CCIT Report #556
September 2005

Blind Source Separation using Relative Newton Method combined with Smoothing Method of Multipliers

Michael Zibulevsky
Department of Electrical Engineering, Technion, Haifa 32000, Israel
E-mail: mzib@ee.technion.ac.il, Phone: +972-4-8294724

November 15, 2004

***Abstract*— We study a relative optimization framework for quasi-maximum likelihood blind source separation and relative Newton method as its particular instance. The structure of the Hessian allows its fast approximate inversion. In the second part we present Smoothing Method of Multipliers (SMOM) for minimization of sum of pairwise maxima of smooth functions, in particular sum of absolute value terms. Incorporating Lagrange multiplier into a smooth approximation of max-type function, we obtain an extended notion of non-quadratic augmented Lagrangian. Our approach does not require artificial variables, and preserves the sparse structure of Hessian. Convergence of the method is further accelerated by the Frozen Hessian strategy. We demonstrate efficiency of this approach on an example of blind separation of sparse sources. The non-linearity in this case is based on the absolute value function, which provides super-efficient source separation.**

***Index Terms*— blind source separation, maximum likelihood, Newton method, augmented Lagrangian, method of multipliers, sparse representations**

I. Introduction

In this work we study quasi-maximum likelihood blind source separation (quasi-ML BSS) [1], [2] in batch mode, without orthogonality constraint. This criterion provides improved separation quality [3], [4], and is particularly useful in separation of sparse sources. We will present optimization methods, which produce quasi-ML BSS efficiently.

A. *Quasi-ML blind source separation (BSS)*

Consider the BSS problem, where an N-channel sensor signal $x(t)$ arises from N unknown scalar source signals $s_i(t)$, $i = 1, .., N$, linearly mixed together by an unknown $N \times N$ matrix A

$$x(t) = As(t). \tag{1}$$

We wish to estimate the mixing matrix A and the N-dimensional source signal $s(t)$. In the discrete time case $t = 1, 2, \ldots, T$ we use matrix notation $X = AS$, where X and S are $N \times T$ matrices with the signals $x_i(t)$ and $s_i(t)$ in the corresponding rows. We also denote the unmixing matrix $W = A^{-1}$.

The author would like to acknowledge support for this project by the Ollendorff Minerva Center and the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission

When the sources are *i.i.d*, stationary and white, the normalized minus-log-likelihood of the observed data X is (see for example [4])

$$L(W; X) = -\log|\det W| + \frac{1}{T}\sum_{i,t} h\big(W_i x(t)\big), \tag{2}$$

where W_i is i-th row of W, $h(\cdot) = -\log f(\cdot)$, and $f(\cdot)$ is the probability density function (pdf) of the sources. Consistent estimator can be obtained by minimization of (2), also when $h(\cdot)$ is not exactly equal to $-\log f(\cdot)$. Such *quasi-ML estimation* is practical when the source pdf is unknown, or is not well-suited for optimization. For example, when the sources are sparse or sparsely representable, the absolute value function or its smooth approximation is a good choice for $h(\cdot)$ [5], [6], [7], [8], [9], [10]. Here we will use a family of convex smooth approximations to the absolute value

$$h_1(c) = |c| - \log(1 + |c|) \tag{3}$$

$$h_\lambda(c) = \lambda h_1(c/\lambda) \tag{4}$$

with λ a proximity parameter: $h_\lambda(c) \to |c|$ as $\lambda \to 0^+$. Widely accepted natural gradient method does not work well when the approximation of the absolute value becomes too sharp. In this work we consider the relative Newton method, which overcomes this obstacle.

The Newton equations considered in this work are similar in part to those obtained by Pham and Garat [1], using different considerations. However, the algorithm given in [1], is not used in practice, because of a possibility of convergence to spurious solutions. We overcome this difficulty using line search and forcing positive definiteness of the Hessian.

Several other Newton-like BSS methods have been studied in the literature. They are based on negentropy approximation with orthogonality constraint [11], cumulant model [12], [13] and joint diagonalization of correlation matrices [14], [15], [16], [17].

The relative Newton method presented here is dedicated to quasi-ML BSS in general (not only to the sparse source case).

B. *Smoothing Method of Multipliers (SMOM) for Sum-Max problems*

In the second part we present a method for minimization of a sum of pairwise maxima of smooth functions, in particular