Space-Time Variant Blind Source Separation with Additive Noise

Ivica Kopriva and Harold Szu

Digital Media RF Laboratory
Department of Electrical and Computer Engineering
The George Washington University
725 23rd Street NW, Washington DC 20052, USA
ikopriva@gwu.edu, SZUH@onr.navy.mil

Abstract. We propose a method for solving linear space-time variant blind source separation (BSS) problem with additive noise, **x=As+n,** on the "pixel-by-pixel" basis i.e. assuming that unknown mixing matrix is different for every space or time location. Solution corresponds with the isothermal-T_0 equilibrium of the free energy $H = U - T_0 S$ contrast function where U represents the input/output energy exchange and S represents the Shannon entropy. Solution of the inhomogeneous equation (data model with additive noise) is obtained by augmenting inhomogeneous equation into homogeneous "noise free" equation. Consequently, data model with additive noise can be solved by algorithm for the noise free space-time variant BSS problems, [1],[2]. We demonstrate the algorithm capability to perfectly recover images from the space variant mixture of two images with additive noise.

1 Introduction

The BSS problem with additive noise and positivity constraints is defined with

$$\mathbf{x}(r) = \mathbf{A}(r)\mathbf{s}(r) + \mathbf{n}(r) \tag{1}$$

where r is generalized coordinate and $\mathbf{x}, \mathbf{s}, \mathbf{n} \in \mathrm{R}^{N}$, $\mathbf{A} \in \mathrm{R}^{N\times N}$ represent data vector, source vector, additive noise vector and mixing matrix respectively, N represents problem dimension and R is a set of real numbers. We have presented in [1], [2], [3], [4] algorithm that solves the BSS problem without additive noise on the "pixel-by-pixel" basis. Hence, we may assume unknown mixing matrix to be space variant. In this paper we formulate an extension of the algorithm presented in [1],[2] to treat the BSS problems with additive noise. Because we have focused our attention on imaging applications the positivity constraints were imposed on the data vector, source vector, noise vector and mixing matrix as $\mathbf{x}, \mathbf{s}, \mathbf{n} \in \mathrm{R}_0^{+N}$, $\mathbf{A} \in \mathrm{R}_0^{+N\times N}$ where R_0^{+} is a set of positive real numbers including zero. In real world applications such as telescope images in astronomy or remotely sensed images the pixel values correspond to intensities and must be positive, [1],[2],[3],[9],[10],[11]. Also mixing matrix itself must be positive if it for example represents point spread function of an imaging system, [13] [16], or spectral reflectance matrix in remote sensing, [3],[11]. Standard BSS approaches, [5],[6],[7],[8], do not take into account these positivity constraints and that can lead to reconstructed images that have areas of negative intensity. The so-called

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 240–247, 2004.

non-negative ICA methods that explicitly take into account positivity constraints are described in [9],[10]. Like other ICA methods they are probabilistic methods and rely on the priors for the source pixels to be mixture of Laplacians with high probability for positive values around zero and zero probability for the negative values. These probabilistic assumptions implicitly assume that unknown mixing matrix is space invariant. We will show how it is possible to apply the same BSS "single-pixel" deterministic method developed for noise free data model, [1],[2] to treat the model with additive noise (1) by augmenting dimensionality of the data model twice. Consequently deterministic algorithm can be used for solving blind space-time variant linear imaging problem with additive noise by selecting among multiple possible solutions the one at the isothermal-T_o equilibrium of the free energy $H = U - T_o S$ where U represents the input/output energy exchange and S represents the Shannon entropy. Derivation of the algorithm is given in Section 2. We demonstrate the algorithm capability to perfectly recover images from the synthetic space variant linear mixture of two images with additive noise in Section 3. Conclusion is given in Section 4.

2 The Algorithm

The inhomogeneous BSS problem is defined by (1). Note that such formulation allows the mixing matrix $\mathbf{A}(r)$ to be space-time variant. The generalized coordinate r can for example represent pixel location $r(p,q)$ in the case of multispectral image [2],[4] or image sequence [16]. We shall keep argument r in the subsequent derivations in order to indicate that BSS problem is formulated on the "pixel-by-pixel" basis. In order to illustrate how to treat space (time)-variant BSS problem with additive noise (1) we shall assume that $\mathbf{n}(r)$ varies extremely rapidly compared with the variations of both $\mathbf{A}(r)$ and $\mathbf{s}(r)$ i.e.

$$\begin{aligned} &\mathbf{A}(r,t) \cong \mathbf{A}(r,t+\Delta t) \\ &\mathbf{s}(r,t) \cong \mathbf{s}(r,t+\Delta t) \\ &\mathbf{n}(r,t) \neq \mathbf{n}(r,t+\Delta t) \end{aligned} \tag{2}$$

Eq.(2) is usual assumption in solving Langevin's equation that describes the Brownian motion of a free particle, [17]. Under assumptions (2) data model (1) can be written in the augmented form that assumes two time measurements as

$$\begin{bmatrix} \mathbf{x}(r,t) \\ \mathbf{x}(r,t+\Delta t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(r,t) & \mathbf{I} \\ \mathbf{A}(r,t) & \mathbf{D}(r,t,\Delta t) \end{bmatrix} \begin{bmatrix} \mathbf{s}(r,t) \\ \mathbf{n}(r,t) \end{bmatrix} \tag{3}$$

where $\mathbf{I}$ is N-dimensional identity matrix and $\mathbf{D}(r,t,\Delta t)$ is a diagonal matrix defined with

$$\mathbf{D}(r,t,\Delta t) = \mathrm{diag}\left\{ \frac{\mathrm{n_i}(r,t+\Delta t)}{\mathrm{n_i}(r,t)} \right\}_{i=1}^{N} \tag{4}$$

and t and $t+\Delta t$ denote two time points at which the measurements are taken. In order to ensure that two sets of measurements are linearly independent the following must hold

$$rank\left(\begin{bmatrix}\mathbf{A}(r,t) & \mathbf{I}\\ \mathbf{A}(r,t) & \mathbf{D}(r,t,\Delta t)\end{bmatrix}\right)=2N \tag{5a}$$

which is fulfilled when

$$\mathrm{n_i}(r,t+\Delta t)\neq \mathrm{n_i}(r,t) \qquad i=1,\ ...,\ N \tag{5b}$$

i.e. noise realizations must be different which is consistent with assumptions (2). In order to fulfill conditions (5) the second measurement at each data vector component at the time point $t+\Delta t$ must be repeated until the following condition is satisfied

$$x_i(r,t)\neq x_i(r,t+\Delta t) \quad i=1,..,N \tag{5c}$$

because by assumption both mixing matrix and source vector remain constant during measurements and according to the augmented data model (3) the only contribution that can change data vector component $x_i(r,t+\Delta t)$ can come from the corresponding noise component $n_i(t+\Delta t)$. If due to the positivity reasons the mixing matrix is parameterized in terms of the mixing angles [1],[2],[3],[4] the augmented data model (3) can be rewritten on the component level for the 2-dimensional case as

$$\begin{bmatrix}x_1(r,t)\\ x_2(r,t)\\ x_1(r,t+\Delta t)\\ x_2(r,t+\Delta t)\end{bmatrix}=\begin{bmatrix}\cos\theta_{11}(r,t) & \cos\theta_{12}(r,t) & 1 & 0\\ \sin\theta_{11}(r,t) & \sin\theta_{12}(r,t) & 0 & 1\\ \cos\theta_{11}(r,t) & \cos\theta_{12}(r,t) & \tan\theta_{13}(r,t,\Delta t) & 0\\ \sin\theta_{11}(r,t) & \sin\theta_{12}(r,t) & 0 & \tan\theta_{14}(r,t,\Delta t)\end{bmatrix}\begin{bmatrix}s_1(r,t)\\ s_2(r,t)\\ n_1(r,t)\\ n_2(r,t)\end{bmatrix} \tag{6}$$

In order to be consistent with data model (1)/(3) the following must hold

$$\tan\theta_{13}(r,t,\Delta t)=\frac{n_1(r,t+\Delta t)}{n_1(r,t)} \quad \tan\theta_{14}(r,t,\Delta t)=\frac{n_2(r,t+\Delta t)}{n_2(r,t)} \tag{7}$$

The augmented data model (3)/(6) can now be solved using the algorithm developed for the noise free model [1],[2]. The price that has to be paid to solve the problem with additive noise is the increased number of unknowns. We show on Fig. 1 the vector diagram representation of the data model with additive noise (1) where the mixing matrix column vectors are $\mathbf{a}_1=[\cos\theta_{11}(r,t) \quad \sin\theta_{11}(r,t)]^{\mathrm{T}}$ and $\mathbf{a}_2=[\cos\theta_{12}(r,t) \quad \sin\theta_{12}(r,t)]^{\mathrm{T}}$ and $\tilde{\mathbf{x}}$ represents the noise free part of the data vector (1).

It has been shown in [1],[2] that solution of the noise free blind space-variant imaging problem can be found from the minimum of the Helmholtz free energy contrast function

$$H(\mathbf{W},\mathbf{s})=U-T_0S=\boldsymbol{\mu}^{\mathrm{T}}\left[\mathbf{Wx}-|\mathbf{s}|\mathbf{s}'\right]+K_BT_0|\mathbf{s}|\sum\nolimits_{i=1}^{N}s_i'\ln s_i'+\ |\mathbf{s}|(\mu_0-K_BT_0)\left(\sum\nolimits_{i=1}^{N}s_i'-1\right) \tag{9}$$

where S in (9) represents Shannon entropy approximated by

$$S=-K_BT_0\sum\nolimits_{i=1}^{N}s_i'\ln s_i'+const\times(\sum\nolimits_{i=1}^{N}s'_i-1) \tag{10}$$

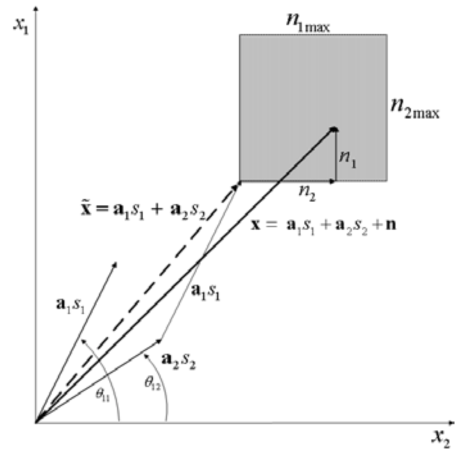

Fig. 1. Vector diagram representation of the 2-D data model (1).

where K_B represents Boltzmann's constant and T_0 represents temperature. They are introduced in (9) due to the dimensionality reasons. Also in (9) $|\mathbf{s}|$ represents L_1 norm of the source vector $\mathbf{s}$, $s_i' = s_i/|\mathbf{s}|$ is the i-th component of the normalized source vector and $\mathbf{W}$ is NxN matrix that approximates inverse of the mixing matrix i.e. $\mathbf{W} \cong \mathbf{A}^{-1}$ and $\boldsymbol{\mu}$ is the vector of Lagrange multipliers. $U = \boldsymbol{\mu}^{\mathrm{T}}\left[\mathbf{Wx} - |\mathbf{s}|\mathbf{s}'\right]$ in (9) represents a linear error energy term and enables generalization of the Shannon maximum entropy S of the closed system to an open system having non-zero input-output energy exchange U. To solve the BSS imaging problem with the positivity constraints we formulate an algorithm, [14],[1],[15], that looks for the global minimum of the error energy function

$$\left(\mathbf{W}^*, |\mathbf{s}|^*\right) = \arg\min\left(\mathbf{Wx} - |\mathbf{s}|\mathbf{s}'\right)^{\mathrm{T}}\left(\mathbf{Wx} - |\mathbf{s}|\mathbf{s}'\right) \tag{11}$$

Either deterministic search or stochastic simulated annealing based search, [1],[14],[15], over the phase space could be used in solving optimization problem (11). For a given doublet $\left(\mathbf{W}^{(l)}, |\mathbf{s}|^{(l)}\right)$, where l denotes iteration index in a solution of problem (11), the MaxEnt-like algorithm, [1],[2], computes the most probable solution for the vector of source probabilities, $\mathbf{s}'^{(l)}$

$$s_j' = \frac{1}{1 + \sum_{\substack{i=1 \\ i \neq j}}^{N} \exp\left(\frac{1}{K_B T_0}\left(\mu_i - \mu_j\right)\right)} = \sigma(\boldsymbol{\mu}) \tag{12}$$

with the Lagrange multipliers $\boldsymbol{\mu}$ learning rule given with [2] as

$$\mu_j^{(k+1)} = \mu_j^{(k)} + \left(\frac{K_B T_0}{s_j'^{(k)}} + \mu_j^{(k)}\right)\left(\mathbf{w}_j^{(l)}\mathbf{x} - |\mathbf{s}|^{(l)} s_j'^{(k)}\right) + \sum_{\substack{i=1 \\ i \neq j}}^{N} \mu_i^{(k)}\left(\mathbf{w}_i^{(l)}\mathbf{x} - |\mathbf{s}|^{(l)} s_i'^{(k)}\right) \tag{13}$$

where k stands for the iteration index related to the Lagrange multipliers learning rule, l stands for the iteration index related to the iterative solution of the optimization problem (11) and $\mathbf{w}_i$ represents the i-th row of the de-mixing matrix $\mathbf{W}$.

3 Simulation Results

To model positivity constraints we have parameterized mixing matrix in terms of the mixing angles as in (6), [1],[2],[3],[4]. Such parameterization reduces a search in higher dimensional parameter space to the first quadrant only and in that sense is an economical representation from the computational complexity standpoint. We illustrate deterministic BSS algorithm on the $N=2$ example of (3). If according to (6) we choose for the particular single pixel case the mixing angles to be $\theta_{11}=5^0$, $\theta_{12}=1^0$, $\theta_{13}=69^0$, $\theta_{14}=60^0$ the model (6) becomes

$$\begin{bmatrix} 350.7711 \\ 100.3941 \\ 268.4710 \\ 107.1244 \end{bmatrix} = \begin{bmatrix} 0.9962 & 0.9998 & 1 & 0 \\ 0.0872 & 0.0175 & 0 & 1 \\ 0.9962 & 0.9998 & 2.6051 & 0 \\ 0.0872 & 0.0175 & 0 & 1.7321 \end{bmatrix} \begin{bmatrix} 54 \\ 154 \\ 143 \\ 43 \end{bmatrix} \tag{14}$$

Fig. 2 shows logarithm of the inverse of the error energy function (11) as a function of angles θ_{11},θ_{14} for the given model (14) when the mixing angles θ_{12} and θ_{13} were kept at the true values. Note the very sharp peak that correspond with the true solution $\theta_{11}=5^0,\theta_{14}=60^0$.

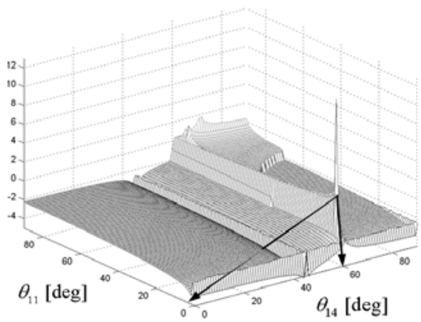

Fig. 2. 2-D plot of the logarithm of the inverse of the error energy (11) in the θ_{11} - θ_{14} domain for data model (14). The other two mixing angles θ_{12} and θ_{13} were assumed to be known.

We now mix two images by mixing matrix that has been changed from pixel to pixel in order to simulate the space variant imaging problem with additive noise, (1). Angles θ_{11} and θ_{12} are changed column wise according to Fig. 3 i.e. for every column index angles were changed for 1^0 and mutual distance between them was 4^0. According to the augmented data model (3)/(6) two measurements per each data channel were assumed to be performed. The angles θ_{13} and θ_{14}, that model the additive noise contribution, were generated randomly. On that way realization of the noise vector at time $t+\Delta t$ was independent from the realization at time t.

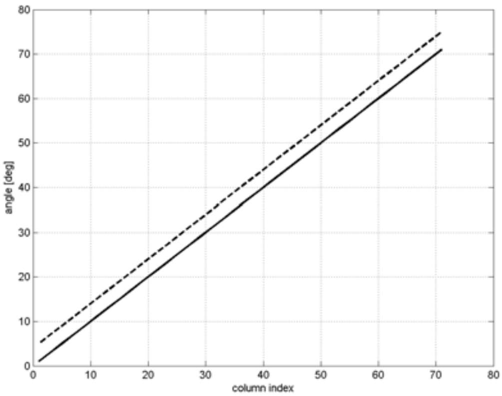

Fig. 3. Change of the mixing angles vs. column index. Solid line – the angle θ_{11}; dashed line – the angle θ_{12} for the mixture given on Fig. 4.

Fig. 4 shows from left to right: two source images, two mixed images without additive noise, two mixed images with additive noise at the time point *t*, two mixed images with additive noise at the time point $t + \Delta t$ and two separated images obtained by using the deterministic BSS algorithm (9)-(13) and the augmented data model (3)/(6). Thanks to the fact that presented algorithm solves the augmented BSS problem on the "pixel-by-pixel" basis the recovery was perfect although the mixing matrix was space variant and the additive noise was present in the model. Results shown on Fig. 4 are obtained by employing exhaustive search in the mixing angle parameter domain. However, another computationally more efficient strategy would be to employ simulated annealing optimization, [1],[14],[15], to look for global minimum of the error energy function (11). We compare our result with two representative ICA methods that were applied on the same mixture shown on Fig. 4. Fig. 5 shows from left to right separation results obtained by the Infomax algorithm, [6], and by the fourth-order cumulant based JADE algorithm. Due to the space variant nature of the mixing matrix both algorithms fail to recover the original images.

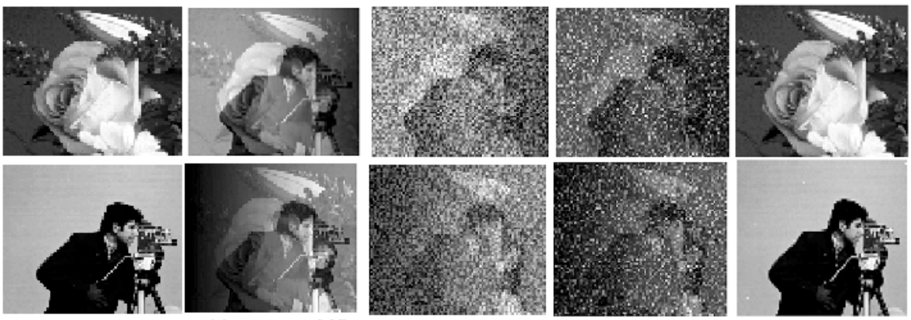

Fig. 4. From left to right: two source images, two mixed images without additive noise, two mixed images with additive noise at the time point *t*, two mixed images with additive noise at the time point $t + \Delta t$ and two separated images obtained by using the deterministic BSS algorithm (9)-(13) and the augmented data model (3)/(6).

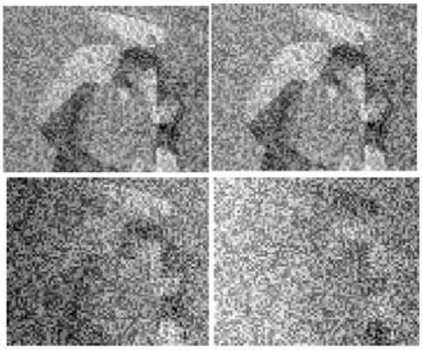

Fig. 5. Source images recovered from the space variant mixture shown on Fig. 4. by the Infomax algorithm, [6], (left) and by the JADE algorithm, [8], (right).

4 Conclusion

The algorithm capable of solving blind linear space-time variant imaging problem with additive noise on the "pixel-by-pixel" basis has been presented. This is accomplished by seeking the global minimum of the free energy contrast function and computing for each pixel the most probable value of the source vector under given macroscopic constraints defined by the data vector. In order to cope with additive noise standard N-dimensional data model has been augmented by one additional measurement per each dimension of the data vector generating the $2N$-dimenisonal "noise free" data model where the additive noise is treated as a source in the extended source vector. It is shown how multiple measurements can be made linearly independent by repeating measurement per each data channel until data channel has different values at the two corresponding time points. The algorithm performance has been demonstrated on the perfect recovery of images from synthetic space variant linear mixture of two images with additive noise. Due to the space variant nature of the mixing matrix the standard ICA algorithms failed to recover the unknown source images.

References

1. H. Szu and I. Kopriva, "Unsupervised Learning with Stochastic Gradient," *submitted* to Neurocomputing.
2. H. Szu and I. Kopriva, "Deterministic Blind Source Separation for Space Variant Imaging," Proc. of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation, ed. S.I. Amari, A.Cichocki, S. Makino, N. Murata, Nara, Japan, April 1-4, 2003, pp. 669-674.
3. H. H. Szu and C. Hsu, "Landsat spectral Unmixing à la superresolution of blind matrix inversion by constraint MaxEnt neural nets," *Proc. SPIE 3078,* 1997, pp.147-160.
4. H. Szu and I. Kopriva, "Comparison of the Lagrange Constrained Neural Network with Traditional ICA Methods," *Proc. of the IEEE 2002 World Congress on Computational Intelligence-International Joint Conference on Neural Networks*, Hawaii, USA, May 17-22, 2002, pp. 466-471.
5. S. Amari, A. Cihocki, H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems,* **8**, MIT Press, 1996, pp. 757-763.

6. A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Comp.* **7**, 1995, pp. 1129-1159.
7. A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Computation,* vol. 9, 1997, pp. 1483-1492.
8. J. F. Cardoso, A. Soulomniac, “Blind beamforming for non-Gaussian signals,” Proc. IEE F, vol. 140, 1993, pp. 362-370.
9. M. Plumbley*, “Algorithms for Nonnegative Independent Component Analysis”,* IEEE Transaction on Neural Networks, Vol.14, No.3, May, 2003, pp.534-543.
10. D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, No. 21, 1999, pp.788-791.
11. L. Parra, C. Spence, P. Sajda, A. Ziehe, K-R. Müller, “Unmixing Hyperspectral Data,” in *Advances in Neural Information Processing Systems (NIPS12),* S. A. Jolla, T. K. Leen and K-R. Müller (eds), MIT Press, 2000.
12. K. Huang, “Statistical Mechanics,” J. Wiley, 1963.
13. H. H. Szu, I. Kopriva, “Artificial Neural Networks for Noisy Image Super-resolution,” Optics Communications, Vol. 198 (1-3), 2001, pp. 71-81.
14. Szu H. and Hartley R., “Fast Simulated Annealing,” *Physical Letters A*, **122**, No. 3, 1987, pp. 157-162.
15. H. Szu and I. Kopriva, “Cauchy Machine for Blind Inversion in Linear Space-Variant Imaging,” Proc. of the 2003 IEEE International Joint Conference on Neural Networks, Portland, OR, Vol. I, July 20-24, 2003, pp. 747-752.
16. I. Kopriva, Q. Du, H. Szu and W. Wasylkiwskyj, “Independent Component Analysis Approach to Image Sharpening in the Presence of Atmospheric Turbulence,” *Optics Communications*, vol. 233 (1-3) pp.7-14, 2004.
17. S. Chandrasekhar, “Stochastic Problems in Physics and Astronomy,” *Review of Modern Physics*, vol. 15, No. 1, January, 1943, pp. 1-87.