Probabilistic Formulation of Independent Vector Analysis Using Complex Gaussian Scale Mixtures

Jason A. Palmer[1,⋆], Ken Kreutz-Delgado[2], and Scott Makeig[1]

[1] Swartz Center for Computational Neuroscience
[2] Department of Electrical and Computer Engineering
University of California San Diego, La Jolla, CA 92093
{jason,scott}@sccn.ucsd.edu, kreutz@ece.ucsd.edu

Abstract. We propose a probabilistic model for the Independent Vector Analysis approach to blind deconvolution and derive an asymptotic Newton method to estimate the model by Maximum Likelihood.

1 Introduction

In this paper we propose a probabilistic model for the complex STFT coefficient vectors, which can be used in Independent Vector Analysis (IVA) [7] to perform blind deconvolution. We use adaptive source densities [11] that are dependent across frequencies [6,12] via complex Gaussian scale mixtures (CGSMs). These densities are shown to be the natural result of the sampling of Fourier coefficients of intermittent sources, in which a source is active only over a random fraction of each time window used to sample the DFT. We derive a Newton method for Maximum Likelihood estimation [2]. The IVA model was developed by Kim, Eltoft, Lee, and others [8,6,7].

Notation: We denote the imaginary number by $\jmath \triangleq \sqrt{-1}$.

2 Densities over Complex Vectors and STFT Coefficients

A probability density defined on a complex vector space is simply a joint density formulated over the real and imaginary parts of the vector. In particular, the probability density function is real valued. Integration over $\mathbf{C}^n$ is equivalent to integration over $\mathbf{R}^{2n}$. However, some real valued functions of complex vectors can be expressed in a simpler form in terms of the complex vectors $\mathbf{z} = \mathbf{x} + \jmath\,\mathbf{y}$ and $\mathbf{z}^* = \mathbf{x} - \jmath\,\mathbf{y}$, rather than the real and imaginary parts $\mathbf{x}$ and $\mathbf{y}$. And optimization of the real valued function may be more conveniently carried out in the complex space $\mathbf{C}^n$ using the Wirtinger calculus, instead of working in $\mathbf{R}^{2n}$ [9].

Fourier coefficients of a stationary time series are complex valued random variables. Stationarity implies that the covariance matrix of the real and imaginary parts has a particular form. Let $\mathbf{x}$ be an n dimensional complex valued random vector. Then if the real-valued random vectors $\mathbf{x}_R$, and $\mathbf{x}_I$ are jointly distributed as,

$$\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{c}_R \\ \mathbf{c}_I \end{bmatrix}, \tfrac{1}{2} \begin{bmatrix} \boldsymbol{\Sigma}_R & -\boldsymbol{\Sigma}_I \\ \boldsymbol{\Sigma}_I & \boldsymbol{\Sigma}_R \end{bmatrix} \right)$$

⋆ This research was partially supported by NSF grants ISS-0613595 and CCF-0830612.

T. Adali et al. (Eds.): ICA 2009, LNCS 5441, pp. 90–97, 2009.

for some complex vector $\mathbf{c}$ and Hermitian positive definite matrix $\boldsymbol{\Sigma}$, then we say $\mathbf{x}$ is *complex multivariate Normal* distributed, and write $\mathbf{x} \sim \mathcal{N}_C(\mathbf{c}, \boldsymbol{\Sigma})$.

The real domain probability density over $(\mathbf{x}_R, \mathbf{x}_I)$ can be written in terms of complex quantities $p(\mathbf{x}) \triangleq p(\mathbf{x}_R, \mathbf{x}_I)$ as,

$$p(\mathbf{x}) = \pi^{-n}(\det \boldsymbol{\Sigma})^{-1} \exp\big(-(\mathbf{x}-\mathbf{c})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{c})\big)$$

In particular, the univariate complex Normal distribution is given by

$$\mathcal{N}_C(x; \mu, \sigma^2) = \pi^{-1}\sigma^{-2}\exp(-\sigma^{-2}|x-\mu|^2)$$

where x_R and x_I are independent $\mathcal{N}(\mu_R, \sigma^2/2)$ and $\mathcal{N}(\mu_I, \sigma^2/2)$ respectively.

Using the central limit theorem, it is shown that for stationary time series $y_i(t)$, the distribution of the Fourier coefficients $y_i(\omega_k)$, $k = 1, \ldots, N$, are independent complex multivariate Gaussian [4]. Specifically we have the following [4, Thm. 4.4.1].

Theorem 1. *Let $\mathbf{x}(t)$ be a strictly stationary real vector time series with absolutely summable cumulant spectra of all orders. Suppose $2\omega_j$, $\omega_j \pm \omega_k \neq 0\,(mod\, 2\pi)$ for $1 \leq j \leq k \leq N$. Let,*

$$\mathbf{d}^{(T)}(\omega) = \frac{1}{T}\sum_{\tau=0}^{T-1} \mathbf{x}(\tau)\exp(-\jmath\omega\tau)$$

Then the random vectors $\mathbf{d}^{(T)}(\omega_k)$, $k = 1, \ldots, N$, are asymptotically independent complex Normal random vectors, $\mathcal{N}_C(\mathbf{0}, \boldsymbol{\Sigma}(\omega_k))$. Also if $\omega = 0\,(mod\, 2\pi)$ then $\mathbf{d}^{(T)}(\omega)$ is asymptotically $\mathcal{N}(\mathbf{c}, \boldsymbol{\Sigma}(0))$, independent of the previous estimates, and if $\omega = \pi\,(\bmod 2\pi)$, then $\mathbf{d}^{(T)}(\omega)$ is asymptotically $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(0))$, independent of the previous estimates.

The density of linear transformations can also be expressed simply in terms of complex quantities. If $\mathbf{b} \sim p_{\mathbf{b}}(\mathbf{b})$, then the density of $\mathbf{x} = \mathbf{A}\mathbf{b}$ is, with $\mathbf{W} \triangleq \mathbf{A}^{-1}$,

$$p_{\mathbf{x}}(\mathbf{x}) = \det\big(\mathbf{W}\mathbf{W}^H\big)\, p_{\mathbf{b}}(\mathbf{W}\mathbf{x})$$

3 Blind Deconvolution and Complex Gaussian Scale Mixtures

In the noiseless blind deconvolution problem [5], we have,

$$\mathbf{x}(t) = \sum_{\tau=0}^{\infty} \mathbf{A}(\tau)\mathbf{s}(t-\tau)$$

where the time series $s_i(t)$, $i = 1, \ldots, n$ are mutually independent. Using the discrete Fourier transform, this is expressed in the frequency domain as,

$$\mathbf{x}(\omega) = \mathbf{A}(\omega)\mathbf{s}(\omega)$$

where $\mathbf{x}(\omega) \triangleq \sum_\tau \mathbf{x}(\tau)\exp(-\jmath\omega\tau)$, etc. We consider the asymptotic distribution of the short time Fourier transform (STFT) coefficients, calculated by,

$$\mathbf{x}(\omega_k, t) \triangleq \mathbf{x}_{:kt} \triangleq \sum_{\tau=0}^{T-1} \mathbf{x}(t-\tau)\exp(-\jmath\,\omega_k\tau)$$

Since the sources $s_i(t)$ are independent, we have that $s_i(\omega_1)$ and $s_j(\omega_2)$ are independent for $i \neq j$, and any ω_1, ω_2. Thus the joint distribution over the sources is a product,

$$p\big(\mathbf{s}(\omega_1), \ldots, \mathbf{s}(\omega_N)\big) = \prod_{i=1}^{n} p\big(s_i(\omega_1), \ldots, s_i(\omega_N)\big)$$

By Theorem 1, for stationary sources, the joint distribution over the frequencies is the product of n univariate complex Gaussians,

$$p\big(s_i(\omega_1), \ldots, s_i(\omega_N)\big) = \prod_{k=1}^{N} \mathcal{N}_C\big(s_i(\omega_k); 0, \sigma_k^2\big)$$

where $\mathcal{N}_C(\mu, \sigma^2)$ is the univariate complex Normal distribution, which is a radially symmetric distribution over the real and imaginary parts of $s_i(\omega_k)$, with variance σ_k^2, which gives the spectral power at frequency ω_k. Theorem 1 holds for stationary time series, but most long term observations of sensor data are non-stationary.

A more realistic assumption is that the time series is stationary over contiguous blocks, switching at discrete (random) times among a set of stationary regimes. For each individual stationary regime, as the window size, and/or sampling rate tend to infinity, the Fourier coefficients will tend to complex Gaussians. In a switching time series, a window may contain two or more stationary regimes. If we let the sampling rate tend to infinity, the individual stationary segments in the window will still be asymptotically complex Normal, and the overall distribution of the coefficients in the window will be a convex sum of the two (or more) complex Normal random variables.

Since each window will contain a random size segment of a given stationary state, the distributions of the Fourier coefficients for each state will be complex Gaussian scale mixtures, with the Fourier coefficients at different frequencies being uncorrelated but dependent. The joint distribution is given by,

$$p\big(s_i(\omega_1), \ldots, s_i(\omega_N)\big) = \int_0^\infty \prod_{k=1}^{N} \mathcal{N}_C\big(s_i(\omega_k); 0, \xi\,\sigma_{i,k}^2\big) f(\xi)\, d\xi$$

The distribution of $\mathbf{s}_i$ is equivalent to the product of a non-negative mixing random variable with a complex Normal random vector, $\mathbf{s}_i = \xi^{1/2}\mathbf{z}_i$, where $\xi \sim p(\xi)$ and $\mathbf{z}_i \sim \mathcal{N}_C(0, \operatorname{diag}(\boldsymbol{\sigma}_i))$, and can be expressed in a form similar to that of real valued Gaussian scale mixtures, where the squared magnitude of the complex variable is substituted for the square of the real variable [12].

Knowing the distribution of the Fourier coefficients of the independent sources, we can calculate the distribution of the observation Fourier coefficients as,

$$p\big(\mathbf{x}(\omega_1), \ldots, \mathbf{x}(\omega_N)\big) = \bigg(\prod_{k=1}^{N} \det \mathbf{W}(\omega_k)\mathbf{W}(\omega_k)^H\bigg) \prod_{i=1}^{n} p_i\big(b_i(\omega_1), \ldots, b_i(\omega_N)\big)$$

The parameters of the unmixing system $\mathbf{W}(\omega)$, as well as the CGSM density models, for each source $i = 1, \ldots, n$ are adapted by Quasi Maximum Likelihood.

3.1 Generalized Inverse Gaussian and Generalized Hyperbolic

The Generalized Inverse Gaussian (GIG) density [3,6,12] is a convenient mixing density. It can be thought of as a combination and generalization of the gamma and inverse gamma distributions. The GIG density has the form,

$$\mathcal{N}^{\dagger}\big(\xi\,;\lambda,\,\tau,\upsilon\big) = \frac{\upsilon^{\lambda}}{2\,K_{\lambda}(\tau)}\,\xi^{\lambda-1}\exp\big(-\tfrac{1}{2}\tau\big((\upsilon\xi)^{-1}+\upsilon\xi\big)\big) \tag{1}$$

for $\xi > 0$, where K_{λ} is the Bessel K function, or modified Bessel function of the second kind. The moments of the Generalized Inverse Gaussian are easily found by direct integration, using the fact that (1) integrates to one,

$$E\{\xi^{a}\} = \upsilon^{a}\frac{K_{\lambda+a}(\tau)}{K_{\lambda}(\tau)} \tag{2}$$

Since υ is a scale parameter and we we will estimate the scale of the complex Normal directly, we will eliminate this redundancy by setting $\upsilon = 1$ in the mixing density.

The generalized hyperbolic distribution [3] is the complex Gaussian scale mixture arising from the GIG mixing density. We write the generalized hyperbolic density in the form,

$$\mathcal{GH}_C(\mathbf{b}_i;\lambda,\tau) = \frac{\tau^{n-\lambda}}{\pi^{n}K_{\lambda}(\tau)}\,\tau(\mathbf{b}_i)^{\lambda-n}K_{\lambda-n}\big(\tau(\mathbf{b}_i)\big) \tag{3}$$

where $\tau(\mathbf{b}_i) \triangleq \sqrt{\tau^2 + 2\tau\|\mathbf{b}_i\|^2}$.

The GIG density is conjugate to the Normal density, so the posterior density of ξ given $\mathbf{b}_i$ is also a GIG density. In particular the parameters of the posterior density of ξ are,

$$\lambda' = \lambda - n, \quad \tau' = \tau(\mathbf{b}_i), \quad \upsilon' = \tau/\tau(\mathbf{b}_i)$$

We can thus find the posterior expectation required for the EM algorithm using (2),

$$E\{\xi_i^{-1}\,|\,\mathbf{b}_i\} = \frac{\tau(\mathbf{b}_i)}{\tau}\,\frac{K_{\lambda-n-1}(\tau(\mathbf{b}_i))}{K_{\lambda-n}(\tau(\mathbf{b}_i))} \tag{4}$$

4 Probabilistic IVA and Maximum Likelihood

We shall assume for simplicity of presentation and implementation that $\omega_k \neq 0(\mathrm{mod}\,\pi)$ for $k = 1, \ldots, N$, so that all the STFT coefficients are all complex valued. Thus for each observation $\mathbf{x}_{::t}$ of STFT coefficients of the data, there is a set of random vectors of independent, zero mean sources $\mathbf{b}_{:kt}$, such that,

$$\mathbf{x}_{:kt} = \mathbf{A}_k\mathbf{b}_{:kt}\,, \quad k = 1, \ldots, N$$

The density of the observations $p(\mathbf{x}_{::t}) \triangleq p(\mathbf{x}_{:0t}, \mathbf{x}_{:1t} \ldots, \mathbf{x}_{:Nt})$ is given by,

$$p(\mathbf{x}_{::t}) = \left(\prod_{k=1}^{N} \det \mathbf{W}_k \mathbf{W}_k^H \right) \prod_{i=1}^{n} q_i\big(\mathbf{b}_{i:t}\big)$$

where $\mathbf{W}_k \triangleq \mathbf{A}_k^{-1}$, and $\mathbf{b}_{:kt} = \mathbf{W}_k \mathbf{x}_{:kt}$, $k = 1, \ldots, N$. The sources are modeled as mixtures of Gaussian Scale Mixtures, extending the real model in [11] to the complex dependent case. We define,

$$y_{ijkt} \triangleq \beta_{ijk}^{1/2} b_{ikt}$$

The source density models are defined as follows,

$$q_i\big(\mathbf{b}_{i:t}\big) = \sum_{j=1}^{m} \alpha_{ij} \left(\textstyle\prod_{k=1}^{N} \beta_{ijk} \right) q_{ij}\big(\mathbf{y}_{ij:t}\,; \tau_{ij}\big)$$

Each q_{ij} is a CGSM parameterized by ν_{ij}.

Thus the density of the observations $\mathbf{x}_{:::} \triangleq \{\mathbf{x}_{::t}\}$, $t = 1, \ldots, T$, is given by $p(\mathbf{x}_{:::}; \Theta) = \prod_{t=1}^{T} p(\mathbf{x}_{::t}; \Theta)$, and the parameters to be estimated are

$$\Theta = \big\{ \mathbf{W}_k, \alpha_{ij}, \beta_{ijk}, \tau_{ij} \big\}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m$, and $k = 1, \ldots, N$.

4.1 Maximum Likelihood

Given the STFT data $\mathbf{X} = \{\mathbf{x}_{:kt}\}$, $k = 1, \ldots, N$, $t = 1, \ldots, T$, we consider the ML estimate of $\mathbf{W}_k = \mathbf{A}_k^{-1}$, $k = 1, \ldots, N$. We employ the EM algorithm to allow the source density at each frequency to be a mixture of (multivariate intra-dependent) CGSMs. Thus we define the random index j_{it} ranging over $\{1, \ldots, m\}$ with probabilities $\alpha_{ij}, j = 1, \ldots, m$, and we let $z_{ijt} = 1$ if $j_{it} = j$, and 0 otherwise. For the joint density of $\mathbf{X}$ and $\mathbf{Z}$, we have,

$$p(\mathbf{X}, \mathbf{Z}) = \prod_{t=1}^{T} \Big(\prod_{k=1}^{N} \det \mathbf{W}_k \mathbf{W}_k^H \Big) \prod_{i=1}^{n} \prod_{j=1}^{m} \alpha_{ij}^{z_{ijt}} \tilde{\beta}_{ij}^{z_{ijt}} q_{ij}\big(\mathbf{y}_{ij:t}\big)^{z_{ijt}}$$

where $q_{ij}(\mathbf{y}_{ij:t})$ is the jth mixture component of the dependent multivariate density model for the STFT coefficients of ith source, $\mathbf{b}_{i:t}$ is the vector of STFT coefficients for source i at time t, and $\tilde{\beta}_{ij} \triangleq \prod_{k=1}^{N} \beta_{ijk}$. We define,

$$f_{ij}(\|\mathbf{y}_{ij:t}\|) \triangleq g_{ij}(\|\mathbf{y}_{ij:t}\|^2) \triangleq -\log q_{ij}(\mathbf{y}_{ij:t})$$

For the log likelihood of the data then (scaled by $1/T$), which is to be maximized with respect to the parameters, we have,

$$L(\mathbf{W}) = \left(\sum_{k=1}^{N} \log\det \mathbf{W}_k \mathbf{W}_k^H \right) - \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{z}_{ijt}\, g_{ij}\big(\textstyle\sum_{\ell=1}^{N} \beta_{ij\ell} |b_{i\ell t}|^2 \big)$$

Note that $|b_{i\ell t}|^2 = (\mathbf{w}_{i\ell}\mathbf{x}_{:\ell t})(\mathbf{w}_{i\ell}\mathbf{x}_{:\ell t})^* = (\mathbf{w}_{i\ell}\mathbf{x}_{:\ell t})(\mathbf{w}_{i\ell}^*\mathbf{x}_{:\ell t}^*)$. The complex gradient, or Wirtinger conjugate derivative of L is thus,

$$\frac{\partial}{\partial \mathbf{W}_k^*} L(\mathbf{W}) \triangleq \mathbf{G}_k = \mathbf{W}_k^{-H} + \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varphi}_{:kt}\mathbf{x}_{:kt}^H \tag{5}$$

for $k = 1, \ldots, N$, where,

$$\varphi_{ikt} \triangleq -\, b_{ikt}\gamma_{ikt}, \quad \gamma_{ikt} \triangleq \sum_{j=1}^{m} \hat{z}_{ijt}\, \beta_{ijk}\, g_{ij}'\big(\textstyle\sum_{\ell=1}^{N}\beta_{ij\ell}|b_{i\ell t}|^2\big) \tag{6}$$

If we multiply (5) by $\mathbf{W}_k^H\mathbf{W}_k$ on the right, we get,

$$\Delta\mathbf{W}_k \propto \left(\mathbf{I} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varphi}_{:kt}\mathbf{b}_{:kt}^H\right)\mathbf{W}_k \tag{7}$$

This linear transformation of the complex gradient is a positive definite, and thus a valid descent direction. The direction (7) is known as the “natural gradient” [1].

4.2 Hessian and Newton Method

If we expand L in a second order Taylor series, $L(\mathbf{W} + d\mathbf{W})$, and solve for the maximizing $d\mathbf{W}$, we arrive at the equation,

$$\mathcal{H}_{11}(d\mathbf{W}_k^*) + \mathcal{H}_{12}(d\mathbf{W}_k) + \sum_{\ell\neq k}\mathcal{H}_{11}'(d\mathbf{W}_\ell^*) + \mathcal{H}_{12}'(d\mathbf{W}_\ell) = -\mathbf{G}_k \tag{8}$$

for $k = 1, \ldots, N$, where $\mathcal{H}_{11}$ denotes two consecutive conjugate derivatives of $\mathbf{W}_k$, $\mathcal{H}_{12}$ denotes a conjugate derivative of $\mathbf{W}_k$ followed by a non-conjugate derivative of $\mathbf{W}_k$, etc. The $\mathcal{H}'$ terms are derivatives with respect to $\mathbf{W}_k$ followed by derivatives with respect to $\mathbf{W}_\ell$ where $\ell \neq k$.

Now, we have,

$$\frac{\partial\varphi_{ikt}}{\partial[\mathbf{W}_k]_{rs}} = -\delta_{ir}\left(\gamma_{ikt} + \psi_{ikt}\right) x_{skt}$$

where δ_{ir} is the Kronecker delta symbol, and,

$$\psi_{ikt} \triangleq |b_{ikt}|^2 \sum_{j=1}^{m} \hat{z}_{ijt}\, \beta_{ijk}^2\, g_{ij}''\big(\textstyle\sum_{\ell=1}^{N}\beta_{ij\ell}|b_{i\ell t}|^2\big)$$

Similarly, we have,

$$\frac{\partial\varphi_{ikt}}{\partial[\mathbf{W}_k]_{rs}^*} = -\delta_{ir}\,\psi_{ikt}\,\big(b_{ikt}/|b_{ikt}|\big)^2\, x_{skt}^*$$

Now let $\mathbf{B}_k = d\mathbf{W}_k$. Then we have,

$$\begin{aligned}\mathcal{H}_{11}(\mathbf{B}_k) + \mathcal{H}_{12}(\mathbf{B}_k^*) \;=\; & -\mathbf{W}_k^{-H}\mathbf{B}_k^H\mathbf{W}_k^{-H} \\ & - \Big\langle \mathrm{diag}(\boldsymbol{\zeta}_{:kt})\mathbf{B}_k^*\mathbf{x}_{:kt}^*\mathbf{x}_{:kt}^H\Big\rangle_T - \Big\langle \mathrm{diag}(\boldsymbol{\gamma}_{:kt} + \boldsymbol{\psi}_{:kt})\mathbf{B}_k\mathbf{x}_{:kt}\mathbf{x}_{:kt}^H\Big\rangle_T\end{aligned}$$

where $\langle \cdot \rangle_T$ denotes the empirical average $\frac{1}{T}\sum \cdot$, and,

$$\zeta_{ikt} \triangleq \psi_{ikt} \left(b_{ikt}/|b_{ikt}|\right)^2$$

The terms involving $\mathcal{H}'_{11}$, etc., tend to zero since the STFT coefficients at different frequencies are orthogonal (though nevertheless dependent) by definition.

Thus (8) reduces to,

$$\mathbf{G}_k = \mathbf{W}_k^{-H}\mathbf{B}_k^H\mathbf{W}_k^{-H} + \left\langle \mathrm{diag}(\boldsymbol{\zeta}_{:kt})\mathbf{B}_k^*\mathbf{x}_{:kt}^*\mathbf{x}_{:kt}^H \right\rangle_T + \left\langle \mathrm{diag}(\boldsymbol{\gamma}_{:kt} + \boldsymbol{\psi}_{:kt})\mathbf{B}_k\mathbf{x}_{:kt}\mathbf{x}_{:kt}^H \right\rangle_T \tag{9}$$

We would like to solve this equation for $\mathbf{B}_k = d\mathbf{W}_k$.

If we define $\tilde{\mathbf{B}}_k \triangleq \mathbf{B}_k\mathbf{W}_k^{-1}$ and $\tilde{\mathbf{G}}_k \triangleq \mathbf{G}_k\mathbf{W}_k^H$, then (9) can be written,

$$\tilde{\mathbf{G}}_k = \tilde{\mathbf{B}}_k^H + \left\langle \mathrm{diag}\big(\boldsymbol{\zeta}_{:kt}\big)\tilde{\mathbf{B}}_k^*\mathbf{b}_{:kt}^*\mathbf{b}_{:kt}^H \right\rangle_T + \left\langle \mathrm{diag}\big(\boldsymbol{\gamma}_{:kt} + \boldsymbol{\psi}_{:kt}\big)\tilde{\mathbf{B}}_k\mathbf{b}_{:kt}\mathbf{b}_{:kt}^H \right\rangle_T \tag{10}$$

We find asymptotically for the diagonal elements,

$$\big[\tilde{\mathbf{G}}_k\big]_{ii} = [\tilde{\mathbf{B}}_k]_{ii}^* + E\big\{\psi_{ikt}|b_{ikt}|^2\big\}[\tilde{\mathbf{B}}_k]_{ii}^* + E\big\{(\gamma_{ikt} + \psi_{ikt})|b_{ikt}|^2\big\}[\tilde{\mathbf{B}}_k]_{ii}$$

The cross terms drop out since the expected value of $\psi_{ikt}b_{skt}b_{ikt}^*$ is zero for $i \neq s$ by the independence and zero mean assumption on the sources. Thus we have,

$$\big[\tilde{\mathbf{G}}_k\big]_{ii} = \big([\tilde{\mathbf{G}}_k]_{ii} - 1 + \eta_{ik}\big)[\tilde{\mathbf{B}}_k]_{ii} + (1 + \eta_{ik})[\tilde{\mathbf{B}}_k]_{ii}^*$$

where we define $\eta_{ik} \triangleq E\{\psi_{ikt}|b_{ikt}|^2\}$. Since $[\tilde{\mathbf{G}}_k]_{ii}$ is real, this equation implies that $[\tilde{\mathbf{B}}_k]_{ii}$ must be real, and,

$$[\tilde{\mathbf{B}}_k]_{ii} = \frac{[\tilde{\mathbf{G}}_k]_{ii}}{[\tilde{\mathbf{G}}_k]_{ii} + 2\eta_{ik}} \tag{11}$$

The equation for the off-diagonal elements in (10) is,

$$[\tilde{\mathbf{G}}_k]_{ij} = [\tilde{\mathbf{B}}_k]_{ji}^* + E\{\zeta_{ikt}\}E\{b_{jkt}^2\}^*[\tilde{\mathbf{B}}_k]_{ij}^* + E\{\gamma_{ikt} + \psi_{ikt}\}E\{|b_{jkt}|^2\}[\tilde{\mathbf{B}}_k]_{ij}$$

For circular complex sources such as the Fourier coefficients of stationary signals, we have $E\{b_{ikt}^2\} = 0$. We can thus solve the following system of equations for $[\tilde{\mathbf{B}}_k]_{ij}$,

$$\big[\tilde{\mathbf{G}}_k\big]_{ij} = [\tilde{\mathbf{B}}_k^*]_{ji} + \kappa_{ik}\sigma_{jk}^2[\tilde{\mathbf{B}}_k]_{ij}$$
$$\big[\tilde{\mathbf{G}}_k^*\big]_{ji} = [\tilde{\mathbf{B}}_k]_{ij} + \kappa_{jk}\sigma_{ik}^2[\tilde{\mathbf{B}}_k^*]_{ji}$$

where we define $\kappa_{ik} \triangleq E\{\gamma_{ikt} + \psi_{ikt}\}$ and $\sigma_{ik}^2 \triangleq E\{|b_{ikt}|^2\}$. Thus,

$$\big[\tilde{\mathbf{B}}_k\big]_{ij} = \frac{\kappa_{jk}\sigma_{ik}^2[\tilde{\mathbf{G}}_k]_{ij} - [\tilde{\mathbf{G}}_k^*]_{ji}}{\kappa_{ik}\kappa_{jk}\sigma_{ik}^2\sigma_{jk}^2 - 1} \tag{12}$$

where $\tilde{\mathbf{G}}_k = \mathbf{I} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varphi}_{:kt}\mathbf{b}_{:kt}^H$. With $\tilde{\mathbf{B}}_k$ defined by (11) and (12), we then put,

$$\Delta\mathbf{W}_k = \tilde{\mathbf{B}}_k\mathbf{W}_k \tag{13}$$

For the EM updates of α_{ij} and β_{ijk}, we have,

$$\alpha_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{ijt}, \quad \beta_{ijk}^{-1} = \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{ijt}\, E\{\xi_{ijt}^{-1}|\mathbf{b}_{i:t}\}\,|b_{ikt}|^2 \tag{14}$$

where the posterior expectation is given by equation (4).

5 Conclusion

We have formulated a probabilistic framework for IVA and developed an asymptotic Newton method to estimate the parameters by Maximum Likelihood with adaptive complex Gaussian scale mixtures. We have extended the IVA model by deriving the model probabilistically using complex Gaussian scale mixtures. We also allow sources to have arbitrary spectra by adapting the spectral power parameters β_{ijk}, and to assume multiple spectral regimes using a mixture of CGSMs in the source model. Here we have concentrated on Generalized Inverse Gaussian mixing density, but the framework can accommodate other CGSM families as well, e.g. Generalized Gaussian or Logistic densities [12]. The Newton method can be extended to handle non-circular sources as well, leading from equation (9) to a block 4×4 Hessian structure [10]. The probabilistic model also allows straightforward extension to an IVA mixture model.

References

1. Amari, S.-I.: Natural gradient works efficiently in learning. Neural Computation 10(2), 251–276 (1998)
2. Amari, S.-I., Chen, T.-P., Cichocki, A.: Stability analysis of learning algorithms for blind source separation. Neural Networks 10(8), 1345–1351 (1997)
3. Barndorff-Nielsen, O., Kent, J., Sørensen, M.: Normal variance-mean mixtures and z distributions. International Statistical Review 50, 145–159 (1982)
4. Brillinger, D.R.: Time Series: Data Analysis and Theory. SIAM, Philadelphia (2001); reproduction of original publication by Holden Day, Inc., San Francisco (1981)
5. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing. John Wiley & Sons, Ltd., West Sussex (2002)
6. Eltoft, T., Kim, T.-S., Lee, T.-W.: Multivariate Scale Mixture of Gaussians Modeling. In: Rosca, J.P., et al. (eds.) ICA 2006. LNCS, vol. 3889, pp. 799–806. Springer, Heidelberg (2006)
7. Kim, T., Attias, H., Lee, S.-Y., Lee, T.-W.: Blind source separation exploiting higher-order frequency dependencies. IEEE Transactions on Speech and Audio Processing 15(1) (2007)
8. Kim, T.-S., Eltoft, T., Lee, T.-W.: Independent vector analysis: An extension of ICA to multivariate components. In: Rosca, J.P., et al. (eds.) ICA 2006. LNCS, vol. 3889, pp. 165–172. Springer, Heidelberg (2006)
9. Kreutz-Delgado, K.: The complex gradient operator and the cr-calculus. Lecture Supplement to ECE275A, Ver. ECE275CG-F05v1.3c, University of California, San Diego, USA (2005), `http://dsp.ucsd.edu/~kreutz/PEI05.html`
10. Li, H., Adali, T.: Stability analysis of complex Maximum Likelihood ICA using Wirtinger calculus. In: Proceedings of the 33rd IEEE International Conference on Acoustics and Signal Processing (ICASSP 2008), Las Vegas, NV, pp. 1801–1804 (2008)
11. Palmer, J.A., Kreutz-Delgado, K., Makeig, S.: Super-gaussian mixture source model for ICA. In: Rosca, J.P., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 854–861. Springer, Heidelberg (2006)
12. Palmer, J.A., Kreutz-Delgado, K., Rao, B.D., Makeig, S.: Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) ICA 2007. LNCS, vol. 4666, pp. 97–104. Springer, Heidelberg (2007)