2004 International Conference on Image Processing (ICIP)

OPTIMAL SPARSE REPRESENTATIONS FOR BLIND SOURCE SEPARATION AND BLIND DECONVOLUTION: A LEARNING APPROACH

Michael M.Bronstein

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

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

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

ABSTRACT

We present a generic approach, which allows to adapt sparse blind deconvolution and blind source separation algorithms to arbitrary sources. The key idea is to bring the problem to the case in which the underlying sources are sparse by applying a sparsifying transformation on the mixtures. We present simulation results and show that such transformation can be found by training. Properties of the optimal sparsifying transformation are highlighted by an example with aerial images.

1. INTRODUCTION

The problem of blind deconvolution (BD) and blind source separation (BSS), arise in numerous image processing applications. A popular solution to these problems is the maximum likelihood (ML) approach which is reported to produce good results [1, 2, 3, 4]. However, the ML methods require in general (at least approximately) knowledge of the statistical distribution of the sources, which is usually hard to model and not well-suited for optimization. Cases where a suitable simple prior on the sources is available analytically are scarce and mostly trivial.

One of such important examples are sparse sources. Blind deconvolution and separation of sparse sources was extensively treated in [5, 3, 6, 4]. As for themselves such sources are of little interest since natural sources are seldom sparse. If, however, the sources can be made sparse by means of some appropriate transformation, it is possible to apply the sparse BD and BSS algorithms.

In this paper we present a *sparsification* approach which allows to adapt sparse BD and BSS algorithms to general sources. The key idea is to transform the mixtures in such a way that the problem would become equivalent to the case in which the underlying sources are sparse. In the general case, the sparsifying transformations must be linear shift invariant (LSI) and can be found by training. We exemplify the sparsification approach on the problem of pure BSS and SISO blind deconvolution.

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

2. PROBLEM FORMULATION

Let us consider a general problem of multiple-input multiple-output (MIMO) 2D BD, and the BSS problem as a particular case[1]. We use tensor notation and Einstein summation convention, according to which each repeated upper and lower indexes are summed over, i.e. $a_i^j b_{jk} = \sum_j a_{ij} b_{jk}$.

In a general noiseless multichannel BD setup, the source images $s_{i,mn}$ (where $i = 1, ..., N$ denotes the source number, and $m = 1, ..., M_X, n = 1, ..., N_X$ are the pixel indexes), pass through a convolutive mixing system denoted by the operator $\mathcal{A} = (a_{ij,mn})$ and form mixtures $x_{i,mn}$ ($i = 1, ..., M$) in the following way:

$$x_{i,mn} = (\mathcal{A}s)_{i,mn} = a_{i,\ mn}^{j} * s_{j,mn}, \qquad (1)$$

where $a_{i,\ mn}^{j} * s_{j,mn} = a_i^{j,m'n'} s_{j,m-m'\ n-n'}$ denotes concolution w.r.t. the indexes m, n and mixing w.r.t. j, and $a_{ii,mn}$ are direct channels and $a_{i\neq j,mn}$ are cross-talk channels. If there is no convolution but only crosstalk (i.e. $a_{ij,mn} = a_{ij}\delta_{mn}$), the model reduces to the BSS case:

$$x_{i,mn} = a_i^j s_{j,mn}, \qquad (2)$$

and the coefficients (a_{ij}) are usually referred to as the *mixing matrix*. In case of a single source, we have a single-input multiple-output (SIMO) model without crosstalk:

$$x_{i,mn} = a_{i,\ mn} * s_{mn}. \qquad (3)$$

Blind deconvolution attempts to find such a deconvolution (or restoration) operator $\mathcal{W} = (w_{ij,mn})$ that once applied on the observations $x_{ij,mn}$, produces an estimate of

[1] Without loss of generality, we consider the 2D case. 1D BD and BSS are particular cases of our formulation.

$s_{i,mn}$ up to a possible shift, scaling and permutation:

$$\hat{s}_{i,mn} = w_{i,\,mn}^{\;j} * x_{j,mn} \approx c_i \cdot s_{\pi_i, m-\Delta_i^1 n-\Delta_i^2}, \tag{4}$$

where c_i are scaling factors, Δ_i^1, Δ_i^2 are integer shifts, and π_i is a permutation of $\{1, ..., N\}$. In the BSS case, the restoration is up to scaling and permutation:

$$\hat{s}_{i,mn} = w_i^{\;j} x_{j,mn} \approx c_i \cdot s_{\pi_i, mn}. \tag{5}$$

The restoration operator is called the *unmixing matrix*. When $N = M$, the unmixing matrix (w_{ij}) can be found by inverting the mixing matrix (a_{ij}). Henceforth, we will assume that $N = M$ and that the identifiability conditions hold.

3. SPARSE BD AND BSS

A popular technique for BSS and BD is the ML approach. ML deconvolution is performed by minimizing the minus-log-likelihood of the data w.r.t. the restoration operator. In the noise-free case, the normalized minus-log-likelihood function of the observed data $x_{i,mn}$ is given by [1]:

$$\ell(x; w) = -\frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log |\det W_{ij}(\theta, \eta)| \, d\theta d\eta + \frac{1}{M_X N_X} \sum_{m,n} \sum_{i=1}^{N} \varphi_i(y_{i,mn}), \tag{6}$$

where $W_{ij}(\theta, \eta) = \sum_{m,n} w_{ij,mn} e^{-\imath(m\theta + n\eta)}$ denotes the discrete space Fourier transform (DSFT) of $w_{ij,mn}$ w.r.t. m, n; $y_{i,mn} = w_{i,\,mn}^{\;j} * x_{j,mn}$ is a source estimate, $\varphi_i(s) = -\log p_i(s)$ and $p_i(s)$ is the probability density function (PDF) of the i-th source. We tacitly assume that all the expressions are well-defined and the source image is real and i.i.d.

Consistent estimator can be obtained by minimizing $\ell(x; w)$ even when $\varphi_i(s_i)$ is not exactly equal to $-\log p_i(s_i)$. Such quasi-ML estimation has been shown to be practical in instantaneous BSS [2, 6] and BD [7, 4] when the source PDF is unknown or not well-suited for optimization. For example, when the source is super-Gaussian (e.g. it is sparse or sparsely representable), a smooth approximation of the absolute value function is a good choice for $\varphi(s)$ [3]. This type of a prior of source distribution is especially convenient for the underlying optimization problem due to its convexity, and results in very accurate deconvolution (or separation). In addition, source sparsity in cases of pure BSS and SIMO BD allows to exploit simple geometric methods for separation [5, 6] or deconvolution [8].

However, natural images arising in the majority of BD and BSS applications can by no means be considered to be neither sparse, nor i.i.d., and thus the sparisity prior is not valid for most natural sources. On the other hand, it is very difficult to model actual distributions of natural images, which are often multi-modal and non-log-concave. This apparent gap between convenient models and real-world signals calls for an alternative approach.

4. SPARSIFICATION

While it is difficult to derive a prior suitable for natural images, it is much easier 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. Sparsifying transformations also have a decorrelation effect, allowing to use simple QML BD and BSS approaches, which assume i.i.d. sources.

The idea of *sparsification* by applying some transformation (e.g. Gabor-, wavelet- or wavelet packet transforms) on the mixtures was successfully exploited in BSS [5, 6]. However, these transformations were derived from empirical considerations. Here we present a method for finding optimal sparsifying transformations.

To begin, let us assume that there exists a known *sparsifying transformation* $\mathcal{T}_S$ that makes the sources sparse. In this case, a BD (or a BSS) algorithm is likely to produce a good estimate of the restoration operator $\mathcal{W}$ since the source properties are in accord with the sparsity prior. The problem is, however, that s is not available, and $\mathcal{T}_S$ can be applied only to the observation x. Hence, it is necessary that the sparsifying transformation commute with the convolutive mixture operation, i.e.

$$\mathcal{A}(\mathcal{T}_S s) = \mathcal{T}_S(\mathcal{A}s) = \mathcal{T}_S x, \tag{7}$$

such that applying the deconvolution (or separation) algorithm on $\mathcal{T}_S x$ is equivalent to deconvolving mixtures resulting from sparse sources $\mathcal{T}_S s$. A family of transformation obeying this property are linear shift-invariant (LSI) transformations[2], which can be described by a convolution kernel t_{mn}:

$$(\mathcal{T}_S s)_{i,mn} = t_{mn} * s_{i,mn} = t^{m'n'} s_{i,m-m'\, n-n'}, \tag{8}$$

such that

$$\mathcal{T}_S x = t^{m''n''} x_{i,m-m''\, n-n''} = t^{m''n''} a_i^{\;j,m'n'} s_{j,m-m''-m'\, n-n''-n'} = a_i^{\;j,m''n''} t^{m'n'} s_{j,m-m'-m''\, n-n'-n''} = (\mathcal{A}(\mathcal{T}_S s))_{i,mn}. \tag{9}$$

Thus, we obtain a general BD (or BSS) algorithm, which is not limited to sparse sources. We first sparsify the mixtures x using $\mathcal{T}_S$ (which has to be found in a way described in Section 4.2), and then apply the sparse BD (or BSS) algorithm to the result $\mathcal{T}_S x$. The obtained restoration operator $\mathcal{W}$ is then applied to the original (non-sparsified) mixtures x to produce the source estimate.

[2] In pure BSS problems, the sparsifying transformation needs to be lin-

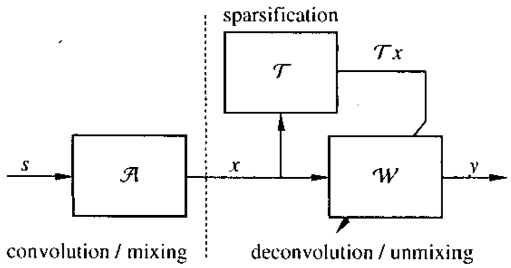


Fig. 1. Scheme of BSS / BD using sparsification.

4.1. Optimum sparsification

It is known that different classes of signals have some "natural" sparse representation. For example, 1D block signals become sparse after applying a discrete derivative; acoustic signals usually appear sparse after performing a short time Fourier transform, etc [6]. However, such a sparsifying transformation is not necessarily the most suitable in case of a general class of signals.

By definition, the sparsifying transformation $\mathcal{T}_S$ must produce a sparse representation of the source; it is obvious that $\mathcal{T}_S$ would usually depend on the sources s, and also, $\mathcal{T}_S$ does not necessarily have to be invertible, since we use it as a pre-processing of the data and hence never need its inverse. A general way to find $\mathcal{T}_S$, is by maximizing some sparsity criterion of $\mathcal{T}_S s$. Particularly, the ℓ_1 norm (the second term in the quasi ML function) can be used as the objective function, i.e.

$$\mathcal{T}_S = \operatorname*{argmin}_t \sum_{i=1}^{N} \sum_{m,n} |t_{mn} * s_{i,mn}| \quad s.t. \quad \|t\|_2^2 = 1, \quad (10)$$

where the constant energy constraint is imposed on t to avoid the trivial (zero kernel) solution. This is an optimization problem with nonlinear convex objective and a non-convex quadratic constraint; one of the most common methods to solve such a problem is by using the penalty method.

In the SISO BD problem, the optimal sparsifying transformation can be found according to $\mathcal{T}_S = \operatorname{argmin}_t \ell(s;t)$, as proposed in [9]. As the result, $w_{mn} = \delta_{mn}$ is a local minimizer of $\ell(\mathcal{T}_S s; w)$, and due to the equivariance property[3], $w_{mn} = (\mathcal{A}^{-1})_{mn}$ is a local minimizer of $\ell(\mathcal{A}\mathcal{T}_S s; w) = \ell(\mathcal{T}_S x; w)$, i.e., the QML estimate of the restoration kernel given the sparsified observation $\mathcal{T}_S x$ is the inverse of $\mathcal{A}$ (see details in [9]).

ear and not necessarily shift-invariant: $(\mathcal{T}s)_{i,mn} = t_{mn}^{m'n'} s_{i,m'n'}$ For example, wavelet packets were used for sparsification in [5, 6].

[3]Equivariance implies that for any invertible $\mathcal{B}$, the estimator $\hat{\mathcal{W}}(x)$ of the restoration operator $\mathcal{W}$ given the observation x, obtained by minimization of $\ell(x,w)$ obeys: $\hat{\mathcal{W}}(\mathcal{B}x) = \mathcal{B}^{-1}\hat{\mathcal{W}}(x)$, i.e. $\mathcal{W}$ form a group [10].

4.2. Learning the optimum sparsification kernel

Since the source images S are not available, computation of the sparsifying kernel by the procedure described above is possible only theoretically. 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, and assume that the unknown sources $s_1, ..., s_N$ belong to $\mathcal{C}$. We can find images $u_1, ..., u_{N_T} \in \mathcal{C}$ and use them to find the optimal sparsifying transformation of $s_1, ..., s_N$. Optimization problem (10) becomes in this case

$$\mathcal{T}_U = \operatorname*{argmin}_t \sum_{i=1}^{N_T} \sum_{m,n} |t_{mn} * u_{i,mn}| \quad s.t. \quad \|t\|_2^2 = 1, \quad (11)$$

i.e. t is required to be the optimal sparsifying kernel for all $u_1, ..., u_{N_T}$ simultaneously. The images $u_1, ..., u_{N_T}$ constitute a *training set*, and the process of finding $\mathcal{T}_U$ is called *training*. Given that the images in the training set are "sufficiently similar" to s, the optimal sparsifying transformation $\mathcal{T}_U$ obtained by training is similar enough to $\mathcal{T}_S$.

5. RESULTS

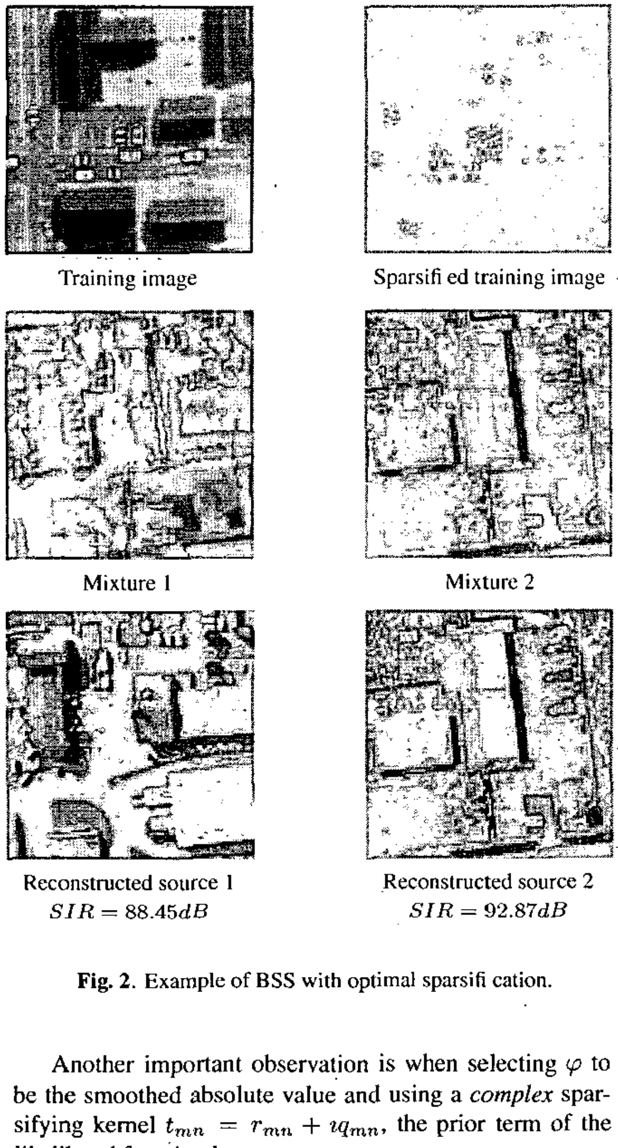
In the first experiment we exemplify the use of sparsification in pure BSS with $N = M = 2$. The sources were two 100×100 aerial images of San Francisco metro area. The training image was a synthetic aerial image drawn in Adobe PhotoShop. The optimal sparsifying LSI transformation was found by solving (11) with MATLAB constrained optimization solver `fmincon` and was very close to the 2D corner detector. Separation was carried out using the relative Newton algorithm [3] with sparsity prior.

The separation results are presented in Figure 2 (the images were normalized). Note the very high signal-to-interference ratio (SIR), significantly higher than typical SIR values obtained using wavelet transform as in [6].

The second experiment demonstrates SISO deconvolution. The source was an aerial image, blurred by a Lorenzian-shaped kernel, typical of atmospheric scattering. The training image was the same as in the first experiment. Blind deconvolution was performed with a 3×3 FIR kernel using the relative Newton algorithm described in [4]. The restoration results are presented in Figure 3.

6. CONCLUSIONS

The proposed approach allows the extension of sparse BD and BSS algorithms to arbitrary sources by using sparsifying transformations. An interesting observation is that in many natural images the optimal sparsifying LSI transformation is the 2D corner detector.



Training image — Sparsified training image

Mixture 1 — Mixture 2

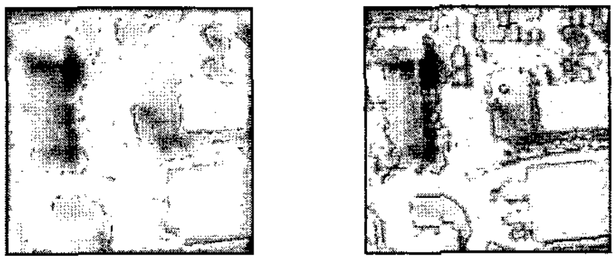
Reconstructed source 1 $SIR = 88.45dB$ — Reconstructed source 2 $SIR = 92.87dB$

Fig. 2. Example of BSS with optimal sparsification.

Another important observation is when selecting φ to be the smoothed absolute value and using a *complex* sparsifying kernel $t_{mn} = r_{mn} + \imath q_{mn}$, the prior term of the likelihood function becomes

$$\sum_{mn} \varphi((t * y)_{mn}) = \sum_{mn} \sqrt{(r * y)^2_{mn} + (q * y)^2_{mn} + \epsilon}, \quad (12)$$

which is a generalization of the discrete 2D *total variation* (TV) norm. The TV norm found to be a successful prior in numerous studies related to signal restoration and denoising (e.g. [11, 12]), is obtained when r_{mn} and q_{mn} are chosen to be discrete derivatives in x- and y-direction.



Blurred source — Reconstructed source $SIR = 15.7dB$

Fig. 3. Example of SISO BD with optimal sparsification.

7. REFERENCES

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

[2] 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.

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

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

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

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

[7] 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.

[8] D. Luengo, I. Santamaría, J. Ibá nez, L. Vielva, and C. Pantaleón, "A fast blind SIMO channel identification algorithm for sparse sources," *IEEE Signal Proc. Letters*, vol. 10, no. 5, pp. 148 – 151, May 2003.

[9] 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.

[10] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, 1996.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[12] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," in *Proc. ONR Workshop*, 1996.