kernel t , that $\tilde{w}(t * x) = ca^{-1}$. Due to the equivariance property

$$\begin{aligned} \tilde{w}(t * x) &= \operatorname{argmin}_w \ell(t * a * s; w) \\ &= a^{-1} * \operatorname{argmin}_w \ell(t * s; w) = a^{-1} * \tilde{w}(t * s). \end{aligned}$$

²In BSS problems, the sparsifying transformation needs to be linear and not necessarily shift-invariant, e.g., wavelet packets were used for sparsification in [18], [23].

TABLE I
SIR, SIR_∞, AND PSIR OF THE OBSERVED IMAGES

Source		SIR [dB]	SIR _∞ [dB]	PSIR [dB]
s_1	Susy	-1.46	7.84	22.83
s_2	Aerial	-1.46	7.84	19.89
s_3	Gabby	4.90	11.55	30.41
s_4	Hubble	3.40	10.65	27.18

Hence, we demand $\tilde{w}(t * s) = c\delta$. Assuming t is invertible and invoking again the equivariance property, we obtain

$$c\delta = \tilde{w}(t * s) = t^{-1} * \tilde{w}(s)$$

from where omitting the scaling factor c

$$t = \tilde{w}(s) = \operatorname{argmin}_w \ell(s; w). \quad (10)$$

In other words, the optimal sparsifying kernel $t_{[s]}$ is obtained by applying the QML BD algorithm to the source image s . We note that although such a formulation does not allow us to obtain noninvertible sparsifying kernels (such as discrete derivatives), it is still able to produce an invertible approximation of such kernels. The kernel t does not necessarily have to be stable or invertible, since it is used as a preprocessing of the data, and, hence, there is never a need to invert it.

Figs. 3 and 4 (first row) show examples of optimal sparsifying transformations of 1-D and 2-D signals. In the 1-D case, a row from a natural image was taken; the optimal sparsifying kernel



Fig. 7. First column: Source images. Second column: Blurred images. Third column: Restoration results using the QML deconvolution approach.

is approximately a discrete derivative. In the 2-D case of a block signal, as expected intuitively, the optimal sparsifying kernel is approximately a corner detector.

Since the source image s is not available, computation of the sparsifying kernel by the above procedure is possible only the-

oretically. However, empirical results indicate that, for images belonging to the same class, the proper sparsifying kernels are sufficiently similar. Let \mathcal{C} denote a class of images (e.g., aerial photos) and assume that the unknown source s belongs to \mathcal{C} . We can take images $s^{(1)}, s^{(2)}, \dots, s^{(N_T)} \in \mathcal{C}$ and use them to find

the optimal sparsifying kernel of s . Optimization problem (10) becomes, in this case

$$t = \underset{t}{\operatorname{argmin}} \left\{ \ell_1(t) + \frac{1}{N_T} \sum_{n=1}^{N_T} \ell_2(s^{(n)} * t) \right\} \quad (11)$$

i.e., t is required to be the optimal sparsifying kernel for all $s^{(1)}, s^{(2)}, \dots, s^{(N_T)}$ simultaneously. The images $s^{(1)}, s^{(2)}, \dots, s^{(N_T)}$ constitute a *training set* and the process of finding such t as *training*. Given that the images in the training set are “sufficiently similar” to t , the optimal sparsifying kernel obtained from (11) is similar enough to $t_{[s]}$.

Being similar to the deconvolution problem, the training problem (11) can be solved efficiently using the relative Newton method presented beforehand. In addition to computational efficiency, the use of relative optimization forms the sparsifying kernel of several low-order kernels, which has a regularizing effect.

In the most general case, the sparsifying kernel can be *complex*: $t = t_R + t_I$. After sparsification with such t , the prior term ℓ_2 of the likelihood function, in which the real absolute value is replaced by the complex one, becomes

$$\sum_{\mathbf{n}} |(Ty)\mathbf{n}| = \sum_{\mathbf{n}} \sqrt{(t_R * y)_{\mathbf{n}}^2 + (t_I * y)_{\mathbf{n}}^2}. \quad (12)$$

The latter is a generalization of the 2-D *total-variation* (TV) norm. The TV norm, which has been found to be a successful prior in numerous studies related to signal restoration and denoising [32]–[34], and was also used by Chan and Wong as a regularization in BD [35], is obtained when t_R, t_I are chosen to be discrete x - and y -directional derivatives.

The proposed optimal sparsification can be also viewed as construction of optimal problem-specific prior of the form

$$\sum_{\mathbf{n}} \varphi(Ty\mathbf{n}) \quad (13)$$

parameterized by \mathcal{T} . In addition to optimizing over \mathcal{T} , one can also find the optimal nonlinear function $\varphi(s)$. Such an approach is related to optimal regularization of MAP-type inverse problems using supervised learning, proposed by Haber and Tenorio [36].

B. Sparsification in the Presence of Noise

In the case of *source noise* (for example, arising due to a low number of emitted photons in an imaging setup), the training can be performed on training images, degraded by the expected level of noise. Also, the function $\varphi(s)$ can be modified in order to account for the noise distribution. For a detailed discussion, see [21]. Fig. 4 (second row) depicts the optimal sparsification of a medical image contaminated by Poisson noise. The optimal

TABLE II
SIR, SIR_{∞} , AND PSIR OF THE RESTORED IMAGES

Source		SIR [dB]	SIR_{∞} [dB]	PSIR [dB]
s_1	Susy	17.80	22.21	27.23
s_2	Aerial	17.04	23.55	25.16
s_3	Gabby	19.32	23.81	40.64
s_4	Hubble	14.51	17.16	33.67

sparsifying kernel (constructed using the relative Newton algorithm), in this case, is a smoothed corner detector.

The problem of *sensor noise* is more complicated. Although the relative Newton algorithm was found to be robust to noise levels of up to about 10 dB [20], the use of sparsification kernels (which are typically high-pass filters, since, by their nature, sparse signals have high-frequency components) is liable to lower the SNR. To cope with this problem, the signal should be smoothed with a low-pass filter f and, afterward, the sparsifying kernel t should be applied. Due to commutativity of the convolution, it is equivalent to carrying out the sparsification with a smoothed kernel $t * f$.

V. SIMULATIONS

The presented approach was tested in three experiments under zero-noise conditions. In the first experiment, the goal was to compare between the performance of fast relative Newton and full relative Newton algorithms. The purpose of the second experiment was to demonstrate the utility of the training approach for finding optimal sparse representations. In the third experiment, we used the sparsification approach to perform deconvolution of natural images. The fourth experiment demonstrates the performance of our method in the presence of Poisson source noise. As a criterion for evaluation of the reconstruction quality, we used the signal-to-interference-ratio (SIR) in sense of the L_2, L_{∞} norms, and the peak SIR (PSIR) in decibel units [21].

A. Deconvolution of Sparse Images

A 101×101 sparse normally distributed image, generated by the MATLAB function `sprandn` with 20% density, was used as the source in the first experiment. The image was convolved with a 3×3 FIR kernel with a slowly decaying inverse (see Fig. 5). Full Newton and fast relative Newton (with a diagonal Hessian approximation) were used to estimate the inverse kernel. $3 \times 3, 5 \times 5, 7 \times 7,$ and 9×9 restoration kernels were used. The smoothing parameter was set to $\lambda = 10^{-2}$. Optimization was terminated when $\|\nabla \ell\|$ reached 10^{-10} . Gradient norms, SIR, and SIR_{∞} were measured as a function of CPU time³ and iteration number.

The experiments demonstrate convergence of both algorithms (Fig. 5). The fast relative Newton converged about ten times faster in terms of SIR, compared with the full Newton algorithm, and provided better restoration by about 2–5 dB for the same

³All algorithms were implemented in MATLAB and executed on an ASUS portable computer with Intel Pentium IV Mobile processor and 640MB RAM.

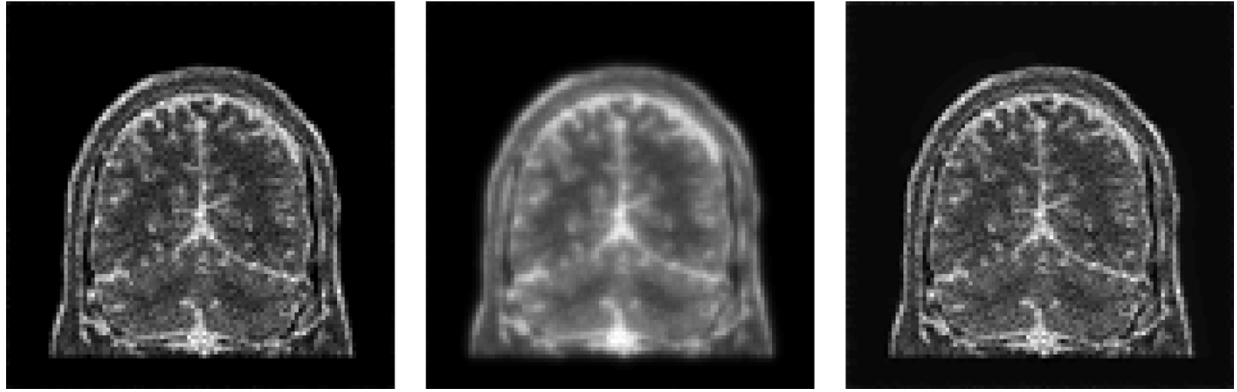


Fig. 8. Left to right: source image contaminated by Poisson noise; blurred noisy source image; deconvolution results.

values of M and N . This improvement is since the effective restoration kernel was of higher order.

B. Training

In the second experiment, a real aerial photo of a factory was used as the source image, and a synthetic one (drawn using PhotoShop) as the training image (Fig. 6). A 3×3 sparsifying kernel is found by training on a single image, then the same kernel is used as a preprocessing for BD applied to a different blurred source image from the same class of images. The source image was convolved with a symmetric FIR 31×31 Lorentzian-shaped blurring kernel. Deconvolution kernel was of size 3×3 . The sparsifying kernel obtained by training was very close to a corner detector. The signal-to-interference ratio in the deconvolution result was SIR= 20.16 dB and $\text{SIR}_\infty = 25.72$ dB.

C. Deconvolution of Natural Images

In the third experiment, four natural source images were used: s_1 (Susy), s_2 (Aerial), s_3 (Gabby), and s_4 (Hubble). Nearly stable Lorentzian-shaped kernels were applied to the corresponding sources. This type of kernels characterizes scattering media, such as biological fluids and aerosols found in the atmosphere [37]. Quality of the degraded images in terms of SIR, SIR_∞ , and PSIR is presented in Table I.

Fast relative Newton step with kernel size set to 3×3 was used in this experiment. The smoothing parameter was set to $\lambda = 10^{-2}$. Corner detector was used as the sparsifying kernel. Optimization was terminated when the gradient norm reached 10^{-10} . Convergence was achieved in 10–20 iterations (about 10 s). The images are depicted in Fig. 7. Restoration quality results in terms of SIR, SIR_∞ , and PSIR are presented in Table II. The use of the total variation prior in the place of ℓ_2 with the optimal sparsifying kernel yielded relatively poor results.

D. Deconvolution in Presence of Source Noise

In the fourth experiment, we performed deconvolution of a source contaminated by multiplicative Poisson noise with maximum SNR of 10 dB. Such a type of noise arises in emission processes (e.g., emission tomography and optical imaging with

low-photon yield). A brain image was used as the source. The noisy source was convolved with a Lorentzian-shaped blurring kernel, resulting in SIR= 3.99 dB, $\text{SIR}_\infty = 11.50$ dB, and SIR= 23.43 dB in the observed image.

The sparsification kernel was obtained by training on a different brain image (see Fig. 4, second row) and had the form of a smoothed corner detector. The restoration resulted in SIR= 12.44 dB, SIR= 16.50 dB, and PSIR= 33.20 dB (Fig. 8).

VI. CONCLUSION

We showed that the QML framework, recently presented in the context of 1-D deconvolution is also useful for BD of images. We presented an extension of the relative optimization approach to QML BD in the 2-D case and studied the relative Newton method as its special case. Simple Hessian structure obtained for near i.i.d. sparse sources allows to perform the relative Newton iteration with the complexity of a gradient descent iteration, leading to the fast relative Newton algorithm. We showed a supervised learning approach for finding optimal sparse representations, in order to yield a general-purpose BD method. A particular class of LSI sparsifying transformations generalizes some previous results such as the total variation prior.

Simulation results demonstrated the efficiency of the proposed method. Potential applications of our approach are in optics, remote sensing, microscopy and biomedical imaging, especially where the SNR is moderate. The method is especially efficient in problems involving slowly decaying (e.g., Lorentzian shaped) blurring kernels, which can be approximately inverted using a kernel with small support. Such kernels are typical of imaging through scattering media.

APPENDIX

GRADIENT AND HESSIAN OF $\ell(x; w)$

The relative Newton algorithm requires the knowledge of the gradient and the Hessian of $\ell(x; w)$. Since the optimization variable w is a $K_M \times K_N$ matrix, the gradient $\nabla \ell$ is also a $K_M \times K_N$ matrix, whereas the Hessian $\nabla^2 \ell$ is a $K_M \times K_N \times K_M \times$

K_N fourth-order tensor. For convenience, we parse the variables columnwise into a $K_M K_N \times 1$ vector

$$\text{vec}(w) = [w_{-M,-N}, \dots, w_{M,-N}, \dots, w_{M,N}]^T$$

and define the gradient and the Hessian of $\ell(x; w)$ as a $K_M K_N \times 1$ vector and a $K_M K_N \times K_M K_N$ matrix, respectively.

The elements of the gradient and the Hessian are given by

$$\frac{\partial \ell}{\partial w_{\mathbf{i}}} = -\mathcal{F}^{-1} \left\{ \hat{w}_{\mathbf{k}}^{-1} \right\}_{-\mathbf{i}} + \frac{1}{M_F N_F} \sum_{\mathbf{n}} \varphi'(y_{\mathbf{n}}) x_{\mathbf{n}-\mathbf{i}}$$

and

$$\begin{aligned} \frac{\partial^2 \ell}{\partial w_{\mathbf{i}} \partial w_{\mathbf{j}}} = & \mathcal{F}^{-1} \left\{ \hat{w}_{\mathbf{k}}^{-2} \right\}_{-(\mathbf{i}+\mathbf{j})} \\ & + \frac{1}{M_F N_F} \sum_{\mathbf{n}} \varphi''(y_{\mathbf{n}}) x_{\mathbf{n}-\mathbf{i}} x_{\mathbf{n}-\mathbf{j}} \end{aligned} \quad (14)$$

respectively. For derivation, see [21]. Computational complexity of the target function ℓ and its gradient is $\mathcal{O}(M_F N_F \log_2 M_F N_F + M_X N_X \log_2 M_X N_X)$, whereas evaluation of the Hessian requires $\mathcal{O}(M_F N_F \log_2 M_F N_F + M N M_X N_X \log_2 M_X N_X)$ operations, assuming efficient implementation involving the FFT.

ACKNOWLEDGMENT

The authors would like to thank S. Pitzanti and G. Sbordone for permission to use their photographs as test images in simulations. They would also like to thank the anonymous referees for constructive comments, which helped them improve the paper.

REFERENCES

- [1] T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1064–1073, 1993.
- [2] M. Bertero and P. Boccacci, "Image restoration methods for the large binocular telescope," *Astron. Astrophys. Suppl. Ser.*, vol. 147, pp. 323–333, 2000.
- [3] J. P. Muller, Ed., *Digital Image Processing in Remote Sensing*. Philadelphia, PA: Taylor & Francis, 1988.
- [4] M. Mignotte and J. Meunier, "Three-dimensional blind deconvolution of SPECT images," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 2, pp. 274–280, Feb. 2000.
- [5] D. Adam and O. Michailovich, "Blind deconvolution of ultrasound sequences using nonparametric local polynomial estimates of the pulse," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 2, pp. 118–131, Feb. 2002.
- [6] T. Wilson and S. J. Hewlett, "Imaging strategies in threedimensional confocal microscopy," in *Proc. SPIE 1245*, 1991, pp. 35–45.
- [7] T. J. Holmes, S. Bhattacharyya, J. A. Cooper, D. Hanzel, V. Krishnamurthy, W. Lin, B. Roysam, D. H. Szarowski, and J. N. Turner, "Light microscopic images reconstructed by maximum likelihood deconvolution," in *Handbook of Biological and Confocal Microscopy*, 2nd ed., J. B. Pawley, Ed. New York: Plenum, 1995.
- [8] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, no. 5, pp. 43–64, May 1996.
- [9] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 1, pp. 58–63, Feb. 1976.

- [10] M. M. Chang, A. M. Tekalp, and A. T. Erdem, "Blur identification using the bispectrum," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2323–2325, Oct. 1991.
- [11] R. Fabian and D. Malah, "Robust identification of motion and outoffocus blur parameters from blurred and noisy images," *CVGIP: Graph. Models Image Process.*, vol. 53, no. 5, pp. 403–412, Jul. 1991.
- [12] E. Thiébaud and J.-M. Conan, "Strict a priori constraints for maximum likelihood blind deconvolution," *J. Opt. Soc. Amer. A*, vol. 12, no. 3, pp. 485–492, 1995.
- [13] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," Tech. Rep., 1996.
- [14] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neur. Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [15] D. Pham and P. Garrat, "Blind separation of a mixture of independent sources through a quasimaximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.
- [16] M. Zibulevsky, "Blind source separation with relative Newton method," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, Nara, Japan, Apr. 2003, pp. 897–902.
- [17] M. Joho and P. Schniter, "Frequency domain realization of a multi-channel blind deconvolution algorithm based on the natural gradient," presented at the 4th Int. Symp. Independent Component Analysis and Blind Signal Separation, Apr. 2003.
- [18] P. Kisilev, M. Zibulevsky, and Y. Zeevi, "A multiscale framework for blind separation of linearly mixed signals," *JMLR*, vol. 4, pp. 1339–1363, 2003.
- [19] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. SPAWC*, Apr. 1997, pp. 101–104.
- [20] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky. (2003) Blind deconvolution with relative newton method. CCIT, Technion, Israel. [Online]. Available: <http://visl.technion.ac.il/bron/michael>
- [21] —, (2003) Quasi Maximum likelihood blind deconvolution of images using optimal sparse representations. CCIT, Technion, Israel. [Online]. Available: <http://visl.technion.ac.il/bron/michael>
- [22] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [23] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition," *Neur. Comput.*, to be published.
- [24] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, "Blind source separation by sparse decomposition," in *Independent Components Analysis: Principles and Practice*, S. J. Roberts and R. M. Everson, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [26] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Separation of reflections via sparse ICA," presented at the IEEE Int. Conf. Image Processing, 2003.
- [27] 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*, Jul. 1997, pp. 1057–1062.
- [28] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electron. Lett.*, vol. 30, no. 17, pp. 1386–1387, 1994.
- [29] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neur. Inf. Process. Syst.*, vol. 8, pp. 757–763, 1996.
- [30] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
- [31] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vis. Res.*, vol. 37, pp. 3311–3325, 1997.
- [32] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.
- [33] P. Blomgren, T. F. Chan, P. Mulet, and C. Wong, "Total variation image restoration: Numerical methods and extensions," presented at the IEEE Int. Conf. Image Processing, 1997.
- [34] T. F. Chan and P. Mulet, "Iterative methods for total variation image restoration," *SIAM J. Numer. Anal.*, to be published.
- [35] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," presented at the ONR Workshop, 1996.
- [36] E. Haber and L. Tenorio, "Learning regularization functionals," *Inv. Probl.*, vol. 19, pp. 611–626, 2003.
- [37] M. Moscoso, J. B. Keller, and G. Papanicolaou, "Depolarization and blurring of optical images by biological tissue," *J. Opt. Soc. Amer. A*, vol. 18, no. 4, pp. 948–960, 2001.



Michael M. Bronstein (S'02) was born in Russia in 1980. He received the B.Sc. degree in electrical engineering (summa cum laude) and the M.Sc. degree in computer science (summa cum laude) from the Technion—Israel Institute of Technology, Haifa, in 2002 and 2005, respectively. He is currently pursuing the Ph.D. degree in computer science at the Technion—Israel Institute of Technology.

As part of his thesis work, he is involved in the 3DFACE research project, which was featured on CNN. He is the co-inventor of two patents (pending)

and 20 publications in leading scientific journals and conferences. His main research interests include three-dimensional face recognition and blind inverse problems.

Mr. Bronstein is a recipient of the Technion Humanities and Arts Department prize (2001), the Kasher prize (2002), the Thomas Schwartz award (2002), the Hershel Rich Technion Innovation award (2003), the Counter Terrorism prize (2003), and the Copper Mountain Conference on Multigrid Methods Best Paper Award (2005). He was an honorary student delegate to the International Achievement Summit of the Academy of Achievement, Washington, DC, in May 2003.



Alexander M. Bronstein (S'02) was born in Russia in 1980. He received the B.Sc. and M.Sc. degrees (both summa cum laude) from the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, in 2002 and 2005, respectively. He is currently pursuing the Ph.D. degree in the Department of Computer Science, Technion—Israel Institute of Technology.

He is involved in the 3DFACE research project, which was featured on CNN. He is the co-inventor of two patents (pending) and 20 publications in leading

scientific journals and conferences. His main research interests include blind inverse problems and three-dimensional face recognition.

He received the Technion Humanities and Arts Department prize (2001), the Kasher prize (2002), the Thomas Schwartz award (2002), the Hershel Rich Technion Innovation award (2003), the Counter Terrorism prize (2003), and the Copper Mountain Conference on Multigrid Methods Best Paper Award (2005). He was an honorary student delegate to the International Achievement Summit of the Academy of Achievement, Washington, DC, in May 2003.



Michael Zibulevsky received the M.Sc. degree in electrical engineering (summa cum laude) from Moscow State University of Railway Engineering (MIIT), Moscow, Russia, and the Ph.D. degree in operation research (nonlinear optimization) from the Technion—Israel Institute of Technology, Haifa.

He is currently with the Department of Electrical Engineering, Technion—Israel Institute of Technology. His research interests include nonlinear optimization and inverse problems in signal/image reconstruction (blind source separation, deconvolu-

tion, tomography, and EEG/MEG).



Yehoshua Y. Zeevi received the Ph.D. degree from the University of California, Berkeley, where he was also a Visiting Scientist at Lawrence Berkeley Laboratory.

He is the Barbara and Norman Seiden Professor of Computer Sciences in the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, where he is also the founder of the Jacobs Center for Communication and Information Technologies (CCIT) and the Ollendorff Minerva Center, and where he served as the Head of

these centers. He was also the Dean of the Faculty of Electrical Engineering from 1994 to 1999. He was a Vinton Hayes Fellow at Harvard University, Cambridge, MA, and has been a regular visitor there. He was also a Visiting Professor at the Massachusetts Institute of Technology, Cambridge, and Rutgers University, New Brunswick, NJ, and a Senior Visiting Scientist at the NTT Research Center, Yokoska, Japan. He is presently a Visiting Professor at Columbia University, New York. He is the co-inventor of many patents and the author of more than 200 publications related to vision and image sciences. He is the Editor-in-Chief of the *Journal of Visual Communication and Image Representation*, published by Academic Press, and the editor of three books. He is also one of the founders of the Adaptive Sensitivity camera that mimics the eye, as well as one of the founders of i Sight, Inc.—a company that developed intelligent cameras.

Dr. Zeevi is a Fellow of the SPIE and the Rodin Academy.