STABILITY ANALYSIS OF COMPLEX MAXIMUM LIKELIHOOD ICA USING WIRTINGER CALCULUS

Hualiang Li and Tülay Adalı

University of Maryland Baltimore County, Baltimore, Maryland 21250
{lihua1, adali}@umbc.edu

ABSTRACT

The desirable asymptotic optimality properties of the maximum likelihood (ML) estimator make it an attractive solution for performing independent component analysis (ICA) as well. Wirtinger calculus is shown to provide an attractive framework for the derivation and analysis of complex-valued algorithms using nonlinear functions, and hence of ICA algorithms as well. Local stability analysis of complex ICA based on ML presents a unique challenge, since in addition to the need for computation of derivatives, the Hessian of a matrix quantity needs to be evaluated, and for the complex case, it assumes a significantly more complicated form than the real-valued case. In this paper, we demonstrate how Wirtinger calculus allows the use of an elegant approach proposed by Amari *et al.* [5] in the analysis, thus enabling the derivation of the conditions for local stability of complex ML ICA. We further study the implications of the conditions for a generalized Gaussian density model.

Index Terms— Independent component analysis, Maximum likelihood, Local stability, Complex analysis

1. INTRODUCTION

Independent component analysis (ICA) for separating complex-valued signals is needed in a number of applications such as medical image analysis, radar, and communications. In ICA, the observed data are typically expressed as a linear combination of independent latent variables such that $\mathbf{x} = \mathbf{A}\mathbf{s}$ where $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$ is the vector of sources, $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ is the vector of observed random variables, and $\mathbf{A}$ is the mixing matrix. We consider the simple case where the number of independent variables is the same as the number of observed mixtures. The main task of the ICA problem is to estimate a separating matrix $\mathbf{W}$ that yields the independent components through $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$.

A number of approaches have been proposed to solve the complex ICA problem[1],[3],[6],[10],[12]. Maximum likelihood estimation offers the desirable large sample optimality properties and is an attractive solution for the ICA problem as it has been for many other estimation problems. Since the likelihood function is real valued and not complex-differentiable, a common approach for deriving the update algorithm and its analysis has been to transform the complex optimization problem to the real domain, which increases the dimensionality of the problem and usually leads to complicated forms. In addition, in order to study the local stability property of the ML ICA algorithms, one needs to calculate the complex Hessians of the likelihood function as a function of the parameter, which is a matrix quantity. Amari *et al.* [5] proposed an elegant approach to the problem through the definition of an invariant quantity that can be used for second-order analysis of ML ICA for the real case. In this paper, we show how Wirtinger calculus [15] allows one to extend this approach for the analysis of ML ICA in the complex domain, and more generally, for second-order study of matrix variables for complex-valued signal processing.

In [2], we have shown how to derive the complex ML ICA algorithm using Wirtinger calculus [15] without the use of any real-complex mappings. In this paper, we use the Wirtinger calculus to derive the complex natural gradient ML ICA algorithm. By extending the results given in [5] to the complex case, we show how to calculate the second-order differentials of the likelihood function where the fourth-rank tensor is involved. The resulting stability conditions are studied using a generalized Gaussian source model that allows the sources to assume both super-Gaussian and sub-Gaussian distribution through the control of a shape parameter.

This work is supported by the NSF grants NSF-CCF 0635129 and NSF-IIS 0612076.

2. COMPLEX GRADIENTS, BRANDWOOD'S RESULT AND WIRTINGER CALCULUS

In this section, we review results on complex gradients, which have not been consistently used in the literature and introduce the proper definitions for vector and matrix gradients. In particular, we underline the fact that complex gradient computations can be greatly simplified and obtained in a fashion similar to the real-valued gradients.

We are interested in functions $g(\cdot)$ that are *cost* functions and thus consider the special case of g: $\mathbb{C} \times \mathbb{C} \to \mathbb{R}$, for which Brandwood states the following result [7]:

Let f: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function of real variables x and y such that $g(z, z^*) = f(x, y)$, where $z = x + iy$ and $i = \sqrt{-1}$. Then,

1) The partial derivative $\partial g/\partial z$ (treating z^* as a constant in g) gives the same result as $1/2\,(\partial f/\partial x - i\partial f/\partial y)$ on substituting for z. Similarly, $\partial g/\partial z^* = 1/2\,(\partial f/\partial x + i\partial f/\partial y)$.

2) A necessary and sufficient condition for f to have a stationary point is that $\partial g/\partial z = 0$. Similarly, $\partial g/\partial z^* = 0$ is also a necessary and sufficient condition.

The two complex symbolic derivatives are actually defined much earlier and known as Wirtinger calculus[15] and, to date, the result has been primarily recognized in the German-speaking engineering community. It has been shown that as long as the real and imaginary parts of a function f are real-differentiable, the two Wirtinger derivatives also exist[15]. As a simple example, consider the function $g(z, z^*) = zz^* = |z|^2 = x^2 + y^2 = f(x, y)$. We have $1/2\,(\partial f/\partial x + i\partial f/\partial y) = x + iy = z$, which we can also evaluate as $\partial g/\partial z^* = z$, i.e., by treating z as a constant in g when calculating the partial derivative.

The results above can easily be extended to vector and matrix gradients. We define $\langle \cdot, \cdot \rangle$ as the scalar inner product between two matrices $\mathbf{W}$ and $\mathbf{V}$ as $\langle \mathbf{W}, \mathbf{V} \rangle = \text{Trace}(\mathbf{V}^H\mathbf{W})$ so that $\langle \mathbf{W}, \mathbf{W} \rangle = \|\mathbf{W}\|^2_{\text{Fro}}$, where the subscript Fro denotes the Frobenius norm. For vectors, the definition simplifies to $\langle \mathbf{w}, \mathbf{v} \rangle = \mathbf{w}^H\mathbf{v}$.

We define $\nabla_{\mathbf{z}} = [\partial/\partial z_1, \partial/\partial z_2, \ldots, \partial/\partial z_N]^T$ for vector $\mathbf{z} = [z_1, z_2, \ldots, z_N]^T$ with $z_k = x_k + jy_k$ in order to write

$$\triangle g = \langle \triangle \mathbf{z}, \nabla_{\mathbf{z}^*} g \rangle + \langle \triangle \mathbf{z}^*, \nabla_{\mathbf{z}} g \rangle = 2\text{Re}\{\triangle \mathbf{z}, \langle \nabla_{\mathbf{z}^*} g \rangle\} \qquad (1)$$

for a function $g(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \rightarrow \mathbb{R}$. It is also important to reiterate that the gradient $\nabla_{\mathbf{z}^*} g$ defines the direction of the maximum rate of change in $g(\cdot, \cdot)$ with respect to $\mathbf{z}$, not $\nabla_{\mathbf{z}} g$.

The extension from the vector gradient to matrix gradient is straightforward. If the function $g(\mathbf{W}, \mathbf{W}^*) : \mathbb{C}^{N\times N} \times \mathbb{C}^{N \times N} \rightarrow \mathbb{R}$ is real-differentiable, we then have the first-order expansion

$$\triangle g = \left\langle \triangle \mathbf{W}, \frac{\partial g}{\partial \mathbf{W}^*} \right\rangle + \left\langle \triangle \mathbf{W}^*, \frac{\partial g}{\partial \mathbf{W}} \right\rangle$$

where $\partial g/\partial \mathbf{W}$ is an $N \times N$ matrix whose (i, j)th entry is the partial derivative of g with respect to w_{ij}. By arranging the matrix gradient into a vector and by using the Cauchy-Schwarz inequality, it is easy to show that the matrix gradient $\partial g/\partial \mathbf{W}^*$ defines the direction of the maximum rate of change g with respect to $\mathbf{W}$.

Based on (1), similar to a scalar function of two real vectors, we can write the second-order Taylor expansion of $g(\mathbf{z}, \mathbf{z}^*)$

$$\triangle g = \langle \triangle \mathbf{z}, \nabla_{\mathbf{z}^*} g \rangle + \langle \triangle \mathbf{z}^*, \nabla_{\mathbf{z}} g \rangle + \frac{1}{2}\left\langle \frac{\partial g}{\partial \mathbf{z} \partial \mathbf{z}^T} \triangle \mathbf{z}, \triangle \mathbf{z}^* \right\rangle + \frac{1}{2}\left\langle \frac{\partial g}{\partial \mathbf{z}^* \partial \mathbf{z}^H} \triangle \mathbf{z}^*, \triangle \mathbf{z} \right\rangle + \left\langle \frac{\partial g}{\partial \mathbf{z} \partial \mathbf{z}^H} \triangle \mathbf{z}^*, \triangle \mathbf{z}^* \right\rangle \qquad (2)$$

as shown in [4], which will be used in the analysis results of this paper as well.

3. COMPLEX MAXIMUM LIKELIHOOD ICA

We first briefly introduce our notation and the relevant preliminaries for application to ICA. The probability density function (pdf) of a complex random variable $X = X_{re} + iX_{im}$ is defined as $p_X(x) \equiv p_{X_{re}X_{im}}(x_{re}, x_{im})$. Expectation of $g(X)$ is given by $E\{g(X)\} = \int\int g(x_{re} + ix_{im}) p_X(x) dx_{re} dx_{im}$ for any measurable function $g : \mathbb{C} \rightarrow \mathbb{C}$. A complex random variable X is circular in the strict-sense if X and $Xe^{j\theta}$ have the same pdf.

The traditional ICA problem is to determine a weight matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ approximates the source $\mathbf{s}$ subject to the permutation and scaling ambiguity. To write the likelihood, we need to consider the mapping $\mathbb{C}^N \rightarrow \mathbb{R}^{2N}$ such that $\bar{\mathbf{y}} = \bar{\mathbf{W}}\bar{\mathbf{x}} = \bar{\mathbf{s}}$ where $\bar{\mathbf{y}} = [\mathbf{y}^T_{re}\mathbf{y}^T_{im}]^T$, $\bar{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_{re} & -\mathbf{W}_{im} \\ \mathbf{W}_{im} & \mathbf{W}_{re} \end{bmatrix}$, $\bar{\mathbf{x}} = [\mathbf{x}^T_{re}\mathbf{x}^T_{im}]^T$ and $\bar{\mathbf{s}} = [\mathbf{s}^T_{re}\mathbf{s}^T_{im}]^T$.

Given T independent samples $\mathbf{x}(t)$, we write the log-likelihood function as[5]

$$l'(\mathbf{y}, \mathbf{W}) = \log |\det(\bar{\mathbf{W}})| + \sum_{k=1}^{m} \log p_{\bar{\mathbf{s}}_k}(\bar{y}_k)$$

where $\bar{\mathbf{s}}_k$ is the kth source $\in \mathbb{R}^2$ with density denoted by p_k. Maximization of l' is equivalent to minimization of l where $l = -l'$. Simple algebraic and differential calculus yields

$$dl = -\text{tr}(d\bar{\mathbf{W}}\bar{\mathbf{W}}^{-1}) + \bar{\psi}^T(\bar{\mathbf{y}})d\bar{\mathbf{y}} \qquad (3)$$

where $\bar{\psi}(\bar{\mathbf{y}})$ is a $2N \times 1$ column vector with components $\bar{\psi}(\bar{\mathbf{y}}) = -[\frac{\partial \log p_1(y_1)}{\partial y_{1,re}} \; \cdots \; \frac{\partial \log p_N(y_N)}{\partial y_{N,re}} \frac{\partial \log p_1(y_1)}{\partial y_{1,im}} \; \cdots \; \frac{\partial \log p_N(y_N)}{\partial y_{N,im}}]$. We write $\log p_{\bar{\mathbf{s}}}(\bar{y}) = \log p_{\mathbf{s}}(y, y^*)$ and using Wirtinger calculus, it is straightforward to show

$$\bar{\psi}^T(\bar{\mathbf{y}})d\bar{\mathbf{y}} = \psi^T(\mathbf{y}, \mathbf{y}^*)d\mathbf{y} + \psi^H(\mathbf{y}, \mathbf{y}^*)d\mathbf{y}^*$$

where $\psi(\mathbf{y}, \mathbf{y}^*)$ is an $N \times 1$ column vector with complex components

$$\psi_k(y_k, y_k^*) = -\frac{\partial \log p_k(y_k, y_k^*)}{\partial y_k}.$$

Defining a $2N \times 2N$ matrix $\mathbf{P} = \frac{1}{2}\begin{bmatrix} \mathbf{I} & i\mathbf{I} \\ i\mathbf{I} & \mathbf{I} \end{bmatrix}$, we obtain

$$\begin{aligned} \text{tr}(d\bar{\mathbf{W}}\bar{\mathbf{W}}^{-1}) &= \text{tr}(d\bar{\mathbf{W}}\mathbf{P}\mathbf{P}^{-1}\bar{\mathbf{W}}^{-1}) \\ = \text{tr}\left\{ \begin{bmatrix} d\mathbf{W}^* & id\mathbf{W} \\ id\mathbf{W}^* & d\mathbf{W} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{W}^* & i\mathbf{W} \\ i\mathbf{W}^* & \mathbf{W} \end{bmatrix}^{-1} \right\} \\ &= \text{tr}(d\mathbf{W}\mathbf{W}^{-1}) + \text{tr}(d\mathbf{W}^*\mathbf{W}^{-*}) \end{aligned}$$

Therefore we can write (3) as

$$\begin{aligned} dl = -\text{tr}(d\mathbf{W}\mathbf{W}^{-1}) - \text{tr}(d\mathbf{W}^*\mathbf{W}^{-*}) + \\ \psi^T(\mathbf{y}, \mathbf{y}^*)d\mathbf{y} + \psi^H(\mathbf{y}, \mathbf{y}^*)d\mathbf{y}^* \end{aligned} \qquad (4)$$

Using $\mathbf{y} = \mathbf{W}\mathbf{x}$ and defining $d\mathbf{Z} = (d\mathbf{W})\mathbf{W}^{-1}$, we obtain

$$d\mathbf{y} = (d\mathbf{W})\mathbf{x} = d\mathbf{W}(\mathbf{W}^{-1})\mathbf{y} = d\mathbf{Z}\mathbf{y}, \qquad d\mathbf{y}^* = d\mathbf{Z}^*\mathbf{y}^*.$$

By treating $\mathbf{W}$ as a constant matrix, the differential matrix $d\mathbf{Z}$ has components dz_{ij} that are linear combinations of dw_{ij} and is a non-integrable differential form. However, this transformation greatly simplifies the expression for the Taylor series expansion without changing the function value. It also provides an elegant approach for the derivation of the natural gradient update for ML ICA[5]. Using this transformation, we can write (4) as

$$\begin{aligned} dl = -\text{tr}(d\mathbf{Z}) - \text{tr}(d\mathbf{Z}^*) + \psi^T(\mathbf{y}, \mathbf{y}^*)d\mathbf{Z}\mathbf{y} + \\ \psi^H(\mathbf{y}, \mathbf{y}^*)d\mathbf{Z}^*\mathbf{y}^* \end{aligned}$$

Therefore, the gradient update rule for $\mathbf{Z}$ is given by

$$\triangle \mathbf{Z} = -\mu \frac{\partial l}{\partial \mathbf{Z}^*} = \mu[\mathbf{I} - \psi^*(\mathbf{y}, \mathbf{y}^*)\mathbf{y}^H]$$

which is equivalent to

$$\triangle \mathbf{W} = \mu[\mathbf{I} - \psi^*(\mathbf{y}, \mathbf{y}^*)\mathbf{y}^H]\mathbf{W} \qquad (5)$$

by using $d\mathbf{Z} = (d\mathbf{W})\mathbf{W}^{-1}$.

Thus the complex score function is defined as $\psi^*(\mathbf{y}, \mathbf{y}^*)$, as in [2], which takes a form very similar to the real case [5], but with the difference that in the complex case the entries in the score function are defined using Wirtinger derivatives.

4. STABILITY OF THE COMPLEX ML ICA UPDATE

The stationary point of the update rule given in (5) satisfies

$$E\{\mathbf{I} - \psi^*(\mathbf{y}, \mathbf{y}^*)\mathbf{y}^H\} = \mathbf{0}. \qquad (6)$$

By the definition of the ICA problem, it is easy to see that $\mathbf{W} = \mathbf{A}^{-1}$ is the solution of (6). To ensure that the ML solution is asymptotically stable, we need to calculate the Hessian term d^2l since

$\mathbf{I} - \psi(\mathbf{y}, \mathbf{y}^*)^* \mathbf{y}^H$ is derived from the gradient dl. From (2), the equilibrium is stable if and only if the expectation of

$$d^2 l = \frac{1}{2}\bigg(\sum \frac{\partial l}{\partial w_{ij} \partial w_{kl}} dw_{ij} dw_{kl} + \sum \frac{\partial l}{\partial w_{ij}^* \partial w_{kl}^*} dw_{ij}^* dw_{kl}^* \bigg) + \sum \frac{\partial l}{\partial w_{ij} \partial w_{kl}^*} dw_{ij} dw_{kl}^*$$

is positive definite[5, 14]. To derive the stability conditions, we can again work on differentials dz_{ij} instead of working with dw_{ij} which helps circumvent the difficulty of working with tensors.

The stability conditions are given in the following theorem. To state the theorem, we define

$$\begin{aligned} \alpha_i &= E\{y_i^2\}, & \beta_i &= E\{|y_i|^2\}, \\ \gamma_i &= E\{\eta_i(y_i, y_i^*)\}, & \delta_i &= E\{\theta_i(y_i, y_i^*)\}, \\ u_i &= E\{y_i^2 \eta_i(y_i, y_i^*)\}, & v_i &= E\{|y_i|^2 \theta_i(y_i, y_i^*)\}, \end{aligned} \tag{7}$$

where $\eta = \partial\psi/\partial y$ and $\theta = \partial\psi/\partial y^*$.

Theorem 1 *The separating solution is a stable equilibrium of the update equation given in (5), if and only if*

$$v_i > 0 \tag{8}$$

$$v_i^2 > |u_i + 1|^2 \tag{9}$$

$$\beta_j \delta_i > 0 \tag{10}$$

$$\frac{(\beta_i\delta_j)^2 - 1 - |\alpha_i\gamma_j|^2}{\beta_i\delta_j} + \frac{(\beta_j\delta_i)^2 - 1 - |\alpha_j\gamma_i|^2}{\beta_j\delta_i} > 0 \tag{11}$$

$$(\beta_i\delta_j\beta_j\delta_i - 1)^2 + |\alpha_i\gamma_j|^2|\alpha_j\gamma_i|^2 - (\beta_j\delta_i)^2|\alpha_i\gamma_j|^2 - (\beta_i\delta_j)^2|\alpha_j\gamma_i|^2 - 2\mathrm{Re}\big(\alpha_i\gamma_j\alpha_j\gamma_i\big) > 0 \tag{12}$$

for all i, j such that $i \neq j$.

The proof of this theorem is given in Appendix A. If all the sources are second-order circular, we have $\alpha_i = 0$, and the last two sufficient stability conditions can be further simplified as

$$\frac{(\beta_i\delta_j)^2 - 1}{\beta_i\delta_j} + \frac{(\beta_j\delta_i)^2 - 1}{\beta_j\delta_i} > 0 \tag{13}$$

$$(\beta_i\delta_j\beta_j\delta_i - 1)^2 > 0 \tag{14}$$

for all i, j such that $i \neq j$. Compared with the real stability conditions given in [5], the stability conditions for complex ML ICA assume more complicated forms even for the circular case. When the sources are noncircular, we have $|\alpha_i| > 0$, thus the conditions are obviously more difficult to be satisfied if we compare the last two stability conditions for both cases.

If $\mathbf{W}$ is constrained to be a unitary matrix, it can be shown that the stability conditions will be greatly simplified as shown for the real case [11].

5. EXAMPLES

As an example, consider the case where the source densities take the form of bivariate generalized Gaussian distribution (GGD), which can be written as

$$p_{GGD}(\mathbf{s}) = a \exp(-[\gamma(\bar{\mathbf{s}} - \mu)^T \mathbf{K}^{-1} (\bar{\mathbf{s}} - \mu)]^c)$$

where $\bar{\mathbf{s}} = [s_{re} s_{im}]^T$ is the mapping of one complex source s in $\mathbb{R}^2$, $a = \frac{c\gamma}{\pi\Gamma(1/c)|\mathbf{K}|^{1/2}}$, $\gamma = \frac{\Gamma(2/c)}{2\Gamma(1/c)}$, $\mathbf{K}$ is the covariance matrix ,

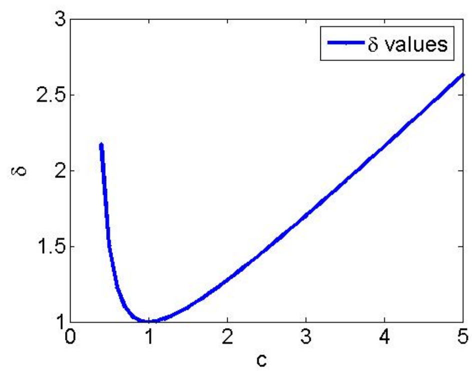

Fig. 1. The stability indicator δ as a function of shape parameter c

μ is the mean vector, and $\Gamma(\cdot)$ is the Gamma function. The bivariate GGD model is completely determined by the shape parameter c, covariance matrix $\mathbf{K}$ and μ. By changing the shape parameter c, we can obtain a family of distributions, which can vary from bivariate Laplacian to bivariate normal for small c values, and to bivariate uniform as c goes to infinity.

Let us assume that the sources are circular and have zero mean to write

$$p_{GGD,circ}(\bar{\mathbf{s}}) = a \exp(-[2\gamma s s^*]^c)$$

since we have $\mathbf{K} = 1/2\mathbf{I}$. Therefore we can easily evaluate

$$\begin{aligned} \psi(s, s^*) &= -\frac{\log p(s, s^*)}{\partial s} = c \cdot [2\gamma s s^*]^{c-1} \cdot 2\gamma s^*, \\ \eta(s, s^*) &= \frac{\partial\psi}{\partial s} = c(c-1)(2\gamma)^c (ss^*)^{c-2}(s^*)^2, \\ \theta(s, s^*) &= \frac{\partial\psi}{\partial s^*} = c^2(2\gamma)^c(ss^*)^{c-1}. \end{aligned}$$

An important indicator of stability, $\delta = E\{\theta(y, y^*)\}$, is plotted in Fig. 1 as a function of the shape parameter c. Since $\beta_i = 1$, the condition $\beta_i\delta_j > 1$ is satisfied for all $c \neq 1$, i.e., for all nongaussian circular GGDs. Hence, the stability conditions given in (10) – (14) are also satisfied.

The two quantities defined in (7) can be calculated as

$$\begin{aligned} u &= \int\int s^2 \eta(s, s^*) p(s, s^*) ds_{re} ds_{im} = c - 1, \\ v &= \int\int s s^* \theta(s, s^*) p(s, s^*) ds_{re} ds_{im} = c. \end{aligned}$$

Therefore the stability condition given in (8) is also satisfied. However, (9) is not satisfied since $v = u + 1$. Thus, the complex ML ICA algorithm does not satisfy the stability conditions completely for this case. However, note that the first quadratic term in (15) is positive definite and the second quadratic term can be calculated as $c(\mathrm{Re}\{dz_{ii}^2 + |dz_{ii}|^2\})$. For any complex number z, it is true that $|z| \geq \mathrm{Re}\{z\}$. Therefore $E\{d^2 l\}$ might take the problematic value, which is zero, only when $d\mathbf{Z}$, or actually $\mathbf{I} - \psi^*(\mathbf{y}, \mathbf{y}^*)\mathbf{y}^H$, is a diagonal matrix with pure imaginary numbers as the diagonal elements.

When the sources are noncircular, we have $|\alpha_i| > 0$, thus the conditions are obviously more difficult to be satisfied if we compare the last two stability conditions for both cases. This is not surprising since for the real case, it has been shown that the stability conditions are not satisfied for some small value of c even when the optimal score function matched to the source distribution is used [8]. Our numerical studies confirm these observations, which we could not include in this paper due to space limitations.

6. CONCLUSIONS

We studied the local stability of complex ML ICA algorithm and derived the sufficient conditions. We demonstrate how to derive and analyze complex ICA algorithm using Wirtinger calculus without the need for the commonly used real-complex mappings. We also study the implication of the stability condition for the class of generalized Gaussian probability density functions.

A. APPENDIX

Proof of Theorem 1

Proof Given
$dl = -\mathrm{tr}(d\mathbf{Z}) - \mathrm{tr}(d\mathbf{Z}^*) + \psi^T(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}\mathbf{y} + \psi^H(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}^*\mathbf{y}^*$
as a function of $(\mathbf{Z},\mathbf{Z}^*,\mathbf{y},\mathbf{y}^*)$, we have $\partial(\mathrm{tr}(d\mathbf{Z}))/\partial\mathbf{Z} = \mathbf{0}$, $\partial(\mathrm{tr}(d\mathbf{Z}^*))/\partial\mathbf{Z}^* = \mathbf{0}$, therefore the second-order differential can be written as

$$\begin{aligned} d^2l &= d(\psi^T(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}\mathbf{y} + \psi^H(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}^*\mathbf{y}^*) \\ &= 2\mathrm{Re}\{d(\psi^T(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}\mathbf{y})\} \\ &= 2\mathrm{Re}\{\mathbf{y}^T d\mathbf{Z}^T\eta(\mathbf{y},\mathbf{y}^*)d\mathbf{y} + \mathbf{y}^T d\mathbf{Z}^T\theta(\mathbf{y},\mathbf{y}^*)d\mathbf{y}^* + \\ &\qquad \psi^T(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}d\mathbf{y}\} \\ &= 2\mathrm{Re}\{\mathbf{y}^T d\mathbf{Z}^T\eta(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}\mathbf{y} + \mathbf{y}^T d\mathbf{Z}^T\theta(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}^*\mathbf{y}^* + \\ &\qquad \psi^T(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}d\mathbf{Z}\mathbf{y}\} \end{aligned}$$

where $\eta(\mathbf{y},\mathbf{y}^*)$ is a diagonal matrix with ith diagonal element $-\partial\log p_i(y_i,y_i^*)/\partial y_i\partial y_i$, $\theta(\mathbf{y},\mathbf{y}^*)$ is a diagonal matrix with ith diagonal element $-\partial\log p_i(y_i,y_i^*)/\partial y_i\partial y_i^*$.

The expectation of the first term is given by

$$\begin{aligned} E\{\mathbf{y}^T d\mathbf{Z}^T\eta(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}\mathbf{y}\} &= \sum E\{y_i dz_{ji}\eta_j(y_j,y_j^*)dz_{jk}y_k\} \\ &= \sum_{j\neq i} E\{y_i^2\}E\{\eta_j(y_j,y_j^*)\}dz_{ji}^2 + \sum_i E\{y_i^2\eta_i(y_i,y_i^*)\}dz_{ii}^2 \\ &= \sum_{j\neq i}\alpha_i\gamma_j dz_{ji}^2 + \sum_i u_i dz_{ii}^2 \end{aligned}$$

where we have used $\mathbf{W} = \mathbf{A}^{-1}$, definitions given in (7), and independence of y_is.

The expectation of the second term of $E\{d^2l\}$ is written as

$$\begin{aligned} E\{\mathbf{y}^T d\mathbf{Z}^T\theta(\mathbf{y},\mathbf{y}^*)d\mathbf{Z}^*\mathbf{y}^*\} &= \sum E\{y_i dz_{ji}\theta_j(y_j,y_j^*)dz_{jk}^* y_k^*\} \\ &= \sum_{j\neq i} E\{|y_i|^2\}E\{\theta_j(y_j,y_j^*)\}|dz_{ji}|^2 + \\ &\qquad \sum_i E\{|y_i|^2\theta_i(y_i,y_i^*)\}|dz_{ii}|^2 \\ &= \sum_{j\neq i}\beta_i\delta_j|dz_{ji}|^2 + \sum_i v_i|dz_{ii}|^2, \end{aligned}$$

and the third term by

$$\begin{aligned} E\{\psi(\mathbf{y},\mathbf{y}^*)^T d\mathbf{Z}d\mathbf{Z}\mathbf{y}\} &= \sum E\{\psi_i(y_i,y_i^*)dz_{ij}dz_{jk}y_k\} \\ &= \sum E\{y_i\psi_i(y_i)\}dz_{ij}dz_{ji} = \sum_{i,j} dz_{ij}dz_{ji} \end{aligned}$$

because $E\{y_i\psi_i(y_i)\} = 1$ at the stationary point.

Now we can write the expectation of the second-order differential as

$$\begin{aligned} E\{d^2l\} &= 2\mathrm{Re}\{\sum_{j\neq i}\alpha_i\gamma_j dz_{ji}^2 + \sum_i u_i dz_{ii}^2 + \sum_{j\neq i}\beta_i\delta_j|dz_{ji}|^2 + \\ &\qquad \sum_i v_i|dz_{ii}|^2 + \sum_{i,j} dz_{ij}dz_{ji}\} \\ &= \sum_{i\neq j}\begin{bmatrix} dz_{ij} & dz_{ji} & dz_{ij}^* & dz_{ji}^* \end{bmatrix}\begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1 \end{bmatrix}\begin{bmatrix} dz_{ij}^* \\ dz_{ji}^* \\ dz_{ij} \\ dz_{ji} \end{bmatrix} \\ &\qquad + \sum_i \begin{bmatrix} dz_{ii} & dz_{ii}^* \end{bmatrix}\mathbf{H}_3\begin{bmatrix} dz_{ii}^* \\ dz_{ii} \end{bmatrix} \end{aligned} \quad (15)$$

where $\mathbf{H}_1 = \begin{bmatrix} \beta_j\delta_i & 0 \\ 0 & \beta_i\delta_j \end{bmatrix}$, $\mathbf{H}_2 = \begin{bmatrix} \alpha_j\gamma_i & 1 \\ 1 & \alpha_i\gamma_j \end{bmatrix}$,
$\mathbf{H}_3 = \begin{bmatrix} v_i & u_i+1 \\ u_i^*+1 & v_i \end{bmatrix}$.

Given a real-valued function $g^*(\mathbf{y},\mathbf{y})$, it can be easily shown that the matrix $\partial^2 g/\partial\mathbf{y}\partial\mathbf{y}^H$ is a Hermitian matrix using Wirtinger calculus[9]. Since the diagonal elements of a Hermitian matrix is real-valued, θ_i is a real-valued function. Thus, the matrices $\mathcal{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1 \end{bmatrix}$ and $\mathbf{H}_3$ are all Hermitian. If the matrices $\mathcal{H}$ and $\mathbf{H}_3$ are all positive-definite, then we have $E\{d^2l\} > 0$ for any $\mathbf{Z}$.

From [13], $\mathcal{H} > 0$ (positive definite) if and only if $\mathbf{H}_1 > 0$ and $\mathbf{H}_1 - \mathbf{H}_2\mathbf{H}_1^{-1}\mathbf{H}_2^* > 0$. The first inequality simply implies (10). The positiveness of the trace and determinant of $\mathbf{H}_1 - \mathbf{H}_2\mathbf{H}_1^{-1}\mathbf{H}_2^*$ give the (13) and (14). Similarly $\mathbf{H}_3 > 0$ if and only (8) and (9) hold. ∎

B. REFERENCES

[1] T. Adalı, T. Kim and V. Calhoun, “Independent component analysis by complex nonlinearities,” in *Proc. ICASSP*, vol. 5, pp. 525–528, Montreal, Canada, May 2004.

[2] T. Adalı and H. Li, “A practical formulation for computation of complex gradients and its application to maximum likelihood,”in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Honolulu, Hawaii, April 2007.

[3] J. Anemüller, T. Sejnowski and S. Makeig, “Complex independent component analysis of frequency-domain electroencephalographic data,” *Neural Networks*, vol. 16, pp. 1311–1323, Nov. 2003.

[4] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, “The constrained total least squares technique and its applications to harmonic superresolution,” *IEEE Trans. Signal Processing*, vol. 39, pp. 1070–1087, May 1991.

[5] S.-I. Amari, T. Chen, and A. Cichocki, “Stability analysis of adaptive blind source separation,” *Neural Networks*, vol. 10, pp. 1345–1351, 1997.

[6] E. Bingham an A. Hyvarinen, “A fast fixed-point algorithm for independent component analysis of complex valued signals,” *Int. J. Neural Systems*., vol. 10, pp. 1–8, Feb. 2000.

[7] D. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *Proc. Inst. Elect. Eng.*, vol. 130, pp. 11–16, Feb. 1983.

[8] S. Choi, A. Cichocki and S.-I. Amari, “Local stability analysis of flexible independent component analysis algorithm,” in *Proc. ICASSP*, vol. 6, pp. 3426–3429, Istanbul, Turkey, June 2000.

[9] A. V. D. Bos, “Complex gradient and Hessian,” *Proc. Inst. Elec. Eng., Vision, Image, Signal Process.*, vol. 141, pp. 380–382, Dec. 1994.

[10] J.-F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proc. Radar Signal Process.*, vol. 140, pp. 362–370, 1993.

[11] J.-F. Cardoso, “On the stability of source separation algorithms,” *Journal of VLSI signal processing systems*, vol. 26, pp. 7–14, Aug. 2000.

[12] J. Eriksson and V. Koivunen, “Complex-valued ICA using second order statistics,” in *Proc. MLSP*, Saõ Luis, Brazil, 2004.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1988.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer 2000.

[15] R. Remmert, *Theory of Complex Functions*, Springer 1991.