

Quasi maximum likelihood MIMO blind deconvolution: super- and sub-Gaussianity vs. consistency

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

Abstract—In this note we consider the problem of MIMO quasi maximum likelihood (QML) blind deconvolution. We examine two classes of estimators, which are commonly believed to be suitable for super- and sub-Gaussian sources. We state the consistency conditions and demonstrate a distribution, for which the studied estimators are unsuitable, in the sense that they are asymptotically unstable.

Index Terms—MIMO, blind deconvolution, blind source separation, quasi maximum likelihood, consistency, super-Gaussian, sub-Gaussian, kurtosis.

I. INTRODUCTION

We consider the problem of MIMO blind deconvolution, in which the observed vector-valued sensor time signal $x(t) = (x_1(t), \dots, x_N(t))^T$ is created from the vector-valued *source signal* $s(t) = (s_1(t), \dots, s_N(t))^T$ passing through a convolutive mixing system defined by the $N \times N$ matrix of impulse responses $a_{ij}(t)$,

$$x_i(t) = \sum_{j=1}^N (a_{ij} * s_j)(t) = \sum_{j=1}^N \sum_{\tau=-\infty}^{\infty} a_{ij}(\tau) s_j(t - \tau).$$

The setup is termed *blind* when only x is accessible, whereas no knowledge on a and s is available. The problem of blind deconvolution aims to find such a deconvolution (or restoration) kernel w_{ijt} , that produces a possibly delayed, scaled and permuted, waveform-preserving estimate of s :

$$\hat{s}_i(t) = \sum_{j=1}^N (w_{ij} * x_j)(t) \approx c_i \cdot s_{\pi_i}(t - \Delta_i),$$

where c_i are scaling factors, π_i is a permutation, and Δ_i are integer shifts. A commonly used assumption is that sources are non-Gaussian.

Let us denote by $W(\theta)$ the matrix of the discrete Fourier transforms of $w_{ij}(t)$, and assume that $\det W(\theta)$ has no zeros on the unit circle, and the source signals are pairwise independent, real and i.i.d. Neglecting edge effects and assuming no noise, the normalized log-likelihood function of the observed signal x is [1]–[3]

$$\ell(x; w) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\det W(\theta)| d\theta - \frac{1}{T} \sum_{i,j=1}^N \sum_{t=0}^{T-1} \varphi_i((w_{ij} * x_i)(t)), \quad (1)$$

where T is the sample size, and $\varphi_i(s) = -\log p_i(s)$, where $p_i(s)$ is the probability density function (PDF) of the i -th source.

In the case of instantaneous (delay-less) mixture case, the normalized log-likelihood function (1) becomes

$$\ell(x; w) = \log |\det W| - \sum_{i,j=1}^N \varphi_i(w_{ij}x_i), \quad (2)$$

Manuscript received October 10, 2003. 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 for Vision and Image Sciences.

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

and estimation of the *unmixing matrix* W is usually referred to as blind source separation (BSS). In the case of single-channel (SISO) case, where no cross-talk is present, (1) reduces to

$$\ell(x; w) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |W(\theta)| d\theta - \frac{1}{T} \sum_{t=0}^{T-1} \varphi((w * x)(t)). \quad (3)$$

A consistent estimator can be obtained by maximizing $\ell(x; w)$ even when $\varphi_i(s)$ are not exactly equal to $-\log p_i(s)$. Such QML estimation has been shown to be practical in instantaneous blind source separation [4]–[7] and blind deconvolution [2], [8], [9] when the source PDF is unknown or not well-suited for optimization. Generally, $\ell(x; w)$ is maximized by gradient-based methods, hence, the main concern is the choice of $\varphi'(s)$.

It is commonly believed, that the knowledge of whether the source is super- or sub-Gaussian (i.e., such that its *kurtosis excess* defined by

$$\kappa = \frac{\mathbf{E}s^4}{\mathbf{E}^2s^2} - 3$$

is either positive or negative, respectively) is sufficient in order to construct a consistent QML estimator. This belief leads to attributing $\varphi'(s)$ either to the class of functions suitable for estimation of super-Gaussian sources, and not suitable for estimation of the sub-Gaussian ones, or vice versa. For example, it is usually assumed (see e.g. [1], [2], [10], [11]) that the choice of the smoothed sign function, e.g.,

$$\varphi'(s) = \tanh(\beta s), \quad (4)$$

for $\beta \geq 1$, leads to a QML estimator suitable for super-Gaussian sources. Another example is the family of functions

$$\varphi'(s) = |s|^\mu \text{sign}(s) \quad (5)$$

with the parameter $\mu > 1$, which is believed to be suitable for sub-Gaussian sources.

In this note, we state the conditions, under which a QML estimator is consistent, and show that generally there is no connection between the sign of kurtosis excess and consistency. We study the estimators obtained from (4), (5), for sources obeying the generalized Cauchy distribution.

II. CONSISTENCY CONDITIONS

For a general choice of $\varphi'_i(s)$'s, the corresponding QML estimator is (asymptotically) consistent if the following conditions hold [12]: there exist a set of positive scaling factors c_i obeying

$$\mathbf{E}\varphi'_i(c_i s_i) c_i s_i = 1 \quad (6)$$

and

$$\mathbf{E}\varphi''_i(s_i) > 0 \quad (7)$$

$$\mathbf{E}\varphi''_i(s_i) \mathbf{E}\varphi''_j(s_j) \mathbf{E}(c_i s_i)^2 \mathbf{E}(c_j s_j)^2 > 1 \quad (8)$$

$$\mathbf{E}\varphi''_i(c_i s_i) (c_i s_i)^2 + 1 > 0, \quad (9)$$

for $i, j = 1, \dots, N$. These conditions are valid when the expected values $\mathbf{E}\varphi''(s)$, $\mathbf{E}\varphi''(s)s^2$, $\mathbf{E}\varphi'(s)$, and $\mathbf{E}s^2$ exist and are bounded. Similar consistency conditions exist in the particular cases of instantaneous blind source separation [13], [14] and SISO blind deconvolution [9]. For derivations and important statistical properties of QML estimators, the reader is referred to the above cited references as well as [14]–[18]. Although QML estimators include no noise model and the consistency conditions are derived assuming the noiseless case, practice shows that QML blind deconvolution and source separation are quite robust to noise [2], [8]–[11].

Observe that under mild assumptions on the source distribution, conditions (7) and (9) are satisfied if $\varphi_i(s)$ is *convex*, due to the

fact that $\varphi_i''(s) > 0$. Demanding convexity of $\varphi_i(s)$ is not very restrictive, since maximization of $\ell(x; w)$ with non-convex $\varphi_i(s)$'s is usually impractical. The essential set of conditions for consistency of the QML estimator can be therefore divided into the *single-channel* terms

$$\mathbf{E}^2 \varphi_i''(s_i) \mathbf{E}^2 (c_i s_i)^2 > 1 \quad (10)$$

and the *cross-talk* terms

$$\mathbf{E} \varphi_i''(s_i) \mathbf{E} \varphi_j''(s_j) \mathbf{E} (c_i s_i)^2 \mathbf{E} (c_j s_j)^2 > 1, \quad (11)$$

where $i = 1, \dots, N$ and $j \neq i$. The former are required for consistent estimation of w_{ii} , whereas the latter are necessary for consistent estimation of the cross-talk kernels w_{ij} .¹

We will henceforth focus our attention on the case where all the sources are identically distributed and the same (convex) $\varphi(s)$ is used for all sources. The underlying QML estimator is consistent if

$$\mathbf{E}^2 \varphi''(s_i) \mathbf{E}^2 (cs)^2 > 1. \quad (12)$$

In the more general case, the cross-talk consistency conditions impose additional restriction on consistency of the QML estimator.

We will now examine the consistency conditions of the estimators obtained by choosing $\varphi(s)$ according to (4) and (5). When $\varphi'(s) = |s|^\mu \text{sign}(s)$, it can be shown that

$$\begin{aligned} c &= ((\mu + 1) \cdot \mathbf{E}|s|^{\mu+1})^{-1/(\mu+1)} \\ \mathbf{E} \varphi''(s) (cs)^2 &= \mu(\mu + 1) c^{\mu+1} \cdot \mathbf{E}|s|^{\mu+1} \\ \mathbf{E} \varphi''(s) &= \mu(\mu + 1) c^{\mu-1} \cdot \mathbf{E}|s|^{\mu-1}. \end{aligned}$$

For $\mu > 1$, consistency condition (12) yields

$$\Delta_s = \frac{\mathbf{E}|s|^{\mu+1}}{\mathbf{E}s^2 \mathbf{E}|s|^{\mu-1}} - \mu < 0. \quad (13)$$

In the particular case when $\mu = 3$, the latter condition becomes $\Delta_s = \kappa < 0$, meaning that the estimator is consistent for sub-Gaussian sources, and inconsistent for the super-Gaussian ones.

When $\varphi'(s) = \tanh(\beta s)$, the consistency condition (12) becomes

$$\Delta_s = 1 - \mathbf{E} \varphi''(s) \cdot \mathbf{E} (cs)^2 < 0. \quad (14)$$

In the case of a general β , derivation of analytic expression of Δ_s is complicated. However, in the limit $\beta \rightarrow \infty$, $\varphi'(s) \rightarrow \text{sign}(s)$, and $\varphi''(s) \rightarrow 2\delta(s)$. Hence, for a large β ,

$$\begin{aligned} \mathbf{E} \varphi''(s) (cs)^2 &= \mathbf{E} \varphi''(s) (cs)^2 \approx c^2 \beta \int_{-1/\beta}^{+1/\beta} sp(s) ds \approx 0 \\ \mathbf{E} \varphi''(s) &= \mathbf{E} \varphi''(s) \approx \beta \int_{-1/\beta}^{+1/\beta} p(s) ds \approx 2p(0), \end{aligned}$$

where c is obtained by substituting $\varphi'(cs)cs \approx \text{sign}(cs)cs$ into equation (6):

$$c \approx \frac{1}{\mathbf{E}|s|}.$$

Therefore, the estimator is consistent if

$$\Delta_s \approx \frac{\mathbf{E}|s|}{2p(0)\sigma^2} - 1 < 0. \quad (15)$$

In the limit $\beta \rightarrow \infty$, the latter condition is exact.

¹For example, the single-channel condition is responsible for the inconsistency when the sources are Gaussian. In the latter case, (11) holds with equality for every $\varphi(s)$ [12], leading to the well-known fact that Gaussian sources can be restored up to a rotation matrix and an all-pass term.

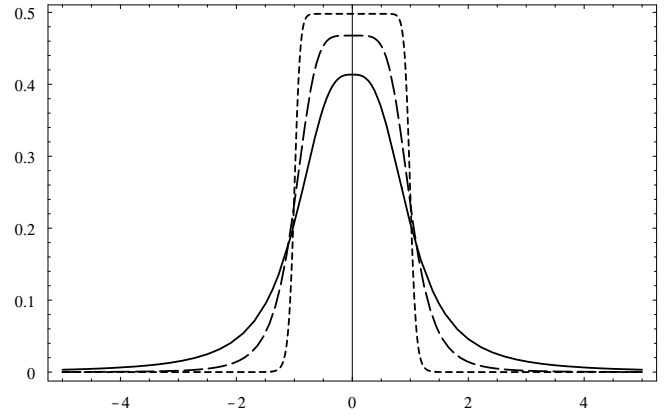


Fig. 1. PDF of the generalized Cauchy distribution for $r = 1$, $a = 1.5$ (solid), $a = 2.5$ (dashed), and $a = 10$ (dotted).

III. THE GENERALIZED CAUCHY DISTRIBUTION

Let us consider a parametric family of distributions with the parameters $a, r > 0$, described by the following PDF:

$$p(s) = \frac{ar^{1-\frac{1}{2a}} \sin\left(\frac{\pi}{2a}\right)}{\pi(|s|^{2a} + r)}$$

(see Figure 1). The parameter r influences the variance of s . For $a = 1$, one gets the Cauchy distribution; for this reason, this family of distributions will be henceforth referred to as the generalized Cauchy distribution.

It can be shown that the p -th moment of $|s|$ exists for $a > \frac{p+1}{2}$, and is given by

$$\mathbf{E}|s|^p = r^{\frac{p}{2a}} \cdot \csc\left(\frac{(p+1)\pi}{2a}\right) \sin\left(\frac{\pi}{2a}\right),$$

where

$$\csc x = \frac{1}{\sin x}$$

is the cosecant function. Particularly, the fourth order moment exists for $a > 2.5$ and the kurtosis excess is given by

$$\kappa(a) = \csc\left(\frac{\pi}{2a}\right) \csc\left(\frac{5\pi}{2a}\right) \sin^2\left(\frac{3\pi}{2a}\right) - 3.$$

$\kappa(a)$ is monotonically decreasing as a function of a and crosses zero for $a \approx 3.3567$ (see Figure 2, solid). This means that the source is super-Gaussian for $2.5 < a < 3.3567$, and sub-Gaussian for $a > 3.3567$.

For $\varphi'(s) = \tanh(\beta s)$, in the limit $\beta \rightarrow \infty$, the consistency condition is given by

$$\begin{aligned} \Delta_s &= \frac{\mathbf{E}|s|}{2p(0)\sigma^2} - 1 \\ &= \frac{\pi}{2a} \csc\left(\frac{\pi}{2a}\right) \csc\left(\frac{\pi}{a}\right) \sin\left(\frac{3\pi}{2a}\right) - 1 < 0, \end{aligned}$$

and is valid for $a > 1.5$. Observe that

$$\frac{d\Delta_s}{da} = \frac{\pi}{4a^2} \left(2 + \cos\left(\frac{\pi}{a}\right)\right) \csc^2\left(\frac{\pi}{2a}\right) \sec^2\left(\frac{\pi}{2a}\right),$$

where

$$\sec x = \frac{1}{\cos x}$$

is the secant function. Since the derivative of Δ_s w.r.t. a is strictly positive, Δ_s is monotonically increasing with a . Δ_s crosses zero at $a \approx 2.3379$ (see Figure 2, dashed). This means that the corresponding

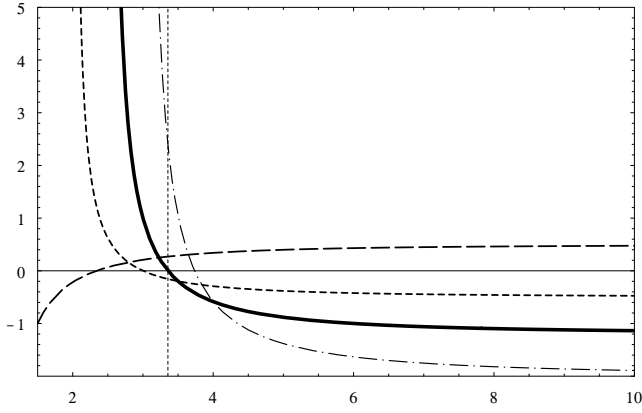


Fig. 2. The value of Δ_s as a function of a for different QML estimators: $\varphi'(s) = \text{sign}(s)$ (dashed), $\varphi'(s) = |s|^2 \text{sign}(s)$ (dotted), $\varphi'(s) = |s|^3 \text{sign}(s)$ (solid), and $\varphi'(s) = |s|^4 \text{sign}(s)$ (dash-dotted). Kurtosis excess κ corresponds to Δ_s is the case $\varphi'(s) = |s|^3 \text{sign}(s)$. The estimator is consistent for $\Delta_s < 0$.

TABLE I
CONSISTENCY REGIONS OF DIFFERENT QML ESTIMATORS

$\varphi'(s)$	Consistency region
$\tanh(s)$	$1.5 < a < 1.8666$
$\tanh(10s)$	$1.5 < a < 1.9344$
$\tanh(\beta s), \beta \rightarrow \infty$	$1.5 < a < 2.3379$
$ s ^2 \text{sign}(s)$	$a > 3$
$ s ^3 \text{sign}(s)$	$a > 3.3567$
$ s ^4 \text{sign}(s)$	$a > 3.7352$

QML estimator is inconsistent for $a > 2.3379$, particularly, the estimator is asymptotically inconsistent for both super- and sub-Gaussian sources. Δ_s was also evaluated numerically for $\beta = 1, 10$ (see Figure 3). Consistency regions of the estimators are presented in Table I.

For $\varphi'(s) = |s|^\mu \text{sign}(s)$, the consistency condition is given by

$$\Delta_s = \frac{\mathbf{E}|s|^{\mu+1}}{\mathbf{E}s^2 \mathbf{E}|s|^{\mu-1}} - \mu = \left(1 + 2 \cos\left(\frac{\pi}{a}\right)\right) \cdot \csc\left(\frac{\pi(\mu+2)}{2a}\right) \sin\left(\frac{\pi\mu}{2a}\right) - \mu < 0,$$

and is valid for $a > 1 + \mu/2$. Observe that

$$\frac{d\Delta_s}{da} = \frac{\pi}{2a^2} \csc\left(\frac{\pi(\mu+2)}{2a}\right) \left(2 + \cos\left(\frac{2\pi}{a}\right) \csc\left(\frac{2\pi}{a}\right) - \mu \left(1 + 2 \cos\left(\frac{\pi}{a}\right)\right) \csc\left(\frac{\pi(\mu+2)}{2a}\right) \sin\left(\frac{\pi}{a}\right)\right)$$

is negative for $a > 1 + \mu/2 > 1.5$ for every $\mu > 1$, and consequently, Δ_s is monotonically decreasing, with zero crossing depending on μ . Consistency regions for some values of μ are summarized in Table I, and the values of Δ_s are plotted as a function of a in Figure 2. Note that for $\mu = 3$, consistency is fully determined by the sign of kurtosis excess. However, this is not true for other values of μ .

IV. CONCLUSION

An important element of an efficient QML blind deconvolution is the choice of the non linear functions $\varphi_i(s)$, which is conditioned, among all, by consistency criteria. It is commonly believed that the

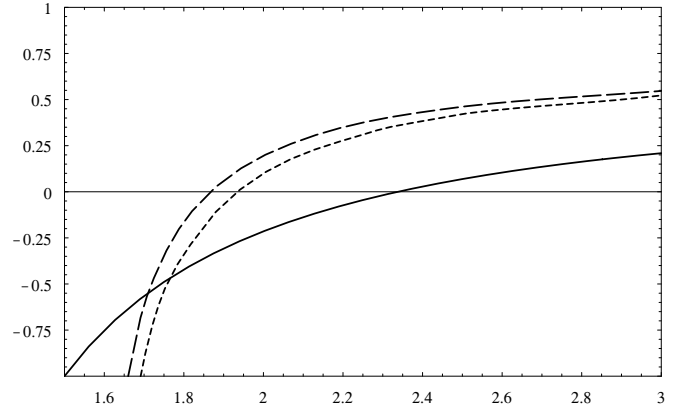


Fig. 3. The value of Δ_s as a function of a for the QML estimator $\varphi'(s) = \tanh(\beta s)$: $\beta = 1$ (dashed), $\beta = 10$ (dotted), and $\beta \rightarrow \infty$ (solid). The estimator is consistent for $\Delta_s < 0$.

knowledge of whether the sources are super- or sub-Gaussian is sufficient for construction of a consistent QML estimator.

We have examined the consistency conditions for two classes of QML estimators, commonly used for super- and sub-Gaussian sources in blind source separation and deconvolution problems. The particular case of the generalized Cauchy distribution was examined. It can be concluded that consistency does not always correspond to the sign of kurtosis excess, which determines whether the source is super- or sub-Gaussian. For example, the choice $\varphi'(s) = \tanh(\beta s)$, which is commonly believed to be suitable for super-Gaussian sources, is inconsistent for such sources. The choice $\varphi'(s) = |s|^2 \text{sign}(s)$, which is known to be suitable for sub-Gaussian sources, is also suitable for some super-Gaussian sources (wherein $a > 3$). The choice $\varphi'(s) = |s|^4 \text{sign}(s)$, known to be suitable for sub-Gaussian sources, is inconsistent for some of such sources (wherein $3.3567 < a < 3.7352$). With the only exception of $\varphi'(s) = |s|^3 \text{sign}(s)$, whose consistency is always determined by the sign of kurtosis excess, other QML estimators require more delicate analysis in order to determine whether they are suitable or not for estimation of super- or sub-Gaussian sources. Generally, the answer is distribution-dependent. The main conclusion from this note is that the non linear functions $\varphi_i(s)$ should be chosen with more caution.

ACKNOWLEDGEMENT

The authors are grateful to the anonymous referees for their constructive critics and valuable comments.

REFERENCES

- [1] 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.
- [2] 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.
- [3] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," 1997.
- [4] 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.
- [5] 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.

- [6] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, “Blind source separation via multinode sparse representation,” in *Proc. NIPS*. MIT Press, 2002.
- [7] P. Kisilev, M. Zibulevsky, and Y. Zeevi, “Multiscale framework for blind source separation,” *JMLR*, 2003, in press.
- [8] A. M. Bronstein, M. 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>
- [9] A. M. Bronstein, M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, “Quasi-maximum likelihood blind deconvolution of images using sparse representations,” Technion, Israel, Tech. Rep., 2003. [Online]. Available: <http://visl.technion.ac.il/bron/alex>
- [10] H. Mathis, M. Joho, and S. Moschytz, “A simple threshold nonlinearity for blind source separation of sub-gaussian signals,” in *Proc. ISCAS*, 2000, pp. 489–492.
- [11] M. Joho and P. Schniter, “Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient,” in *Proc. ICA2003*, April 2003.
- [12] 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>
- [13] S.-I. Amari, T.-P. Chen, and A. Cichocki, “Stability analysis of learning algorithms for blind source separation,” *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.
- [14] J.-F. Cardoso, “Blind signal separation statistical principles,” *Proc. IEEE. Special issue on blind source separation*, vol. 9, no. 10, pp. 2009–2025, 1998.
- [15] Y. Bresler and A. Macovski, “Exact maximum likelihood parameter estimation of superimposed exponential signals in noise,” *IEEE Trans. on Ac., Speech and Sig. Proc.*, vol. 34, no. 6, pp. 1081–1089, 1986.
- [16] A. Gorokhov and J.-F. Cardoso, “Equivariant blind deconvolution of MIMO-FIR channels,” in *Proc. SPAWC*, 1997, pp. 489–492.
- [17] P. Stoica, J. Li, and T. Soederstroem, “On the inconsistency of iqml,” *Sig. Proc.*, vol. 56, no. 2, pp. 185–190, 1997.
- [18] J.-F. Cardoso, “On the stability of source separation algorithms,” *J. VLSI Sig. Proc. Sys.*, vol. 26, no. 1/2, pp. 7–14, 2000.



Michael M. Bronstein (M’02) was born in Russia in 1980. He received the B.Sc. in Electrical Engineering (summa cum laude) in 2002 and the M.Sc. in Computer Science (summa cum laude) in 2005, both from the Technion – Israel Institute of Technology. He is currently pursuing the Ph.D. degree in Computer Science. 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) and the Copper Mountain Conference on Multigrid Methods Best Paper Award (2005). He was a honorary student delegate to the International Achievement Summit of the Academy of Achievement (Washington D.C.) in May 2003. Michael Bronstein is involved in the 3DFACE research project, which was featured in the CNN news.



Alexander M. Bronstein (M’02) was born in Russia in 1980. He received the B.Sc. (2002) and the M.Sc. (2005) degrees (both summa cum laude) from the Department of Electrical Engineering, Technion – Israel Institute of Technology. He is currently a Ph.D. student at the Department of Computer Science, Technion. 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) and the Copper Mountain Conference on Multigrid Methods Best Paper Award (2005). He was a honorary student delegate to the International Achievement Summit of the Academy of Achievement (Washington D.C.) in May 2003. Alexander Bronstein is a member of the 3DFACE research project, which was featured in the CNN news.



Michael Zibulevsky received the M.Sc. degree in Electrical Engineering from MIT, 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).