SOURCE ESTIMATION IN NOISY SPARSE COMPONENT ANALYSIS

Hadi Zayyani[1], Massoud Babaie-Zadeh[1], Christian Jutten[2]

1) Electrical engineering department, Sharif University of Technology, Tehran, Iran.
2) Laboratoire des Images et des Signaux (LIS), Institut National Polytechnique de Grenoble (INPG), France.
Email addresses: zayyani2000@yahoo.com, mbzadeh@yahoo.com, Christian.Jutten@.inpg.fr

ABSTRACT

In this paper, a new algorithm for Sparse Component Analysis (SCA) in the noisy underdetermined case (i.e., with more sources than sensors) is presented. The solution obtained by the proposed algorithm is compared to the minimum l^1-norm solution achieved by Linear Programming (LP). Simulation results show that the proposed algorithm is approximately 10 dB better than the LP method with respect to the quality of the estimated sources. It is due to optimality of our solution (in the MAP sense) for source recovery in noisy underdetermined sparse component analysis in the case of spiky model for sparse sources and Gaussian noise.

Index Terms— Sparse component analysis, MAP estimation, spiky model, EM algorithm

1. INTRODUCTION

Blind source separation (BSS) consists in recovering m unknown source signals from n known mixtures of them, when little details are available about the sources and the mixing system. For example, usually the only assumption about the sources is their statistical independence. Here, we only consider the noisy linear instantaneous model:

$$x(t) = As(t) + n(t) \qquad (1)$$

where $s(t)$ and $x(t)$ are the vectors containing sources and mixtures, A is the unknown $n \times m$ mixing matrix and $n(t)$ is the measurement noise vector with covariance matrix $\sigma_n^2 I$. In the underdetermined case, i.e. where the number of sources is larger than the number of mixtures ($m > n$), estimating the mixing matrix is not sufficient to recover the sources, since the mixing matrix is not invertible. Then it appears that source estimation requires prior information on the sources for allowing their reconstruction. One such prior

is the source sparsity. Source sparsity means that only a few samples of the sources are significantly non zero. Combining sparsity with the assumption of source independence shows that, at each instant, only a few number of sources have significant values (say they are active), and most of them arealmost zero (say they are inactive). This assumption about sources leads to a semi-blind source separation framework that is called sparse component analysis (SCA) [1-9]. It has been shown in [1] that, if an underdetermined linear equation has a sufficiently sparse solution then this solution is unique. The purpose of this paper is to find this unique solution.

Sparse component analysis in the underdetermined case usually consists of two steps. The first step is the estimation of the mixing matrix which can be solved by clustering approaches [2-3] or other methods [4-5]. The second step is the estimation of sources assuming that the mixing matrix is known [6]. Also, some methods estimate the mixing matrix and sources simultaneously [7]. In this paper we only focus on the source estimation and assume that the mixing matrix is already known.

Various methods are used for separating sparse sources. Minimization of the l^1 norm, which is the optimum solution assuming that the sources are Laplacian [10], can be implemented with linear programming (LP) and is the most successful method to date [1], [4]. In [6] an iterative detection-estimation algorithm is used to solve this problem. Moreover, the Mixture of Gaussian (MoG) model for sparse sources with utilization of Expectation-Maximization (EM) algorithm for maximum likelihood (ML) estimation of the sources has been used in [7].

In this paper, we assume sparse sources satisfy spiky model, and we will find an optimum (with actual known source parameters) and a near-optimum (with estimated parameters) estimate of the sources. Our method consists of three estimation steps which are introduced in section 3. Our approach is not practical for large number of sources (for example larger than 14) because of the high complexity of the method. However, our method is the optimum solution (in the MAP sense) when the estimated parameters are equal to the real source parameters. So the results of our approach

can provide bounds for performance comparisons with other approaches. Simulation results in a special case show a divergence of approximately 10dB between LP solutions (one of the most successful solutions) and our optimum approach.

The paper is organized as follows. In section 2, the spiky model of sparse sources and other notations are introduced. In section 3, our algorithm, which is a Bayesian approach, is presented with details. Section 4 shows the simulation results. Finally, section 5 concludes the paper.

2. THE SPIKY MODEL

We assume p is the probability of sources being inactive ($p \approx 1$). In the inactive case the source sample is zero and in the active case the source sample has a Gaussian distribution. We refer to this model as "the spiky model", which is nothing but a special case of the Bernoulli-Gaussian model with the variance of inactive samples being zero. This model has been also used in [9] as a model for sparse sources. This model lets us to deal with additive Gaussian noise. So the probability density of the sources is:

$$p(s_i) = p\delta(s_i) + (1-p)N(0,\sigma_r^2) \qquad i=1,...,m \qquad (2)$$

where p is the probability of inactivity of a source and σ_r^2 is the variance of active source samples. In this case any source sample can be written as the product of a binary variable (with binomial distribution) and an amplitude with a Gaussian distribution:

$$s_i = q_i r_i \qquad s = diag(q).r \qquad i=1..m$$

$$q_i = \begin{cases} 1 & s_i = active\ with\ probability\ 1-p \\ 0 & s_i = inactive\ with\ probability\ p \end{cases} \qquad (3)$$

We refer the vector q as the source activity vector. The probability of any known source activity vector is:

$$p(q) = (1-p)^{n_a}(p)^{(m-n_a)} \qquad (4)$$

where n_a is the number of active sources.

3. OUR OPTIMUM ALGORITHM

The three steps of our algorithm are shown in Fig.1. The first step is the estimation of the parameters including source parameters (p,σ_r) and noise parameter (σ_n). The second step is the estimation of the source activity vector and the final step is the estimation of sources.

3.1. Estimation of parameters

To estimate the source parameters, we utilize second and fourth order moments of one of the mixtures. Then, using (1), one derive relations between the mixture moments and the source moments:
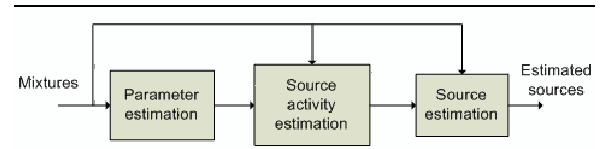

Fig. 1. Block diagram of our algorithm

$$x = \sum_i a_i s_i + n = y + n$$

$$E\{x^2\} \approx E\{y^2\} = (\sum_i a_i^2)E\{s_i^2\} \qquad (5)$$

$$E\{x^4\} \approx E\{y^4\} = (\sum_i a_i^4)E\{s_i^4\} + (\sum_{i,j}\frac{4!}{2!2!}a_i^2 a_j^2)E^2\{s_i^2\}$$

In the above equations, we have assumed that the noise term is small, i.e. σ_n^2 is small in comparison to other terms. From above equations and by assuming that all sources have equal moments (due to identical parameters) and the a_i coefficients are known, the source moments can be computed. Especially, concerning second and fourth order moments, one obtained the following relations between the source parameters and the source moments:

$$m_2 = E\{s^2\} = (1-p)\sigma_r^2 \qquad (6)$$

$$m_4 = E\{s^4\} = 3(1-p)\sigma_r^4 \qquad (7)$$

From the above two equations, we can deduce the two source parameters:

$$\sigma_r \approx \sqrt{\frac{m_4}{3m_2}} \qquad (8)$$

$$p \approx 1 - \frac{m_2^2}{3m_4} \qquad (9)$$

To estimate the noise variance (σ_n), we use the EM algorithm of [11], which gives a maximum likelihood estimation of parameters (centers, probabilities, and variances) of a Mixture of Gaussian (MoG) distribution. Since each s_i ($i=1..m$) can be either active or not, a mixture (as in the first row of (5)) which is a weighted sum of m terms, can be modeled by a mixture of 2^m Gaussian components. Since p is close to 1, one can neglect powers of $(1-p)$ greater than two, which leads to the following approximation of noisy sparse MoG:

$$p(x) \approx (p)^m N(0,\sigma_n^2) + (p)^{m-1}(1-p)\sum_{j=1}^m N(0,a_j^2\sigma_r^2 + \sigma_n^2)$$

$$(10)$$

The above equation shows that each mixture has a MoG distribution and the smallest variance of these Gaussians is the variance of noise (σ_n). Consequently, for estimating σ_n, we apply the EM algorithm of [11] to one of the mixture signals (note that our simulations show that applying the algorithm of [11] for estimating other parameters of the sources instead of (8) and (9) will not give in a good estimation result).

3.2. Estimation of the source activity vector

To estimate the source activity vector (q), we utilize the MAP-detector which maximizes the following posterior probability:

$$MAP : \underset{q}{Max} \quad p(q \mid x) \qquad (11)$$

Using the Bayes theorem, we can write:

$$p(q \mid x) = \frac{p(q)p(x \mid q)}{p(x)} \qquad (12)$$

In the above equation $p(x)$ is fixed and, for any source activity vector $p(q)$, is given by (4) after the first step of the algorithm. The vector x given q has a Gaussian distribution with the following conditional covariance:

$$Q_q = E\{xx^T \mid q\} = AV_qA^T + \sigma_n^2 I \qquad (13)$$

where V_q is the conditional covariance of the sources and can be stated as:

$$V_q = E\{ss^T \mid q\} = \sigma_r^2 Q \quad Q = diag(q) \qquad (14)$$

So the prior probability $p(x \mid q)$ has the following Gaussian distribution:

$$p(x \mid q) = \frac{1}{\sqrt{\det(2\pi Q_q)}} \exp(\frac{-1}{2} x^T Q_q^{-1} x) \qquad (15)$$

So the MAP detector maximizes the $p(q)p(x \mid q)$ over all the 2^m cases for all source activity vectors and selects the optimum choice for it.

3.3. Estimation of the sources

By knowing the source activity vector, the estimation of the source samples only requires the estimation of amplitudes (r_i's) or amplitude vector (r). Since the vector r has a Gaussian distribution, its MAP estimate is equal to its linear least square (LLS) estimate [12]. The LLS estimate of r by knowing the source activity vector and the mixture vector has the following form [12]:

$$\hat{r}_{LLS} \mid_q = E(r \mid x,q) = E(rx^T \mid q)(E(xx^T \mid q))^{-1} x \qquad (16)$$

The computation of the two terms, $E(rx^T \mid q)$ and $E(xx^T \mid q)$, results in the following equations:

$$E(rx^T \mid q) = \sigma_r^2 QA^T \qquad (17)$$

$$E(xx^T \mid q) = \sigma_r^2 AQA^T + \sigma_n^2 I \qquad (18)$$

So we have:

$$\hat{r} = \sigma_r^2 QA^T (\sigma_r^2 AQA^T + \sigma_n^2 I)^{-1} x \qquad (19)$$

4. SIMULATION RESULTS

In this section, we will examine the performance of our algorithm. In our experiments, the mixing matrix A is generated randomly by drawing each of its m columns from a uniform distribution on the unit sphere. We will examine the parameter estimation step and the source estimation step separately. The overall performance of our algorithm, for exact parameters as well as for estimated parameters, will be presented and compared with the LP-solution that is one of the most popular solutions for source estimation in the underdetermined sparse component analysis [1], [4].

4.1. Parameter estimation

In this experiment, we present the results of the estimation of the parameters of the sources (p, σ_r) and the parameter of the noise (σ_n) with the algorithm described in section 3.1. We selected 8 sources $p = 0.9$ with $\sigma_r = 1$ and we assumed $\sigma_n = .01$. Furthermore, we used 100 iterations for the EM algorithm. The estimation errors are measured by the normalized standard deviation of three errors.

We repeated our simulation 50 times with new randomly chosen mixing matrix and sparse sources. Also we repeated it for the length of the signals equals to N=1000, 2000, 3000, 4000, 5000. Finally, we computed the averaged normalized error for three parameters. Figure 2 shows the results.

4.2. Overall source estimation

In this experiment, we compare our algorithm (briefly called MAP) with the LP method. For investigating the effect of parameter estimation error, our algorithm was simulated both with exact and estimated parameters.

We selected $N = 1000, p = 0.9, \sigma_r = 1$ and $\sigma_n = .01$. In our simulations $m = 8$ sources are mixed and n mixtures are obtained. The value of n varies between 3 and 7. As for the previous experiment, we repeated our simulation 50 times with randomly chosen mixing matrix and sparse sources. The averaged results (SNR) for three cases (LP, MAP with exact and estimated parameters) are depicted in Fig. 3.

The results show that the SNR of our algorithm with exact parameters is approximately 10 dB better than the LP method. It is due to the optimality of our solution, and due
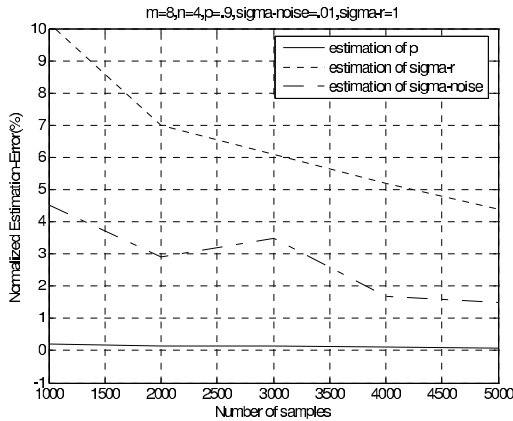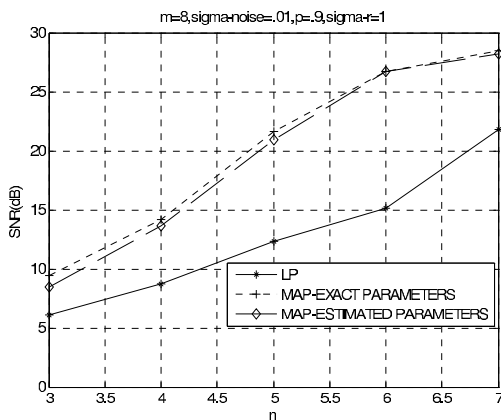
Fig. 2. Results of parameter estimation

Fig. 3. Results of the source estimation

to the fact that the equivalency of minimizing the l^1 norm and finding the sparse solution has been guaranteed just for large number of sources and sensors [1]. Moreover, the averaged SNR of our algorithm with estimated parameters is very close to the averaged SNR with exact parameters (up to 2dB degradation). It shows that the algorithm is not so sensitive to parameter estimation errors.

5. CONCLUSION

In the special case of our simulation, our algorithm, that is based on the MAP detection and MAP estimation, results in a better estimation of noisy mixtures of spiky sources than the LP method (the SNR improvement is approximately 10 dB). It is the main advantage of our method due to optimality (in MAP sense) of our algorithm (in the case of real parameters) or near-optimality of our algorithm (in the case of estimated parameters). Moreover, we showed

experimentally that our algorithm is not very sensitive to parameter estimation errors. Unfortunately, the complexity of our algorithm is its main disadvantage. This complexity comes from the method for maximizing (11), which requires an exhaustive search of 2^m different states. However, although our algorithm is not of high practical interest for large number of sources and mixtures, the (near-)optimality of our algorithm suggests that it can provide a bound convenient for comparing the performance of other methods. Looking for a simpler and faster method for maximizing the posterior probability is one of the main issues of future works.

6. REFERENCES

[1] D.L. Donoho, "For most large underdetermined systems of linear equations the minimal l^1 norm solution is also the sparsest solution," *Tech. rep.*, 2004.
[2] M. Zibulevsky, B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," Neural Computation, vol.13, pp. 863-882, 2001.
[3] M. Babaie-Zadeh, C. Jutten, A. Mansour, "Sparse ICA via cluster-wise PCA," Neurocomputing, vol.69, pp. 1458-1466, August 2006.
[4] Y. Li, S.I. Amari, A. Cichoki, "Underdetermined blind source separation based on sparse representation," IEEE Trans. Signal Processing, vol.54, No.2, pp. 423-437, Feb 2006.
[5] P.G. Georgiev, F.J. Theis, A. Cichoki, "Sparse component analysis and blind source separation of underdetermined mixtures," IEEE Trans. Neural Networks, vol.16, pp. 992-996, July 2005.
[6] A.A. Amini, M. Babaie-zadeh, C. Jutten, "Fast sparse decomposition by iterative detection-estimation," Submitted to IEEE Trans. Signal processing, 2007.
[7] M. Davies, N. Mitianoudis, "Simple mixture model for sparse overcomplete ICA," IEE Proc. Visual image and signal processing, vol.151, pp. 35-43, Feb 2004.
[8] R. Gribonval, S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges," ESANN'06, pp. 323-330, 2006.
[9] L. Vielva, D. Erdogmus, J. C. Principe "Underdetermined blind source separation using a probabilistic source sparsity model," Proceeding of international workshop on Independent Component Analysis and blind signal separation (ICA'01), pp. 675-679, Dec 2001.
[10] M.S. Lewicki, T.J. Sejnowski, "Learning overcomplete representations," Neural Comp., vol. 12, pp. 337-365, 2000.
[11] C. Tomasi, "Estimating Gaussian mixture density with EM –a tutorial," [online]. Available: http://www.cs.duke.edu/~tomasi/software/ tomasi.pdf.
[12] Anderson, B.D., and J.B. Moore, *Optimal filtering*, Prentice Hall, 1979.