Blind deconvolution using the relative Newton method

Alexander M. Bronstein, Michael M. Bronstein, and Michael Zibulevsky

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

Abstract. We propose a relative optimization framework for quasi maximum likelihood blind deconvolution and the relative Newton method as its particular instance. Special Hessian structure allows its fast approximate construction and inversion with complexity comparable to that of gradient methods. The use of rational IIR restoration kernels provides a richer family of filters than the traditionally used FIR kernels. Smoothed absolute value and the smoothed deadzone functions allow accurate and robust deconvolution of super- and sub-Gaussian sources, respectively. Simulation results demonstrate the efficiency of the proposed methods.

1 Introduction

Blind deconvolution problem appears in various applications related to acoustics, optics, geophysics, communications, control, etc. In the general setup of the single-channel blind deconvolution, the observed sensor signal x is created from the *source signal* s passing through a causal convolutive system

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

with impulse response a and additive sensor noise u. The setup is termed *blind* if only x is accessible, whereas no knowledge on a, s and u is available. The problem of blind deconvolution aims to find such a deconvolution (or restoration) kernel w, that produces a possibly delayed waveform-preserving source estimate $\hat{s}_n = (w * x)_n \approx c \cdot s_{n-\Delta}$, where c is a scaling factor and Δ is an integer shift. Equivalently, the *global system response* $g = a * w$ should be approximately a Kroenecker delta, up to scale factor and shift. A commonly used assumption is that s is non-Gaussian.

Many blind deconvolution 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 [1–5]. Many of these methods use batch mode calculations and usually suffer from high computational complexity. Conversely, a wide class of the so-called *Bussgang-type* algorithms estimate directly the inverse kernel $W(z) = A^{-1}(z)$ by minimizing some functional using gradient descent iterations. 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 [6]. A review of these algorithms can be found in [7].

In their fundamental work, Amari *et al.* [8] introduced an iterative time-domain blind deconvolution algorithm based on the natural gradient learning, which was originally used in context of blind source separation [9–11] and became very attractive due to the so-called *uniform performance property* [11]. The natural gradient algorithm estimates directly the restoration kernel and allows real-time processing. Efficient frequency-domain implementation was presented in [12].

Natural gradient demonstrates significantly higher performance compared to gradient descent. In this work, we present a blind deconvolution algorithm based on the relative Newton method, which brings further acceleration. The relative Newton algorithm was originally proposed in the context of sparse blind source separation in [13, 14]. We utilize special Hessian structure to derive a fast version of the algorithm with complexity comparable to that of gradient methods. We focus our attention on a batch mode single-channel blind deconvolution algorithm with FIR restoration kernel and outline the use of IIR kernels. We use the smoothed absolute value for deconvolution of super-Gaussian sources, and propose the smoothed deadzone linear function for sub-Gaussian sources.

2 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 [8]

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

where $y = w * x$ is a source estimate; $\varphi(s) = -\log p\left(s\right)$, where $p(s)$ is the probability density function (PDF) of the source s. We assume that w is an FIR kernel supported on $n = -N, ..., N$, and denote its length by $K = 2N+1$. We will also assume without loss of generality that s is zero-mean. Cost function (2) can be also derived using negative joint entropy and information maximization considerations. 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 using the FFT.

Consistent estimator can be obtained by minimizing $\ell(x;w)$ even when $\varphi(s)$ is not exactly equal to $-\log p\left(s\right)$. Such *quasi ML* estimation has been shown to be practical in instantaneous blind source separation when the source PDF is unknown or not well-suited for optimization [13]. The choice of $\varphi(s)$ and the consistency conditions of the QML estimator are discussed in Section 5.

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

$$g_i = -q_{-i} + \frac{1}{T}\sum_{n=0}^{T-1}\varphi'(y_n)\, x_{n-i}, \tag{3}$$

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

$$H_{ij} = r_{-(i+j)} + \frac{1}{T}\sum_{n=0}^{T-1} \varphi''(y_n)\, x_{n-i}x_{n-j}, \tag{4}$$

where r_n is the inverse DFT of W_k^{-2} (for derivation see [15]). Both the gradient and the Hessian can be evaluated efficiently using FFT.

3 Relative optimization

Here we introduce a relative optimization framework for blind deconvolution. 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 [14] in the context of blind source separation.

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. 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 $g^{(k-1)} = a * w^{(0)} * ... * w^{(k-1)}$, since the update in Step 5 does not depend explicitly on a, but on the currents 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.

3.1 Fast relative Newton step

A Newton iteration can be used in Step 4 of the relative optimization algorithm, yielding very fast convergence. However, its practical use is limited to small values of N and T, 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 $\nabla^2 \ell(x;\delta)$ 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 ones on the anti-diagonal.*

Proof. Substituting $w=\delta$, $x=cs$ and $y=\delta*x=cs$ into $\ell(x;w)$ in (4), one obtains

$$H_{ij}=\delta_{i+j}+\frac{1}{T}\sum_{n=0}^{T-1}\varphi''(cs_n)\, cs_{n-i}\, cs_{n-j}.$$

For a large sample size T, the sum approaches the corresponding expectation value. Invoking the assumption that s is zero-mean i.i.d., the off-diagonal and off-anti-diagonal elements of H vanish. □

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

$$\begin{pmatrix} H_{-k,-k} & 1 \\ 1 & H_{kk} \end{pmatrix}\begin{pmatrix} d_{-k} \\ d_k \end{pmatrix}=-\begin{pmatrix} g_{-k} \\ g_k \end{pmatrix} \tag{5}$$

for $k=1,...,K$, and an additional equation

$$H_{00}\, d_0=-g_0. \tag{6}$$

In order to guarantee decent direction and avoid saddle points, we force positive definiteness of the Hessian by inverting the sign of negative eigenvalues in system (5) and forcing small eigenvalues to be above some positive threshold. Computation of the Hessian approximation involves evaluation of its main diagonal only, which is of the same order as gradient computation. Approximate solution of the Newton system requires $\mathcal{O}(N)$ operations.

4 IIR restoration kernels

When the convolution system $A(z)$ has zeros close to the unit circle, the restoration kernel $W(z)$ has to be long in order to achieve good restoration quality. Therefore, when $W(z)$ is parameterized by the set of FIR coefficients $w_{-N},...,w_N$, the number of parameters to be estimated is large. Under such circumstances, it might be advantageous to use a rational IIR restoration kernel of the form

$$W(z)=\frac{h_{-N}z^{N}+...+h_N z^{-N}}{(1+b_1z^{-1}+...+b_Mz^{-M})\,(1+c_1z+...+c_Lz^{L})},$$

parameterized by $h_{-N},...,h_N$, $b_1,...,b_M$ and $c_1,...,c_L$. The asymptotic Hessian of $\ell(x;h,b,c)$ 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 sparse structure depicted in Figure 1 (right) [16]. 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. In both cases, approximate Hessian evaluation and Newton system solution have the complexity of a gradient descent iteration.

Fig. 1. Hessian structure 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.

5 The choice of $\varphi(s)$

The choice of $\varphi(s)$ is limited first of all by the QML estimator consistency (or asymptotical stability) conditions, which guarantee that $w = a^{-1}$ is a stable minimum of $\ell(x;w)$ in the limit $T \to \infty$ [16].

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 [17, 18]. We use the following function [14]

$$\varphi_\lambda^{\mathrm{ABS}}(s) = |s| - \lambda \log\left(1 + \frac{|s|}{\lambda}\right), \tag{7}$$

which in the limit $\lambda \to 0^+$ yields an asymptoticall stable QML estimator if $\mathbb{E}|s| < 2\sigma^2 p(0)$, where $\sigma^2 = \mathbb{E}s^2$ [16]. In the particular case of strictly *sparse* sources, i.e. such sources that take the value of zero with some non-zero probability, *super-efficiency* is achieved in the limit $\lambda \to 0^+$ and in absence of noise [16].

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

$$\varphi_\mu^{\mathrm{PWR}}(s) = |s|^\mu \tag{8}$$

with the parameter $\mu > 2$ is usually a good choice for $\varphi(s)$. This function yields an asymptotically stable estimator for $\mathbb{E}|s|^{\mu+2} < (\mu+1)\sigma^2\ \mathbb{E}|s|^\mu$, which for the particular choice of $\mu = 4$ corresponds to negative kurtosis excess [16]. An increase of μ usually yields better performance. However, it is obvious that large values of μ imply high sensitivity to outliers due to the high powers. As a remedy, we propose to replace the power function with the *deadzone linear* function of the form

$$\varphi_\mu^{\mathrm{DZ}}(s) = \mu \cdot \max\left\{|s| - 1, 0\right\}, \tag{9}$$

which is often used for regression, data fitting and estimation [19]. 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

$$\varphi_{\lambda,\mu}^{\mathrm{DZ}}(s) = \frac{\mu}{2}\left(\varphi_\lambda^{\mathrm{ABS}}(s-1) + \varphi_\lambda^{\mathrm{ABS}}(s+1)\right), \tag{10}$$

where the parameter λ controls the smoothness.

When the source PDF is compactly supported (e.g. digital communication signals), both the power function and the smoothed deadzone linear function yield super-efficient estimators in the limit $\mu \to \infty$. When in addition the source signal takes the values at the extremal points of the interval, s_{ext}, with some non-zero probability ρ, the use of the smoothed deadzone linear function achieves super-efficiency with $\lambda \to 0^+$ and *finite* μ. In the latter case, the estimator is asymptotically stable if $\mu\rho > 1$ and $2\sigma^2 \max\left\{(\mu\rho - 1)^2, 1\right\} > s_{\mathrm{ext}}^2 \lambda\mu\rho$ [16].

6 Numerical results

The convolution system was modelled by the empirically measured digital microwave channel impulse response from [20]. Two 10^4 samples long 2-level PAM and sparse normal i.i.d. processes were used as inputs. Input SNRs from 10 to 100 dB were tested. 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 algorithm [12]. In both the power function with $\mu = 4$ was used for the PAM signal, whereas for the sparse source the smoothed absolute value with $\lambda = 10^{-2}$ was used in the relative Newton algorithm and the exact absolute value was used in the FDBD algorithm. In case of the PAM signal, performance was also compared to CMA with $p = 2$. Figure 2 (left) presents the restoration SIR averaged over 10 independent Monte-Carlo runs, as a function of the input SNR (95% confidence intervals are indicated on the plot). For SNR higher than 20 dB, the block relative Newton algorithm demonstrates an average improvement of about 4 dB compared to other methods for the PAM sources and about 7 dB for the sparse sources. Good restoration quality is obtained for SNR starting from 10 dB. Figure 2 (right) depicts the convergence of the compared algorithms, averaged over 10 independent runs with input SNR set to 20 dB.

Figure 3 (left) shows the SIR for the PAM source, averaged over 20 independent Monte-Carlo runs, wherein $\varphi(s)$ is chosen as the power function and the smoothed deadzone linear function. 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). Unlike the power function, the proposed smoothed deadzone linear function appears to yield higher performance and demonstrates negligible sensitivity to outliers.

Advantages of an IIR restoration kernel can be seen in Figure 3 (right), which depicts the SIR for the sparse source, averaged over 10 Monte-Carlo runs, as a function of the number of optimization variables for different assignments of the degrees of freedom to restoration kernel numerator and denominator. A practically ideal SIR was achieved by the all-pole IIR kernel starting from 8 degrees of freedom. Additional simulation results can be found in [15, 18].

7 Conclusion

We have presented a relative optimization framework for QML single channel blind deconvolution and studied the relative Newton method as its particular instance. Diagonal-

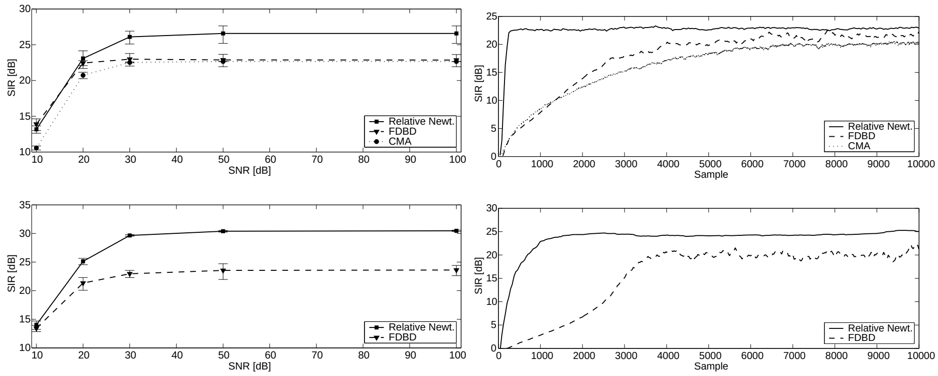

Fig. 2. Left: average SIR as a function of input SNR; right: average convergence in terms of SIR for input SNR of 20 dB. Top: 2-level PAM source; bottom: sparse source.

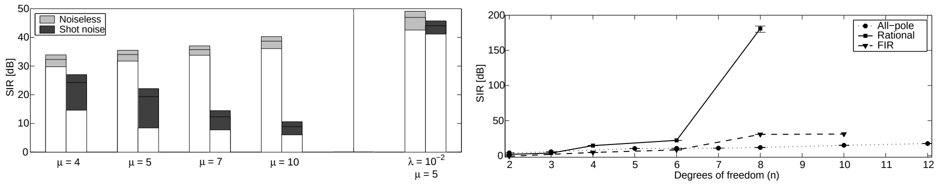

Fig. 3. Left: Average restoration SIR for the power function (left), and the smoothed deadzone linear function (two rightmost bars), with and without the presence of shot noise. Right: SIR as a function of degrees of freedom for different restoration kernel configurations.

anti-diagonal structure of the Hessian in the proximity of the solution allowed 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 often allow to reduce the optimization problem size. We also propose the use of the deadzone linear function for sub-Gaussian sources, which is significantly less sensitive to outliers than the commonly used non-linearities, and achieves super-efficient estimation in the absence of noise.

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. We are currently working on extending the presented approach to the multichannel and complex cases.

Acknowledgment

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.

References

1. Tong, L., Xu, G., Kailath, T.: Blind identification and equalization based on secondorder statistics: A time domain approach. IEEE Trans. Inform. Theory **40** (1994) 340–349
2. Gurelli, M., Nikias, C.: EVAM: An eigenvectorbased algorithm for multichannel bind deconvolution of input colored signals. IEEE Trans. Signal Processing **43** (1995) 134–149
3. Xu, G., Liu, H., Tong, L., Kailath, T.: Least squares approach to blind channel identification. IEEE Trans. Sig. Proc. **43** (1995) 2982–2993
4. Hua, Y.: Fast maximum likelihood for blind identification of multiple FIR channels. IEEE Trans. Sig. Proc. **44** (1996) 661–672
5. Gorokhov, A., Loubaton, P., Moulines, E.: Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach. In: Proc. ICASSP. Volume 5. (1996) 2415–2418
6. Godard, D.N.: Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE Trans. Commun. **28** (1980) 1867–1875
7. Chi, C.Y., Chen, C.Y., Chen, C.H., Feng, C.C.: Batch processing algorithms for blind equalization using higher-order statistics. IEEE Sig. Proc. Magazine (2003) 25–49
8. 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
9. Cichocki, A., Unbehauen, R., Rummert, E.: Robust learning algorithm for blind separation of signals. Electronics Letters **30** (1994) 1386–1387
10. Amari, S.I., Douglas, S.C., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. Advances in Neural Information Processing Systems **8** (1996) 757–763
11. Cardoso, J.F., Laheld, B.: Equivariant adaptive source separation. IEEE Trans. Sig. Proc. **44** (1996) 3017–3030
12. Joho, M., Mathis, H., Moschytz, G.S.: On frequency-domain implementations of filtered-gradient blind deconvolution algorithms. In: Proc. Asilomar Conf. Signals, Syst., Comput. (2002)
13. 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
14. Zibulevsky, M.: Sparse source separation with relative Newton method. In: Proc. ICA2003. (2003) 897–902
15. Bronstein, A.M., Bronstein, M., Zibulevsky, M.: Blind deconvolution with relative Newton method. Technical Report 444, Technion, Israel (2003)
16. Bronstein, A.M., Bronstein, M.M., Zibulevsky, M.: Relative optimization for blind deconvolution. IEEE Sig. Proc. (2004) Submitted. [Online] http://visl.technion.ac.il/bron/alex.
17. Zibulevsky, M., Pearlmutter, B.A., Bofill, P., Kisilev, P.: Blind source separation by sparse decomposition. In Roberts, S.J., Everson, R.M., eds.: Independent Components Analysis: Princeiples and Practice. Cambridge University Press (2001)
18. Bronstein, A.M., Bronstein, M.M., Zibulevsky, M., Zeevi, Y.Y.: Quasi maximum likelihood blind deconvolution of images using optimal sparse representations. Technical Report 455, Technion, Israel (2003)
19. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2003)
20. Giannakis, G.B., Halford, S.D.: Blind fractionally spaced equalization of noisy FIR channels: Direct and adaptive solutions. IEEE Trans. Sig. Proc. (**45**)