[illegible]

Jinwen Ma[1,3], Zhe Chen[2], and Shun-ichi Amari[1]

[1] [illegible]

[2] [illegible] 351-0198, [illegible]

[3] [illegible] 100871, [illegible]

{jwma, zhechen, amari}@brain.riken.jp

Abstract. [illegible]

[illegible]

Independent component analysis (ICA) is a powerful tool for blind signal processing and has remained as an intense research subject in the literature. One of important application of ICA is used for blind source separation where the source signals are assumed to be independent and non-Gaussian. In particular, consider a conventional ICA problem which assumes an instantaneous linear mixing model: $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ denotes the mixing matrix; $\mathbf{s} \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^m$ correspond to the n-dimensional source vector and m-dimensional mixture vector, respectively. The goal of ICA is to seek a demixing matrix, $\mathbf{W} \in \mathbb{R}^{n \times m}$, applied to the mixture vector $\mathbf{x}$:

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}(\mathbf{A}\mathbf{s}) = (\mathbf{W}\mathbf{A})\mathbf{s} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ corresponds to the unmixed signal vector. When the sources in $\mathbf{s}$ are statistically independent, it is hoped that the recovered $\mathbf{y}$ is also componentwise independent, that is,

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 838–845, 2006.

$$q(\mathbf{y}) = \prod_{i=1}^{n} q(y_i), \tag{2}$$

where $q(\cdot)$ denotes the probability density function. Generally and unless stated otherwise, it is assumed in the paper that $m = n$ and the square mixing matrix $\mathbf{A}$ is invertible.

The study on the ICA problem can be traced back to Tong, Inouye & Liu [1] who showed that $\mathbf{y}$ recovers the sources $\mathbf{s}$ up to scaling and permutation ambiguity when y_i $(i = 1, \ldots, n)$ become componentwise independent and at most one of them is Gaussian. Later on, Comon [2] further formalized the problem under the name ICA. Since then, the ICA problem has been widely studied from di erent perspectives by many researchers (e.g., [3]-[7]). In particular, one of essential goal to exploit the independence in parallel is to minimize the following objective function, or the so-called "minimum mutual information (MMI)":

$$D(\mathbf{W}) = -H(\mathbf{y}) - \sum_{i=1}^{n} \int p \quad (y_i; \mathbf{W}) \log p_i(y_i) dy_i, \tag{3}$$

where $H(\mathbf{y}) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$ represents the entropy of $\mathbf{y}$, $p_i(y_i)$ denotes the predetermined model probability density function (pdf) that is implemented to approximate the marginal pdf of $\mathbf{y}$, and $p \quad (y_i; \mathbf{W})$ denotes the joint probability distribution on $\mathbf{y} = \mathbf{W}\mathbf{x}$. In the literature, how to choose the model pdf's is an important issue for the ICA problem. It is known that, with each model pdf $p_i(y_i)$ predefined, this MMI method works only in the cases where the components of $\mathbf{y}$ are either all super-Gaussians [4] or all sub-Gaussians [5].

For the cases where sources contain both super-Gaussian and sub-Gaussian signals in an unknown manner, it was suggested that each model pdf $p_i(y_i)$ should be flexibly adjustable and be learned together with demixing matrix $\mathbf{W}$. In fact, the learning of $p_i(y_i)$ can be done by adapting the parameters in a finiteb.20pro,((

one-bit-matching condition guarantees a feasible solution of the ICA problem by *globally* maximizing the simplified objective function (to be defined later in Section 2) derived from Eq.(3). However, this result is rather restrictive in that it is generally difficult to obtain a feasible solution of the ICA problem by searching the global maximum of the objective function. As a matter of fact, it is more significant to study the local separation property of the ICA problem under the one-bit-matching condition that the sources can be separated by *locally* maximizing that objective function in the same setting. Along this direction, Ma, Liu & Xu [17] already proved that all the local maxima of the formulated objective function correspond to the feasible solutions of the ICA problem in the two-source mixing setting.

In this paper, we further investigate the formulated objective function in the general case. Specifically, we prove that there always exist many local maxima of the objective function that correspond to the stable feasible solutions of the ICA problem (i.e., the stable solutions of a local searching algorithm on the objective function) in the general case under the one-bit-matching condition. Moreover, in ceratin situation under the one-bit-matching condition, there also exist some local minima of the objective function that correspond to the stable feasible solutions of the ICA problem with mixed super- and sub-Gaussian sources. That is, the successful separation can be obtained via locally minimizing the objective function under the one-bit-matching condition in such a case with mixed super- and sub-Gaussian sources.

The rest of the paper is structured as follows. We first formulate the objective function and introduce a leema in section 2. Section 3 presents the main results of two theorems. We conclude briefly in section 4.

For discussion simplicity, we assume that the source, mixed, and recovered signals are all whitened and thus $\mathbf{W}$ and $\mathbf{A}$ are both orthonormal. When the skewness and kurtosis statistics are considered and when the non-Gaussian sources have nonzero kurtosis statistics, under the zero skewness assumption for all the model pdf's, the objective function derived from Eq.(3) can be simplified as follows [16]:

$$J(\mathbf{R}) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^4 \nu_j^s k_i^m, \tag{4}$$

where $\mathbf{R} = (r_{ij})_{n\times n} = \mathbf{WA}$ is an orthonormal matrix to be estimated (the reason that we optimize $\mathbf{R}$ instead of $\mathbf{W}$ is for convenience of analysis); ν_j^s denotes the kurtosis of the source s_j, and k_i^m is a constant with the same algebraic sign as the kurtosis ν_i^m of the model pdf.

For the purpose of clarity, we define a matrix $\mathbf{K}$ by

$$\mathbf{K} = (k_{ij})_{n\times n}, \quad k_{ij} = \nu_j^s k_i^m. \tag{5}$$

By that we may rewrite (4) as

$$J(\mathbf{R}) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^4 \nu_j^s k_i^m = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^4 k_{ij}. \tag{6}$$

Under the one-bit-matching condition, with the help of certain permutation we can always obtain $k_1^m \geq \cdots \geq k_p^m > 0 > k_{p+1}^m \geq \cdots \geq k_n^m$ and $\nu_1^s \geq \cdots \geq \nu_p^s > 0 > \nu_{p+1}^s \geq \cdots \geq \nu_n^s$, which will be considered as the one-bit-matching condition in this paper.

It has been proved in [16] that the global maximization of Eq.(6) under the one-bit-matching condition can only be approachable by setting $\mathbf{R}$ as an identity matrix up to certain permutation and sign indeterminacy. That is, the global maximization of Eq.(6) will recover the original sources up to sign and permutation indeterminacies if the one-bit-matching condition is satisfied. In the two-source mixing case, i.e., $n = 2$, it has been further proved in [17] that the local maxima of $J(\mathbf{R})$ are also only reachable by the permutation matrices up to sign indeterminacy under the one-bit-matching condition. In the following, we will prove that there exist many local maxima of $J(\mathbf{R})$ that correspond to the stable feasible solutions of the ICA problem. Moreover, in certain cases where both super- and sub-Gaussian sources coexist, some minima of $J(\mathbf{R})$ also correspond to the stable feasible solutions of ICA problem. Before doing so, we introduce one lemma as follows.

Lemma 1. *Suppose that $F(\mathbf{x})$ ($\mathbf{x} \in \mathbb{R}^m$) is a twice differentiable scalar function under the following constraints:*

$$C_i(\mathbf{x}) = 0, \qquad i = 1, 2, \cdots, k. \tag{7}$$

Construct a Lagrange function with a Lagrange multiplier set $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$, i.e., $L(\mathbf{x}, \boldsymbol{\lambda}) = F(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i C_i(\mathbf{x})$, and assume that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a solution of the system of the equalities that all the derivatives of $L(\mathbf{x}, \boldsymbol{\lambda})$ with respect to the variables of $\mathbf{x}$ and the Lagrange multipliers λ_i are equal to zeros. It is also assumed that these $\nabla C_i(\mathbf{x}^*)$ are linearly independent. If for any nonzero vector $\boldsymbol{q} \neq 0$ under the constraints $\mathbf{q}^T \nabla\ C_i(\mathbf{x}^*) = 0$ for $i = 1, 2, \cdots, k$, we have

$$\mathbf{q}^T \nabla^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*)\mathbf{q} < 0 \ \ (or > 0), \tag{8}$$

then $\mathbf{x}^*$ is a local maximum (or local minimum) of $F(\mathbf{x})$ under the constraints.

Lemma 1 is a well-known mathematical result in optimization theory; its proof can be found in [18].

With the above background, we are ready to investigate the local maximization of objective function $J(\mathbf{R})$ defined in (6), where $\mathbf{R}$ is a permutation matrix up to sign indeterminacy (namely, as a special orthonormal matrix). We consider the general optimization problem of maximizing $J(\mathbf{R})$ with a fixed matrix $\mathbf{K}$ and $\mathbf{R}\mathbf{R}^T = \mathbf{I}$.

In order to solve this constrained optimization problem, we introduce a set of Lagrange multipliers $\boldsymbol{\lambda} = \{\lambda_{ij} : i \leq j\}$ and construct the Lagrange objective function:

$$L(\mathbf{R}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}^4 k_{ij} + \sum_{i=1}^{n} \sum_{j=i}^{n} \lambda_{ij} \Big(\sum_{l=1}^{n} r_{li} r_{lj} - \delta_{ij} \Big), \tag{9}$$

where δ_{ij} denotes the Kronecker function. By derivation, we have

$$\frac{\partial L(\mathbf{R}, \boldsymbol{\lambda})}{\partial r} = 4k \;\; r^3 + \sum_{=1}^{1} r \;\; \lambda \;\; + 2\lambda \;\; r \;\; + \sum_{= \; +1} r \;\; \lambda \;\; ; \tag{10}$$

$$\frac{\partial L(\mathbf{R}, \boldsymbol{\lambda})}{\partial \lambda} = \sum_{=1} r \;\; r \;\; - \delta \;\; . \tag{11}$$

Given $\boldsymbol{\lambda}$, we define a new matrix $\mathbf{U} = (u_{ij})_{n \times n}$ as

$$u \;\; = \begin{cases} \lambda \\ \end{cases}$$

Theorem 1. *If* $\mathbf{R}^*$ *is a permutation matrix up to sign indeterminacy and* $k_{ij} > 0$ *at all the positions where* $|r^*_{ij}| = 1$*, it corresponds to a local maximum of the objective function* $J(\mathbf{R})$*.*

Proof: For convenience, we vectorize the $n \times n$ matrix $\mathbf{R}$ into an $n^2 \times 1$ vector

$$vec[\mathbf{R}] = [r_{11}, r_{21}, \cdots, r_{n1}, r_{12}, r_{22}, \cdots, r_{n2}, \cdots, r_{1n}, r_{2n}, \cdots, r_{nn}]^T \in \mathbb{R}^{n^2}.$$

Correspondingly, we may also construct a nonzero $n^2 \times 1$ vector $\mathbf{q}$

$$\mathbf{q} = [q_{11}, q_{21}, \cdots, q_{n1}, q_{12}, q_{22}, \cdots, q_{n2}, \cdots, q_{1n}, q_{2n}, \cdots, q_{nn}]^T \in \mathbb{R}^{n^2}.$$

Taking the derivative of Eq.(13) yields

$$\frac{\partial^2 L(\mathbf{R}, \boldsymbol{\lambda})}{\partial r_{ij} \partial r_{i'j'}} = \delta_{(i,j),(i',j')}[12 k_{ij} r_{ij}^2 + u_{jj}], \tag{16}$$

where $\delta_{(i,j),(i',j')}$ denotes the Kronecker function such that it equals to 1 if $(i', j') = (i, j)$ (namely, $i = i'$ and $j = j'$) and zero otherwise. It follows from Eq.(14) that

$$\mathbf{U} = -4\mathbf{R}^T\mathbf{B}. \tag{17}$$

When $\mathbf{R} = \mathbf{R}^*$ is a permutation matrix up to sign indeterminacy, $\mathbf{U}^*$ (associated with $\boldsymbol{\lambda}^*$) will be a diagonal matrix. By the condition that $k_{ij} > 0$ at every $|r^*_{ij}| = 1$, it follows that u^*_{jj}

submatrices $\mathbf{K}_{11}$ and $\mathbf{K}_{22}$, their corresponding k_{ij} are all positive. Therefore, these permutation matrices (up to sign indeterminacy) are all local maxima of $J(\mathbf{R})$. Clearly, there are $p!(n-p)!$ such permutation matrices. For $0 \leq p \leq n$, the number of these permutation matrices is fairly large. Therefore, there always exists many local maxima of $J(\mathbf{R})$ that correspond to the stable feasible solutions of the ICA problem. In other words, the ICA problem has many stable feasible solutions under the one-bit-matching condition via locally maximizing the objective function $J(\mathbf{R})$.

In the similar context, we can prove the following theorem.

Theorem 2. *If $\mathbf{R}^*$ is a permutation matrix up to sign indeterminacy and $k_{ij} < 0$ at all the positions where $|r^*_{ij}| = 1$, it corresponds to a local minimum of the objective function $J(\mathbf{R})$.*

Remark 2. According to Theorem 2 and under the one-bit-matching condition, if the nonzero elements of a permutation matrix are all in the submatrices $\mathbf{K}_{12}$ and $\mathbf{K}_{21}$, it is a local minimum of $J(\mathbf{R})$. That is, it is possible that the local minimum of the objective function can be a feasible solution of the ICA problem, which actually explains why a local gradient-descent search of the objective function can also lead to a feasible solution of the ICA problem in certain scenarios. However, this kind of permutation matrix can only exist in the special case where $n = 2p$ (i.e., half super-Gaussian and half sub-Gaussian).

Moreover, since the condition (8) is also necessary for a local optimum solution (maximum or minimum) of the constrained function we can conclude that if the numbers of positive and negative k_{ij} at the positions where $|r^*_{ij}| = 1$ are both greater than 1, $\mathbf{R}^*$ will be a saddle point of the objective function $J(\mathbf{R})$. Clearly, such a permutation matrix generally exists and also corresponds to a feasible solution of the ICA problem with mixed super- and sub-Gaussian sources; however, this solution is always unstable.

To sum up the above results, we have established that under the one-bit-matching condition, there always exist many stable feasible solutions of the ICA problem via locally maximizing the objective function (6); in the meanwhile, there may exist some unstable feasible solutions of the ICA problem; in addition, there may exist local minima of $J(\mathbf{R})$ that correspond to the stable feasible solutions in the cases of mixed super- and sub-Gaussian sources.

In this paper, we have analyzed the feasible solutions of the ICA problem under the one-bit-matching condition. By mathematical analysis, we have proved that there always exist many stable feasible solutions of the ICA problem under the one-bit-matching condition. In the meanwhile, under the one-bit-matching condition, there may exist some unstable feasible solutions of the ICA problem; moreover, there may exist local minima of $J(\mathbf{R})$ corresponding to the stable feasible solutions of the ICA problem with mixed super- and sub-Gaussian sources.

1. o ., o ., .: for - r s r b s on of p s o r s. r s. ro ss , 41(7)(1993) 2461-2470
2. o o .: o o s s o ?. ro ss , 36(1994) 287-314
3. foss . o b o .: A b s r o of s o r s: A o ro . ro ss , 45(1995) 59-83
4. A. o s .: A for o - o ro o b s r - o b o o o . r o o , 7(1995) 1129-1159
5. A r ., o A., .: A r or for b s - r o of s o r s. A s r for o ro ss s s, 8(1996) 757-763.
6. .: A r r s: s o r ,s b , s o s. . f (.), ro . of . or s o o A r or s, . 97-103, 1998.
7. r o s o . .: fo oo for s o r s r o . ro ss rs, 4(1999) 112-114
8. ., . ., ., A r .: o o ss b for o - or ro r of s . ro . 1997 . o o f. r or s, , . 1821-1826, 1997.
9. ., . ., A r .: r r r r b s A or . ro o , 22(1998) 69-80
10. ., . ., A r .: r r r s s o o r s r- o b of r r A o r . . f (.) Proc. I&ANN'98, . 39-45, 1998.
11. ro .: A r r s o o o - s s or s. r o o , 10(1998) 2103-2114
12. r s o . ob r s : o o s s: A b o r- orr fo ro . r o o , 11(1999) 1957-1983
13. . b r .: A o s r or for o o s s. r o o , 13(2001) 677-689
14. o ., ., .: A r s r r o for o o s s (A). ro o , 68(2005) 267-272
15. . . .: ob o o r s s o for o - or o o s s ro . ro o - , 30(2000) 79-102
16. . ., . ., .: -b - o r for o o s s. r o o , 16(2004) 383-399
17. . ., .: A f r r r s o A o -b - o r . r o o , 17(2005) 331-334
18. b . .: A o r ro . r - oo o - , or , 1972.
19. A., Ar s . A., .: o r of or s or- o o s r s. A . r A . A ., 20(1998) 303-353