

Relative Optimization for Blind Deconvolution

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

Abstract—We propose a relative optimization framework for quasi maximum likelihood (QML) blind deconvolution and the relative Newton method as its particular instance. Special Hessian structure allows fast Newton system construction and solution, resulting in a fast-convergent algorithm with iteration complexity comparable to that of gradient methods. We also propose the use of rational IIR restoration kernels, which constitute a richer family of filters than the traditionally used FIR kernels. We discuss different choices of non-linear functions suitable for deconvolution of super- and sub-Gaussian sources, and formulate the conditions, under which the QML estimation is stable. Simulation results demonstrate the efficiency of the proposed methods.

Index Terms—blind deconvolution, Newton method, natural gradient, maximum likelihood.

I. INTRODUCTION

BLIND deconvolution (BD) appears in various applications related to acoustics, optics, geophysics, communications, control, etc. In communications, the term *blind channel equalization* is more common, as the main interest lies in retrieving the data transmitted over a dispersive communication channel [1]–[4]. In control, BD is usually known as *blind identification*, since the main goal is to obtain a model of the system [5]–[7], whereas in acoustics, optics and geophysics the term *blind deconvolution* is more adequate, since the goal is to “undo” the influence of a system by finding its stable inverse.

In the general setup of the single-channel BD, the observed sensor signal x is created from the *source signal* s passing through a causal convolutive system

$$x_n = \sum_{k=0}^{\infty} a_k s_{n-k} + u_n, \quad (1)$$

with impulse response a and additive sensor noise u . The setup is termed *blind* if only x is available, whereas a , s and u are unknown. BD aims to find such a deconvolution (restoration) kernel w , that produces a possibly delayed waveform-preserving estimate of s :

$$\tilde{s}_n = \sum_{k=0}^{\infty} w_k x_{n-k} \approx c \cdot s_{n-\Delta}, \quad (2)$$

where c is an arbitrary scaling factor and Δ is an integer shift. Equivalently, the *global system response* should be

Manuscript received October 10, 2003; revised February 16, 2004; final revision June 18, 2004. This research has been supported by the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission, and by the Ollendorff Minerva Center.

A. M. Bronstein, M. M. Bronstein and M. Zibulevsky are with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: alexbron@iee.org; bronstein@iee.org; mzub@ee.technion.ac.il).

approximately a Krönecker delta, up to scale factor and shift:

$$g_n = (a * w)_n \approx c \cdot \delta_{n-\Delta}. \quad (3)$$

A commonly used assumption is that s is non-Gaussian i.i.d.

The majority of BD methods described in literature focus on estimating the impulse response of the convolution system $A(z)$ from the observed signal x using a causal finite length (FIR) model and then determining the source signals from this estimate [6], [8]–[11]. Many of these methods use batch mode calculations and usually suffer from high computational complexity.

A wide class of the so-called *Bussgang* algorithms estimate directly the inverse kernel $W(z) = A^{-1}(z)$ by minimizing some cost function using gradient descent iterations. The cost function is usually based on high-order statistics or derived from maximum likelihood or information maximization considerations. These methods usually operate in the time domain and the gradient is usually derived by applying some non-linearity to the correlation of the observed signal and the estimated source. One of the most popular algorithms in this class is the constant modulus algorithm (CMA) proposed by Godard [3]. A review of these algorithms can be found in [12].

In their fundamental work, Amari *et al.* [13] introduced an iterative time-domain quasi maximum likelihood BD algorithm based on natural gradient learning, which was originally used in the context of blind source separation (BSS) [14]–[16] and became very attractive due to *uniform performance* [16], [17]. The natural gradient algorithm estimates directly the restoration kernel and allows real-time processing. In [18], a generalization of the algorithm for multichannel case was presented. Efficient frequency-domain implementations were derived in [19], [20].

The natural gradient algorithm demonstrates significantly higher performance compared to standard gradient descent. In this work, we present a BD algorithm based on the relative Newton method, which brings further acceleration. The relative Newton algorithm was originally proposed in the context of BSS in [21], [22]. We utilize special Hessian structure to derive a fast version of the algorithm with iteration complexity comparable to that of gradient methods. Our algorithm constructs an expanding restoration filter, permitting to estimate long restoration kernels using cheap iterations.

The main focus of this work is a batch mode single-channel BD algorithm with FIR restoration kernel; yet, we also outline the possibilities of block-wise processing and the use of IIR kernels, and discuss the extensions to the multichannel case. The smoothed absolute value is used as the nonlinearity for deconvolution of super-Gaussian sources, and the smoothed deadzone linear function is proposed for the sub-Gaussian ones.

II. QML BLIND DECONVOLUTION

Under the assumption that the restoration kernel $W(z)$ is strictly stable, and the source signal is real and i.i.d., the normalized minus-log-likelihood function of the observed signal x in the noise-free case is [13], [18], [23], [24]

$$\ell(x; w) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \log |W(e^{i\theta})| d\theta + \frac{1}{T} \sum_{n=0}^{T-1} \varphi(y_n), \quad (4)$$

where T is the sample size, $y = w * x$ is a source estimate, $\varphi(s) = -\log p(s)$, and $p(s)$ is the probability density function (PDF) of the source s . We assume that w is an FIR kernel supported on $n = -N, \dots, N$, and denote its length by $K = 2N + 1$. We will also assume that s is zero-mean. Cost function (4) can be also derived using negative joint entropy [13], [18] and information maximization [25] considerations.

Consistent estimator can be obtained by minimizing $\ell(x; w)$ even when $\varphi(s)$ is not exactly equal to $-\log p(s)$. Such *quasi* ML (QML) estimation has been shown to be practical in instantaneous BSS when the source PDF is unknown or not well-suited for optimization [21]. The choice of $\varphi(s)$ and the stability conditions of the QML estimator are discussed in Section IV.

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

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log |W(e^{i\theta})| d\theta \approx \frac{1}{N_F} \sum_{k=0}^{N_F-1} \log |W_k|, \quad (5)$$

where W_k are the DFT coefficients of w zero-padded to N_F . The approximation error vanishes as N_F grows to infinity. It is convenient to choose N_F to be an integer power of 2, to allow the use of FFT for efficient computation. For convenience, we will henceforth refer to the approximate target function as to $\ell(x; w)$.

The gradient of $\ell(x; w)$ w.r.t. w_i is given by

$$\frac{\partial \ell}{\partial w_i} = -q_{-i} + \frac{1}{T} \sum_{n=0}^{T-1} \varphi'(y_n) x_{n-i}, \quad (6)$$

where q_n is the inverse DFT of W_k^{-1} . The Hessian of $\ell(x; w)$ is given by

$$\frac{\partial^2 \ell}{\partial w_i \partial w_j} = r_{-(i+j)} + \frac{1}{T} \sum_{n=0}^{T-1} \varphi''(y_n) x_{n-i} x_{n-j}, \quad (7)$$

where r_n is the inverse DFT of W_k^{-2} (for derivation see [26]). Both the gradient and the Hessian can be evaluated efficiently using FFT.

III. RELATIVE OPTIMIZATION

Here we introduce a relative optimization framework for BD. The main idea of relative optimization is to iteratively produce source signal estimate and use it as the observed signal at the next iteration. Similar approach was explored in [22] in the context of BSS.

Relative optimization algorithm

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

The restoration kernel estimate at k -th iteration is $\hat{w} = w^{(0)} * \dots * w^{(k)}$, 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. An arbitrary restoration kernel can be constructed in this manner. In real application, it might be necessary to limit the filter length to some maximum order, which can be done by cropping w after each update. The relative optimization algorithm has uniform performance, i.e. its step at iteration k depends only on $g^{(k-1)} = a * w^{(0)} * \dots * w^{(k-1)}$, since the update in Step 5 does not depend explicitly on a , but on the current global system response only.

When the input signal is very long, it is reasonable to partition the input into blocks and estimate the restoration kernel for the current block using the data of the previous block and the previous restoration kernel estimate. We refer to this method as to the *block relative optimization algorithm*. It might also be useful in cases when the input signal x is produced as a result of s passing through a slowly varying convolution system.

A. Relative Newton method

A Newton iteration can be used in Step 4 of the relative optimization algorithm, yielding very fast convergence. In the standard Newton method (see e.g. [27], [28]), the descent direction d at each iteration is given by solution of the linear system

$$Hd = -g, \quad (8)$$

where $g = \nabla_w \ell$ and $H = \nabla_w^2 \ell$ are the gradient and Hessian of $\ell(x; w)$, respectively. In order to guarantee descent direction, positive definiteness of the Hessian is usually forced by using modified Cholesky factorization, which requires about $\frac{1}{3}K^3 + K^2$ operations [27]. Having the Newton direction $d^{(k)}$, the new iterate $w^{(k+1)}$ is given by

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

where α is the step size determined by either exact line search

$$\alpha^{(k)} = \operatorname{argmin} \ell(x; w^{(k)} + \alpha^{(k)} d^{(k)}),$$

or by backtracking line search [22], [27]–[29] (the latter was used in our implementation). The use of line search guarantees monotonic decrease of the objective function at every iteration. It should be noted that when the gradient norm becomes very small (say, below 10^{-5}), computational inaccuracies make the line search inefficient. For this reason, we used the Newton direction as is (i.e. chose $\alpha = 1$) when the gradient norm dropped below 10^{-5} . Relative optimization algorithm using the Newton step will be termed henceforth as the *relative Newton method*.

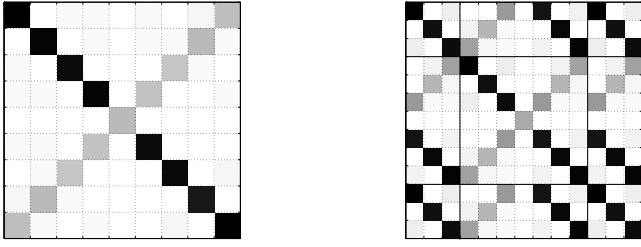


Fig. 1. Structure of the Hessian at the solution point for FIR restoration kernel with $N = 3$ (left) and IIR restoration kernel with $N = M = L = 3$ (right). White represents near-zero elements.

B. Fast relative Newton step

Practical use of the relative Newton step is limited to small values of N and T , 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. In the proximity of the solution point, $x^{(k)} \approx cs$, hence $\nabla^2 \ell(x; \delta)$ evaluated at each relative Newton iteration becomes approximately $\nabla^2 \ell(cs; \delta)$. For a sufficiently large sample size (in practice, $T > 10^2$), the following approximation holds:

Proposition 1: The Hessian $\ell(cs; \delta)$ has an approximate diagonal-anti-diagonal structure, with unit anti-diagonal.

Proof: Substituting $w = \delta$, $x = cs$ and $y = \delta * x = cs$ into $\ell(x; w)$ in (7), one obtains

$$\frac{\partial^2 \ell}{\partial w_i \partial w_j} = \delta_{i+j} + \frac{1}{T} \sum_{n=0}^{T-1} \varphi''(cs_n) cs_{n-i} cs_{n-j}.$$

For a large sample size T , the sum approaches the corresponding expectation value. Invoking the assumption that s is zero-mean i.i.d., one obtains

$$\begin{aligned} \frac{\partial^2 \ell}{\partial w_i \partial w_j} &\approx \delta_{i+j} + \mathbf{E} \{ \varphi''(cs_n) cs_{n-i} cs_{n-j} \} \\ &= \begin{cases} \mathbf{E} \varphi''(cs)(cs)^2 + 1 & : i = j = 0, \\ \mathbf{E} \varphi''(cs) \cdot \mathbf{E}(cs)^2 & : i = j \neq 0, \\ 1 & : i = -j \neq 0, \\ 0 & : \text{otherwise,} \end{cases} \end{aligned}$$

where \mathbf{E} denotes the expectation operator. ■

Typical Hessian structure is depicted in Figure 1 (left). Under this approximation, the Newton system (8) separates to K systems of linear equations of size 2×2

$$\begin{pmatrix} H_{-k,-k} & 1 \\ 1 & H_{kk} \end{pmatrix} \begin{pmatrix} d_{-k} \\ d_k \end{pmatrix} = - \begin{pmatrix} g_{-k} \\ g_k \end{pmatrix} \quad (9)$$

for $k = 1, \dots, K$, and an additional equation

$$H_{00} d_0 = -g_0. \quad (10)$$

In order to guarantee descent direction and avoid saddle points, we force positive definiteness of the Hessian by inverting the sign of negative eigenvalues λ_k^1, λ_k^2 in system (9) and forcing small eigenvalues to be above some positive threshold, say, $\epsilon = 10^{-8} \cdot \max \{ |\lambda_k^1|, |\lambda_k^2| \}$. Computation of the Hessian approximation involves evaluation of its main diagonal only,

which is of the same order of complexity as gradient computation. Approximate solution of the Newton system requires $\mathcal{O}(N)$ operations.

Due to the separable structure of the problem, the fast relative Newton step can be performed by updating each time a different triplet of coordinates w_{-k}, w_0, w_k in the vector of optimization variables w . This implies solution of a small problem with 3 optimization variables, using the fast relative Newton method. This idea can be simply generalized for more general symmetric blocks of coordinates of the vector w . A similar block-coordinate update has been successfully used in QML BSS [29].

IV. THE CHOICE OF $\varphi(s)$

The choice of $\varphi(s)$ is limited first of all by the QML estimator stability conditions. Here we address *asymptotic* stability only. In order to obtain an asymptotically stable estimator of the restoration kernel w , $w = \delta$ must be a strict minimum point of $\ell(cs; w)$ in the limit $T \rightarrow \infty$.

Proposition 2: The QML estimator of the restoration kernel w obtained by minimization of $\ell(x; w)$ is asymptotically stable if [30]

$$\mathbf{E} \varphi''(cs)(cs)^2 + 1 > 0 \quad (11)$$

$$\mathbf{E} \varphi''(cs) > 0 \quad (12)$$

$$\mathbf{E} \varphi''(cs) \cdot \mathbf{E}(cs)^2 > 1, \quad (13)$$

where the scaling factor c obeys the equation

$$\mathbf{E} \varphi'(cs)cs = 1. \quad (14)$$

Remark: that the expectations are evaluated w.r.t. the true source PDF.

Proof: $w = \delta$ is a strict minimum point of $\ell(cs; w)$ in the limit $T \rightarrow \infty$ if the asymptotic gradient $\mathbf{E} \nabla \ell(cs; \delta)$ vanishes and the asymptotic Hessian $\mathbf{E} \nabla^2 \ell(cs; \delta)$ from Proposition 1 is positive definite. The former holds trivially when (14) is satisfied, whereas the latter holds if conditions (11)-(13) are satisfied. ■

Note that any *convex* $\varphi(s)$ satisfies conditions (11) and (12) under mild assumptions on the source PDF, hence, (13) becomes essentially the only stability condition. In this paper we discuss three choices of convex $\varphi(s)$: the smoothed absolute value, the power function and the smoothed deadzone linear function.

A. Smoothed absolute value

When the source is super-Gaussian, e.g. sparse (sources common in seismology), or sparsely representable, a smooth approximation of the absolute value function usually obeys the asymptotic stability conditions [31]–[33]. The typical choice of $\varphi(s)$ in this case is

$$\varphi_\lambda^{\text{ABS}}(s) = -\frac{\lambda}{2} \log \left(\tanh^2 \left(\frac{s}{\lambda} \right) - 1 \right),$$

so that $\varphi'_\lambda(s) = \tanh(s/\lambda)$. The scalar λ acts as a smoothing parameter, yielding $\varphi_\lambda^{\text{ABS}}(s) \rightarrow |s|$ in the limit $\lambda \rightarrow 0^+$. Other possible choices are $\varphi_\lambda^{\text{ABS}}(s) = \sqrt{s^2 + \lambda^2}$ and [22]

$$\varphi_\lambda^{\text{ABS}}(s) = |s| - \lambda \log \left(1 + \frac{|s|}{\lambda} \right) \quad (15)$$

(see Figure 2, left). The latter choice was found especially suitable for Newton-based optimization due to its *self-concordance* up to a factor [28], [34]. It can be shown that in the limit $\lambda \rightarrow 0^+$, the asymptotical stability condition (13) reduces to

$$\mathbf{E}|s| < 2p(0)\mathbf{E}s^2.$$

B. Power function

In case of sub-Gaussian sources, common in digital communications, the family of power functions

$$\varphi_\mu^{\text{PWR}}(s) = |s|^\mu \quad (16)$$

with the parameter $\mu > 2$ is usually a good choice for the non-linearity (see Figure 2, center). It can be shown that the asymptotical stability conditions reduce to

$$\mathbf{E}|s|^{\mu+2} < (\mu+1)\mathbf{E}s^2 \mathbf{E}|s|^\mu,$$

which for the particular choice of $\mu = 4$ corresponds to negative kurtosis excess. An increase of μ usually yields better performance; for example, when sources are uniformly distributed, $\varphi_\mu^{\text{PWR}}(s)$ approaches the minus log-PDF in the limit $\mu \rightarrow \infty$. However, it is obvious that large values of μ imply high sensitivity to outliers due to the high powers.

C. Smoothed deadzone linear function

As a remedy to the sensitivity to outliers, we propose to replace the power function with the *deadzone linear* function of the form

$$\varphi_\mu^{\text{DZ}}(s) = \mu \cdot \max\{|s| - 1, 0\}, \quad (17)$$

which is often used in regression, data fitting and estimation [28]. This function has linear increase with controllable slope μ , and is known to have low sensitivity to outliers compared to the power function. Up to an additive constant, the deadzone linear function can be smoothly approximated by

$$\varphi_{\lambda,\mu}^{\text{DZ}}(s) = \frac{\mu}{2} (\varphi_\lambda^{\text{ABS}}(s-1) + \varphi_\lambda^{\text{ABS}}(s+1)) \quad (18)$$

(see Figure 2, right), where the parameter λ controls the smoothness.

In the limit $\lambda \rightarrow 0^+$, the smoothed deadzone linear function (18) yields an asymptotically stable QML estimator for μ satisfying the following coupled equation and inequality w.r.t. μ and c :

$$\begin{aligned} \mu &> \frac{1}{2c \cdot \mathbf{E}s^2 (p(c^{-1}) + p(-c^{-1}))} \\ 1 &= \mu c \int_{|cs| \geq 1} |s| p(s) ds. \end{aligned}$$

In the particular case of sources with compactly supported PDF (e.g. digital communication signals), which take the extreme values with non-zero probability, a more explicit condition can be obtained. Let us denote by s_{ext} the extreme value that s takes (either positive or negative, including the case when the distribution is symmetric), and let $\rho = P(|s| = s_{\text{ext}})$. Then, for a sufficiently small λ , (14) becomes

$$1 = \mathbf{E}\varphi'(cs)cs \approx \rho \cdot \varphi'(cs_{\text{ext}}).$$

Since $\varphi'(s)$ is monotonic and $\varphi'(s) \rightarrow \mu$ for $s \rightarrow \infty$, one obtains $\mu > \rho^{-1}$. For example, if the source is a symmetric N -level pulse amplitude modulation (PAM) signal, $\mu > \frac{N}{2}$ must hold.

For a sufficiently small λ ,

$$\begin{aligned} \varphi'(cs_{\text{ext}}) &\approx \frac{\mu}{2} \left(1 + \frac{cs_{\text{ext}} - 1}{\lambda + |cs_{\text{ext}} - 1|} \right) \\ \varphi''(cs_{\text{ext}}) &\approx \frac{\mu\lambda}{2(\lambda + |cs_{\text{ext}} - 1|)^2}. \end{aligned} \quad (19)$$

Substituting $\rho \cdot \varphi'(cs_{\text{ext}}) \approx 1$ and expressing φ'' in terms of φ' yields $\varphi''(cs_{\text{ext}}) \approx 2 \max\{(\mu\rho - 1)^2, 1\} / 2\lambda\mu\rho^2$, from where, using $cs_{\text{ext}} \approx 1$, the asymptotical stability condition can be derived:

$$\begin{aligned} 1 &< \mathbf{E}\varphi''(cs)\mathbf{E}(cs)^2 \approx \rho \varphi''(cs_{\text{ext}}) \cdot c^2 \mathbf{E}s^2 \\ &\approx \frac{2\sigma^2 \max\{(\mu\rho - 1)^2, 1\}}{s_{\text{ext}}^2 \lambda \mu \rho}. \end{aligned} \quad (20)$$

V. SUPER-EFFICIENT ESTIMATION

In [33], it was shown that in the noise-free case, the asymptotic variance of the estimation error Δw is given by

$$\begin{aligned} \text{var}\Delta w_k &\approx \\ &\frac{(\mathbf{E}\varphi'^2(cs) (\mathbf{E}^2\varphi''(cs)\mathbf{E}^2(cs)^2 + 1) - 2\mathbf{E}\varphi''(cs)) \mathbf{E}(cs)^2}{T (\mathbf{E}^2\varphi''(cs)\mathbf{E}^2(cs)^2 + 1)^2} \end{aligned}$$

for $k \neq 0$, and

$$\text{var}\Delta w_0 \approx \frac{\mathbf{E}\varphi'^2(cs)(cs)^2 - 1}{T(\mathbf{E}\varphi''(cs)(cs)^2 + 1)^2}.$$

Let us now consider the particular case of strictly *sparse* sources, i.e. such sources that take the value of zero with some non-zero probability $\rho > 0$. An example of such distribution is the Gauss-Bernoulli (sparse normal) distribution [33]. When $\varphi(s)$ is chosen according to (15), $\varphi'_\lambda(s) \rightarrow \text{sign}(s)$ and $\varphi''_\lambda(s) \rightarrow 2\delta(s)$ as $\lambda \rightarrow 0^+$. Hence, for a sufficiently small λ ,

$$\mathbf{E}\varphi''(cs) \approx \frac{1}{\lambda} \int_{-\lambda/c}^{+\lambda/c} p(s) ds \approx \frac{\rho}{\lambda},$$

whereas $\mathbf{E}\varphi'^2(cs)$ and c are bounded. Consequently, for $k \neq 0$

$$\text{plim}_{T \rightarrow \infty} T \cdot \text{var}\Delta w_k \leq \frac{\mathbf{E}\varphi'^2(cs)}{\mathbf{E}^2\varphi''(cs)\mathbf{E}(cs)^2} \leq \text{const} \cdot \lambda^2,$$

where plim denotes the probability limit. Observe that

$$\lim_{\lambda \rightarrow 0^+} \text{plim}_{T \rightarrow \infty} T \cdot \text{var}\Delta w_k \leq \lim_{\lambda \rightarrow 0^+} \text{const} \cdot \lambda^2 = 0,$$

that is, the QML estimator is *super-efficient* in the limit $\lambda \rightarrow 0^+$, assuming absence of noise.

Similarly, for sources whose PDF is compactly supported, it can be shown that the choice of the power function (16) or the smoothed deadzone linear function (18) yields a super-efficient estimator in the limit $\mu \rightarrow \infty$. When in addition the source signal takes the values at the extremal points of the interval with some non-zero probability ρ , the use of the smoothed deadzone linear function (18) with $\lambda \rightarrow 0^+$ and *finite* $\mu > \rho^{-1}$ yields a super-efficient estimator, since $\varphi''(cs_{\text{ext}})$ in (19) grows to infinity as λ approaches zero.

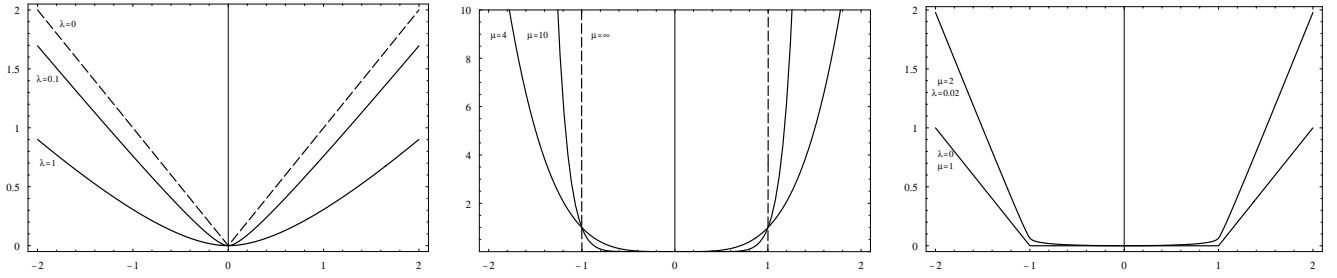


Fig. 2. Different choices of $\varphi(s)$: smoothed absolute value (15) (left), power function (16) (center), and the smoothed deadzone linear function (18) up to an additive constant (right), for different values of λ and μ .

Sequential optimization

For extremely small values of λ or extremely large values of μ , optimization of the objective function $\ell(x; w)$ becomes numerically difficult. To avoid this problem, one can start with a relatively large value of λ (or small value of μ) and gradually reduce (or increase) it on each iteration of the relative optimization algorithm. This *sequential optimization* algorithm has been previously used in the BSS problem, wherein it has demonstrated very high accuracy [22]. As an alternative, the smoothing method of multipliers proposed in [35] can be used.

VI. GENERALIZATIONS

A. IIR restoration kernels

For consistency of relative optimization, it is favorable that the restoration kernel w for a group, such that the action of w^k can be undone by w^{k+1} . This may be satisfied with rational IIR restoration filters of the form

$$W(z) = \frac{h_{-N}z^N + \dots + h_N z^{-N}}{(1 + b_1 z^{-1} + \dots + b_M z^{-M})(1 + f_1 z + \dots + f_L z^L)},$$

parameterized by the ARMA coefficients $h_{-N}, \dots, h_N, b_1, \dots, b_M$ and f_1, \dots, f_L . Additional advantage of using an ARMA parametrization of w appears when the convolution system a has zeros close to the unit circle, and the restoration kernel w has to be long in order to achieve good restoration quality.

Restoration kernel produced at the k -th iteration of the relative Newton algorithm is of the following form:

$$\hat{W}^{(k)}(z) = \frac{H^{(0)}(z) \cdot \dots \cdot H^{(k)}(z)}{B^{(0)}(z) \cdot \dots \cdot B^{(k)}(z) \cdot F^{(0)}(z) \cdot \dots \cdot F^{(k)}(z)}.$$

Since the kernels found at first iterations can be arbitrary, a necessary condition that the zeros and the poles of $W^{(0)}(z)$ can be cancelled by the poles and the zeros of the subsequent factors. This implies that the kernel $W^{(k)}(z)$ has to be invertible for every k , i.e. $H^{(k)}(z)$, $B^{(k)}(z)$ and $F^{(k)}(z)$ have no roots outside the unit circle. When $L = M = N$, the subspace of filters of the form $\frac{H(z)}{B(z)F(z)}$ with stable non-zero $H(z)$, $B(z)$ and $F(z)$ form a convolutional group. In order to force stability of the restoration kernel in the relative optimization algorithm, line search should be restricted to its stability region, e.g. by checking the value of the log

determinant of the Toeplitz matrices associated with h , b and f [26], [33].

The asymptotic Hessian of $\ell(x; h, b, f)$ with respect to these coefficients, evaluated at $w = \delta$ (i.e., all the coefficients, except $h_0 = 1$ are set to zero) and $x = cs$ has the following sparse structure:

$$\frac{\partial^2 \ell}{\partial b_i \partial b_j} = \frac{\partial^2 \ell}{\partial f_i \partial f_j} \approx \begin{cases} \mathbf{E} \varphi''(cs) \cdot \mathbf{E}(cs)^2 & : i = j, \\ 0 & : otherwise, \end{cases}$$

$$\frac{\partial^2 \ell}{\partial b_i \partial f_j} \approx \begin{cases} \mathbf{E} \varphi'(cs)(cs) & : i = j, \\ 0 & : otherwise, \end{cases}$$

and

$$\frac{\partial^2 \ell}{\partial b_i \partial h_j} = \frac{\partial^2 \ell}{\partial f_i \partial h_j} \approx \begin{cases} \mathbf{E} \varphi''(cs) \cdot \mathbf{E}(cs)^2 + \mathbf{E} \varphi'(cs)(cs) & : i = j = 0, \\ \mathbf{E} \varphi''(cs) \cdot \mathbf{E}(cs)^2 & : i = j > 0, \\ \mathbf{E} \varphi'(cs)(cs) & : i = -j > 0, \\ 0 & : otherwise, \end{cases}$$

where the indices i, j are in the appropriate ranges. The derivative $\frac{\partial^2 \ell}{\partial h_i \partial h_j}$ has the same form as $\frac{\partial^2 \ell}{\partial w_i \partial w_j}$ in Proposition 1. Typical Hessian structure is depicted in Figure 1 (right). Approximate Newton system solution can be carried out using an analytical expression for the regularized inverse of the structured Hessian. Another possibility is to consider techniques for solution of sparse symmetric systems. For example, one can use sparse modified Cholesky factorization for direct solution, or alternatively, conjugate gradient-type methods, possibly preconditioned by incomplete Cholesky factor, for iterative solution. In both cases, Cholesky factor is often not as sparse as the original matrix, but it becomes sparser, when appropriate reordering is applied before factorization. Approximate Hessian evaluation and Newton system solution have the complexity of a gradient descent iteration.

B. The MIMO case

In the MIMO case, the multichannel convolution model involves *crossstalk*,

$$x_{in} = \sum_{j,k} a_{ijk} s_{j,n-k},$$

where x_{in} is the i -th observed channel and s_{jn} is the j -th source. Restoration is performed using a matrix of filters w_{ijk} :

$$y_{in} = \sum_{j,k} w_{ijk} x_{j,n-k}.$$

The normalized minus log likelihood function becomes [13]

$$\ell(x; w) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\det W(e^{i\theta})| d\theta + \frac{1}{T} \sum_{i,n} \varphi_i(y_{in}),$$

where $W(e^{i\theta})$ is a matrix of the DFTs of w_{ijk} taken w.r.t. the index k .

In [30], it was shown that in the MIMO case, the Hessian of $\ell(x; w)$ for $x_i = c_i s_i$ and $w_{ijk} = \delta_{ij} \delta_k$ has the following approximate sparse structure:

$$\frac{\partial^2 \ell}{\partial w_{ijk} \partial w_{i'j'k'}} \approx \begin{cases} \mathbf{E} \varphi_i''(c_i s_i) (c_i s_i)^2 + 1 & : i = i' = j = j', k = k' = 0, \\ \mathbf{E} \varphi_i''(c_i s_i) \cdot \mathbf{E} (c_j s_j)^2 & : i = i', j = j', k = k' \neq 0, \\ 1 & : i = j', j = i', k = -k', \\ 0 & : \text{otherwise.} \end{cases}$$

Due to this sparse structure, the use of the fast relative Newton method is feasible. The block-coordinate update might be also done by updating each time the symmetric sets of coefficients $w_{ij,-k}, w_{ji,-k}, w_{ijk}, w_{jik}$, or more general blocks containing them.

C. The 2D case

The fast relative Newton method can be generalized for deconvolution of images (see [33]). Direct estimation of the restoration kernel is especially advantageous for blurs arising from scattering, whose point spread functions can be inverted using a small FIR restoration kernel. It must be noted, however, that unlike the 1D case, an arbitrary 2D kernel can not be generally factorized into smaller 2D kernels as the relative Newton method suggests. This implies that the restoration kernel obtained using relative optimization is usually suboptimal to the restoration kernel of the same size obtained using standard optimization methods. However, in practice, the achieved performance is very good.

VII. NUMERICAL RESULTS

Simulation results are presented to evaluate the performance of the proposed algorithms. Signal-to-interference ratio (SIR) is used as the restoration quality measure. Additional results can be found in [26], [33].

A. Deconvolution of a PAM source

The source signal was a 10^4 samples long i.i.d. 2-level PAM process. The empirically measured digital microwave channel impulse response from [36] was used to model the convolution system. Input SNRs of 10, 20, 30, 40 and 100 dB were used. FIR restoration kernel with 33 coefficients was adapted in a block-wise manner, using blocks of length 33. The block fast relative Newton algorithm was compared to Joho's FDBD

natural gradient-based algorithm [20] (in both cases, the power function was used with $\mu = 4$) and to CMA with $p = 2$.

Figure 3 (top) presents the restoration SIR averaged over 10 independent Monte-Carlo runs, as a function of the input SNR. For SNR higher than 20 dB, the block relative Newton algorithm demonstrates an improvement of about 4 dB compared to other methods. Good restoration quality is obtained for SNR starting from 10 dB. In absence of noise, the relative Newton algorithm performs very close to the theoretical SIR bound of 30.51 dB, achieving SIR of about 29 dB. Figure 4 (top) depicts the convergence of the compared algorithms, averaged over 10 independent runs with input SNR set to 20 dB. In Figure 5 (top), the average convergence time (time in samples required for the SIR to achieve 90% of its final value) is depicted for different algorithms. The block relative Newton algorithm converges in average about 10 times faster compared to FDBD and CMA.

In our MATLAB implementation on a 700MHz Pentium III PC, the block fast relative Newton algorithm requires about 0.6354 ± 0.0431 msec per input sample, compared to 0.0438 ± 0.0435 msec/sample for the FDBD algorithm and 0.1605 ± 0.0298 msec/sample for CMA.

B. The deadzone linear function vs. the power function

The use of the power function (16) with different values of μ , and the smoothed deadzone linear function (18) was compared for deconvolution of a 2000 samples long 2-level PAM source, degraded by the channel from the previous experiment. The comparison was performed both in the absence of noise, and in the presence of shot noise (sparse normal noise with 0.1% density, which introduced outliers into the signal). Figure 6 shows the SIR, averaged over 20 independent Monte-Carlo runs, for different choices of $\varphi(s)$. In the noiseless case, an increase of μ leads to better performance, as expected. However, in the presence of shot noise, the performance drops dramatically for large μ 's due to sensitivity to outliers. The proposed smoothed deadzone linear function appears to yield higher performance in the noiseless case and demonstrates negligible sensitivity to outliers (Figure 6, rightmost bars).

C. Deconvolution of a sparse source

The previous experiment was repeated for a 10^4 samples long i.i.d. sparse normal source with 20% density of the non-zero samples. The block fast relative Newton algorithm was compared to the FDBD algorithm. In the first case, $\varphi(s)$ was chosen according to (15) with $\lambda = 10^{-2}$, whereas in the second the exact absolute value was used. Figure 3 (bottom) presents the average restoration SIR as a function of the input SNR. For SNR higher than 20 dB, the block relative Newton algorithm demonstrates an improvement of about 7 dB compared to other methods. Good restoration quality is obtained for SNR starting from 10 dB. Figure 4 (bottom) depicts the convergence of the compared algorithms, averaged over 10 independent runs, with input SNR set to 20 dB. In Figure 5 (bottom), the average convergence time is depicted for different algorithms. The block relative Newton algorithm converges in average about 5-10 times faster compared to FDBD.

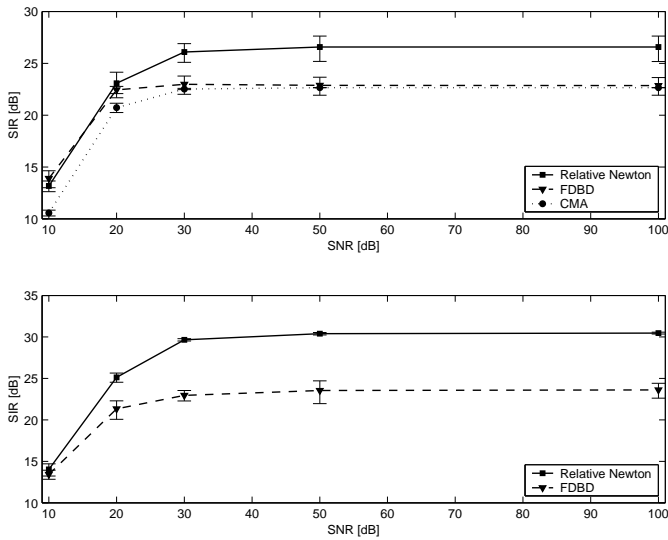


Fig. 3. Average restoration quality (in terms of SIR) as a function of input SNR for the 2-level PAM source (top) and the sparse source (bottom). 95% confidence intervals are indicated.

D. Rational restoration kernel

Advantages of an IIR restoration kernel are demonstrated in the following experiment. Input signal was generated by filtering a 1000 samples long sparse normal i.i.d. process by the kernel from the previous experiments. Sequential optimization algorithm was used with λ decreasing from 1 to 10^{-10} with the rate of 0.1 per iteration. Three restoration kernel configurations were tested: FIR (numerator only), causal all-pole IIR (causal denominator b only) and a rational kernel with equal length numerator and causal denominator. Figure 7 depicts the SIR, averaged over 10 Monte-Carlo runs, as a function of the number optimization variables for different assignments of the degrees of freedom to restoration kernel numerator and denominator. SIR higher than 20 dB was obtained for the all-pole IIR kernel starting from 6 degrees of freedom, for the rational kernel starting from 8 degrees of freedom, and for the FIR kernel starting only from 16 degrees of freedom. A practically ideal SIR was achieved by the all-pole IIR kernel starting from 8 degrees of freedom.

VIII. CONCLUSION

We have presented a relative optimization framework for QML single channel BD and studied the relative Newton method as its particular instance. Expanding kernel permits construction of long restoration kernels with cheap iterations. Diagonal- anti-diagonal structure of the Hessian in the proximity of the solution allows to derive a fast version of the relative Newton algorithm, with iteration complexity comparable to that of gradient methods. Additionally, we introduced rational restoration kernels, which constitute a richer and more flexible family of filters than the traditionally used FIR kernels, and often allow to reduce the optimization problem size. We also proposed the use of the deadzone linear function for sub-Gaussian sources, which is significantly less sensitive to

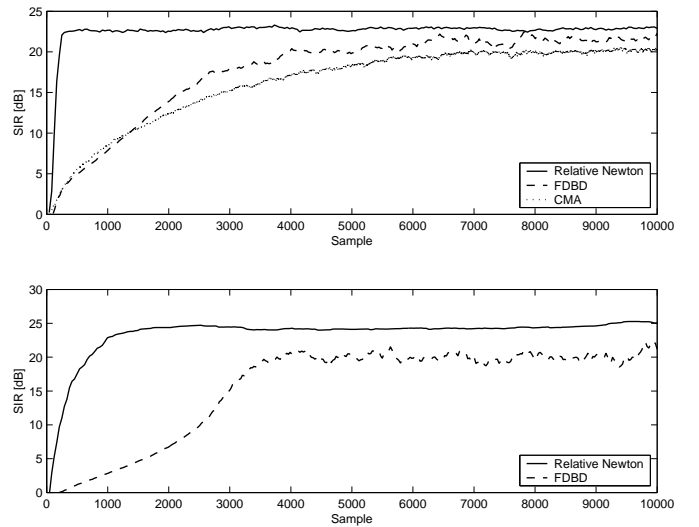


Fig. 4. Average convergence (in terms of SIR) as a function of time in samples for the 2-level PAM source (top) and the sparse source (bottom). Input SNR is set to 20 dB.

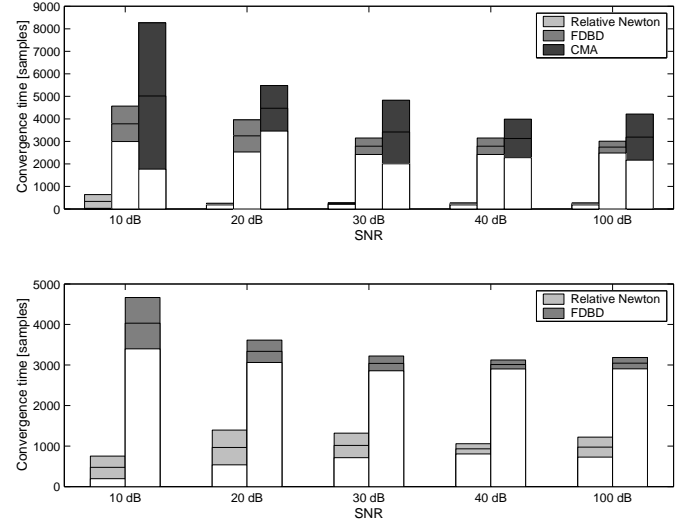


Fig. 5. Average convergence time in samples as a function of input SNR for the 2-level PAM source (top) and the sparse source (bottom). 95% confidence intervals are indicated.

outliers than the commonly used non-linearities, and achieves super-efficient estimation in the absence of noise.

Relative Newton method appears to be effective for deconvolution of sparse signals, when the smoothed absolute value function is used. Despite the natural signals and images are often not sparse, they can be brought to a sparse form by means of sparsifying transformation [33]. Such a transformation can be constructed either from knowledge of the source nature, or by means of learning on a set of representative sources. Similar sparsifying approach based on wavelet-type representations was successfully used in BSS [32].

In simulation studies with super- and sub-Gaussian sources, the proposed methods exhibited very fast convergence and higher accuracy compared to the state-of-the-art approaches such as CMA and natural gradient-based QML algorithms. Our

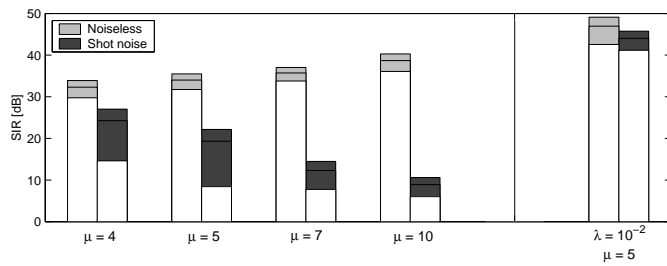


Fig. 6. Average restoration SIR for different choices of $\varphi(s)$: the power function (16) with different values of μ (left), and the smoothed deadzone linear function (18) (two rightmost bars), with and without the presence of shot noise. 95% confidence intervals are indicated.

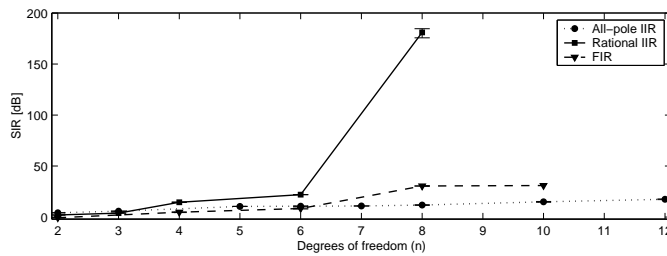


Fig. 7. SIR as a function of degrees of freedom for different restoration kernel configurations. 95% Confidence intervals are shown.

algorithm yields plausible restoration quality in low to medium noise conditions. Possible applications are in acoustics and communications, especially where high accuracy and fast convergence are required. We are currently working on extending the presented approach to the multichannel and complex cases.

REFERENCES

- [1] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Computers*, pp. 679–682, 1975.
- [2] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. 25, no. 3, pp. 385–399, 1980.
- [3] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, 1980.
- [4] J. R. Treichler and B. G. Agee, "A new approach to the multi-path correction of constant modulus signals," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 31, no. 2, pp. 331–344, 1983.
- [5] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. 25th Asilomar Conf. Signal, Syst., Comput.*, 1991.
- [6] G. Xu, H. Liu, L. Tong, and T. Kailath, "Least squares approach to blind channel identification," *IEEE Trans. Sig. Proc.*, vol. 43, no. 12, pp. 2982–2993, 1995.
- [7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. Sig. Proc.*, vol. 43, pp. 516–525, 1995.
- [8] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on secondorder statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 340–349, 1994.
- [9] M. Gurelli and C. Nikias, "EVAM: An eigenvectorbased algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 134–149, 1995.
- [10] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Sig. Proc.*, vol. 44, no. 3, pp. 661–672, 1996.

- [11] A. Gorokhov, P. Loubaton, and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach," in *Proc. ICASSP*, vol. 5, 1996, pp. 2415–2418.
- [12] C.-Y. Chi, C.-Y. Chen, C.-H. Chen, and C.-C. Feng, "Batch processing algorithms for blind equalization using higher-order statistics," *IEEE Sig. Proc. Magazine*, pp. 25–49, January 2003.
- [13] 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.
- [14] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electronics Letters*, vol. 30, no. 17, pp. 1386–1387, 1994.
- [15] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.
- [16] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, 1996.
- [17] S. Douglas, "On equivariant adaptation in blind deconvolution," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, November 2002.
- [18] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. SPAWC*, April 1997, pp. 101–104.
- [19] R. H. Lambert, "Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures," Ph.D. dissertation, University of Southern California, 1996.
- [20] M. Joho, H. Mathis, and G. S. Moschytz, "On frequency-domain implementations of filtered-gradient blind deconvolution algorithms," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, November 2002.
- [21] 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.
- [22] M. Zibulevsky, "Blind source separation with relative newton method," in *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, Apr. 2003, pp. 897–902.
- [23] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," 1997.
- [24] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Wiley, 2003.
- [25] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [26] A. M. Bronstein, M. Bronstein, and M. Zibulevsky, "Blind deconvolution with relative Newton method," Technion, Israel, Tech. Rep. 444, October 2003. [Online]. Available: <http://visl.technion.ac.il/bron/alex>
- [27] D. P. Bertsekas, *Nonlinear Programming (2nd edition)*. Athena Scientific, 1999.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.
- [29] A. M. Bronstein, M. Bronstein, and M. Zibulevsky, "Block-coordinate relative Newton method for blind source separation," *Sig. Proc.*, 2004, to appear.
- [30] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Asymptotic performance analysis of MIMO blind deconvolution," Technion, Israel, Tech. Rep., January 2004. [Online]. Available: <http://visl.technion.ac.il/bron/alex>
- [31] 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.
- [32] 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 University Press, 2001.
- [33] 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.
- [34] Y. E. Nesterov and A. S. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
- [35] M. Zibulevsky. (2002) Smoothing method of multipliers for sum-max problems. [Online]. Available: <http://iew3.technion.ac.il/~mcib>
- [36] G. B. Giannakis and S. D. Halford, "Blind fractionally spaced equalization of noisy FIR channels: Direct and adaptive solutions," *IEEE Trans. Sig. Proc.*, vol. 45.



Alexander M. Bronstein (M'02) was born in Russia in 1980. He received the B.Sc. degree (summa cum laude, 2002) from the Department of Electrical Engineering, Technion – Israel Institute of Technology, where he is currently a M.Sc student. His main research interests include blind inverse problems and 3D face recognition.

Alexander Bronstein received the Technion Humanities and Arts Department prize (2001), the Kasher prize (2002), the Thomas Schwartz award (2002), the Hershel Rich Technion Innovation award and the Counter Terrorism prize (2003). He was invited as a honorary student delegate to the International Achievement Summit of the Academy of Achievement (Washington D.C.) in May 2003. He is a member of the Advisory Committee of the US Biometrics Council. Alexander Bronstein is a member of the 3DFACE research project, which was featured in the CNN news. He is the co-inventor of 2 patents (pending) and 20 publications in leading scientific journals and conferences.



Michael M. Bronstein (M'02) was born in Russia in 1980. He received the B.Sc. in Electrical Engineering (summa cum laude) from the Technion – Israel Institute of Technology in 2002. He is currently pursuing the M.Sc in Computer Science, Technion. His main research interests include 3D face recognition and blind inverse problems.

Michael 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 and the Counter Terrorism prize (2003). He was a honorary student delegate to the International Achievement Summit of the Academy of Achievement (Washington D.C.) in May 2003. He is a member of the Advisory Committee of the US Biometrics Council. As part of his thesis work, Michael Bronstein is involved in the 3DFACE research project, which was featured in the CNN news. He is the co-inventor of 2 patents (pending) and 20 publications in leading scientific journals and conferences.



Michael Zibulevsky received the M.Sc. degree in Electrical Engineering from MIIT, Moscow, and the Ph.D. degree in Operation Research (Nonlinear Optimization) from the Technion – Israel Institute of Technology. Currently he is with the Department of Electrical Engineering at the Technion. His area of interests includes nonlinear optimization and inverse problems in signal/image reconstruction (blind source separation, deconvolution, tomography, EEG/MEG).