ICA WITH MULTIPLE QUADRATIC CONSTRAINTS

Xuejun Liao and Lawrence Carin

Department of Electrical and Computer Engineering
Duke University
Durham, NC 27708-0291, USA

ABSTRACT

The independent component analysis (ICA) with a single quadratic constraint on each source signal or column of the mixing matrix is extended to the case of multiple quadratic constraints. The criterion of Joint Approximate Diagonalization of Eigen-matrices (JADE) is used to measure the statistical independence. A new algorithm is derived to maximize the JADE criterion subject to the multiple quadratic constraints, using the augmented Lagrangian method. The extension offers the freedom to design various combinations of quadratic constraints. Examples include simultaneously constraining a source signal and the corresponding column of the mixing matrix, and two-sided constraints on the source signals or columns of the mixing matrix. Example results are provided to demonstrate the effectiveness of the algorithm.

1. INTRODUCTION

Independent Component Analysis (ICA) [1-2] is an important technique to solve the problem of blind source separation when the sources can be assumed to be mutually independent. ICA usually works in a totally blind way, employing no additional information about the sources or the mixing process other than the independence assumption. However, there exist situations where such additional information is available or circumstances under which it is desirable to perform ICA in a region of interest, instead of in the complete signal space. For example, ICA can be used to learn the functional genomic units (FGU) from DNA microarray signals [7-8]. This is learning from data. On the other hand, knowledge about FGUs can also be learned from books or by consulting with human experts. This is learning from experts. Can we learn FGUs simultaneously from data and experts? Motivated by this and other examples, we introduced in [3] the idea of incorporating quadratic constraints into ICA and developed algorithms to implement ICA subject to a single quadratic constraint on each source signal or column of the mixing matrix. The choice of quadratic constraints is due to their ability to code the (normalized) correlation between a source signal (or the corresponding column of the mixing matrix) and a constraining vector, which offers the versatility of representing the expert knowledge about FGUs or the signal subspace that is of interest.

Though the single constraint offers the device of employing additional information in ICA, it has limitations in the freedom of constructing the constraints. In this paper we extend the work in [3] to the case where multiple quadratic constraints are used in place of the single constraint. The extension offers the freedom to design various combinations of quadratic constraints. Examples include simultaneously constraining a source signal and the corresponding column of the mixing matrix, two-sided constraints on the source signals or columns of the mixing matrix, and combinations of both.

The criterion of Joint Approximate Diagonalization of Eigen-matrices (JADE) is used to measure the statistical independence. A new algorithm is derived to maximize the JADE criterion subject to the multiple quadratic constraints, employing the augmented Lagrangian method. Example results are provided to demonstrate the effectiveness of the algorithm.

2. PROBLEM FORMULATION

The ICA with multiple quadratic constraints is formulated as

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{k=1}^{N} s_k(t)\mathbf{a}_k \tag{1-A}$$

$$\text{subject to} \quad \mathbf{a}_k^H \mathbf{F}_{k,m} \mathbf{a}_k \le 0,\ k = 1,2,\cdots,N,\ m = 1,2,\cdots,M_1 \tag{1-B}$$

$$\mathbf{s}_k \mathbf{G}_{k,m} \mathbf{s}_k^H \le 0,\ k = 1,2,\cdots,N,\ m = 1,2,\cdots,M_2 \tag{1-C}$$

where in (1-A), $\mathbf{y}(t) = [y_1(t)\ y_2(t)\ \cdots y_{N_0}(t)]^{\mathrm{T}}$ are N_0 dimensional zero-mean and stationary mixed-signals, $\mathbf{s}(t) = [s_1(t)\ s_2(t)\ \cdots s_N(t)]^{\mathrm{T}}$ are N dimensional statistically independent source signals with at most one of its components Gaussian, $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_N]$, with $\mathbf{a}_k = [a_{1k}\ a_{2k}\ \cdots\ a_{N_0 k}]^{\mathrm{T}}$, is the mixing matrix with full column rank, and t is an index for time or other relevant variables. In (1-B,C), $\mathbf{s}_k = [s_k(1)\ s_k(2)\ \cdots\ s_k(T)]$ with T the number of samples in t, $\mathbf{F}_{k,m}$ with $k = 1,2,\cdots,N$ and $m = 1,2,\cdots,M_1$, and $\mathbf{G}_{k,m}$ with $k = 1,2,\cdots,N$ and $m = 1,2,\cdots,M_2$, are assumed, respectively, to be $N_0 \times N_0$ and $T \times T$ Hermitian constraining matrices, and the superscript H denotes conjugate transpose.

To solve (1), we start with deriving its whitened form. Let the whitened form of (1-A) be [2]

$$\mathbf{z}(t) = \mathbf{U}\mathbf{s}(t) \tag{2}$$

where

$$\mathbf{z}(\mathrm{t}) = \mathbf{M}\mathbf{y}(\mathrm{t}) \tag{3}$$

are the whitened mixed signals, $\mathbf{M}$ is a $N \times N_0$ whitening matrix of full row rank, and

$$\mathbf{U} = \mathbf{M}\mathbf{A} \tag{4}$$

is the unitary factor of **A**. **M** can be estimated from **y**(t) via eigenvalue decomposition [2] or singular value decomposition [1]. Let $\mathbf{U}=[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_N]$, and it follows from (4)

$$\mathbf{a}_k = \mathbf{M}^{\#}\mathbf{u}_k,\quad k=1,2,\cdots,N \tag{5}$$

where the superscript # denotes pseudo inverse. Similarly it follows from (2)

$$\mathbf{s}_k = \mathbf{u}_k^H \mathbf{Z} \tag{6}$$

where $\mathbf{s}_k$ is as defined in (1-C) and **Z**=[**z**(1) **z**(2) … **z**(*T*)]. Substituting (5) into (1-B), and (6) into (1-C), we obtain for $k=1,2,\cdots,N$,

$$\mathbf{u}_k^H \mathbf{B}_{k,m}\mathbf{u}_k \le 0,\quad m=1,2,\cdots,M \tag{7}$$

where $M = M_1 + M_2$, and

$$\mathbf{B}_{k,m} = \begin{cases} (\mathbf{M}^{\#})^H \mathbf{F}_{k,m}\mathbf{M}^{\#}, & \text{when } m=1,2,\cdots,M_1 \\ \mathbf{Z}\mathbf{G}_{k,m-M_1}\mathbf{Z}^H, & \text{when } m=M_1+1, M_1+2,\cdots,M \end{cases} \tag{8}$$

(2) subject to (7) gives the whitened form of (1).

The solution of (2) subject to (7) is based on constrained maximization of the JADE criterion, which is defined as [2]

$$JADE(\mathbf{U}) = \sum_{k=1}^{N}\sum_{r=1}^{N^2} |\mathbf{u}_k^H \mathbf{Q}_z(\mathbf{D}_r)\mathbf{u}_k|^2 \tag{10}$$

where $\mathbf{D}_r$, *r*=1, 2, …, N^2, constitute a set of orthonormal bases for the space of *N*×*N* matrices, $\mathbf{Q}_z(\mathbf{D}_r)$ is the cumulant matrix defined element-wise as

$$[\mathbf{Q}_z(\mathbf{D}_r)]_{ij} = \sum_{p,q=1}^{N} cum[z_i(t), z_j^*(t), z_q(t), z_p^*(t)]\, d_{pq}^{(r)} \tag{11}$$

$cum(\cdot)$ denotes the cumulant, * denotes complex conjugate, and $d_{pq}^{(r)}$ is the (p,q)-th element of $\mathbf{D}_r$.

Now, with the JADE criterion, solving (2) subject to (7) becomes a problem of maximizing *JADE*(**U**), or equivalently, minimizing −*JADE*(**U**), subject to (7), the solution of which is developed in section 3.

3. ALGORITHM

In the subsequent sections, we abbreviate $\mathbf{Q}_z(\mathbf{D}_r)$ in (10) as $\mathbf{Q}_r$ for notational simplicity and assume that all eigen vectors discussed have unit L_2 norm. Our algorithm solves for each column of **U** separately. Each time we find a column, say $\mathbf{u}_k$, that minimizes $-\sum_r^{N^2} |\mathbf{u}_k^H \mathbf{Q}_r \mathbf{u}_k|^2$ under the constraint that $\mathbf{u}_k$ has unit norm and is orthogonal to all columns found previously. Specifically, we have a constrained optimization problem

$$\mathbf{u}_k = \arg\min(-\sum_r^{N^2} |\mathbf{u}_k^H \mathbf{Q}_r \mathbf{u}_k|^2) \tag{12-A}$$

$$\text{s.t.}\quad \mathbf{u}_k^H \mathbf{u}_k = 1 \tag{12-B}$$

$$\mathbf{u}_k^H \mathbf{u}_l = 0,\quad l=1,2,\cdots,k-1 \tag{12-C}$$

$$\mathbf{u}_k^H \mathbf{B}_{k,m}\mathbf{u}_k \le 0,\quad m=1,2,\cdots,M \tag{12-D}$$

We use the orthogonal-projection-based idea [4] to eliminate the constraints in (12-C). Denote the *N*-by-*N* identity matrix as $\mathbf{I}_N$. Let $\mathbf{U}_{k-1}=[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{k-1}]$. Then $\mathbf{U}_{k-1}^H \mathbf{U}_{k-1} = \mathbf{I}_{k-1}$, which is a result of $\mathbf{u}_1$, $\mathbf{u}_2$, …, $\mathbf{u}_{k-1}$ satisfying (12-B,C). Denote $\Omega_{k-1} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{k-1}\}$, and $\Omega_{k-1}^{\perp}$ the orthogonal complement of Ω_{k-1}. We construct

$$\tilde{\mathbf{P}}_k = \mathbf{I}_N - \mathbf{U}_{k-1}(\mathbf{U}_{k-1}^H\mathbf{U}_{k-1})^{-1}\mathbf{U}_{k-1}^H = \mathbf{I}_N - \mathbf{U}_{k-1}\mathbf{U}_{k-1}^H \tag{13}$$

which is the matrix of the orthogonal projection onto $\Omega_{k-1}^{\perp}$. The orthogonalization of columns of $\tilde{\mathbf{P}}_k$ gives $\mathbf{P}_k$, which satisfies $\mathbf{P}_k^H\mathbf{P}_k = \mathbf{I}_{N-k+1}$. Let $\mathbf{u}_k = \mathbf{P}_k\mathbf{w}_k$. Clearly $\mathbf{u}_k \in \Omega_{k-1}^{\perp}$ and therefore satisfies the constraints in (12-C). Moreover, the constraint (12-B) implies that $(\mathbf{P}_k\mathbf{w}_k)^H\mathbf{P}_k\mathbf{w}_k = 1$, which, using $\mathbf{P}_k^H\mathbf{P}_k = \mathbf{I}_{N-k+1}$, is reduced to $\mathbf{w}_k^H\mathbf{w}_k = 1$. Thus, we have a dual problem to (12) in $\Omega_{k-1}^{\perp}$ as

$$\mathbf{u}_k = \mathbf{P}_k\mathbf{w}_k \tag{14-A}$$

$$\mathbf{w}_k = \arg\min(-\sum_r^{N^2} |\mathbf{w}_k^H\mathbf{P}_k^H\mathbf{Q}_r\mathbf{P}_k\mathbf{w}_k|^2) \tag{14-B}$$

$$\text{s.t.}\quad \mathbf{w}_k^H\mathbf{w}_k = 1 \tag{14-C}$$

$$\mathbf{w}_k^H\mathbf{C}_{k,m}\mathbf{w}_k \le 0,\quad m=1,2,\cdots,M \tag{14-D}$$

where

$$\mathbf{C}_{k,m} = \mathbf{P}_k^H\mathbf{B}_{k,m}\mathbf{P}_k \tag{15}$$

We now solve (14-B,C,D) using the Lagrangian multiplier method. We do a partial elimination of the inequality constraints (14-D) via means of penalty, in accordance to which, the augmented Lagrangian function for (14-B,D) is [6]

$$\begin{aligned} L_c(\mathbf{w}_k,\{\mu_m\}) = &-\sum_{r=1}^{N^2}[|\mathbf{w}_k^H\mathbf{P}_k^H\mathbf{Q}_r\mathbf{P}_k\mathbf{w}_k|^2 \\ &+\frac{1}{2c}\sum_{m=1}^{M}[(\max\{0,\mu_m + c\mathbf{w}_k^H\mathbf{C}_{k,m}\mathbf{w}_k\})^2 - \mu_m^2] \end{aligned} \tag{16}$$

where *c* is a positive penalty parameter. With the inequality constraints (14-D) eliminated, the problem of (14-B,C,D) can be approximated as

$$\mathbf{w}_k = \min L_c(\mathbf{w}_k,\{\mu_m\}) \tag{17-A}$$

$$\text{s.t.}\quad \mathbf{w}_k^H\mathbf{w}_k = 1 \tag{17-B}$$

the Karush-Kuhn-Tucker (KKT) optimality conditions for which are (17-B) together with

$$\frac{\partial L_c(\mathbf{w}_k,\{\mu_m\})}{\partial \mathbf{w}_k^*} = \lambda\frac{\partial}{\partial \mathbf{w}_k^*}(\mathbf{w}_k^H\mathbf{w}_k - 1) \tag{18}$$

where we have used the definition in [5] for differentiation with respect to complex vectors. (18) is evaluated to give

$$\begin{aligned} [-\boldsymbol{\Gamma}_k(\mathbf{w}_k) + &\frac{1}{2c}\sum_{m=1}^{M}(\mathbf{w}_k^H\mathbf{R}_{k,m}(c,\mu_m,\mathbf{w}_k)\mathbf{w}_k) \\ &\times \mathbf{R}_{k,m}(c,\mu_m,\mathbf{w}_k)]\mathbf{w}_k = \lambda\mathbf{w}_k \end{aligned} \tag{19}$$

with the matrix function $\boldsymbol{\Gamma}_k(.)$ defined as

$$\begin{aligned} \boldsymbol{\Gamma}_k(\boldsymbol{\beta}) = \frac{1}{2}\sum_{r=1}^{N^2}[&(\boldsymbol{\beta}^H\mathbf{P}_k^H\mathbf{Q}_r\mathbf{P}_k\boldsymbol{\beta})\mathbf{P}_k^H\mathbf{Q}_r^H\mathbf{P}_k \\ &+(\boldsymbol{\beta}^H\mathbf{P}_k^H\mathbf{Q}_r^H\mathbf{P}_k\boldsymbol{\beta})\mathbf{P}_k^H\mathbf{Q}_r\mathbf{P}_k] \end{aligned} \tag{20}$$

and $\mathbf{R}_{k,m}(.,.,.)$ defined as

$$\mathbf{R}_{k,m}(c,\mu,\boldsymbol{\beta}) = \begin{cases} \mu\mathbf{I} + c\mathbf{C}_{k,m}, & \text{if } \mu + c\boldsymbol{\beta}^H\mathbf{C}_{k,m}\boldsymbol{\beta} > 0 \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{21}$$

Equations (17-B) and (19) imply that at the stationary point $\mathbf{w}_k$ is an eigenvector of

$$\tilde{\mathbf{\Gamma}}_k(c,\{\mu_m\},\mathbf{w}_k) \overset{Def}{=} -\mathbf{\Gamma}_k(\mathbf{w}_k) + \frac{1}{2c}\sum_{m=1}^{M}(\mathbf{w}_k^H\mathbf{R}_{k,m}(c,\mu_m,\mathbf{w}_k)\mathbf{w}_k)\mathbf{R}_{k,m}(c,\mu_m,\mathbf{w}_k) \quad (22)$$

with the associated eigenvalue

$$\lambda = L_c(\mathbf{w}_k,\{\mu_m\}) \quad (23)$$

which is equal to the objective function in (17-A). Therefore the optimal eigenvector $\mathbf{w}_k$ for (17) should be associated with the smallest eigenvalue of $\tilde{\mathbf{\Gamma}}_k(c,\{\mu_m\},\mathbf{w}_k)$, to minimize the objective function. We utilize the alternating eigen-search algorithm [3] to solve the eigenvalue problem of (19) subject to (17-B).

It is known from the Lagrangian multiplier theory [6] that when $\{\mu_m\}$ are close to the true Lagrangian multipliers of (14-B,D), or the penalty parameter c approaches infinity, then unconstrained minimization of (16) well approximates (14-B,D). However, a penalty parameter approaching infinity will lead to severe ill-conditioning for most minimization algorithms. Fortunately, a whole theory called the augmented Lagrangian method [6] exists for iteratively adjusting $\{\mu_m\}$ such that $\{\mu_m\}$ will converge to the true Lagrangian multipliers of (14-B,D), with a sufficiently large but not infinite penalty parameter c. Thus, upon convergence of $\{\mu_m\}$, the eigenvector solution of (17) well approximate the solution of (14-B,C,D).

Therefore, we have the following overall method to solve (14-B,C,D). Use the augmented Lagrangian method [6] to produce a sequence $\{\mu_m^{(i)}\}$. For any fixed i, use the alternating eigen-search algorithm [3] to solve (17). Upon convergence of $\{\mu_m^{(i)}\}$, the solution of (17) approximates the solution of (14-B,C,D). The overall algorithm is given as:

Algorithm 1: Augmented Lagrangian with Alternating Eigen-search (ALAE)

Step 1: Define the convergence parameters $\varepsilon_\lambda \geq 0$ and $\varepsilon_\mu \geq 0$, and a constant $\alpha>1$. Choose a moderate initial penalty parameter $c^{(1)}>0$. Initialize the multipliers $\mu_m^{(1)} > 0$, $m=1,2,\ldots,M$. Initialize $\mathbf{w}_k^{(0)} \in C^{N-k+1}$ such that $\mathbf{w}_k^{(0)} \neq \mathbf{0}$, and let $\mathbf{w}_k^{(0)} = \mathbf{w}_k^{(0)} \big/ \sqrt{(\mathbf{w}_k^{(0)})^H\mathbf{w}_k^{(0)}}$. Let i=1.

Step 2: Let $\mathbf{w}_k^{(i)(0)} = \mathbf{w}_k^{(i-1)}$ and let $j=1$.

Step 3: Find the minimum eigenvalue, denoted $\lambda^{(j)}$, of $\tilde{\mathbf{\Gamma}}_k(c^{(i)},\{\mu_m^{(i)}\},\mathbf{w}_k^{(i)(j-1)})$, and denote the associated eigenvector with unit L_2 norm as $\mathbf{w}_k^{(i)(j)}$.

Step 4: Check convergence of $\lambda^{(j)}$. If $|\lambda^{(j)} - \lambda^{(j-1)}| \leq \varepsilon_\lambda$, let $\mathbf{w}_k^{(i)} = \mathbf{w}_k^{(i)(j)}$ and go to Step 5; otherwise, let $j=j+1$ and go back to Step 3.

Step 5: Update the multipliers. Let $\mu_m^{(i+1)} = \max\{0, \mu_m^{(i)} + c^{(i)}(\mathbf{w}_k^{(i)})^H\mathbf{C}_{k,m}\mathbf{w}_k^{(i)}\}$, $m=1,2,\ldots,M$.

Step 6: Update the penalty parameter. Let $c^{(i+1)} = \alpha c^{(i)}$.

Step 7: Check convergence of the multipliers. If $|\mu_m^{(i+1)} - \mu_m^{(i)}| / |\mu_m^{(i)} - \mu_m^{(i-1)}| > 1$ for any $1 \leq m \leq M$, exit the algorithm and return "no feasible solution"; otherwise, if $|\mu_m^{(i)} - \mu_m^{(i-1)}| \leq \varepsilon_\mu$ for $m=1,2,\ldots,M$, exit the algorithm and return $\mathbf{w}_k^{(i)}$ as the solution; otherwise, let $i=i+1$ and go back to Step 2.

4. APPLICATION AND EXAMPLE RESULTS

ICA has been used to model narrow-band antenna array signals [2] and DNA microarray signals [7-8]. The quadratic constraints (1-B,C), when applied to these two types of signals, have interesting interpretations from their respective fields. For the antenna array signals, $\mathbf{F}_{k,m}$'s specify the desired range of direction-of-arrivals (DOA) of the sources and $\mathbf{G}_{k,m}$'s specify the approximate envelopes of the source signals. For DNA microarray signals, $\mathbf{F}_{k,m}$'s and $\mathbf{G}_{k,m}$'s are used to encode the expert knowledge on the constituent genes in each functional genomic unit (FGU) and the responses of each FGU to the experimental conditions, respectively. Here a FGU is defined as a set of genes that operate collectively to effect a biological function [8]. Because of the space limit, we only present here the example results on DNA microarray signals.

We now briefly explain how we construct $\mathbf{F}_{k,m}$'s and $\mathbf{G}_{k,m}$'s. Assume $\mathbf{y}(t)$ in (1) are the DNA microarray signals, then $|a_{i,k}|$ defines the membership of the i-th gene belonging to the k-th FGU and $s_k(t)$ is the response (expression levels) of the k-th FGU to the t-th experimental condition [8]. From the expert knowledge (human experts, books, Gene Ontology (GO) [8], etc.), we can build an approximate and incomplete FGU subsuming a subset $\{i_1,i_2,...,i_L\}$ of its constituent genes. Let the approximate memberships of these partial genes be $|\tilde{a}_{ik}|$, $i=i_1,i_2,...,i_L$. Construct a constraining vector $\mathbf{w}_k$ to be of the same dimension as $\mathbf{a}_k$ and with its i-th element equal to $|\tilde{a}_{ik}|^2$ for $i=i_1,i_2,...,i_L$ and zero for others. Let $\mathbf{w}_k$ to be normalized to unit L_2 norm. The constraint on $\mathbf{a}_k$ is constructed as

$$corr_1 = \mathbf{a}_k^H diag(\mathbf{w}_k)\mathbf{a}_k \Big/ \mathbf{a}_k^H\mathbf{a}_k = \sum_{i=1}^{N_0} w_{ik}|a_{ik}|^2 \Big/ \sum_{i=1}^{N_0}|a_{ik}|^2 \geq \rho \quad (24)$$

with $corr_1$ the non- normalized correlation between $\mathbf{w}_k$ and $|\mathbf{a}_k|^2 \overset{Def.}{=} [|a_{1k}|^2, |a_{2k}|^2, \cdots, |a_{N_0k}|^2]^{\mathrm{T}}$, and ρ the threshold. Comparing (1-B) and (24), we have

$$\mathbf{F}_k = \rho\mathbf{I}_{N_0} - diag(\mathbf{w}_k) \quad (25)$$

The construction of $\mathbf{G}_{k,m}$'s can be implemented in a similar way. From expert knowledge, we build the approximate responses $\tilde{s}_k(t)$, $t=t_1,t_2,...,t_{L'}$, of a GFU to a subset of experimental conditions $\{t_1,t_2,...,t_{L'}\}$. Construct $\mathbf{w}_k$ to be a vector of the same dimension as $\mathbf{s}_k$ and whose t-th element equal to $\tilde{s}_k(t)$ for $t=t_1,t_2,...,t_{L'}$ and zero for others. Let $\mathbf{w}_k$ be normalized to unit L_2 norm. The constraint on $\mathbf{s}_k$ is constructed as

$$corr_2 = |\mathbf{s}_k\mathbf{w}_k^H|^2 / \|\mathbf{s}_k\|^2 = \mathbf{s}_k\mathbf{w}_k^H\mathbf{w}_k\mathbf{s}_k^H / \mathbf{s}_k\mathbf{s}_k^H \geq \rho \quad (26)$$

with $corr_2$ the square of normalized correlation between $\mathbf{w}_k$ and $\mathbf{s}_k$, and ρ the threshold. Comparing (1-C) and (26), we have

$$\mathbf{G}_k = \rho\mathbf{I}_T - \mathbf{w}_k^H\mathbf{w}_k \quad (27)$$

We simulate the true responses $\mathbf{s}(t)=[s_1(t)\ s_2(t)\ s_3(t)]^{\mathrm{T}}$, $t=1,2,\ldots,300$, of three FGUs to 300 experimental conditions and a 126-by-3 mixing-matrix $\mathbf{A}$. $\mathbf{s}(t)$ are then mixed by $\mathbf{A}$ to yield the expression levels $\mathbf{y}(t)$ of 126 genes. Various levels of noises are added to $\mathbf{y}(t)$. The expression levels are in logarithm, with zero denoting the expression levels under the normal conditions, and positive and negative values denoting increases and decreases, respectively, in expression levels, due to the effects of experiments.

The ICA and constrained ICA are performed on the noisy gene expression levels (microarray signals) **y**(t) to extract the FGUs' experimental responses **s**(t) and the mixing matrix **A**. The threshold ρ in the constrained ICA is adaptively chosen to achieve the best match between $|\mathbf{a}_k|^2$ or $s_k(t)$ and their respective constraining vector $\mathbf{w}_k$, with the match quantitatively measured by $corr_1$ or $corr_2$. The errors of the estimates of **A** and **s**(t) are averaged over 200 Monte Carlo runs to yield the mean squared errors (MSE), which are plotted as a function of signal to noise ratio (SNR) in Fig.1 and Fig.2, respectively.

The constraining vectors $\mathbf{w}_k$'s are designed such that they contain 3 peaks of the true $|\mathbf{a}_k|^2$ or $s_k(t)$, with the magnitudes randomly changed by 20% from the corresponding true ones. There are more than 10 peaks in each simulated $|\mathbf{a}_k|^2$ and $s_k(t)$, on average. Therefore the expert knowledge covers less than 30% of the total information. This amount of expert knowledge, small though it is, contribute significantly to the improvement in the estimates of both **A** and **s**(t), compared with the ICA results (which do not use the expert knowledge), as seen from Fig. 1 and Fig. 2. It is also shown in these figures that the ICA with double constraints simultaneously imposed on $|\mathbf{a}_k|^2$ and $s_k(t)$ achieve the top performance. This is natural because the double constraints provide expert knowledge on both $|\mathbf{a}_k|^2$ and $s_k(t)$. The performances of the ICA with a single constraint on $|\mathbf{a}_k|^2$ or $s_k(t)$ stand in the middle, as the single constraint provides only half the expert knowledge that is provided by the double constraints.

5. CONCLUSIONS

We have extended the singly quadratically constrained ICA to multiply quadratically constrained ICA. Using the JADE criterion as the measure of statistical independence, we have derived a new algorithm to perform the multiply quadratically constrained ICA, based on the augmented Lagrangian method. As an application example, we implemented the scheme of learning functional genomic units simultaneously from data and from experts, via the constrained ICA. Numerical results show that significant improvements in the estimates of both **A** and **s**(t) can be obtained from the use of expert knowledge and that double constraints achieve the top performance as a result of their ability of coding more expert knowledge than the single constraint.

6. REFERENCES

[1] P. Comon, "Independent component analysis, A new concept?", *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming for non-Gaussian Signals," *Proc. Inst. Elect. Eng.* F, vol. 40, pp. 362–370, 1993.

[3] X. Liao and L. Carin, "A New Algorithm for Independent Component Analysis With or Without Constraints', *Proc. of the Second IEEE Sensor Array and Multichannel Signal Processing Workshop*, Rosslyn, VA, August 4-6, 2002

[4] J. Zhang et al., " Design of Unitary Precoders for ISI Channels", *Proc. of ICASSP 2002*, Vol. 3, p.p. 2265 -2268

[5] Haykin, S., *Adaptive Filter Theory*, Prentice Hall, 3rd edition, 1996

[6] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2nd Edition, 1999

[7] X. Liao et al., "ICA and PLS Modeling for Functional Analysis and Drug Sensitivity for DNA Microarray Signals', *Proc. of ICASSP 2002*, Vol. 4, p.p. 3880 –3883,

[8] Lin, S. M. et al., "Using Functional Genomic Units to Corroborate User Experiments with the Rosetta Compendium", in Johnson K. F., Lin, S. M., (eds), "Methods of Microarray Data Analysis (volume II)," Kluwer Academic Publishing, Norwell, MA, 2002.

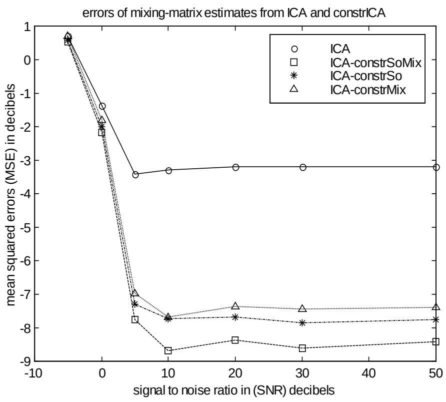

Fig. 1 Mean squared errors (MSE) of the estimates of **A** as a function of signal to noise ratio (SNR). All values are shown in decibels. Circle: ICA; Square: ICA with simultaneous constraints on $|\mathbf{a}_k|^2$ and $s_k(t)$; Star: ICA with single constraint on $s_k(t)$; Triangle: ICA with single constraint on $|\mathbf{a}_k|^2$.

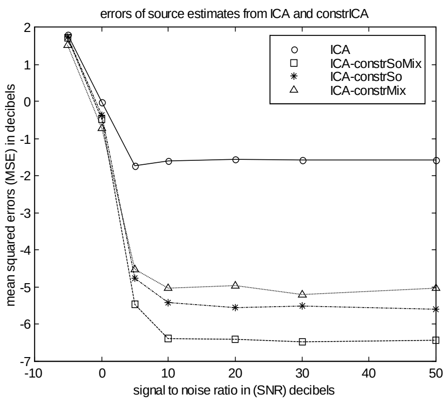

Fig. 2 Mean squared errors (MSE) of the estimates of **s**(t) as a function of signal to noise ratio (SNR). All values are shown in decibels. Circle: ICA; Square: ICA with simultaneous constraints on $|\mathbf{a}_k|^2$ and $s_k(t)$; Star: ICA with single constraint on $s_k(t)$; Triangle: ICA with single constraint on $|\mathbf{a}_k|^2$.