Optimal Sparse Representations for Blind Deconvolution of Images

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

Technion - Israel Institute of Technology, Department of Electrical Engineering,
32000 Haifa, Israel
{alexbron, bronstein}@ieee.org, {mzib, zeevi}@ee.technion.ac.il

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 in modelling the log probability density function, which is suitable for sparse sources. We propose a method of sparsification, which allows blind deconvolution of sources with arbitrary distribution, and show how to find optimal sparsifying transformations by training.

1 Introduction

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

According to the convolution model, the observed sensor image X is created from the *source image S* passing through an LSI system characterized by the point spread function A, $X = A * S$. We assume that the action of A is invertible (at least approximately), i.e. there exists some other kernel W such that $A * W \approx \delta$. This assumption holds well especially in the case of blurring kernels resulting from scattering (such kernels are usually Lorenzian-shaped and their inverse can be approximated by small FIR kernels). The aim of BD is to find such a *deconvolution (restoration)* kernel W that produces an estimate $\tilde{S}$ of S up to integer shift and scaling factor: $\hat{S}_{mn} = (W * X)_{mn} \approx c \cdot S_{m-\Delta_M, n-\Delta_N}$.

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

2 QML blind deconvolution

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

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

where $\varphi(s) = -\log p_s(s)$, $p_s(s)$ stands for the source probability density function (PDF), $M_X \times N_X$ is the observation sample size, and $\mathcal{F}W(\xi, \eta)$ denotes the Fourier transform of W_{mn}. We will henceforth assume that W is an FIR kernel, supported on $[-M, ..., M] \times [-N, ..., N]$. Cost functions similar to (1) were also obtained in the 1D case using negative joint entropy and information maximization considerations [5]. In practice, it is difficult to evaluate the first term of $\ell(X; W)$ containing the integral. However, it can be approximated with any desired accuracy using FFT.

Source images arising in most applications have usually multi-modal non-log-concave distributions. These are difficult to model and are not suitable for optimization. However, consistent estimator of S can be obtained by minimizing $\ell(X; W)$ even when $\varphi(s)$ is not exactly equal to $-\log p_S(\cdot)$. Such *quasi-ML estimation* has been shown to be practical in instantaneous blind source separation [6, 3, 7] and blind deconvolution of time signals [4]. For example, when the source is super-Gaussian (sparse), a smooth approximation of the absolute value function is a good choice for $\varphi(s)$ [8, 9]. Although natural images are usually far from being sparse, they can be transformed into a space of a sparse representation. We will therefore focus our attention on modelling super-Gaussian distributions using a family of convex smooth functions

$$\varphi_\lambda(s) = |s| - \lambda \log \left(1 + \frac{|s|}{\lambda}\right) \tag{2}$$

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

The gradient of $\ell(X; W)$ w.r.t W_{ij} is given by (for derivation see [10]):

$$\frac{\partial \ell}{\partial W_{ij}} = -Q_{-i,-j} + \frac{1}{M_X N_X} \sum_{m,n} \varphi'(Y_{mn}) \, X_{m-i,n-j}, \tag{3}$$

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

$$\frac{\partial^2 \ell}{\partial W_{ij} \partial W_{kl}} = \frac{1}{M_X N_X} \sum_{m,n} \varphi''(Y_{mn}) \, x_{m-i,n-j} x_{m-k,n-l} + R_{-(i+j),-(k+l)}, \tag{4}$$

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

3 The fast relative Newton method

A fast relative optimization algorithm for blind source separation, based on the Newton method was introduced in [3]. In [4] it was used for BD of 1D signals. Here we use the

relative optimization framework for BD of images. The main idea of relative optimization is to iteratively produce an estimate of the source signal and use it as the observed signal at the subsequent iteration:

Relative optimization algorithm

1. Start with initial estimates of the restoration kernel $W^{(0)}$ and the source $X^{(0)} = W^{(0)} * X$.
2. For $k = 0, 1, 2, ...$, 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 $\hat{W} = W^{(0)} * ... * W^{(k)}$, 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. It can be seen easily that the relative optimization algorithm has uniform performance, i.e. its step at iteration k depends only on $A * W^{(0)} * ... * W^{(k-1)}$.

Step 4 can be carried out using any unconstrained optimization algorithm. Particulary, it was found that a single Newton step can be used, yielding very fast convergence. However, its use is limited to small values of M, N and M_X, N_X due to the complexity of Hessian construction, and solution of the Newton system. This complexity can be significantly reduced if special Hessian structure is exploited. Near the solution point, $X^{(k)} \approx cS$, hence $\ell(X; \delta)$ evaluated at each relative Newton iteration becomes approximately $\ell(cS; \delta)$. For a zero-mean i.i.d. source and sufficiently large sample size (in practice, $M_X N_X > 10^2$), the Hessian has an approximately diagonal-anti-diagonal form with ones on the anti-diagonal [10]. Using this approximation, only the main diagonal of the Hessian matrix has to be evaluated at each iteration, and the solution of the Newton system $\nabla^2 \ell d = -\nabla \ell$ separates into the set of 2×2 systems of the form

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

for $(i, j) \neq \mathbf{0}$, and an additional equation $\nabla \ell_{00} = -\nabla^2 \ell_{0000} \, d_{00}$. We will henceforth refer to this approximate Newton step as to the *fast relative Newton method*, since its complexity is of the same order as that of the gradient-based methods.

4 Optimal sparse representations of images

The QML framework presented in Section 2 is valid for sparse sources; this type of a prior of source distribution is especially convenient since the prior term in the underlying optimization problem is convex. In addition, deconvolution of sparse sources is reported to be very accurate. However, natural images arising in the majority of BD applications can by no means be considered to be sparse in their native space of representation (usually, they are sub-Gaussian), and thus such a prior is not valid for "real-life"

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 a simple model and the real world calls for an alternative approach. In this section, we show how to overcome this problem using sparse representation.

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. The idea of *sparsification* was successfully exploited in BSS [8, 7, 11, 10]. It was shown in [10] that even such simple transformation as a discrete derivative can make the image sparse. However, most of these transformations were derived from empirical considerations. Here we present a criterion for finding optimal sparsifying transformations.

Let 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 accord 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. Hence, it is necessary that the sparsifying transformation commute with the convolution operation, i.e. $(\mathcal{T}S) * A = \mathcal{T}(S * A) = \mathcal{T}X$, such that applying $\mathcal{T}$ to X is equivalent to applying it to S. Obviously, $\mathcal{T}$ must be a shift-invariant (SI) transformation.[1]

Using the most general nonlinear form of $\mathcal{T}$, we have a wide class of sparsifying transformations. An important example is a family of SI transformations of the form $(\mathcal{T}S)_{mn} = \sqrt{(T_1 * S)^2_{mn} + (T_2 * S)^2_{mn}}$, where T_1, T_2 are some convolution kernels. After sparsification with $\mathcal{T}$, the prior term of the likelihood function becomes

$$\sum_{m,n} |(\mathcal{T}Y)_{mn}| = \sum_n \sqrt{(T_1 * Y)^2_{mn} + (T_2 * Y)^2_{mn}}, \qquad (5)$$

which is a generalization of the 2D *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 [12–14], and was also used by Chan and Wong as a regularization in BD [1], is obtained when T_1, T_2 are chosen to be discrete x- and y-directional derivatives.

For simplicity, we limit our attention in this study to linear shift-invariant (LSI) transformations, i.e. $\mathcal{T}$ that can be represented by convolution with a *sparsifying kernel* $\mathcal{T}S = T * S$. 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 in Section 4.1), and then apply the sparse BD algorithm on the result $X * T$. The obtained restoration kernel W is then applied to X to produce the source estimate.

An important practical issue is how to find the kernel T. By definition T must produce a sparse representation of the source; it is obvious that T would usually depend on S, and also, T does not necessarily have to be stable, since we use it as a pre-processing of the data and hence never need its inverse. Let assume that the source S is given

---

[1] In BSS problems, the sparsifying transformation needs to be linear and not necessarily shift-invariant, e.g. wavelet packets were used for sparsification in [8, 7].

(this is, of course, impossible in reality; the issue of what to use instead of S will be addressed in Section 4.1). It is desired that the unity restoration kernel δ_{mn} be a local minimizer of the QML function, given the transformed source $S * T$ as an observation, i.e.: $\nabla \ell(\delta_{mn}; S * T) = 0$. Informally, this means that $S * T$ optimally fits the sparsity prior (at least in local sense). Due to the equivariance property, $\nabla \ell(\delta_{mn}; S * T) = 0$ is equivalent to $\nabla \ell(T; S) = 0$. In other words, we can define the following optimization problem:

$$\min_{T} \ell(T; S), \tag{6}$$

whose solution is the optimal sparsifying kernel for S. This problem is equivalent to the problem solved for deconvolution itself. The log-spectrum term in $\ell(T; S)$ eliminates the trivial solution $T = 0$.

4.1 Finding the sparsifying kernel by training

Since the source image S is not available, computation of the sparsifying kernel by the procedure described before is possible only theoretically. However, empirical results indicate that for images belonging to the same class, the proper sparsifying kernels are sufficiently similar.

Let $\mathcal{C}_1$ denote a class of images, e.g. human faces, and assume that the unknown source S belongs to $\mathcal{C}_1$. We can find find images $S^{(1)}, S^{(2)}, ..., S^{(N_T)} \in \mathcal{C}_1$ and use them to find the optimal sparsifying kernel of S. Optimization problem (6) becomes in this case

$$\min_{T} \left\{ -\frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log |\mathcal{F}T(\xi, \eta)| \ d\xi d\eta + \frac{1}{M_X N_X} \sum_{i=1}^{N_T} \sum_{m,n} \varphi((T * S^{(i)})_{mn}) \right\},$$

i.e. T is required to be the optimal sparsifying kernel for all $S^{(1)}, S^{(2)}, ..., S^{(N_T)}$ simultaneously. The images $S^{(1)}, S^{(2)}, ..., S^{(N_T)}$ 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 S, the optimal sparsifying kernel obtained from training is similar enough to T_S.

5 Simulation results

The QML-based deconvolution approach was tested on simulated data under zero-noise conditions. 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 dB units [10]. In the first test, a real aerial photo of a factory was used as the source image, and a synthetic one (drawn using PhotoShop) as the training image (Figure 1). A 3×3 sparsifying kernel is found by training on a single image, then the same kernel is used as a pre-processing 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 Lorenzian-shaped blurring kernel. Deconvolution kernel was of size
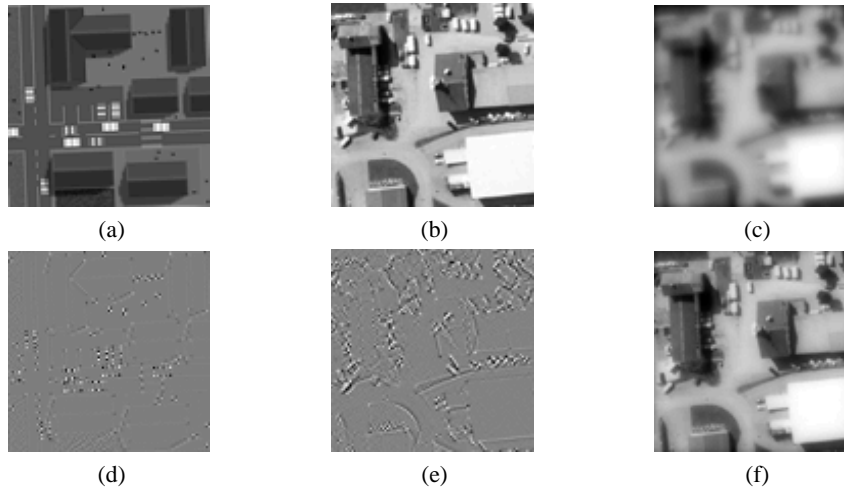
Fig. 1. (a) training synthetic image, (b) source aerial image S, (c) blurred image $S * A$, (d) sparsified training image, (e) sparsified source, (f) restored image.

Table 1. SIR, SIR$_\infty$ and PSIR of the restored images.

Source		SIR [dB]	SIR$_\infty$ [dB]	PSIR [dB]
S_1	Susy	17.7994	22.2092	22.6132
S_2	Aerial	17.0368	23.5482	9.6673
S_3	Gabby	19.3249	23.8109	29.8316
S_4	Hubble	14.5152	17.1552	19.8083

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 $SIR = 20.1561$ dB, $SIR_\infty = 25.7228$ dB.

In the second test, four natural source images were used: S_1 (Susy), S_2 (Aerial), S_3 (Gabby) and S_4 (Hubble) (Figure 2, top). Nearly-stable Lorenzian-shaped kernels were used to model the convolution system. This type of kernels characterizes scattering media, such as biological fluids and aerosols found in the atmosphere [15]. The observed images are depicted in Figure 2 (middle). Fast relative Newton step with kernel size set to 3×3 was used in this experiement. 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 sec). The restored images are depicted in Figure 2 (bottom). Restoration quality results in terms of SIR, SIR$_\infty$ and PSIR are presented in Table 1.
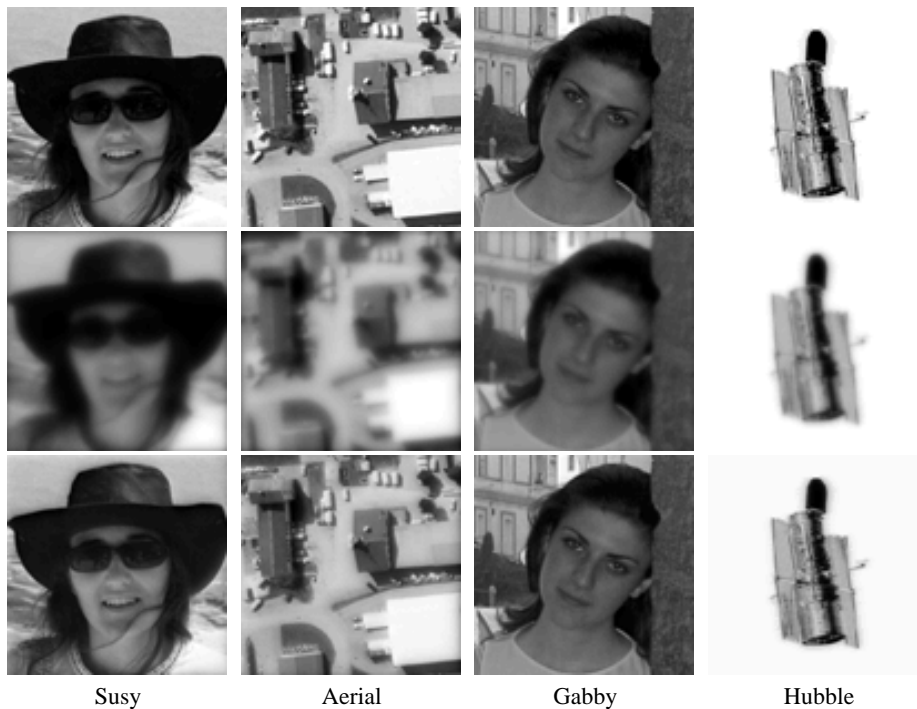
| Susy | Aerial | Gabby | Hubble |

Fig. 2. Top: source images used in the simulations; middle: blurred images (observations); bottom: restored images.

6 Conclusion

The QML framework, recently presented in the context of 1D deconvolution [4] is also attractive for BD of images. We presented an extension of the relative optimization approach to QML BD in the 2D case and studied the relative Newton method as its special case. Similarly to previous works addressing deconvolution in other spaces (e.g. [16]) and our studies of using sparse representation in the context of BBS, in BD the sparse prior appears very efficient as well. We showed a training approach for finding optimal sparse representations, yielding a general-purpose BD method. A particular class of LSI sparsifying transformations generalizes some previous results such as the total variation prior [12–14]. We also showed how optimal sparsifying transformations can be found by training.

Simulation results demonstrated the efficiency of the proposed methods. Although we have limited our attention to noiseless BD, it is important to emphasize that the sparsification framework is applicable to the noisy case as well. Sparsifying kernels are typically high-pass filters, since by their very nature sparse signals have high-frequency components. Such kernels have the property of amplifying noise – thus in case when the signal is contaminated by additive noise, using such kernels in undesired. To cope

with the problem of noise, the signal should be smoothed with a low-pass filter F and afterwards 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 $T * F$.

Potential applications of our approach are in optics, remote sensing, microscopy and biomedical imaging, especially where the SNR is moderate. This approach is especially accurate and efficient in problems involving slowly-decaying (e.g. Lorenzian-shaped) kernels, which can be approximately inverted using a kernel with small support. Such kernels are typical of imaging through scattering media.

References

1. Chan, T.F., Wong, C.K.: Total variation blind deconvolution. (IEEE. Trans. Image Proc.) To appear.
2. Kaftory, R., Sochen, N.A., Zeevi, Y.Y.: Color image denoising and blind deconvolusion using the beltramy operator. In: Proc. 3rd Intl. Symposium on Image and Sig. Proc. and Anal. (2003) 1–4
3. Zibulevsky, M.: Sparse source separation with relative Newton method. In: Proc. ICA2003. (2003) 897–902
4. Bronstein, A.M., Bronstein, M., Zibulevsky, M.: Blind deconvolution with relative Newton method. Technical Report 444, Technion, Israel (2003)
5. Amari, S.I., Cichocki, A., Yang, H.H.: Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach. In: Proc. SYSID. (1997) 1057–1062
6. Pham, D., Garrat, P.: Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach. IEEE Trans. Sig. Proc. **45** (1997) 1712–1725
7. Kisilev, P., Zibulevsky, M., Zeevi, Y.: Multiscale framework for blind source separation. JMLR (2003 (in press))
8. Zibulevsky, M., Pearlmutter, B.A.: Blind source separation by sparse decomposition. Neural Computation **13** (2001)
9. Zibulevsky, M., Kisilev, P., Zeevi, Y.Y., Pearlmutter, B.A.: Blind source separation via multinode sparse representation. In: Proc. NIPS. (2002)
10. Bronstein, A.M., Bronstein, M., Zeevi, Y.Y., Zibulevsky, M.: Quasi-maximum likelihood blind deconvolution of images using sparse representations. Technical report, Technion, Israel (2003)
11. Lewicki, M.S., Olshausen, B.A.: A probabilistic framework for the adaptation and comparison of image codes. J. Opt. Soc. Am. A **16** (1999) 1587–1601
12. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
13. Blomgren, P., Chan, T.F., Mulet, P., Wong, C.: Total variation image restoration: numerical methods and extensons. In: Proc. IEEE ICIP. (1997)
14. Chan, T.F., Mulet, P.: Iterative methods for total variation image restoration. SIAM J. Num. Anal **36** (1999)
15. Moscoso, M., Keller, J.B., Papanicolaou, G.: Depolarization and blurring of optical images by biological tissue. J. Opt. Soc. Am. A **18** (2001) 948–960
16. Banham, M.R., Katsaggelos, A.K.: Spatially adaptive wavelet-based multiscale image restoration. IEEE Trans. Image Processing **5** (1996) 619–634