Blind Separation of Multichannel Biomedical Image Patterns by Non-negative Least-Correlated Component Analysis

Fa-Yu Wang[1,2], Yue Wang[2], Tsung-Han Chan[1], and Chong-Yung Chi[1]

[1] National Tsing Hua University, Hsinchu, Taiwan 30013 ROC
[2] Virginia Polytechnic Institute and State University, Arlington, VA 22203 USA
wangfaa@ms16.hinet.net, yuewang@vt.edu, d935608@oz.nthu.edu.tw,
cychi@ee.nthu.edu.tw
http://www.ee.nthu.edu.tw/cychi

Abstract. Cellular and molecular imaging promises powerful tools for the visualization and elucidation of important disease-causing biological processes. Recent research aims to simultaneously assess the spatial-spectral/temporal distributions of multiple biomarkers, where the signals often represent a composite of more than one distinct source independent of spatial resolution. We report here a blind source separation method for quantitative dissection of mixed yet correlated biomarker patterns. The computational solution is based on a latent variable model, whose parameters are estimated using the non-negative least-correlated component analysis (nLCA) proposed in this paper. We demonstrate the efficacy of the nLCA with real bio-imaging data. With accurate and robust performance, it has powerful features which are of considerable widespread applicability.

1 Introduction

Multichannel biomedical imaging promises simultaneous imaging of multiple biomarkers, where the pixel values often represent a composite of multiple sources independent of spatial resolution. For example, in vivo multispectral imaging exploits emissions from multiple fluorescent probes, aiming at discriminating often overlapped spatial-spectral distributions and reducing background autofluorescence [1, 2]. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) utilizes various molecular weight contrast agents to investigate tumor microvascular status and then obtain the information about the therapeutic effect under anti-angiogenic drugs. However, due to the heterogeneous nature of tumor microvessels associated with different perfusion rate, DCE-MRI measured signals are the mixture of the permeability images corresponding to fast perfusion and slow perfusion. Further examples include dynamic positron emission tomography, and dynamic optical molecular imaging [1].

The major efforts for computational separation of composite biomarker distributions are: supervised spectrum unmixing [2], a priori weighted subtraction [3], parametric compartment modeling and independent component analysis (ICA) [4, 5]. The major limitations associated with the existing methods are

J.C. Rajapakse, L. Wong, and R. Acharya (Eds.): PRIB 2006, LNBI 4146, pp. 151–162, 2006.

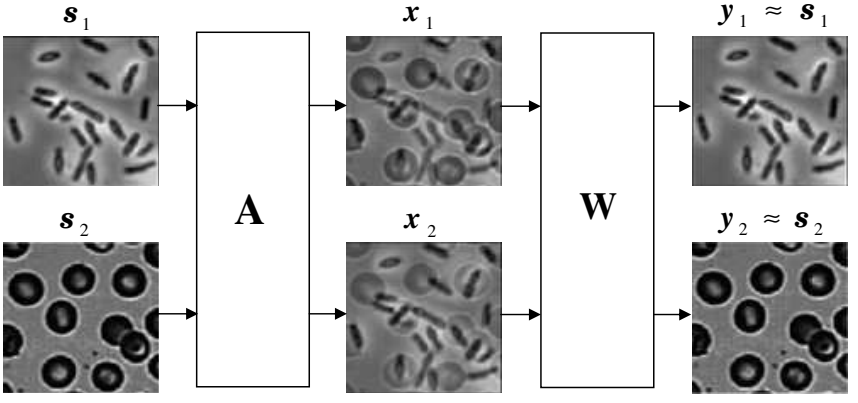

Fig. 1. Block diagram of 2×2 mixing and demixing systems

the inability of acquiring in vivo spectra of the probes (e.g., individual physiological conditions and microenvironment, pH, temperature, oxygen, blood flow, etc.) [2] and the unrealistic assumptions about the characteristics of the unknown sources and mixing processes (e.g., source independence, model identifiability, etc.) [5, 6]. Our goal and effort therefore, is to develop a novel blind source separation (BSS) method that is able to separate correlated or dependent sources under non-negativity constraints [7]. This new BSS method is called the non-negative least-correlated component analysis (nLCA) whose principle and applications will be reported in detail here.

In the next section, we present the nLCA, including model assumptions, theory and computational methods for blind separation of non-negative sources with a given set of observations of a non-negative mixing system. In Section 3, we demonstrate the efficacy of the nLCA by two experiments (human face images and DCE-MRI analysis) and its performance superior to some existing algorithms followed by a discussion of future research.

2 Non-negative Least-Correlated Component Analysis

As shown in Fig. 1, consider a 2×2 non-negative mixing system with the input signal vector $\mathbf{s}[n] = (s_1[n], s_2[n])^T$ (e.g., images of two different types of cells) and the output vector

$$\mathbf{x}[n] = (x_1[n], x_2[n])^T = \mathbf{A}\mathbf{s}[n] \tag{1}$$

where the superscript 'T' denotes the transpose of a matrix or vector and $\mathbf{A} = \{a_{ij}\}_{2\times 2}$ is the unknown non-negative mixing matrix. The blind source separation problem is to find a demixing matrix $\mathbf{W}$ from the given measurements $\mathbf{x}[n]$, $n = 1, 2, ..., L$, such that

$$\mathbf{y}[n] = (y_1[n], y_2[n])^T = \mathbf{W}\mathbf{x}[n] = \mathbf{W}\mathbf{A}\mathbf{s}[n] \approx \mathbf{P}\mathbf{s}[n] \tag{2}$$

i.e., $\mathbf{WA} \simeq \mathbf{P}$ (permutation matrix). Alternatively, let

$$\boldsymbol{x}_i = (x_i[1], x_i[2], ..., x_i[L])^T,\ i = 1,\ 2, \qquad (i\text{th observation}) \tag{3}$$

$$\boldsymbol{s}_i = (s_i[1], s_i[2], ..., s_i[L])^T,\ i = 1,\ 2, \qquad (i\text{th unknown source}) \tag{4}$$

$$\boldsymbol{y}_i = (y_i[1], y_i[2], ..., y_i[L])^T,\ i = 1,\ 2. \qquad (i\text{th extracted source}) \tag{5}$$

Then the observations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and the extracted sources $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ can be expressed as

$$\begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \end{bmatrix} = \mathbf{A} \begin{bmatrix} \boldsymbol{s}_1^T \\ \boldsymbol{s}_2^T \end{bmatrix}, \tag{6}$$

$$\begin{bmatrix} \boldsymbol{y}_1^T \\ \boldsymbol{y}_2^T \end{bmatrix} = \mathbf{W} \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \end{bmatrix}, \tag{7}$$

respectively.

For ease of later use, the correlation coefficient and the angle between $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are defined as

$$\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\boldsymbol{s}_1^T \boldsymbol{s}_2}{\|\boldsymbol{s}_1\| \cdot \|\boldsymbol{s}_2\|}, \tag{8}$$

$$\theta(\boldsymbol{s}_1, \boldsymbol{s}_2) = \cos^{-1}(\rho(\boldsymbol{s}_1, \boldsymbol{s}_2)), \tag{9}$$

respectively, where $\|\boldsymbol{s}_i\|$ $(= (\boldsymbol{s}_i^T \boldsymbol{s}_i)^{1/2})$ is the norm of $\boldsymbol{s}_i$. Next, let us present the assumptions and the associated theory and methods of the nLCA, respectively.

2.1 Model Assumptions

Let us make some assumptions about the sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, and the mixing matrix $\mathbf{A}$ as follows:

- $(\mathcal{A}1)$ $\boldsymbol{s}_1 \succeq \mathbf{0}$ and $\boldsymbol{s}_2 \succeq \mathbf{0}$ (i.e., $s_1[n] \geq 0$, $s_2[n] \geq 0$ for all n), and the two distinct sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are linearly independent (i.e., $\boldsymbol{s}_2 \neq \alpha \boldsymbol{s}_1$ where $\alpha \neq 0$).
- $(\mathcal{A}2)$ $\mathbf{A} \succeq \mathbf{0}$ (i.e., all the entries of $\mathbf{A}$ are non-negative).
- $(\mathcal{A}3)$ $\mathbf{A}$ is full rank (i.e., nonsingular).
- $(\mathcal{A}4)$ $\mathbf{A} \cdot \mathbf{1} = \mathbf{1}$, where $\mathbf{1} = (1, 1)^T$ (i.e., the sum of the entries of each row of $\mathbf{A}$ is equal to unity).

Assumptions $(\mathcal{A}1)$ and $(\mathcal{A}2)$ hold valid in biomedical imaging applications [6] where all the sources and all the entries of the mixing matrix are non-negative, and meanwhile $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are allowed to be correlated (i.e., $\widetilde{\boldsymbol{s}}_1^T \widetilde{\boldsymbol{s}}_2 / L \neq 0$ where $\widetilde{\boldsymbol{s}}_i = (s_i[1] - \mu_i, s_i[2] - \mu_i, ..., s_i[L] - \mu_i)^T$ in which $\mu_i = \sum_{n=1}^{L} s_i[n]/L$). Assumptions $(\mathcal{A}1)$ and $(\mathcal{A}3)$ imply that the two observations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are linearly independent vectors (i.e., $\boldsymbol{x}_2 \neq \alpha \boldsymbol{x}_1$ where $\alpha \neq 0$). Note that the assumption that sources are mutually statistically independent is a fundamental assumption made by most ICA algorithms that requires $\widetilde{\boldsymbol{s}}_1^T \widetilde{\boldsymbol{s}}_2 / L = 0$ (uncorrelated).

2.2 Theory and Methods

The proposed nLCA is supported by the following theorem.

Theorem 1 (Correlation Increase Theorem). Under Assumptions ($\mathcal{A}1$) and ($\mathcal{A}2$), $\rho(\boldsymbol{x}_1, \boldsymbol{x}_2) \geq \rho(\boldsymbol{s}_1, \boldsymbol{s}_2)$ (or $\theta(\boldsymbol{x}_1, \boldsymbol{x}_2) \leq \theta(\boldsymbol{s}_1, \boldsymbol{s}_2)$) as shown in Fig. 2.

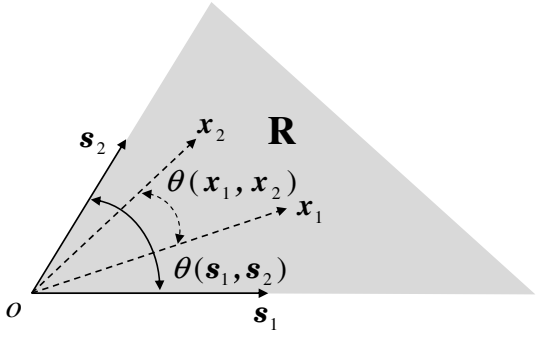

Fig. 2. Observation vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ which are inside the shaded region $\mathbf{R}$ formed by source vectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$

The proof of Theorem 1 is given in Section 4.1. Theorem 1 implies that linear non-negative mixing of non-negative sources leads to increase of the correlation coefficient. Based on Theorem 1, a straightforward method, referred as Method 1, is to design the demixing matrix $\mathbf{W}$ by reducing the correlation coefficient $\rho(\boldsymbol{y}_1, \boldsymbol{y}_2)$ (i.e., maximizing the angle $\theta(\boldsymbol{y}_1, \boldsymbol{y}_2)$) of the extracted sources, subject to the two constraints $\mathbf{y}[n] \succeq \mathbf{0}$ (i.e., $\mathbf{y}[n] \approx \mathbf{Ps}[n]$) and $\mathbf{W} \cdot \mathbf{1} = \mathbf{1}$ (due to $\mathbf{W}(\mathbf{A} \cdot \mathbf{1}) = \mathbf{P} \cdot \mathbf{1} = \mathbf{1}$).

Method 1 (A direct method)
The demixing matrix $\mathbf{W}$ is obtained as

$$\mathbf{W}^\star = \arg\min_{\mathbf{W}} \rho(\boldsymbol{y}_1, \boldsymbol{y}_2)$$

subject to $\mathbf{y}[n] \succeq \mathbf{0}$, $\forall n$ (i.e., $\boldsymbol{y}_1 \succeq \mathbf{0}$ and $\boldsymbol{y}_2 \succeq \mathbf{0}$) and $\mathbf{W} \cdot \mathbf{1} = \mathbf{1}$.

Finding the optimum demixing matrix $\mathbf{W}^\star$ is apparently a nonlinear and non-convex optimization problem. Fortunately, a closed-form solution for $\mathbf{W}^\star$ can be shown to be

$$\mathbf{W}^\star = \begin{bmatrix} \frac{-\tan\phi(\mathbf{x}[k_1])}{1-\tan\phi(\mathbf{x}[k_1])} & \frac{1}{1-\tan\phi(\mathbf{x}[k_1])} \\ \frac{-\tan\phi(\mathbf{x}[k_2])}{1-\tan\phi(\mathbf{x}[k_2])} & \frac{1}{1-\tan\phi(\mathbf{x}[k_2])} \end{bmatrix} \tag{10}$$

where

$$\tan\phi(\mathbf{x}[k_1]) = x_2[k_1]/x_1[k_1] = \max_n\{x_2[n]/x_1[n]\}, \tag{11}$$

$$\tan\phi(\mathbf{x}[k_2]) = x_2[k_2]/x_1[k_2] = \min_n\{x_2[n]/x_1[n]\}. \tag{12}$$

The proof of (10) is given in Section 4.2. Next let us present an indirect method.

The second method is to estimate the mixing matrix $\mathbf{A}$ from the given observations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, based on the following theorem.

Theorem 2. Suppose that Assumptions ($\mathcal{A}1$) and ($\mathcal{A}2$) hold true, and that there exist $\mathbf{s}[l_1] = (s_1[l_1] \neq 0, 0)^T$ and $\mathbf{s}[l_2] = (0, s_2[l_2] \neq 0)^T$ for some l_1 and l_2. Let

$$\phi(\mathbf{x}[k_1]) = \max\{\phi(\mathbf{x}[l_1] = \mathbf{A}\mathbf{s}[l_1]), \phi(\mathbf{x}[l_2] = \mathbf{A}\mathbf{s}[l_2])\},$$
$$\phi(\mathbf{x}[k_2]) = \min\{\phi(\mathbf{x}[l_1]), \phi(\mathbf{x}[l_2])\}.$$

Then $0 \leq \phi(\mathbf{x}[k_2]) \leq \phi(\mathbf{x}[n]) \leq \phi(\mathbf{x}[k_1]) \leq \pi/2, \forall n$.

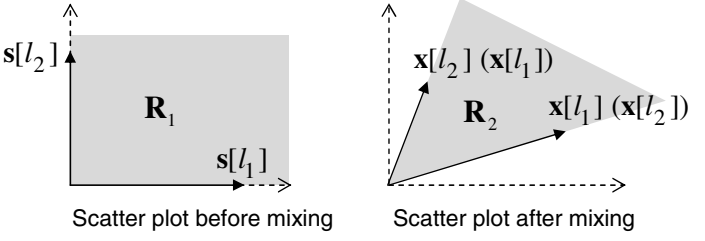

Fig. 3. Scatter plot coverage ($\mathbf{R}_1$) of two sources and that ($\mathbf{R}_2$) of the two associated observations

The proof of Theorem 2 is given in Section 4.3. By Theorem 2, if $\mathbf{s}[l_1] = (s_1[l_1] \neq 0, 0)^T$ and $\mathbf{s}[l_2] = (0, s_2[l_2] \neq 0)^T$ for some l_1 and l_2 are two points each on one edge of the scatter plot of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, a two-dimensional plot of $\mathbf{s}[n]$, $n = 1, ..., L$, (i.e., the shaded region $\mathbf{R}_1$ in Fig. 3). Then the associated $\mathbf{x}[l_1]$ and $\mathbf{x}[l_2]$ will also be on each of the two edges of the scatter plot of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, (i.e., the shaded region $\mathbf{R}_2 \subseteq \mathbf{R}_1$ in Fig. 3), respectively. In view of this observation, the unknown mixing matrix $\mathbf{A}$ can be easily solved from

$$\tan\phi(\mathbf{x}[l_1]) = x_2[l_1]/x_1[l_1] = a_{21}/a_{11}, \tag{13}$$
$$\tan\phi(\mathbf{x}[l_2]) = x_2[l_2]/x_1[l_2] = a_{22}/a_{12}, \tag{14}$$
$$\mathbf{A} \cdot \mathbf{1} = \mathbf{1}, \tag{15}$$

and the solutions for a_{11} and a_{21} are given by

$$a_{11} = \frac{1 - \tan\phi(\mathbf{x}[l_2])}{\tan\phi(\mathbf{x}[l_1]) - \tan\phi(\mathbf{x}[l_2])}, \tag{16}$$
$$a_{21} = \frac{(1 - \tan\phi(\mathbf{x}[l_2]))\tan\phi(\mathbf{x}[l_1])}{\tan\phi(\mathbf{x}[l_1]) - \tan\phi(\mathbf{x}[l_2])}, \tag{17}$$

which together with (15) lead to the solution for $\mathbf{A}$. The above procedure for estimating $\mathbf{A}$, referred as to Method 2, is summarized as follows:

Method 2 (An indirect method)
Find $\tan\phi(\mathbf{x}[k_1])$ and $\tan\phi(\mathbf{x}[k_2])$ using (11) and (12), and set $l_1 = k_1$ and $l_2 = k_2$. Obtain a_{11} and a_{21} using (16) and (17), respectively, and then obtain $a_{12} = 1 - a_{11}$ and $a_{22} = 1 - a_{21}$. Finally, obtain $\mathbf{W} = \mathbf{A}^{-1}$.

Let us conclude this section with the following two remarks.

Remark 1. The condition that $\mathbf{s}[l_1] = (s_1[l_1] \neq 0, 0)^T$ and $\mathbf{s}[l_2] = (0, s_2[l_2] \neq 0)^T$ for some l_1 and l_2 exist as stated in Theorem 2 guarantees that the estimate $\mathbf{A}$ obtained by Method 2 is existent and unique up to a column permutation of $\mathbf{A}$. Under the same condition, one can easily prove that $\mathbf{W}^\star\mathbf{A} = \mathbf{P}$ (see (10)), implying the existence and uniqueness of the estimate $\mathbf{A}$ for Method 1 up to a column permutation of $\mathbf{A}$. However, this condition may not be perfectly satisfied but approximately satisfied in practical applications, i.e., $\mathbf{s}[l_1] = (s_1[l_1] \neq 0, s_2[l_1] \simeq 0)^T$ and $\mathbf{s}[l_2] = (s_1[l_2] \simeq 0, s_2[l_2] \neq 0)^T$ for some l_1 and l_2. For example, non-overlapping region (for which $\mathbf{s}[n] = (s_1[n] \neq 0, s_2[n] \simeq 0)^T$ and $\mathbf{s}[n] = (s_1[n] \simeq 0, s_2[n] \neq 0)^T$) in the spatial distribution of a fast perfusion and a slow perfusion source images on brain MRI [8] is usually higher than 95%. The estimated sources $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ turn out to be approximations of the original sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Remark 2. The proposed nLCA is never limited by Assumption ($\mathcal{A}4$). As $\mathbf{A}\cdot\mathbf{1} \neq \mathbf{1}$, the mixing model given by (1) can be converted into the following model:

$$\widetilde{\mathbf{x}}[n] = \mathbf{D}_1\mathbf{x}[n] = (\mathbf{D}_1\mathbf{A}\mathbf{D}_2)\mathbf{D}_2^{-1}\mathbf{s}[n] = \widetilde{\mathbf{A}}\widetilde{\mathbf{s}}[n] \tag{18}$$

where $\mathbf{D}_1 = \mathrm{diag}\{1/\sum_n x_1[n], 1/\sum_n x_2[n]\}$ and $\mathbf{D}_2 = \mathrm{diag}\{\sum_n s_1[n], \sum_n s_2[n]\}$ (2×2 diagonal matrices), $\widetilde{\mathbf{A}} = \mathbf{D}_1\mathbf{A}\mathbf{D}_2$ for which Assumptions ($\mathcal{A}2$), ($\mathcal{A}3$) and ($\mathcal{A}4$) ($\widetilde{\mathbf{A}}\cdot\mathbf{1} = \mathbf{1}$) are satisfied, and the sources are $\widetilde{\mathbf{s}}[n] = \mathbf{D}_2^{-1}\mathbf{s}[n]$ (instead of $\mathbf{s}[n]$) for which Assumption ($\mathcal{A}1$) is also satisfied.

3 Experiments and Discussion

So far, we have described the theory behind nLCA, and have presented two nLCA methods to separate composite biomarker distributions. We shall now illustrate the efficacy of the proposed nLCA using mixtures of real multichannel images.

The first experiment reports the effectiveness of the nLCA for mixtures of two correlated human face images taken from the benchmarks in [11]. For the second experiment, the proposed nLCA is applied to DCE-MRI of breast cancer, where the two source images correspond to the permeability distributions of a fast perfusion image and a slow perfusion image in the region of interest [5]. In each of the two experiments, 50 randomly independent mixtures were generated and then processed using Method 2, and three existing algorithms, FastICA [9], non-negative ICA (nICA) [6], and non-negative matrix factorization (NMF) [10] for performance comparison. The average of the error index E1 [8] over the 50 independent runs was calculated as the performance index, where

$$\mathrm{E1} = \sum_{i=1}^{2}\left[\left(\sum_{j=1}^{2}\frac{|\widetilde{p}_{ij}|}{\max_k\{|\widetilde{p}_{ik}|\}}\right)-1\right]+\sum_{j=1}^{2}\left[\left(\sum_{i=1}^{2}\frac{|\widetilde{p}_{ij}|}{\max_k\{|\widetilde{p}_{kj}|\}}\right)-1\right] \quad (19)$$

where $\widetilde{p}_{ij}$ denotes the (i,j)-element of $\widetilde{\mathbf{P}} = \mathbf{WA}$. Note that the value of E1 is smaller for $\widetilde{\mathbf{P}}$ closer to a permutation matrix.

The averaged E1 associated with the proposed nLCA, and FastICA, nICA, NMF for the human face experiment are displayed in Table 1. One can see from this table that the proposed nLCA performs best, the nICA second, the FastICA third, and the NMF fourth.

In order to further illustrate the performance insights of the four non-negative source separation algorithms under test. A typical set of results of the human face experiment is displayed in Fig. 4, including two original source images $\boldsymbol{s}_1$, $\boldsymbol{s}_2$ and the associated scatter plot, observations $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and the associated scatter plot, the extracted source images $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and the associated scatter plots obtained by the four algorithms. Some observations from Fig. 4 are as follows. The scatter plots shown in Figs. 4(a) and 4(b) are similar to those shown in Fig. 3 and thus consistent with Theorem 2. The scatter plot associated with the proposed nLCA shown in Fig. 4(c) resembles that shown in Fig. 4(a) much better than the other scatter plots.

As previously mentioned, DCE-MRI provides temporal mixtures of heterogeneous permeability distributions corresponding to slow and fast perfusion rates. The second experiment is to separate the two perfusion distributions from their mixtures. The averaged E1 associated with the proposed nLCA, and FastICA, nICA, NMF are also displayed in Table 1. Moreover, a typical set of results of the DCE-MRI experiment corresponding to those shown in Fig. 4 is displayed in Fig. 5. Again, the same conclusions obtained from the human face experiment apply to the DCE-MRI experiment, i.e., the proposed nLCA outperforms the other three algorithms. These experimental results demonstrate the efficacy of the proposed nLCA.

Next let us discuss why the proposed nLCA performs better than FastICA, nICA, NMF. FastICA and nICA are statistical BSS algorithms under the assumption of non-Gaussian independent sources for the former, and the assumption of non-negative uncorrelated and well-grounded sources (i.e., probability $\mathrm{P_r}\{s_i[n] < \delta\} > 0$ for any $\delta > 0$) for the latter. However, the sources in the above experiments are correlated as in many other biomedical imaging appli-

Table 1. The performance (averaged E1) of the proposed nLCA and FastICA, nICA and NMF for the human face and DCE-MRI experiments

Method		nLCA	FastICA	nICA	NMF
Averaged E1	Human face experiment	0.082	0.453	0.369	0.640
	DCE-MRI experiment	0	0.162	0.114	0.432

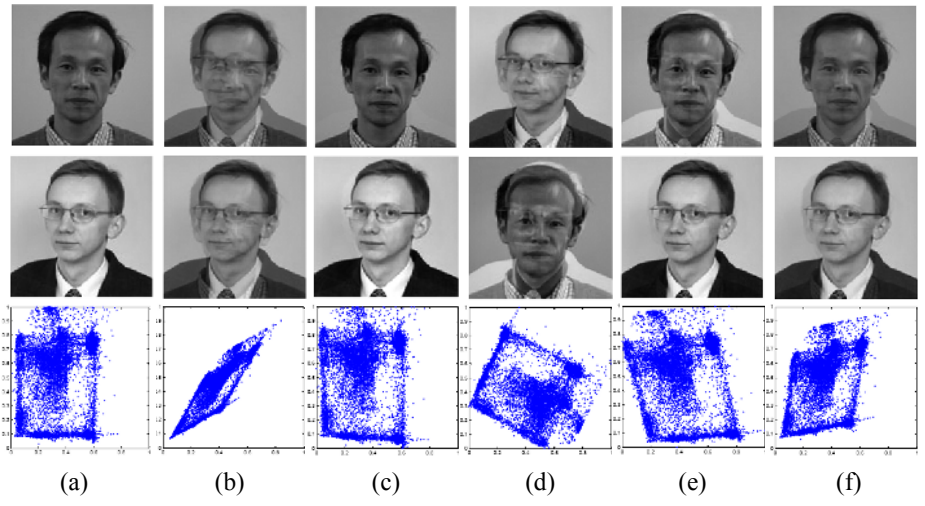

Fig. 4. Human face images (top and middle rows) and associated scatter plots (bottom row) for (a) the sources, (b) the observations and the extracted sources obtained by (c) nLCA, (d) FastICA, (e) nICA, and (f) NMF

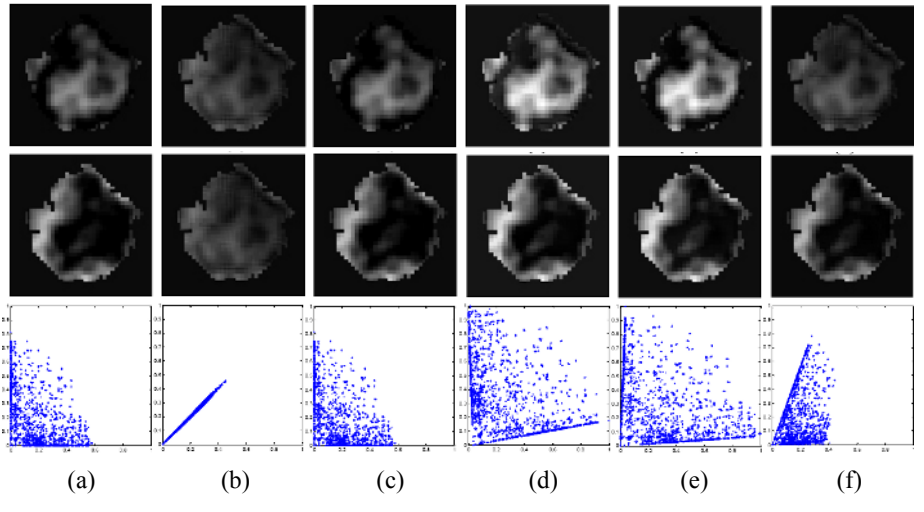

Fig. 5. DCE-MRI images (top and middle rows) and associated scatter plots (bottom row) for (a) the sources (permeability) corresponding to the slow (top plot) and fast (middle plot) perfusion, (b) the observations and the extracted sources obtained by (c) nLCA, (d) FastICA, (e) nICA, and (f) NMF

cations, implying that the source independence assumption made by FastICA and nICA is not satisified. On the other hand, both the proposed nLCA and NMF are algebraic approaches basically under the same realistic assumptions (i.e., Assumptions ($\mathcal{A}1$), ($\mathcal{A}2$) and ($\mathcal{A}3$)). However, the proposed nLCA has op-

timum and closed-form solutions (either for the demixing matrix **W** or for the mixing matrix **A**) for Methods 1 and 2, but NMF is an iterative algorithm which may provide a local optimum solution. Therefore, the proposed nLCA is computationally efficient and outperforms the other three algorithms in the above experiments.

We believe that the proposed nLCA is a promising method for blind separation of multiple biomarker patterns. We would expect it to be an effective image formation tool applicable to many other multichannel biomedical imaging modalities [1]. Extension of the proposed nLCA to the case of more than two sources and its performance in the presence of measurement noise are currently under investigation.

Acknowledgments. This work was supported partly by the National Science Council (R.O.C.) under Grant NSC 94-2213-E-007-026 and partly by the National Institutes of Health under Grant EB000830.

References

1. H. R. Herschman, "Molecular imaging: looking at problems, seeing solutions," Science, vol. 302, pp. 605-608, 2003.
2. M. Zhao, M. Yang, X.-M. Li, P. Jiang, E. Baranov, S. Li, M. Xu, S. Penman, and R. M. Hoffman, "Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing salmonella typhimurium," Proc. Natl. Acad. Sci., vol. 102, pp. 755-760, 2005.
3. S. G. Armato, "Enhanced visualization and quantification of lung cancers and other diseases of the chest," Experimental Lung Res., vol. 30, pp. 72-77, 2004.
4. Y. Wang, J. Zhang, K. Huang, J. Khan, and Z. Szabo, "Independent component imaging of disease signatures," Proc. IEEE Intl. Symp. Biomed. Imaging, Washington DC, July 7-10, 2002, pp. 457-460.
5. Y. Wang, R. Srikanchana, P. Choyke, J. Xuan, and Z. Szabo, "Computed simultaneous imaging of multiple functional biomarkers," Proc. IEEE Intl. Symp. Biomed. Imaging, Arlington, VA, April 15-18, 2004, pp. 225-228.
6. E. Oja and M. Plumbley, "Blind separation of positive sources by globally convergent gradient search," Neural Computation, vol. 16, pp. 1811-1825, 2004.
7. Y. Wang, J. Xuan, R. Srikanchana, and P. L. Choyke, "Modeling and reconstruction of mixed functional and molecular patterns," Intl. J. Biomed. Imaging, 2005 in press.
8. Y. Zhou, S. C. Huang, T. Cloughesy, C. K. Hoh, K. Black, and M. E. Phelps, "A modeling-based factor extraction method for determining spatial heterogeneity of Ga-68 EDTA kinetics in brain tumors," IEEE Trans. Nucl. Sci., vol. 44, no. 6, pp. 2522-2527, Dec. 1997.
9. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley, 2001.
10. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, pp. 788-791, Oct. 1999.
11. Andrzej Cichocki and Shun-ichi Amari, *Adaptive Blind Signal and Image Processing*. John Wiley and Sons, Inc., 2002.

4 Appendix

4.1 Proof of Theorem 1

Let $\mathcal{V}$ be a 2-dimensional vector space spanned by the linearly independent vectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ for which $0 \leq \rho(\boldsymbol{s}_1, \boldsymbol{s}_2) < 1$. Let $\boldsymbol{u}_1 = \boldsymbol{s}_1/\|\boldsymbol{s}_1\|$, and $\boldsymbol{u}_2$ (which can be obtained via Gram-Schmidt orthogonalization) be a set of orthonormal basis vectors of $\mathcal{V}$. Then any vector in $\mathcal{V}$ can be represented in terms of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, as shown in Fig. 6.

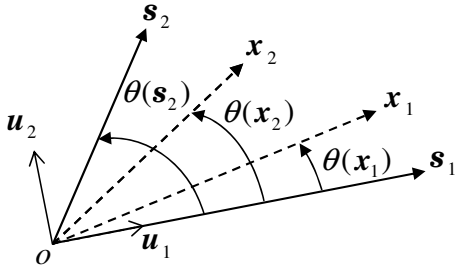

Fig. 6. Source vectors and observation vectors after non-negative mixing

Let $\theta(\boldsymbol{v})$ denote the angle between $\boldsymbol{u}_1$ and $\boldsymbol{v} \in \mathcal{V}$. Then

$$\begin{aligned}\boldsymbol{x}_1 &= a_{11}\boldsymbol{s}_1 + a_{12}\boldsymbol{s}_2 \\ &= a_{11}\|\boldsymbol{s}_1\|\boldsymbol{u}_1 + a_{12}\|\boldsymbol{s}_2\|[\cos(\theta(\boldsymbol{s}_2))\boldsymbol{u}_1 + \sin(\theta(\boldsymbol{s}_2))\boldsymbol{u}_2] \\ &= [a_{11}\|\boldsymbol{s}_1\| + a_{12}\|\boldsymbol{s}_2\|\cos(\theta(\boldsymbol{s}_2))]\boldsymbol{u}_1 + [a_{12}\|\boldsymbol{s}_2\|\sin(\theta(\boldsymbol{s}_2))]\boldsymbol{u}_2\end{aligned}$$

which implies that

$$0 \leq \tan\theta(\boldsymbol{x}_1) = \frac{\sin(\theta(\boldsymbol{s}_2))}{\frac{a_{11}\|\boldsymbol{s}_1\|}{a_{12}\|\boldsymbol{s}_2\|} + \cos(\theta(\boldsymbol{s}_2))} \leq \tan\theta(\boldsymbol{s}_2),$$

i.e., $0 \leq \theta(\boldsymbol{x}_1) \leq \theta(\boldsymbol{s}_2)$. Similarly, one can prove $0 \leq \theta(\boldsymbol{x}_2) \leq \theta(\boldsymbol{s}_2)$. Therefore, $|\theta(\boldsymbol{x}_2) - \theta(\boldsymbol{x}_1)| \leq \theta(\boldsymbol{s}_2)$ and $\rho(\boldsymbol{x}_1, \boldsymbol{x}_2) = \cos(\theta(\boldsymbol{x}_2) - \theta(\boldsymbol{x}_1)) \geq \rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \cos(\theta(\boldsymbol{s}_2))$.

4.2 Proof of (10)

Consider a 2-dimensional plane of (ω_1, ω_2). The constraint of $\omega_1 + \omega_2 = 1$ includes the following two cases:

Case 1: $w_{11} + w_{12} = 1$ for $(\omega_1 = w_{11}, \omega_2 = w_{12})$.
Case 2: $w_{21} + w_{22} = 1$ for $(\omega_1 = w_{21}, \omega_2 = w_{22})$.

As shown in Fig. 7, all the points on the line segment $\overline{AB}$ satisfy $\omega_1 + \omega_2 = 1$ and $\omega_1 x_1[n] + \omega_2 x_2[n] \geq 0$, where the coordinates of the points A and B are given by

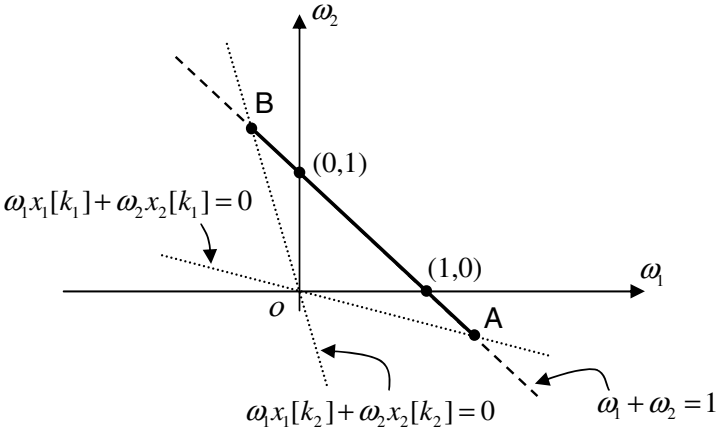

Fig. 7. Feasible region of ω_1 and ω_2 (same for both Cases 1 and 2) satisfying $\omega_1+\omega_2 = 1$ and $\omega_1 x_1[n] + \omega_2 x_2[n] \geq 0$ for all n

$$\boldsymbol{\omega}_A = \left(\frac{-x_2[k_1]}{x_1[k_1] - x_2[k_1]}, \frac{x_1[k_1]}{x_1[k_1] - x_2[k_1]} \right),$$
$$\boldsymbol{\omega}_B = \left(\frac{-x_2[k_2]}{x_1[k_2] - x_2[k_2]}, \frac{x_1[k_2]}{x_1[k_2] - x_2[k_2]} \right).$$

Note that $\omega_1 x_1[n] + \omega_2 x_2[n] = y_1[n]$ for Case 1 and $\omega_1 x_1[n] + \omega_2 x_2[n] = y_2[n]$ for Case 2.

Consider the vector space $\mathcal{V}$ as presented in the proof of Theorem 1 in Section 4.1. Both of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ must be on the line passing $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ due to the constraints $a_{11} + a_{12} = 1$ and $a_{21} + a_{22} = 1$, respectively. Moreover, both of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ must be on the line passing $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ due to the constraints $w_{11} + w_{12} = 1$ and $w_{21} + w_{22} = 1$, respectively. Decreasing $\rho(\boldsymbol{y}_1, \boldsymbol{y}_2)$ is equivalent to increasing the angle between the vectors $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ as shown in Fig. 8, i.e., both w_{11} and w_{22} must be positive and meanwhile w_{12} and w_{21} must be negative. Minimum $\rho(\boldsymbol{y}_1, \boldsymbol{y}_2)$ corresponds to the values of w_{ij} with maximum $|w_{ij}|$. In other words, the optimum solution for (w_{11}, w_{12}) corresponds to either the point A or the point B in Fig. 7, so does the optimum solution for (w_{21}, w_{22}). Therefore, the optimum demixing matrix is given by $\mathbf{W}^\star = \left[\boldsymbol{\omega}_A^T, \boldsymbol{\omega}_B^T\right]^T$ which can be easily proven to be the one given by (10).

4.3 Proof of Theorem 2

Because of $\phi(\mathbf{s}[l_1]) = 0$, $\phi(\mathbf{s}[l_2]) = \pi/2$, one can easily see

$$\tan\phi(\mathbf{x}[n]) = \frac{x_2[n]}{x_1[n]} = \frac{a_{21}s_1[n] + a_{22}s_2[n]}{a_{11}s_1[n] + a_{12}s_2[n]} = \frac{a_{21} + a_{22}\tan\phi(\mathbf{s}[n])}{a_{11} + a_{12}\tan\phi(\mathbf{s}[n])}. \quad (20)$$

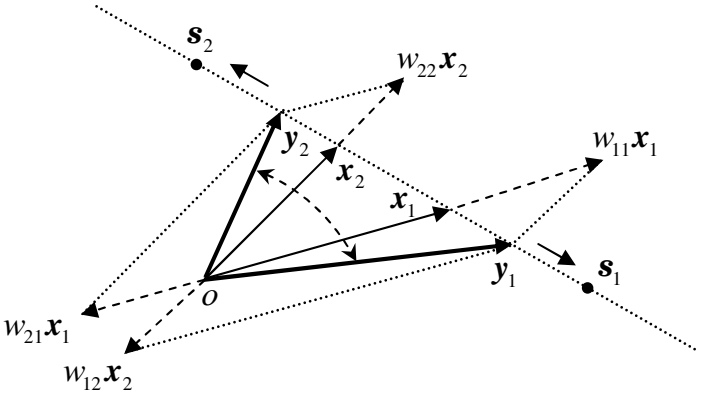

Fig. 8. Vector diagram of source signals ($\boldsymbol{s}_i$), observations ($\boldsymbol{x}_i$) and extracted signals ($\boldsymbol{y}_i$)

By (20), (13) and (14), one can easily obtain

$$\tan\phi(\mathbf{x}[n]) - \tan\phi(\mathbf{x}[l_1]) = \frac{\det(\mathbf{A})\tan\phi(\mathbf{s}[n])}{a_{11} + a_{12}\tan\phi(\mathbf{s}[n])}, \tag{21}$$

$$\tan\phi(\mathbf{x}[n]) - \tan\phi(\mathbf{x}[l_2]) = \frac{-\det(\mathbf{A})/a_{12}}{a_{11} + a_{12}\tan\phi(\mathbf{s}[n])}, \tag{22}$$

where $\det(\mathbf{A}) = a_{11}a_{22} - a_{21}a_{12}$. One can easily infer, from (21) and (22), that $\phi(\mathbf{x}[l_1]) \leq \phi(\mathbf{x}[n]) \leq \phi(\mathbf{x}[l_2])$ if $\det(\mathbf{A}) \geq 0$, and $\phi(\mathbf{x}[l_2]) \leq \phi(\mathbf{x}[n]) \leq \phi(\mathbf{x}[l_1])$ if $\det(\mathbf{A}) \leq 0$, implying $\phi(\mathbf{x}[k_2]) \leq \phi(\mathbf{x}[n]) \leq \phi(\mathbf{x}[k_1]), \forall n$. Thus, we have completed the proof.