COMBINING INDEPENDENT COMPONENT ANALYSIS WITH GEOMETRIC INFORMATION AND ITS APPLICATION TO SPEECH PROCESSING

Wenyi Zhang and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093-0409, USA
Email: w3zhang@ucsd.edu, brao@ece.ucsd.edu

ABSTRACT

In this paper, we propose two approaches for combining geometric information with ICA algorithm to solve permutation problem under the scenario where a rough information about the direction of the desired source is known. The first approach is a new blind extraction algorithm with a soft quadratic geometric constraint. The desired source is guaranteed to be conveyed to the output with little distortion by the quadratic constraint and the negentropy maximization criterion is used to ensure that the other sources get suppressed at the output. The second approach employs a quadratic geometric test as a post-processing step to pickup the desired source after ICA processing. An advantage of the proposed two approaches is that they do not require accurate knowledge of the number of sources in the mixtures to recover the desired source, in contrast, other geometric ICA approaches usually fail if the number of sources is not known accurately.

***Index Terms*—** Independent component analysis, geometric information, direction of arrival, quadratic constraint

1. INTRODUCTION

Independent component analysis (ICA) is a statistical method for extracting independent components from a group of mixtures [1]. For convolutive mixtures, frequency domain ICA approach is mostly used since it simplifies the problem into instantaneous mixing problem in every frequency bin and can be solved therein by simple instantaneous mixing ICA algorithms [2]. However the intrinsic scaling and permutation ambiguities need to be addressed in applying the frequency domain ICA approach.

ICA assumes no knowledge about the mixing process except the independence between sources. However, sometimes extra information is available and can be utilized to aid the ICA process. For example, in microphone array speech processing, speech's temporal structure or geometric information with the array may be employed to solve the permutation problem. Many recent works have been developed to combine ICA with geometric information to solve the permutation problem (see [4, 5, 6, 7] and refs therein). In [4, 5], beam pattern of the ICA processor is utilized to figure out the directions of the sources to solve the permutation problem. These methods become too complicated and are not robust when the number of sources exceed 2. Parra and Alvino proposed the geometrically constrained (or initialized) ICA algorithm [6], but accurate source number is required and correct permutation is not guaranteed. Knaak and Araki proposed an ICA algorithm with a hard linear geometric constraint [7]. However, accurate source number is mandatory for the algorithm to perform properly.

In microphone array speech processing, the geometry of the array and rough information about the direction of the desired signal may be known a priori, for instance, the direction of the desired signal may be assumed to be the broadside direction for a linear array, or be acquired by some direction of arrival (DOA) estimation algorithms [3]. In this paper, We propose two approaches for combining DOA information of the desired source with ICA algorithm to solve permutation problem. The first approach is a new blind extraction algorithm with a soft quadratic geometric constraint. Given a rough estimate of the direction of the desired source, the proposed algorithm will extract the desired source from the mixed signals. The quadratic constraint restricts the weighted square error between the desired and actual response of the processor over a small spatial uncertainty region chosen to deal with look direction uncertainty. Thereby the desired source is guaranteed to be conveyed to the output with little distortion and the negentropy maximization criterion is used to ensure that the other sources get suppressed at the output. This method solves the permutation problem in the frequency domain ICA approach since the desired signal is extracted consistently across all the frequency bins. The second approach employs a quadratic geometric test as a post-processing step to pickup the desired source after ICA processing. In every frequency bin, the ICA algorithm separates instantaneously mixed source signals, then the quadratic geometric test will pick up the desired source. An advantage of the proposed two approaches is that they do not require accurate knowledge of the number of sources in the mixtures to recover the desired source, in contrast, other geometric ICA approaches usually fail if the number of sources is not known accurately.

In Section 2, we describe the signal model, the complex FastICA algorithm and the linear constrained ICA algorithm. Section 3 discusses the proposed two approaches combining ICA with geometric information. The simulation results and discussion are presented in Section 4.

2. BACKGROUND

2.1. Complex FastICA

Hyvarinen and Oja proposed a fast fixed point algorithm (FastICA) for solving the real variable instantaneous mixing ICA problem [1]. FastICA maximizes the negentropy of the output $y_i, i = 1..N$, subject to the constraints that all the $y_i, i = 1..N$ are uncorrelated and

This material is based upon work supported by the RESCUE project under National Science Foundation Award Number 0331690 and UC Micro Grants 07-034 and 08-065 sponsored by Qualcomm Inc.

have unit variances. Bingham and Hyvarinen extended it to complex variables and developed a complex FastICA algorithm [8]. The mathematical problem to be solved is,

$$\min_{\mathbf{w}_j, j=1..N} \sum_{j=1}^{N} J_G(\mathbf{w}_j), \quad s.t. \quad E\{(\mathbf{w}_k^H \mathbf{x})(\mathbf{w}_j^H \mathbf{x})^*\} = \delta_{jk}$$

where $\delta_{jk} = 1$ for $j = k$ and 0 otherwise. $\mathbf{w}_j$ is M-dimensional complex weight vector. $\mathbf{x}$ is the observation vector. The contrast function is defined as,

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \tag{1}$$

where $G : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a smooth even function. For example, $G(z) = \log(a + z)$ is a good choice for speech separation task.

2.2. ICA with a linear geometric constraint

Knaak, Araki and Makino studied the similarities between adaptive beamforming and ICA, and proposed a geometrically constrained ICA algorithm (CICA) [7]. CICA combined a linear look direction constraint with the ICA criterion.

$$\arg\min_{\mathbf{w}} E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \equiv \arg\min_{\mathbf{t}} E\{G(|\mathbf{t}^H \mathbf{z}|^2)\} \tag{2a}$$

$$s.t. \quad \mathbf{w}^H \mathbf{a} = \mathbf{t}^H V \mathbf{a} = 1. \tag{2b}$$

where $\mathbf{w}^H = \mathbf{t}^H V$, $\mathbf{w}$ is the weight vector for original observed data $\mathbf{x}$, and $\mathbf{t}$ is the weight vector for the sphered data $\mathbf{z}$. V is the sphering matrix determined by principle component analysis (PCA). $\mathbf{a}$ is the estimated steering vector for the desired source. The algorithm is initialized with the MVDR beamformer [3] on the sphered data,

$$\mathbf{t}_0 = V\mathbf{a} \tag{3}$$

3. COMBINE ICA WITH GEOMETRIC INFORMATION

3.1. ICA with quadratic geometric constraint

One shortcoming of the CICA algorithm is that it requires accurate knowledge of the number of the sources in the mixtures. Also the performance of the CICA algorithm depends heavily on the initialization of the weight vector. A careful chosen initialization weight vector ensures the algorithm converge to the desired source. However, such initialization weight vector is available only when the source number is accurately known and the PCA preprocessing is properly done based on that.

Assuming a M sensor, N source system. When the PCA preprocessing is performed based on accurate source number, we effectively have a mixing system with a square mixing matrix, i.e. $M' = N$, where M' is the effective sensor number. As is known, ICA is closely related to beamforming [4, 5, 7]. When all the original sources have similar variances, the MVDR beamformer $\mathbf{t}_0$ (3) on the sphered data is close to the optimum ICA solution because of the limited free spatial dimensions. Thereby, the CICA is inclined to converge to the right solution. However, when the PCA preprocessing is performed based on an overestimated source number, we are effectively working on a mixing system with more sensors than sources, i.e. $M' > N$. In this case, there is enough free spatial dimensions in the weight vector such that the MVDR beamformer $\mathbf{t}_0$ (Eq.(3)) on the sphered data is far from the right solution (Here we assume the estimated steering vector $\mathbf{a}$ is close to but not exactly the true steering vector for the desired source since DOA error always exists in real applications). Thereby, the CICA algorithm may not converge to the right solution and it may amplify an undesired source and suppress the desired source. Furthermore, the CICA algorithm fails when the PCA preprocessing is performed based on underestimated source number. These drawbacks are confirmed in the experiments. Another possibility is to initialize the CICA algorithm with the Delay-and-Sum beamformer. However, experiments show that this still can not ensure the CICA algorithm to converge to the right solution consistently.

In this section, we propose a new ICA algorithm with a quadratic geometric constraint which combines the geometric information of the array with the ICA criterion. Instead of using a hard linear constraint as in CICA algorithm, here we propose to use a soft quadratic constraint which can accommodate uncertainty in the look direction information. In [9], a quadratic constraint is used to ensure the robustness of the adaptive beamformer in the look direction. The quadratic constraint restricts the weighted square error between desired and actual beam pattern of the beamformer over a small spatial region chosen to deal with look direction uncertainty. This error item can be written as,

$$\begin{aligned} e^2 &= \int_{\overline{\theta}-\Delta\theta}^{\overline{\theta}+\Delta\theta} f(\theta) \left| \mathbf{w}^H \mathbf{a}(\theta,\omega) - \mathbf{w}_d^H \mathbf{a}(\theta,\omega) \right|^2 d\theta \\ &= (\mathbf{w} - \mathbf{w}_d)^H \int_{\overline{\theta}-\Delta\theta}^{\overline{\theta}+\Delta\theta} f(\theta) \mathbf{a}(\theta,\omega) \mathbf{a}(\theta,\omega)^H d\theta (\mathbf{w} - \mathbf{w}_d) \\ &= (\mathbf{w} - \mathbf{w}_d)^H \Phi (\mathbf{w} - \mathbf{w}_d), \end{aligned}$$

$$\text{with} \quad \Phi = \int_{\overline{\theta}-\Delta\theta}^{\overline{\theta}+\Delta\theta} f(\theta) \mathbf{a}(\theta,\omega) \mathbf{a}(\theta,\omega)^H d\theta.$$

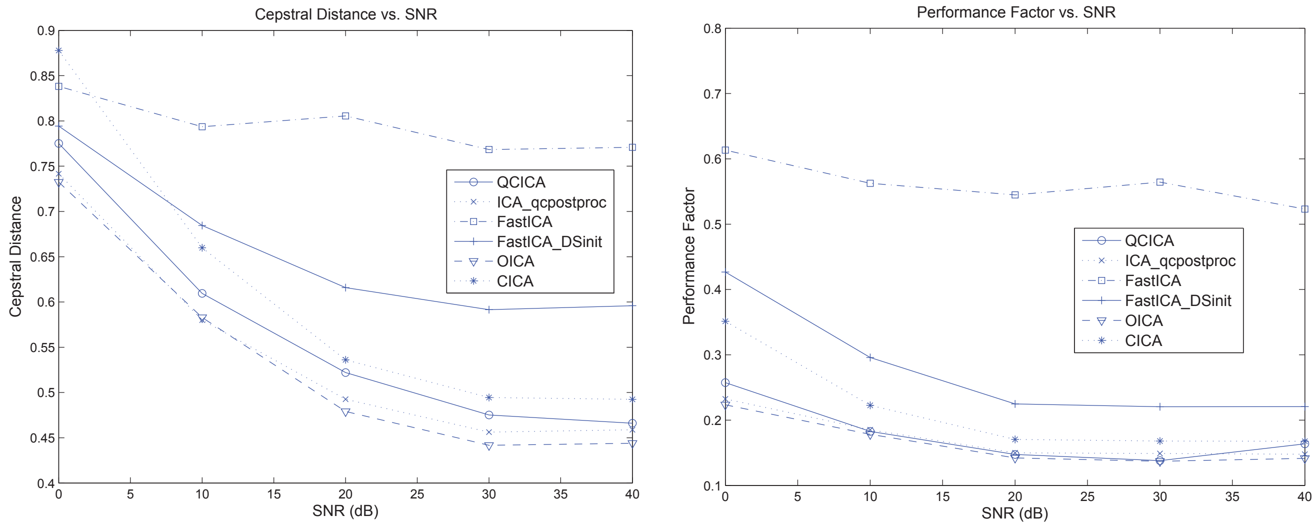
$\overline{\theta}$ is the assumed look direction, $\Delta\theta$ is a measure of uncertainty in the assumed look direction, $f(\theta)$ is a spatial weighting function, ω is a fixed frequency and $\mathbf{a}(\theta, \omega)$ is the array steering vector. Φ is a positive definite constraint matrix which can be calculated by either mathematical integration or by numerical techniques. $\mathbf{w}$ is the beamformer's weight vector of interest. $\mathbf{w}_d^H \mathbf{a}(\theta, \omega)$ is the desired response in the direction θ, and it is expressed as the inner product between a desired beamformer's weight vector $\mathbf{w}_d$ and steering vector $\mathbf{a}(\theta, \omega)$ to simplify computation. Generally, Delay-and-Sum beamformer is used as the desired beamformer because of its robustness in the look direction. Then the quadratic constraint is written as,

$$(\mathbf{w} - \mathbf{w}_d)^H \Phi (\mathbf{w} - \mathbf{w}_d) \le \varepsilon$$

We propose to combine this quadratic constraint with the negentropy maximization criterion. The new optimization problem is stated as,

$$\min_{\mathbf{w}} E\{G(|\mathbf{w}^H \mathbf{x}|^2)\}, \quad s.t. \quad (\mathbf{w} - \mathbf{w}_d)^H \Phi (\mathbf{w} - \mathbf{w}_d) \le \varepsilon \tag{5}$$

The quadratic constraint will ensure the solution of interest has a flat response close to 1 in the uncertainty region. When the desired source lies in the uncertainty region, it will be conveyed to the output with little distortion. The negentropy maximization criterion then will suppress the undesired sources to ensure distribution of the output be as far as possible from the Gaussian distribution. In other words, maximizing the negentropy, as an ICA criterion, will converge to recover one of the N sources, while the quadratic geometric constraint will ensure it converge to recover the desired source. The little distorted conveyance of the desired source to the output is an attractive attribute. It avoids the scaling ambiguity intrinsic in ICA. This is most useful for frequency domain ICA approach, enabling



(a) Cepstral distance vs. SNR

(b) Mean performance factor vs. SNR

Fig. 1. The performance of various ICA algorithms vs. SNR (2 sources exist)

proper scaling in every frequency bin and escaping the re-scaling headache in conventional frequency domain ICA approach. Define

$$\tilde{\mathbf{w}} = \Phi^{1/2}(\mathbf{w} - \mathbf{w}_d)$$
$$f(\tilde{\mathbf{w}}) = E\{G(|(\Phi^{-1/2}\tilde{\mathbf{w}} + \mathbf{w}_d)^H \mathbf{x}|^2)\}.$$

Problem (5) can be written as,

$$\min_{\tilde{\mathbf{w}}} f(\tilde{\mathbf{w}}), \quad s.t. \quad \|\tilde{\mathbf{w}}\|^2 \leq \varepsilon$$

We use the following iterative conjugate complex gradient and projection method to solve the above optimization problem (see [9] and refs therein), where μ is a step size parameter.

$$\tilde{\mathbf{w}}_k = P[\tilde{\mathbf{w}}_{k-1} - \mu \nabla_{\tilde{\mathbf{w}}} f(\tilde{\mathbf{w}})]$$
$$P(\tilde{\mathbf{w}}) = \begin{cases} \tilde{\mathbf{w}}, & \|\tilde{\mathbf{w}}\| \leq \sqrt{\varepsilon} \\ \tilde{\mathbf{w}}\frac{\sqrt{\varepsilon}}{\|\tilde{\mathbf{w}}\|}, & \|\tilde{\mathbf{w}}\| > \sqrt{\varepsilon}. \end{cases}$$

3.2. Use geometric test as post-processing to ICA

The second approach for combining ICA with geometric information is to employ a quadratic geometric test as a post-processing step to pick up the desired source after ICA processing. In every frequency bin, the ICA algorithm separates instantaneously mixed source signals, then the quadratic geometric test will pick up the desired source.

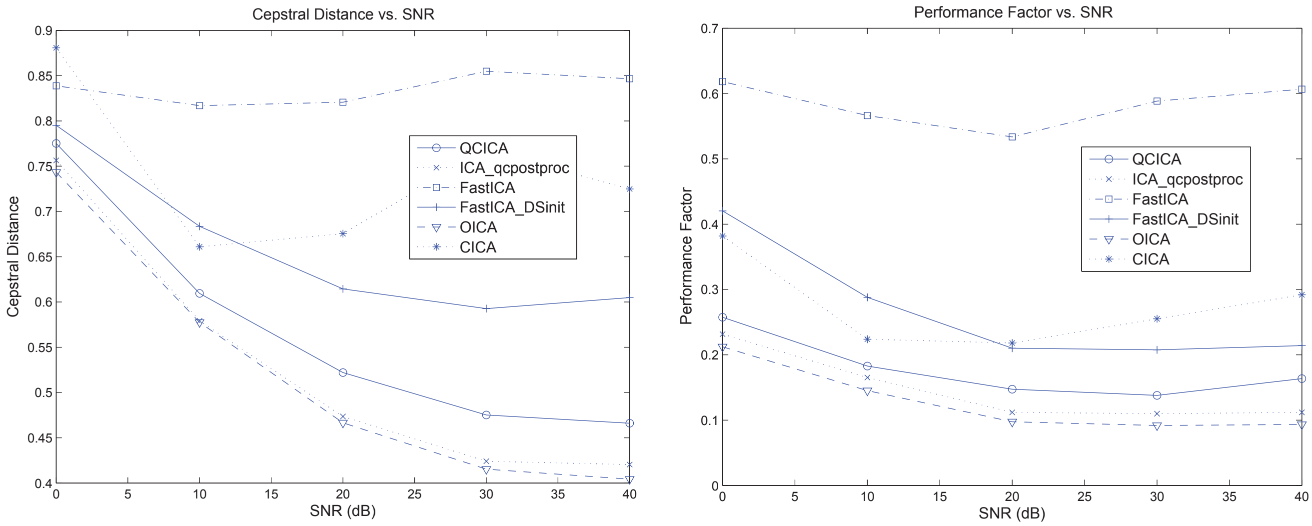
Assuming a M sensor, N source system. Suppose the estimated source number is N' (N' is not necessary to be N but should satisfies $N' \geqslant N$). Suppose the ICA weight vectors are $\mathbf{w}_i, i = 1, .., N'$. We can calculate a set of response scores, $\mathbf{w}_i^H \Phi \mathbf{w}_i, i = 1, .., N'$ (Φ is defined in sec.3.1) based on a rough direction information of the desired source. Because of the similarity between ICA and beamforming, the ICA weight vector which recover the desired source should have a response close to 1 around the desired source's direction, while other ICA weight vectors should have a response close to 0 around that direction. Consequently, one of the response scores should be close to 1 while others should be close to 0. We will pick the ICA weight vector which yields the biggest response score to recover the desired source.

4. SIMULATION

We provide examples on microphone array speech processing to compare the performances of various algorithms. The image method is used to generate artificial room impulse response. Simulated room dimension is $[8, 5, 3.5]m$. We simulate an 8 element uniform linear array with 4cm inter-microphone spacing. In the experiments, it is always assumed the look direction is the broadside direction of the array, i.e. $0°$. Every source signal is a speech wave signal. The frequency domain ICA approach is employed. The performance of various algorithms is measured by the average performance factor across all frequency bins and the cepstral distance between the recovered signal's spectrum and the original desired source's spectrum. The performance factor is defined as, $\sum_i \frac{|p_i|}{\max_j |p_j|} - 1$, where $\mathbf{p} = \mathbf{w}^H A$, A is the mixing matrix on one frequency bin. The performance factor measures how the algorithm enhance the desired source and suppress interference signals on one frequency bin. The cepstral distance is used because it is a perceptual metric commonly used in speech processing to measure distortion. Not only does it account for the interference and noise level, but it also detect spectrum shape distortion. Thereby both permutation and scaling problem are taken into consideration by a single metric.

We use the following notation for each algorithm.

- QCICA: ICA with quadratic geometric constraint (sec.3.1)
- ICA_qcpostproc: Use geometric test as post-processing (sec.3.2)
- FastICA: FastICA with random initialization (sec.2.1)
- FastICA_DSinit: FastICA with Delay-and-Sum beamformer as initialization.
- OICA: optimum ICA, the permutation problem is solved manually assuming we know the correct permutation.



(a) Cepstral distance vs. SNR

(b) Mean performance factor vs. SNR

Fig. 2. The performance of various ICA algorithms vs. SNR (2 sources exist, assumed source number is 3)

- CICA: ICA with linear geometric constraint (sec.2.2)

Example 1

There are two sources exist in this scenario, one desired source, the other interference signal. The desired source and the interference signal comes from direction 5° and 45° respectively. The assumed look direction is broadside, i.e. 0°, which means a 5° look direction error. Fig.1 demonstrates the performance of various algorithms versus SNR (signal to white noise ratio).

Assuming the number of sources is known accurately. The FastICA algorithm use PCA as a preprocessing to make the mixing matrix square. When FastICA use a random initialization, the performance is bad. This is caused by the different permutation in different frequency bins. To give an example, we observe 60 wrong permutations in a total of 128 frequency bins in one sample experiment. When FastICA use Delay-and-Sum beamformer as initialization, the performance improves but still not good enough. Some frequency bin may still shown permutation problem. The CICA use PCA preprocessing as well. It shows better performance than FastICA_DSinit. When the PCA preprocessing is not done properly with the right source number, the CICA algorithm was found to totally fail. The OICA corrects the permutation problem manually and can be taken as a baseline for the ICA algorithms. The proposed QCICA use all 8 channels in the optimization and the Delay-and-Sum beamformer is used for initialization. It does not use PCA preprocessing. Its performance is close to OICA's performance. We observe that the proposed ICA_qcpostproc has almost the same performance as the OICA algorithm.

Example 2:

In this example, the scenario is the same as in **Example 1** except that the assumed source number is 3. In other words, the number of sources is overestimated. Consequently, PCA preprocessing will employ 3 dimensions. Fig.2 demonstrates the performance versus SNR. The experiment results illustrate the performances of QCICA and ICA_qcpostproc are not affected much by the wrong information about the number of sources while CICA fails under such scenario.

Example 3:

This example demonstrates the performances versus SNR when two interference signals exist. The second interference signal comes from direction -60°. All the other settings are the same as those in **Example 1**. The experiment results are consistent with the those shown in **Example 1** and **2**. The plots are not shown here because of limited space.

5. REFERENCES

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

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[3] Harry L. Van Trees, *Optimum Array Processing*, Wiley-Interscience, 2002.

[4] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *ICASSP*, 2000.

[5] H. Saruwatari, S. Kurita, etc. "Blind source separation combining independent component analysis and beamforming," *EURASIP*, vol. 2003, no. 11, pp. 1135–1146, 2003.

[6] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[7] M. Knaak, S. Araki, and S. Makino, "Geometrically constrained independent component analysis," *IEEE Trans. on Audio and Speech Processing*, vol. 15, no. 2, pp. 715–726, Feb. 2007.

[8] E. Bingham and A. Hyvarinen, "A fast fixed point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Networks*, vol. 10, no. 1, pp. 1–8, 2000.

[9] W. Zhang and B. Rao, "Robust broadband beamformer with diagonally loaded constraint matrix and its application to speech recognition," in *ICASSP*, May 2006, pp. 785–788.