INDEPENDENT VECTOR ANALYSIS WITH A MULTIVARIATE GENERALIZED GAUSSIAN SOURCE PRIOR FOR FREQUENCY DOMAIN BLIND SOURCE SEPARATION

Yanfeng Liang, Syed Mohsen Naqvi, Jonathon A. Chambers

Advanced Signal Processing Group, Electronic and Electrical Engineering Department
Loughborough University, Loughborough, Leicester, UK
{y.liang2, s.m.r.naqvi, j.a.chambers}@lboro.ac.uk

ABSTRACT

The independent vector analysis (IVA) algorithm can theoretically avoid the permutation problem in frequency domain blind source separation by using a multivariate source prior to retain the dependency between different frequency bins of each source. In this paper, a new multivariate generalized Gaussian distribution is adopted as the source prior which can exploit fourth order inter-frequency correlation, and therefore better preserve the dependency between different frequency bins to achieve an improved separation performance as compared with the original IVA algorithm. Separation performances are compared by simulation studies when using different source priors, and the experimental results confirm that IVA with the new source prior can consistently achieve improved separation performance.

Index Terms— independent vector analysis, multivariate generalized Gaussian distribution, fourth order inter-frequency correlation

1. INTRODUCTION

Blind source separation (BSS) has been widely researched in the signal processing research community, and focuses on how to extract individual signals from observed mixed signals. It can be widely used in various application fields [1]. One of the most famous application is attempting to solve the classical cocktail party problem, which was first described by Colin Cherry in 1953 [2]. The target is to mimic in a machine the ability of a human to separate one speaker from a mixture of sounds.

Solving the machine cocktail party problem requires the design of a method to focus on the desired speech signal while suppressing or ignoring all the other competing speech sounds [3]. In signal processing, independent component analysis (ICA) is the central tool for solving this problem [4]. In a real room environment, due to the reflections of the ceiling, the floor and the walls, the problem becomes convolutive blind source separation (CBSS) [5]. The time domain methods are generally not appropriate for solving the CBSS problem because of the computational complexity [1]. Then the frequency domain methods are proposed to address the CBSS problem [6]. However, the permutation ambiguity is inherent to the BSS problem, and many methods have been proposed to solve the problem by introducing the source position information or the structure of the source signal [7].

Independent vector analysis (IVA) is a frequency domain method to solve the convolutive blind source separation problem (CBSS), which can theoretically avoid the permutation problem by exploiting certain statistical inter-dependency between frequency bins within each source vector, while removing the dependency between different sources [8]. The IVA method adopts a dependent multivariate spherically symmetric Laplace (SSL) distribution as the source prior, instead of a univariate distribution used by traditional CBSS approaches. However, the form of the multivariate source prior should not always be fixed. Recently, selecting the most appropriate multivariate source prior to improve the separation performance has become a research focus.

In [9], a family of l^p-norm-invariant sparse probability density functions (pdf) is used as the source prior; then the separation performance of maximum likelihood type independent vector analysis (ML-IVA) algorithms is compared. The experimental results indicate that the spherical symmetry pdf, i.e. p=2, is suitable for modeling speech. The sparseness parameter is also discussed, and it is claimed that the best separation can be obtained when the sparseness parameter is around 7.

In this paper, we adopt a new multivariate generalized Gaussian distribution as the source prior, which also belongs to the family of l^2-norm-invariant sparse probability density functions, and the sparseness parameter is chosen to be $\frac{3}{2}$. This new source prior introduces fourth order terms between different frequency bins for each source vector to construct a stronger dependent structure and thereby improve the separation performance, as compared with the original IVA method which chooses the spherically symmetric Laplace (SSL) distribution as the source prior. Moreover, the experimental results show the separation performance when using the new source prior can consistently achieve improved separation performance. The IVA algorithm with the source prior whose sparseness parameter is around 7 is found not to be robust in that it not always increases the separation performance.

The paper is organized as follows, in Section 2 the original independent vector analysis algorithm is introduced. Then IVA with the proposed source prior is described in Section 3, and the advantage of the new source prior is also analyzed. Experimental results are shown in Section 4, and the relation to prior work is finally discussed in Section 5.

2. INDEPENDENT VECTOR ANALYSIS

In a real room environment, due to reverberation, there are many paths between the microphones and the sources, which is modelled as a convolutive case. In order to reduce the computational cost of the time domain methods, separation of such convolutive problems is generally approached in the frequency domain. Thus, the noise free model in the frequency domain is described as:

$$\mathbf{x}^{(k)} = \mathbf{H}^{(k)}\mathbf{s}^{(k)} \tag{1}$$

$$\hat{\mathbf{s}}^{(k)} = \mathbf{W}^{(k)}\mathbf{x}^{(k)} \tag{2}$$

where $\mathbf{x}^{(k)} = [x_1^{(k)}, \ldots, x_m^{(k)}]^T$, $\hat{\mathbf{s}}^{(k)} = [\hat{s}_1^{(k)}, \ldots, \hat{s}_n^{(k)}]^T$ and $\mathbf{s}^{(k)} = [s_1^{(k)}, \ldots, s_n^{(k)}]^T$ are the observed signal vector, estimated signal vector and source signal vector in the frequency domain respectively, and $(\cdot)^T$ denotes vector transpose. The index $k = 1, 2, \ldots, K$ denotes the k-th frequency bin, and K is the number of frequency bins; m is the number of microphones and n is the number of sources. $\mathbf{H}^{(k)}$ and $\mathbf{W}^{(k)}$ are the mixing matrix and the unmixing matrix respectively. In this paper, we assume that the number of sources is the same as the number of microphones, i.e. $m = n$.

Independent vector analysis is proposed as a frequency domain solution which can theoretically avoid the permutation problem by preserving the dependency between different frequency bins of each vector source while maximizing the independence between the vector sources. The IVA method adopts the Kullback-Leibler divergence between the joint probability density function $p(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n)$ and the product of marginal probability density functions of the individual source vectors $\prod q(\hat{\mathbf{s}}_i)$ as the cost function [8].

$$\begin{aligned} J &= KL(p(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n) || \prod q(\hat{\mathbf{s}}_i)) \\ &= const - \sum_{k=1}^{K} log|det(\mathbf{W}^{(k)})| - \sum_{i=1}^{n} E[logq(\hat{\mathbf{s}}_i)] \end{aligned} \tag{3}$$

where $E[\cdot]$ denotes the statistical expectation operator, $det(\cdot)$ is the matrix determinant operator. The dependency between different source vectors should be removed but the interrelationships between the components of each vector can be retained, when the cost function is minimized. The inter-frequency dependency is modelled by the probability density function of the source.

For traditional CBSS approaches, the scalar Laplacian distribution is widely used for the source prior. However, the resultant nonlinear score function is a univariate function, which is not designed to preserve the dependency between different frequency bins for each source. IVA exploits a multivariate spherically symmetric Laplace (SSL) distribution as the source prior, which can be written as

$$q(\mathbf{s}_i) \propto exp\Big(-\sqrt{(\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1} (\mathbf{s}_i - \boldsymbol{\mu}_i)}\Big) \tag{4}$$

where $(\cdot)^\dagger$ denotes the Hermitian transpose, μ_i and Σ_i^{-1} are respectively the mean vector and inverse covariance matrix of the i-th source. When the cost function (3) is minimized by the gradient descent method, the nonlinear score function contained in the update equations is derived according to the source prior [8]. We assume that the mean vector is a zero vector. The covariance matrix is taken to be a diagonal matrix due to the orthogonality of the Fourier bases, which implies that each frequency bin sample is uncorrelated with the others. As such, the nonlinear score function can be obtained as:

$$\varphi^{(k)}(\hat{\mathbf{s}}_i) = -\frac{\partial logq(\hat{s}_i^{(1)} \cdots \hat{s}_i^{(k)})}{\partial \hat{s}_i^{(k)}} = \frac{\hat{s}_i^{(k)} / (\sigma_i^{(k)})^2}{\sqrt{\sum_{k=1}^{K} \left| \frac{\hat{s}_i^{(k)}}{\sigma_i^{(k)}} \right|^2}} \tag{5}$$

where $\sigma_i^{(k)}$ denotes the standard deviation of the ith source at the kth frequency bin. It is a multivariate function, and the dependency between the frequency bins can thereby be retained in learning.

3. INDEPENDENT VECTOR ANALYSIS WITH NEW SOURCE PRIOR

The nonlinear score function is used to retain the inter-frequency dependency, and it is claimed in [8] that the nonlinear score function above in (5) could be replaced. An improved multivariate nonlinear function could be designed to achieve better separation performance. Because the nonlinear function is derived based on the probability density distribution of the source, a more appropriate source prior will help to improve the separation performance.

3.1. Multivariate generalized Gaussian distribution

The univariate generalized Gaussian distribution takes the form

$$q(s_i) \propto exp\Big(-\Big(\frac{|s_i - \mu|}{\alpha}\Big)^\beta\Big) \tag{6}$$

where α is the scale parameter and β is the shape parameter. If α is chosen properly, when $\beta = 2$, it becomes the Gaussian distribution, and when $\beta = 1$, it is the Laplace distribution.

We extend it to the multivariate case; the multivariate generalized Gaussian distribution has the form

$$q(\mathbf{s}_i) \propto exp\Big(-\Big(\frac{\sqrt{(\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1} (\mathbf{s}_i - \boldsymbol{\mu}_i)}}{\alpha}\Big)^\beta\Big) \tag{7}$$

when $\alpha = 1$ and $\beta = 1$, it is the multivariate Laplace distribution, i.e. spherically symmetric Laplace, which is adopted as the source prior for the original IVA algorithm.

One important property of the univariate generalized Gaussian distribution is, the smaller the β, the heavier are the tails. This property can also extend to the multivariate case.

3.2. New source prior for independent vector analysis

In this paper, we therefore propose a new multivariate distribution as the source prior for independent vector analysis which takes the form

$$q(\mathbf{s}_i) \propto exp\Big(- \sqrt[3]{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger} \Sigma_i^{-1} (\mathbf{s}_i - \boldsymbol{\mu}_i)} \Big) \tag{8}$$

which also preserves the inter-relation between different frequency domain components of each source vector. Both the new proposed source prior and the source prior adopted by original IVA belong to the family of multivariate generalized Gaussian distributions. When $\alpha = 1$ and $\beta = 1$, it is the original source prior, and when $\alpha = 1$ and $\beta = \frac{2}{3}$, it is the new proposed source prior. Because of the property of the multivariate generalized Gaussian distribution, as β is reduced, the heavier will be the tails. Thus, the new proposed source prior has heavier tails than the original one, which suggests more robustness to outliers. Due to the time-variability of speech, it is important to make the source prior robust to outliers to achieve a better performance.

When this new source prior is used to derive the nonlinear score function with the same assumption when we derive equation (5), we obtain

$$\varphi^{(k)}(\hat{\mathbf{s}}_i) = \frac{\hat{s}_i^{(k)} / (\sigma_i^{(k)})^2}{\sqrt[3]{[(\hat{\mathbf{s}}_i - \boldsymbol{\mu}_i)^{\dagger} \Sigma_i^{-1} (\hat{\mathbf{s}}_i - \boldsymbol{\mu}_i)]^2}} \tag{9}$$

where $\Sigma_i^{-1} = diag[\frac{1}{(\sigma_i^{(1)})^2}, \cdots, \frac{1}{(\sigma_i^{(K)})^2}]$. If we expand the equation under the cubic root, it can be written as:

$$[(\hat{\mathbf{s}}_i - \boldsymbol{\mu}_i)^{\dagger} \Sigma_i^{-1} (\hat{\mathbf{s}}_i - \boldsymbol{\mu}_i)]^2 = \sum_{k=1}^{K} \left| \frac{\hat{s}_i^{(k)}}{\sigma_i^{(k)}} \right|^4 + \sum_{u \neq v} c_{uv} |\hat{s}_i^{(u)}|^2 |\hat{s}_i^{(v)}|^2 \tag{10}$$

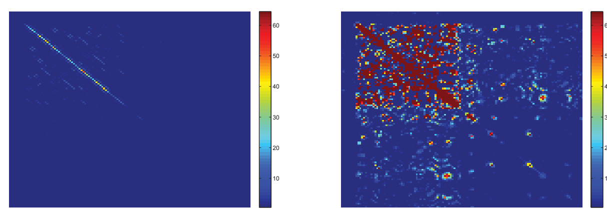
which contains cross items $\sum_{u \neq v} c_{uv} |\hat{s}_i^{(u)}|^2 |\hat{s}_i^{(v)}|^2$, and c_{uv} is a scalar constant between the u-th and v-th frequency bins. These terms are related to the fourth order relationships between different components for each source vector, represent the level of interdependency between different frequency bins. Thus, this new multivariate nonlinear function can provide a more informative model of the dependency structure. Moreover it can better describe the speech model.

Such fourth order relationships of speech signals here not been used in the original IVA. We will show an example of the second order relationships and the fourth order relationships inherent to a particular speech signal "si1010.wav" from the TIMIT database [10], with 8 kHz sampling frequency and 1024 DFT length. The length of this speech is three seconds approximately. Fig. 1(a) is part of the image display of elements of the covariance matrix formed by sample interrelationships between the elements of the signal vector, which is correspondent to the low frequency bins. We can hardly observe any information correspondent to the high frequency bins due to the limited energy. Therefore, we only show part of the image. We can see that only the diagonal has significant second order relationship information. This is because the Fourier transform is an orthogonal based transform.

In order to exploit the fourth order information, we construct a fourth order matrix

$$\begin{pmatrix} E[(s_i^{(1)})^2 (s_i^{(1)})^2] & \cdots & E[(s_i^{(1)})^2 (s_i^{(K)})^2] \\ \vdots & \ddots & \vdots \\ E[(s_i^{(K)})^2 (s_i^{(1)})^2] & \cdots & E[(s_i^{(K)})^2 (s_i^{(K)})^2] \end{pmatrix} \tag{11}$$

Fig. 1(b) is part of this fourth order matrix, which is also correspondent to the same low frequency bins as Fig. 1(a). It is evident that there are fourth order relationships throughout the matrix not only on the diagonal. Thus, such fourth order relationships should be exploited to help separation.



(a) (b)

Fig. 1. Second order and fourth order inter frequency relationships information of the speech signal "si1010.wav", x and y dimensions correspond to frequency bins 1 to 128 of 512.

3.3. Comparison with other source prior models

In Lee's paper [9], the source priors suitable for IVA are discussed. The source prior is described as:

$$q(\mathbf{s}_i) \propto exp(-\|\mathbf{s}_i\|_p)^{\frac{1}{r}} = exp\Big(- \sum_k |s_i^{(k)}|^p \Big)^{\frac{1}{pr}} \tag{12}$$

the parameter r is the sparseness parameter. It has been shown that the spherical symmetry assumption is suitable for modeling the frequency components of speech, i.e. $p = 2$. And it is suggested that the best separation performance can be achieved when r is around 7 .

Our new proposed source prior also belongs to this family. If we choose $p = 2$ to make it spherically symmetric, and choose $r = \frac{3}{2}$, the proposed source prior can be obtained. Our experimental results have found that the improvement of performance is not robust when r is around 7, so Lee's evaluation in [9] appears misleading, however the ML-IVA which adopts our new source prior can consistently achieve improved separation performance.

4. EXPERIMENTS

In this experiment, we used the TIMIT dataset [10]. Each speech signal was approximately seven seconds long. The image method [11] was used to generate the room impulse responses, and the size of the room was $7 \times 5 \times 3m^3$. The DFT length was 1024, and the reverberation time RT60 = 200ms. We used a 2×2 mixing case, for which the microphone positions are [3.48, 2.50, 1.50]m and [3.52, 2.50, 1.50]m respectively. The sampling frequency was 8kHz. The separation performance was evaluated objectively by the signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR) [12].

We chose two different speech signals randomly from the TIMIT dataset and convolved them into two mixtures. Then the orignal IVA method, the ML-IVA method with our proposed source prior and ML-IVA with Lee's source prior where the sparseness parameter $r = 7$, were all used to separate the mixtures respectively. Then we changed the source positions to repeat the simulation. For every pair of speech signals, three different azimuth angles for the sources relative to the normal to the microphone array were set for testing, these angles were selected from 30, 45, 60 and -30 degrees. After that, we chose another pair of speech signals to repeat the above simulations. In total, we used ten different pairs of speech signals, and repeated the simulation 30 times at different positions. Table 1 shows the average separation performance for each pair of speech signals in terms of SDR and SIR in dB.

Table 1. Separation performance comparison (SDR/SIR)

	original	proposed	Lee's
mixture 1	12.27/14.08	12.90/14.84	4.74/5.62
mixture 2	18.13/19.57	18.47/19.86	18.34/19.81
mixture 3	8.88/10.72	11.83/13.74	11.41/13.19
mixture 4	15.57/16.98	16.92/18.46	5.95/7.16
mixture 5	18.10/20.14	18.69/20.47	15.44/16.94
mixture 6	18.81/20.30	19.58/20.98	3.71/4.35
mixture 7	15.94/17.88	16.59/18.40	8.63/10.73
mixture 8	15.29/19.88	15.75/20.41	16.03/20.61
mixture 9	18.58/20.75	19.05/20.89	17.35/18.80
mixture 10	18.80/20.28	19.31/20.60	0.78/1.48

The experimental results show clearly that IVA with the proposed source prior can consistently improve the separation performance. However, for the IVA with Lee's source prior, the separation improvement is not consistent, in some cases there is even no separation such as mixtures 1, 6 and 10. Even though it can achieve better separation than original IVA, it is still no better than the proposed method. Only for mixture 8, does it achieve the best separation performance. We also found similar results with the sparseness parameter around 7. Therefore, the IVA with the proposed source prior is the best method, which can consistently achieve better separation performance. The average SDR improvement and SIR improvement are approximately 0.9dB and 0.8dB respectively.

In the second experiment, we tested the robustness of the IVA with the proposed source prior in different reverberant room environments. We selected two speech signals from the TIMIT dataset randomly and convolved them into two mixtures. The azimuth angles for the sources relative to the normal to the microphone array were set as 60 and -30 degrees. Both the original IVA and the proposed method were used to separate the mixtures. The results are shown in Fig 2, which show the separation performance comparisons in different reverberant environments. Fig. 2(a) and Fig. 2(b) show the SDR and SIR comparison respectively. They indicate that the proposed algorithm can consistently improve the separation performance in different reverberant environments.

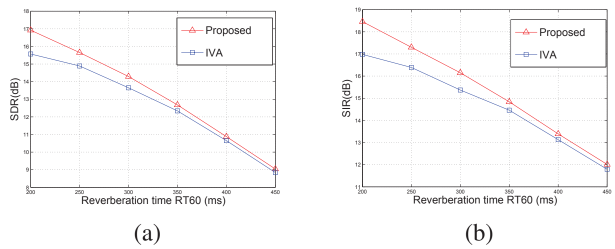


Fig. 2. Separation comparison between original and proposed IVA algorithms as a function of reverberation time.

5. RELATION TO PRIOR WORK

This paper focuses on the design of the source prior for independent vector analysis (IVA). The source prior for independent vector analysis is important because the nonlinear score function used to keep the inter-frequency dependency is derived based on the probability density function of the source. Originally, the spherically symmetric Laplace distribution is widely used as the source prior for IVA [8][13], adaptive step size IVA [14], fast fixed-point IVA [15], auxiliary function based IVA [16] and audio video based IVA [17]. However, this source prior is not necessarily the best form. A better source prior which can exploit other relationships between different frequency bins is still needed. In paper [18], a chain-like source model is introduced. The analysis of the selection of the source prior is also discussed in [9]. Recently in [19], the multivariate Gaussian model is proposed as the source prior, which can exploit the second order correlation. However, second order correlation is minimal for frequency domain BSS. In this paper, we proposed a fresh source prior which belongs to the family of multivariate generalized Gaussian distributions. The new source prior can introduce the fourth order terms between different frequency bins in the score function to better preserve the inter-frequency dependency. Moreover, it can also be adopted for all kinds of IVA methods above, which has been done but not presented in this paper. The experimental results show that the IVA method with the new source prior can consistently achieve improved separation performance.

6. REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: learning algorithms and applications*, Wiley, 2003.

[2] C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of The Acoustical Society of America*, vol. 25, pp. 975–979, 1953.

[3] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, pp. 1875–1902, 2005.

[4] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, 2001.

[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, 2009.

[6] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 320–327, 2000.

[7] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook on Speech Processing and Speech Communication*, 2007.

[8] T. Kim, H. Attias, S. Lee, and T. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech and Language processing*, vol. 15, pp. 70–79, 2007.

[9] I. Lee and T. W. Lee, "On the assumption of spherical symmetry and sparseness for the frequency-domain speech model," *IEEE Trans. on Audio, Speech and Language processing*, vol. 15, pp. 1521–1528, 2007.

[10] J. S. Garofolo et al., "TIMIT acoustic-phonetic continuous speech corpus," in *Linguistic Data Consortium*, Philadelphia, 1993.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America.*, vol. 65, no. 4, pp. 943–950, 1979.

[12] E. Vincent, C. Fevotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1462–1469, 2006.

[13] T. Kim, "Real-time independent vector analysis for convolutive blind source separation," *IEEE Transactions on Circuits and Systems I*, vol. 57, pp. 1431–1438, 2010.

[14] Y. Liang, S.M. Naqvi, and J. Chambers, "Adaptive step size indepndent vector analysis for blind source separation," in *17th International Conference on Digital Signal Processing*, Corfu, Greece, 2011.

[15] I. Lee, T. Kim, and T.-W. Lee, "Fast fixed-point independent vector analysis algorithms for convolutive blind source separation," *Signal Processing*, vol. 87, pp. 1859–1871, 2007.

[16] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *2011 IEEE WASPAA*, New Paltz, USA, 2011.

[17] Y. Liang, S.M. Naqvi, and J. Chambers, "Audio video based fast fixed-point independent vector analysis for multisource separation in a room environment," *EURASIP Journal on Advances in Signal Processing*, vol. 2012:183, 2012.

[18] I. Lee, G.-J. Jang, and T.-W. Lee, "Independent vector analysis using densities represented by chain-like overlapped cliques in graphical models for separation of convolutedly mixed signals," *Electronic Letters*, vol. 45, pp. 710–711, 2009.

[19] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis," *IEEE Trans. on Signal Processing*, vol. 60, pp. 1672–1682, 2012.