ON THE SOLUTION OF CIRCULAR AND NONCIRCULAR COMPLEX KLD-ICA IN THE PRESENCE OF NOISE

Benedikt Loesch and Bin Yang

Institute of Signal Processing and System Theory, University of Stuttgart
Email: \{benedikt.loesch,bin.yang\}@iss.uni-stuttgart.de

ABSTRACT
This paper aims at studying the solution of linear independent component analysis (ICA) based on the Kullback-Leibler divergence (KLD) for a linear noisy mixing model in the determined case. The derivation is done by using a perturbation analysis valid for small noise variance. We study the noncircular complex and the circular complex case. We show that for a wide range of the shape parameter of the generalized Gaussian distribution (GGD) and of the noncircularity index, the signal-to-interference-plus-noise ratio (SNIR) of KLD-based ICA is close to that of linear minimum mean squared error (MMSE) estimation.

Index Terms— Blind source separation, Independent component analysis, Minimum mean square error, Kullback-Leibler divergence, Perturbation analysis, noncircular complex

1. INTRODUCTION
In the last decades, independent component analysis (ICA) has been studied intensively and has found many applications. Usually, a linear noiseless mixing model is assumed and the separated signals are obtained using a linear transform of the observed signals. Although noisy mixing models have been studied quite early, for example in the context of contrast-based approaches [1], many publications still consider a noiseless mixing model for simplicity. The presence of noise leads to a bias in the estimation of the mixing matrix. Douglas et al. introduced measures in [2] to reduce this bias. Cardoso showed in [3] that for the noisy case the performance of source separation depends on the distribution of the sources, the signal-to-noise (SNR) ratio and the mixing matrix. Hyvärinen shown in [4] that in the noisy case, the ML estimate of the signals is a nonlinear function of the observations. Koldovsky et al. drew in [5, 6] parallels between MMSE estimation and ICA for the real data case. They derived the bias of several variants of the FastICA algorithm from the MMSE solution.

In this paper, we focus on KLD-based ICA with equal number of sources and sensors (determined case). As shown by many different authors, KLD is (up to constant terms) identical to mutual information, the well-known information maximization (INFOMAX) principle and for the noiseless case to ML estimation [1]. In many applications such as audio source separation in the frequency domain or telecommunications, signals are complex and not necessarily circular. General conditions regarding identifiability, uniqueness and separability can be found in [7]. Algorithms for complex ML-ICA have been studied for the noiseless case in [8]. However, practical applications have to deal with the noisy case. In [9], we derived the theoretical ICA solution for the noisy determined case and real signals. In this paper, we extend the results to circular and noncircular complex signals.

In the estimation of the demixing matrix, one has to distinguish three different factors:
1. Bias of the demixing matrix from the inverse mixing matrix: Often a bias of an estimator is considered to be unwanted, but in the case of noisy ICA the bias of the demixing matrix from the inverse mixing matrix actually leads to a reduced noise level in the separated signals and hence it can be considered to be desired.
2. Variance of the estimated inverse mixing matrix in the noiseless case due to blind estimation and randomness of the sources. This variance can be lower bounded by the Cramér-Rao bound for ICA derived for the real case in [10, 11] and for the circular complex case in [12].
3. Increased variance of the demixing matrix due to noise. In this paper, we study only the first factor, i.e. the bias of the demixing matrix from the inverse mixing matrix.

2. SIGNAL MODEL
We assume the linear noisy mixing model
\[ x = As + v \] (1)
where \( x \in \mathbb{C}^N \) are \( N \) linear combinations of \( N \) original signals \( s \in \mathbb{C}^N \) with additive noise \( v \in \mathbb{C}^N \). We make the following assumptions:
1. The mixing matrix \( A \in \mathbb{C}^{N \times N} \) is deterministic and invertible.
2. \( s = [s_1, \ldots, s_N]^T \in \mathbb{C}^N \) are \( N \) independent random variables with zero mean and unit variance (after scaling the rows of \( A \) suitably). \( \gamma_i = E[s_i^2] \) can be thought as a non-circularity index of \( s_i \) with \( 0 \leq |\gamma_i| \leq 1 \). Any zero mean complex random variable can be scaled such that \( \gamma_i \) is real, i.e. \( 0 \leq \gamma_i \leq 1 \) [8, 13], which we assume in the following. The probability density functions (pdfs) \( q(s_i) \) of \( s_i \) can be different. \( q(s_i) \) is \( 3 \) times continuously differentiable with respect to \( s_i \) and \( s_i^* \) in the sense of Wirtinger derivatives which will be introduced in section 3.1. All expectations required in this paper exist.
3. \( v = [v_1, \ldots, v_M]^T \in \mathbb{C}^N \) are \( N \) random variables with zero mean and the covariance matrix \( E[vv^H] = \sigma^2 R_v \).
\[ \sigma^2 = \frac{1}{N} tr \left( E[vv^H] \right) \] is the average variance of \( v \) and \( tr(R_v) = N \). \( R_v = \frac{1}{N} E[vv^T] \) is the normalized pseudo-covariance matrix. \( R_v = 0 \) if \( v \) is circular complex. The pdf of \( v \) is arbitrary but assumed to be symmetric, i.e. \( q(v) = q(-v) \). This implies \( E(\prod_{i=1}^N v_i^{k_i}(v_i^*)^{k_i}) = 0 \) for \( \sum_{i=1}^N (k_i + \bar{k_i}) \) odd.
4. \( s \) and \( v \) are independent.

The task of ICA is to demix the signals \( x \) by a linear transform \( W \in \mathbb{C}^{N \times N} \)
\[ y = Wx = WAs + Wv \] (2)
such that \( y \) is "as close to \( s \)" as possible according to some metric.
3. KLD-BASED ICA

In this paper, we focus on the ICA solution based on the KLD

$$D_{pq}(W) = \int p_{y}(y;W) \log \frac{p_{y}(y;W)}{q(y)} \, dy$$

(3)

where $p_{y}(y;W)$ is the pdf of $y$. It depends on the pdf of observation $x$, i.e. on the pdf of the original source signals $s$ and noise $v$, as well as on the demixing matrix $W$. $q(s) = \prod_{i=1}^{N} q_{i}(s_{i})$ is the assumed pdf of the original signals. We assume that we have perfect knowledge about the distribution of the original signals and $q(s)$ is identical to the true pdf $q^{0}(s)$ of $s$. The KLD is known to have the following properties:

- $D_{pq}(W) \geq 0$ for any $p_{y}(y;W)$ and $q(y)$.
- $D_{pq}(W) = 0$ if $p_{y}(y;W) = q(y)$.

This means that minimizing the KLD with respect to $W$ is equivalent to maximizing the pdf of the demixed sources $Z_{i}$ as similar as possible. To achieve this, we assume $q(s) = \prod_{i=1}^{N} q_{i}(s_{i})$, minimizing KLD corresponds to making $y$ as independent as possible and the individual $q_{i}$ to have a pdf as close as possible to $q_{i}(s_{i})$. The ICA solution $W_{ICA}$ for the demixing matrix based on KLD is given by

$$W_{ICA} = \arg \min_{W} D_{pq}(W).$$

(4)

In the following, we will first derive the ICA solution for the general (noncircular) complex case. The circular complex case and the real case are discussed as two special cases.

3.1. General noncircular complex case

The KLD cost function of a complex demixing matrix $W$ is a function of the real and imaginary part of $W$. Using the Wirtinger calculus [13, 14], we can also write it as a function of $W$ and $W^*$:

$$D_{pq}(W,W^*) = \int p_{y}(y^{*},W;W^{*}) \log \frac{p_{y}(y^{*};W;W^{*})}{q(y^{*})} \, dy.$$  

(5)

Before we continue with the derivation of the ICA solution, we provide a short summary of the Wirtinger calculus. Given $Z = X + jY \in \mathbb{C}^{N \times M}$, $X, Y \in \mathbb{R}^{N \times M}$, and a real scalar cost function $f(Z, Z^*) = \hat{f}(X, Y) \in \mathbb{R}$. Instead of calculating the derivatives of $\hat{f}(\cdot)$ with respect to $X, Y$, the Wirtinger calculus calculates the partial derivatives of $f(Z, Z^*)$ with respect to $Z$ and $Z^*$, treating $Z$ and $Z^*$ as two independent variables [13, 14]. Let $\frac{\partial f}{\partial Z} \in \mathbb{C}^{N \times M}$ and $\frac{\partial f}{\partial Z^*} \in \mathbb{R}^{N \times M}$. It can be shown:

1. The partial derivatives of $f(\cdot)$ with respect to $Z$ and $Z^*$ are

$$\frac{\partial \hat{f}}{\partial Z} = \frac{1}{2} \left( \frac{\partial f}{\partial X} - j \frac{\partial f}{\partial Y} \right), \quad \frac{\partial \hat{f}}{\partial Z^*} = \frac{1}{2} \left( \frac{\partial f}{\partial X} + j \frac{\partial f}{\partial Y} \right).$$

(6)

2. The stationary point of $f(\cdot)$ and $\hat{f}(\cdot)$ is given by

$$\frac{\partial \hat{f}}{\partial X} = 0 \quad \text{and} \quad \frac{\partial \hat{f}}{\partial Y} = 0 \quad \Rightarrow \quad \frac{\partial f}{\partial X} = 0 \quad \Rightarrow \quad \frac{\partial \hat{f}}{\partial Z^*} = 0.$$  

(7)

3. The direction of steepest descent of $f(\cdot)$ is given by $-\frac{\partial f}{\partial Z^*}$ and not $-\frac{\partial \hat{f}}{\partial Z^*}$. Coming back to the KLD cost function in (5), its derivative $\frac{\partial D_{pq}(W,W^*)}{\partial W}$ is given by

$$\frac{\partial D_{pq}(W,W^*)}{\partial W} = -W^{H}H + E \left[ \varphi^{*}(y,y^*)x^{H} \right].$$

(8)

where $\varphi(y,y^*) = [\varphi_{1}(y_{1},y_{1}^*), \cdots, \varphi_{N}(y_{N},y_{N}^*)]^{T}$ and $\varphi_{i}(s_{i},s_{i}^*) = -\frac{\partial \log q_{i}(s_{i},s_{i}^*)}{\partial s_{i}}$. The derivative $\frac{\partial \hat{f}}{\partial \hat{f}}$ is also defined using the Wirtinger calculus.

A necessary condition for minimizing $D_{pq}(W,W^*)$ at $W = W_{ICA}$ is

$$\frac{\partial D_{pq}(W,W^*)}{\partial W} \bigg|_{W=W_{ICA}} = 0 \quad \text{or}$$

$$E(\varphi(y_{ICA},y_{ICA}^{*})x_{ICA}^{T}) = I.$$  

(9)

with $y_{ICA} = W_{ICA}x = W_{ICA}s + W_{ICA}v = \tilde{y} + W_{ICA}v$. The properties of the ICA solution are:

- $W_{ICA} = A^{-1}$ if $\sigma^2 = 0$ (no noise) and $q(s) = q^{0}(s)$.
- To compute $W_{ICA}$, we do not need to know $A$ or $s$, but the pdf $q(s) = \prod_{i=1}^{N} q_{i}(s_{i})$ is required. $q_{i}(s_{i})$ needs to be non-Gaussian or all sources must have different noncircularity indices.
- No permutation ambiguity if $q_{i}(s_{i}) \neq q_{j}(s_{j}) \forall i \neq j$.
- There is no scaling ambiguity if $q_{i}(s_{i}) = q_{i}^{0}(s_{i})$ is known $\forall i$. Only a phase ambiguity remains if $q_{i}(s_{i})$ is circular.

As shown in the appendix, the ICA solution for the general noncircular complex case can be derived approximately using a two-step perturbation analysis for low noise and is given by

$$W_{ICA} = (I + \sigma^{2}C)A^{-1} + O(\sigma^{4}).$$

(10)

The elements of $C$ can be obtained from Eq. (41) and (43). If $q(s)$ is symmetric in the real or imaginary part of $s$, they are given by Eq. (42) and (44).

For comparison, we consider the linear MMSE estimator

$$W_{MMSE} = A^{H} \left( AA^{H} + \sigma^{2}R_{v} \right)^{-1}$$

$$= \left[ 1 - \sigma^{2}R_{v}^{-1} \right] A^{-1} + O(\sigma^{4}).$$

(12)

where the last line is a first-order Taylor series expansion in $\sigma^{2}$ and $R_{v}^{-1} = A^{-1}R_{v}A^{-H}$. Comparing (12) with (10) we see that $W_{ICA}$ and $W_{MMSE}$ are similar if $C \approx -R_{v}^{-1}$.

3.2. Circular complex case

A complex random vector $s$ is circular if $s$ and $se^{j\theta}$ have the same pdf for any $\theta$ [13]. We assume now that the source signals $s$ and the noise $v$ are circular. Hence, both the noncircularity indices of the sources $\gamma$ and the pseudo-covariance matrix $R_{v}$ are zero. As a consequence, (42) and (44) simplify to

$$C_{ii} = -\frac{\kappa_{i} + \lambda_{i}}{\rho_{i} + \delta_{i}} [R_{v}]_{ii} \in \mathbb{R},$$

$$C_{ij} = \frac{\kappa_{j} - 1}{\kappa_{i}\kappa_{j} - 1} [R_{v}]_{ij} \in \mathbb{C} \quad (i \neq j).$$

(13)

3.3. Real case

For real signals and noise, we have

$$\gamma_{i} = 1, \quad R_{v} = \bar{R}_{v}.$$  

In the derivation of $W_{ICA}$ we have considered Taylor series expansions of $\varphi(y)$ using Wirtinger derivatives. The Wirtinger derivatives $\partial/\partial s$ and $\partial/\partial s^{*}$ of $\varphi(s) \in \mathbb{R}$ are now identical (see (6)) and hence

$$\tilde{\xi}_{i} = \kappa_{i}, \quad \rho_{i} = \delta_{i}, \quad \lambda_{i} = \omega_{i} = \tau_{i}.$$  

(15)

Furthermore, the Wirtinger derivatives of $\varphi(s) \in \mathbb{R}$ are identical to the real derivatives except for a factor of $\frac{1}{2}$ (see (6)). Hence it holds

$$\kappa_{i} = \frac{\tilde{\kappa}_{i}}{2}, \quad \rho_{i} = \frac{\tilde{\rho}_{i}}{2}, \quad \lambda_{i} = \frac{\tilde{\lambda}_{i}}{4}.$$  

(16)
where $\hat{c}_i$, $\hat{\phi}_i$ and $\hat{\lambda}_i$ are defined using real derivatives of $\varphi(s)$. Using (14)-(16), we get from (42) and (44)
\begin{align*}
C_{ii} &= -\frac{\hat{c}_i + \frac{1}{2}\hat{\lambda}_i}{1 + \hat{\phi}_i} \mathbf{R}_{-1},
C_{ij} &= \frac{\hat{c}_j}{\hat{c}_i \hat{\lambda}_j - 1} \mathbf{R}_{-1} \quad (i \neq j),
\end{align*}
which corresponds to the results in [9].

4. RESULTS FOR COMPLEX GGD

Since the real case has already been studied in [9], we focus on the complex case here. Due to space limitations, we study KLD-ICA only for $N = 3$ sources with spatially white Gaussian noise with $E[|\nu|^2] = \sigma_\nu^2$ and the square mixing matrix $A = [a_{mn}]$, where $a_{mn} = e^{-j \pi m \sin \theta_n}$ and $\theta_n = -60^\circ, 0^\circ, 60^\circ$. [15] proposed the signal-to-interference-plus-noise ratio (SINR) to evaluate separation performance of noisy source separation. For spatially uncorrelated noise, SINR for a given demixing matrix $W$ is calculated as
\begin{equation}
\text{SINR} = \frac{1}{N} \sum_{i=1}^{N} \frac{|WA|_{ii}^2}{\sum_{j \neq i} |WA|_{ij}^2 + \sigma^2 \sum_j |W|_{ij}^2}.
\end{equation}

The term $|WA|_{ii}^2$ reflects the power of the desired source $i$ in the demixed signal $y_i$. $\sum_{j \neq i} |WA|_{ij}^2$ corresponds to the power of the interfering signals $j \neq i$ in the demixed signal $y_i$. $\sigma^2 \sum_j |W|_{ij}^2$ is the noise power in the demixed signal $y_i$.

The signal-to-noise ratio is defined as $1/\sigma^2$, i.e. it denotes the SNR before mixing. It is known that among all linear demixing matrices $W$, $W_{\text{MMSE}}$ from (11) is the one which maximizes the SINR. We compare the SINR of the theoretical ICA solution $W_{\text{ICA}}$ given in (10), the estimate $W_{\text{ICA}}$ obtained by running KLD-based ICA using $L$ samples and the MMSE solution $W_{\text{MMSE}}$ from (11). The ICA algorithm is initialized with $W = I$ and performs gradient descent using the relative gradient [1], i.e. postmultiplies the gradient of KLD (8) by $W^H W$. We normalize each row of the relative gradient, resulting in an adaptive step size for each source. In the derivation of the theoretical solution $W_{\text{ICA}}$, we evaluated all expectations exactly. Hence $W_{\text{ICA}}$ only accounts for the bias from $A^{-1}$ but not for estimation variance. $W_{\text{ICA}}$ contains both factors.

In the following, all sources are GGD with the same shape parameter $c_1 = c$. The pdf of a noncircular complex GGD with zero mean and $E[|\nu|^2] = 1$ can be written as [16] $\varphi(s, s^*) = \frac{c_\nu}{\pi \Gamma(1/c) \Gamma(1-\gamma/c)} \exp \left(-\frac{\gamma/c}{\pi \Gamma(1/c) \Gamma(1-\gamma/c)} \right)$,
where $\gamma = E[|\nu|^2]$ and $\alpha = \Gamma(2/c)/(\Gamma(1/c))$. A change of the shape parameter $c > 0$ varies the form of the pdf from super-Gaussian ($c < 1$) to sub-Gaussian ($c > 1$). For $c = 1$, the pdf is Gaussian. By integration in polar coordinates, it can be shown that
\begin{align*}
\kappa &= E[\theta(s)] = \int \frac{\partial \varphi}{\partial s} \varphi^0(s) ds = \frac{c_\nu^2 \Gamma(2/c)}{(1 - \gamma/c)^2 (1/c)},
\delta &= E[\theta(s) s] = \frac{\partial \varphi}{\partial s} s \varphi^0(s) ds = \frac{2c_\nu^2 (1 - c) \gamma^2}{(1 - \gamma/c)^2},
\rho &= E[\varphi(s) s] = \frac{\partial \varphi}{\partial s} s \varphi^0(s) ds = -\frac{2 - 2s^2 + c(3\gamma^2 - 2)}{2(1 - \gamma/c)^2},
\xi &= E[\varphi(s)] = \int \frac{\partial \varphi}{\partial s^*} \varphi^0(s) ds = -\gamma/c,
\lambda &= E[\varphi(s)] = \int \frac{\partial^2 \varphi}{\partial s \partial s^*} \varphi^0(s) ds = (c - 1) \kappa/c,
\end{align*}
where $\varphi(s)$ is a complex GGD and $\varphi^0(s)$ is the pdf of the noncircular complex GGD with zero mean and $E[|\nu|^2] = 1$.

\begin{equation}
\omega = E[|\nu|^2] = \int \frac{\partial^2 \varphi}{\partial s^2} s \varphi^0(s) ds = -\frac{3}{2} (c - 1) \gamma/c,
\end{equation}
\begin{equation}
\tau = E[|\nu|^2] = \int \frac{\partial^2 \varphi}{\partial s^2} s \varphi^0(s) ds = -\frac{1}{2} (c - 1) \gamma/c.
\end{equation}

4.1. Circular complex case

For a circular complex GGD, $\gamma = 0$ and hence we get $\kappa = \frac{c_\nu^2 \Gamma(2/c)}{\pi \Gamma(1/c)^2}$, $\rho = \frac{2c_\nu^2}{2(1 - \gamma/c)^2}$. For $c = 1$, $\rho = 0$. Hence, $E[|\nu|^2]$ remains constant except for this region, the SINR decreases in the vicinity of $c = 1$. Note that $\kappa$ becomes large if sources are noncircular.

\begin{equation}
\omega = E[|\nu|^2] = \int \frac{\partial^2 \varphi}{\partial s^2} s \varphi^0(s) ds = (c - 1) \kappa/c,
\end{equation}
\begin{equation}
\tau = E[|\nu|^2] = \int \frac{\partial^2 \varphi}{\partial s^2} s \varphi^0(s) ds = (c - 1) \kappa/c.
\end{equation}

4.2. Noncircular complex case

First, we study the performance with circular noise, i.e. $R_\nu = I$ and $R_\nu = 0$, and SINR of 10 dB. The SINR of the MMSE solution $W_{\text{MMSE}}$ is 12.4 dB. Fig. 2 shows that for a wide range of the shape parameter $c$ and the noncircularity index $\gamma$, the theoretical ICA solution $W_{\text{ICA}}$ achieves an SINR close to that of MMSE. Comparing Fig. 2 (a) and (b), we note that the contour plot for the simulation using $L = 10^4$ samples differs from the contour plot for the theoretical ICA solution. The reason is that for noncircular sources with the same noncircularity index $\gamma = \gamma$, the estimation variance increases for $c \rightarrow 1$. Hence, in the simulation the SINR decreases in the vicinity of $c = 1$. However, Fig. 2 (b) shows that even with a limited sample size KLD-ICA can still achieve SINR performance close to that of MMSE except for $c \approx 1$.

Now, we consider the case where sources are noncircular complex with $\gamma_1 = 0.5, \gamma_{2,3} = 0.5 \pm \Delta \gamma$ and the noise $v$ is noncircular with $R_v = I$ and $R_v = 0.5 \cdot I$, i.e. $\gamma_v = 0.5$. Fig. 3 shows decreasing SINR values for $c \rightarrow 1$ and $\Delta \gamma \rightarrow 0$ since in that region $|R_{\text{ICA}}|$ (44) becomes large if sources or noise are noncircular. However, except for this region, the SINR of the theoretical ICA solution (Fig. 3 (a) is still close to that of MMSE. The form of the contour plot for the simulation (Fig. 3 (b) is similar to that of the theoretical solution but shows slightly lower SINR performance especially for $c \approx 1$ and small $\Delta \gamma$. This is again due to increasing estimation variance for $c \rightarrow 1$ and small $\Delta \gamma$. Nevertheless, the performance obtainable in simulations can still be considered good as long as $c$ is not close to 1 or $\Delta \gamma$ is sufficiently large.
In summary, the results in Sect. 4.1 and 4.2 have shown that
• in many cases the theoretical solution \( W_{\text{ICA}} \) of KLD-ICA can achieve an SINR close to the optimum attainable by any linear demixing matrix \( W \) which is given by \( W = W_{\text{MMSE}} \)
• for sources following a GGD, \( W_{\text{ICA}} \) obtained by running KLD-ICA with a finite amount of samples \( L \) can achieve an SINR close to that of \( W_{\text{MMSE}} \) except for (nearly) Gaussian sources with similar noncircularity indices.

Although we assumed that we perfectly know the distributions of the sources, other approaches such as ICA-EBM [17] exist which do not require such knowledge. Simulation results using ICA-EBM show similar SINR performance as KLD-ICA.

5. CONCLUSION AND FUTURE WORK

We have derived an analytic expression for the demixing matrix of KLD-based ICA for the low noise regime. In this paper, we have considered the general noncircular complex determined case. The solution for the circular complex and real case can be derived as special cases. Although the KLD and MMSE solutions differ, linear demixing based on these two criteria yields demixed signals with similar SINR in many cases. In practice, however, not only the bias studied in this paper but also the variance of the estimate are important for SINR.

APPENDIX A. DERIVATION OF (10)

Here we derive an analytic expression for \( W_{\text{ICA}} \) in the presence of noise by using a perturbation analysis. Motivated by \( W_{\text{ICA}} = A^{-1} + \sigma^2 B + O(\sigma^4) \), we assume that \( W_{\text{ICA}} \) can be written as \( W_{\text{ICA}} = A^{-1} + \sigma^2 B + O(\sigma^4) \) and derive \( B \) by a two-step perturbation analysis:

1. Taylor series approximation of \( E(\varphi(y)y^T) \) in (9) at \( y = \hat{y} = W_{\text{ICA}}A s \).
2. Taylor series approximation of the result of the above step by exploiting \( W_{\text{ICA}} = A^{-1} + \sigma^2 B + O(\sigma^4) \) and \( \hat{y} = s + \sigma^2 B A s + O(\sigma^4) = s + \sigma^2 B A s + O(\sigma^4) \) with \( C = B A \) and \( b = C s = [b_1, \ldots, b_N]^T \).

In this way, we determine explicitly the deviation \( \sigma^2 B \) of \( W_{\text{ICA}} \) from the inverse solution \( A^{-1} \).

The general Taylor series expansion of \( \varphi(y) \equiv \varphi(y, y^*) \) is given in (26) on the next page with \( \eta(y, y^*) = \frac{\partial \varphi}{\partial y} \), \( \theta(y, y^*) = \frac{\partial \varphi}{\partial y^*} \), \( \nu(y, y^*) = \frac{\partial^2 \varphi}{\partial y \partial y^*} \), \( \xi(y, y^*) = \frac{\partial^2 \varphi}{\partial y^2} \) and \( \epsilon(y, y^*) = \frac{\partial^4 \varphi}{\partial y^4} \). To simplify notation, we will drop the dependence of \( \varphi(y), \eta(y), \theta(y), \nu(y), \xi(y), \epsilon(y) \) on \( y^* \) and keep only the dependence on \( y \) in the following.

For the first Taylor series, we expand \( \varphi(y) \) at \( y = \hat{y} = W_{\text{ICA}}A s \). By using a vectorized form of (26) with \( \Delta y = W_{\text{ICA}}A s \) we get (27) on the next page. The vectors \( \eta, \theta, \nu, \xi, \epsilon \) contain the elements of \( \eta_i = \frac{\partial \varphi_i}{\partial y_i} \), \( \theta_i = \frac{\partial \varphi_i}{\partial y^*_i} \), \( \nu_i = \frac{\partial^2 \varphi_i}{\partial y_i \partial y^*_i} \), \( \xi_i = \frac{\partial^2 \varphi_i}{\partial y^*_i} \), \( \epsilon_i = \frac{\partial^4 \varphi_i}{\partial y^4} \), respectively. \( \Delta y \) is a diagonal matrix whose diagonal elements are those of the vector \( y \). \( \text{Diag}(Z) \) sets all off-diagonal elements of the matrix \( Z \) to zero. Multiplying (27) by \( y^* \) and taking the expectation, we get (28) from (9).
\[ \varphi(y, y^*) = \varphi(\hat{y}, \hat{y}^*) + \frac{\partial \varphi}{\partial y} \Delta y + \frac{\partial^2 \varphi}{\partial y^2} \Delta y^* + \frac{1}{2} \left( \frac{\partial^2 \varphi}{\partial y^2 T} (\Delta y)^2 + \frac{\partial^2 \varphi}{\partial y^2 T} (\Delta y^*)^2 \right) + \frac{\partial \varphi}{\partial y} \Delta y 
\] 
\[ \rho \cdot C \]

\[ \text{Since } ij = y, y \in \mathbb{S}, \text{i.e. } (39) \]

\[ \text{we then get } \kappa E \]

\[ \| \mathbf{E} \| \]

\[ \text{and } (40) \]

\[ \text{For the remaining terms in (28), we use the following properties: } \]

\[ \text{For } i \neq j, E [\nu(y_i) y_i], E [\xi_i y_i] \text{ and } E [\epsilon_i y_i] \text{ are all } O(\sigma^2) \text{ since } y_i = s_i + O(\sigma^2) \text{ and } s_i \text{ and } y_i \text{ are independent. } \]

\[ \text{With the same reasoning, we get } E [\eta(y_i)], E [\theta(y_i)], E [\nu_i y_i], E [\epsilon_i y_i] = \omega_i + O(\sigma^2), E [\xi_i y_i] = \tau_i + O(\sigma^2) \text{ and } E [\epsilon_i y_i] = \lambda_i + O(\sigma^2). \]

\[ \text{As defined in Sec. 2, } E[Y^H] = \sigma^2 \mathbf{R}_s \text{ and } E[\mathbf{v}^T] = \sigma^2 \mathbf{R}_v. \text{ Using these properties and (37), we get (29) from (28). } \]

\[ \text{Since } \mathbf{W} = \mathbf{A}^{-1} + O(\sigma^2), \text{ we defined the transformed noise covariance matrix } \mathbf{R}_{-1} = \mathbf{W}_r \mathbf{W}^H = \mathbf{A}^{-1} \mathbf{R}_s \mathbf{A}^{-T} + O(\sigma^2). \text{ Note that } \mathbf{R}_{-1} = \mathbf{R}_{-1} \text{ and } \mathbf{R}_{-1}^T = \mathbf{R}_{-1}. \text{ For the diagonal elements } C_{ii} \text{ we get from (29) } \]

\[ \rho_i C_{ii} + \delta_i C_{ii} + C_{ii} = -(\kappa_i + \lambda_i) [\mathbf{R}_{-1}]_{ii} \]

\[ \text{If } q(s, s^*) \text{ is symmetric in the real part } \mathbf{R} \text{ or imaginary part } \mathbf{R}^* \text{ of } s \text{, i.e. } q(\mathbf{R}, \mathbf{R}^*) = q(\mathbf{R}^*, \mathbf{R}), \text{ or } q(\mathbf{R}, -\mathbf{R}^*) = q(\mathbf{R}^*, -\mathbf{R}), \text{ the parameters } \kappa_i, \rho_i, \delta_i, \lambda_i, \xi_i, s_i \text{, and } \tau_i \text{ are real. For } p_i + 1 \neq 0, \text{ we then get } \]

\[ \text{For the o-diagonal elements } C_{ij} \text{ we get from (29) } \]

\[ \gamma_{ij} C_{ij} + \kappa_i C_{ij} + C_{ij} = -(\kappa_i + \lambda_i) [\mathbf{R}_{-1}]_{ij} \]

\[ \text{If } q(s, s^*) \text{ is symmetric in the real or imaginary part of } s \text{, and if } (\gamma_{ij} + \kappa_i \gamma_{ij} + \kappa_j \gamma_{ij}) \neq 1 \text{ and } (\gamma_{ij} \gamma_{ij} + \kappa_i \gamma_{ij} - \kappa_j \gamma_{ij}) \neq 1, \text{ we obtain the real and imaginary part of } C_{ij} \text{ from (43) as } \]

\[ \Re C_{ij} = - \frac{\kappa_j}{\rho_i + 1 - \delta_i} \left( \frac{\xi_i}{\rho_i + 1 - \delta_i} \right) \]

\[ \Im C_{ij} = - \frac{\kappa_j}{\rho_i + 1 - \delta_i} \left( \frac{\xi_i}{\rho_i + 1 - \delta_i} \right) \]

\[ \text{REFERENCES} \]


