Modeling and Estimation of Dependent Subspaces with Non-radially Symmetric and Skewed Densities

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

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

Abstract. We extend the Gaussian scale mixture model of dependent subspace source densities to include non-radially symmetric densities using Generalized Gaussian random variables linked by a common variance. We also introduce the modeling of skew in source densities and subspaces using a generalization of the Normal Variance-Mean mixture model. We give closed form expressions for subspace likelihoods and parameter updates in the EM algorithm.

1 Introduction

The Gaussian scale mixture representation can be extended to vector subspaces to yield a model of non-affine dependency, sometimes referred to as "variance dependency" [8]. Hyvärinen [8,9] has recently proposed such a model for Independent Subspace Analysis of images. A similar approach is developed by Eltoft, Kim et al. [11,6], which is referred to as Independent Vector Analysis (IVA). In the IVA model, the EM algorithm is used with a particular case of the multivariate Gaussian scale mixture involving a Gamma mixing density.

In [10] a method is proposed for convolutive blind source separation in reverberative environments using a frequency domain approach with sources having variance (scale) dependency across frequencies. Typically in the frequency domain approach to blind deconvolution, the "permutation problem" arises when the signals are unmixed separately at each frequency. Due to the permutation ambiguity inherent in ICA [1], the frequency components of each source are output in arbitrary order at each frequency, and some matching heuristic must be employed to reconstruct the complete spectra of the sources. The IVA model allows the modeling of dependency of the frequency components of sources, while maintaining the mutual independence of the sources.

Variance dependency also arises in EEG/MEG analysis. In particular, the electromagnetic signals generated by muscles in the scalp, face, and ears will commonly activate together in various facial expressions. In this case, the individual muscle signals are not related or dependent in phase, but their variance increases and decreases together as the components are activated and deactivated together. Variance dependency may also exist among cortex regions that are simultaneously active in certain contexts.

The densities employed in models proposed previously for speech use only a particular dependent subspace density model, which may limit the flexibility of the model in application to more general domains such as communications and biological signal processing. We propose a general method for constructing multivariate Gaussian scale mixtures, giving an example of a multivariate dependent logistic density.

We also propose a scale mixture of Generalized Gaussians model, in which a generalized Gaussian random vector with independent components, is multiplied by a common scalar variance parameter, which is distributed Generalized Inverse Gaussian. This yields a generalization of the *generalized hyperbolic density* of Barndorff-Nielsen [3].

Finally we show how to use the Normal variance-mean mixtures to model skew in dependent subspaces. The location and "drift" parameters can be updated in closed form using the EM algorithm and exploiting the conditional Gaussianity and closed form formula for the posterior moment in terms of derivatives of the multivariate density function.

2   General Dependent Gaussian Scale Mixtures

In this section we show how general dependent multivariate densities can be derived using scalar Gaussian scale mixtures,

$$x = \xi^{1/2} z$$

where z is a standard Normal random variable, and ξ is a non-negative random variable.

2.1   Example Densities

Examples of Gaussian scale mixtures include the *generalized Gaussian* density, which has the form,

$$\mathcal{GG}(x; \rho) \;=\; \frac{1}{2\Gamma(1 + \frac{1}{\rho})} \, e^{-|x|^\rho}$$

It is a Gaussian scale mixture for $0 < \rho \leq 2$. The scale mixing density is related to a positive alpha stable density of order $\rho/2$.

The *generalized Cauchy* has the form,

$$\mathcal{GC}(x; \alpha, \nu) \;=\; \frac{\alpha \Gamma(\nu + 1/\alpha)}{2\Gamma(\nu)\Gamma(1/\alpha)} \frac{1}{(1 + |x|^\alpha)^{\nu + 1/\alpha}}$$

The Generalized Cauchy is a Gaussian scale mixture for $\nu > 0$ and $0 < \alpha < 2$. The scale mixing density is related to the Gamma density.

The *generalized Logistic*, also called the symmetric Fisher's z distribution [3], has the form,

$$\mathcal{GL}(x; \alpha) \;=\; \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2} \frac{e^{-\alpha x}}{(1 + e^{-x})^{2\alpha}}$$

The Generalized Logistic is a Gaussian scale mixture for all $\alpha > 0$. The scale mixing density is related to the Kolmogorov-Smirnov distance statistic [2,3,7].

2.2 Multidimensional Analogues

If x is distributed according to the Gaussian scale mixture density $p(x)$, then,

$$p(\sqrt{x}) = \frac{1}{(2\pi)^{1/2}} \int_0^\infty \xi^{-1/2} e^{-\frac{1}{2}\xi^{-1}x} p(\xi)d\xi \tag{1}$$

We can construct a random vector by multiplying the same scalar random variable $\xi^{1/2}$ by a Gaussian random vector,

$$\mathbf{x} = \xi^{1/2}\mathbf{z}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For the density of $\mathbf{x}$ we then have,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \xi^{-d/2} e^{-\frac{1}{2}\xi^{-1}\|\mathbf{x}\|^2} p(\xi)d\xi$$

If ξ^{-1} is a Gamma random variable, then the density of $\mathbf{x}$ can be written in terms of the modified Bessel function of the second kind [6].

In general, taking the kth derivative of both sides of (1), we find,

$$\frac{d^k}{dx^k}p(\sqrt{x}) = \frac{(-2)^{-k}}{(2\pi)^{1/2}} \int_0^\infty \xi^{-k-1/2} e^{-\frac{1}{2}\xi^{-1}x} p(\xi)d\xi$$

Thus, if d is odd, then with $k = (d-1)/2$,

$$\pi^{-(d-1)/2}(-D)^{(d-1)/2}p(\sqrt{x}) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \xi^{-d/2} e^{-\frac{1}{2}\xi^{-1}x} p(\xi)d\xi$$

and we can write the density of $p(\mathbf{x})$

$$d \text{ odd}: \qquad p(\mathbf{x}) = \pi^{-(d-1)/2}(-D)^{(d-1)/2}p(\sqrt{x})\big|_{x=\|\mathbf{x}\|^2} \tag{2}$$

For even d, the density of $p(\mathbf{x})$ can be written formally in terms of the Weyl fractional derivative. However as the fractional derivative is is not generally obtainable in closed form, we consider a modification of the original univariate scale density $p(\xi)$,

$$\tilde{p}(\xi) = \frac{\xi^{-1/2}p(\xi)}{\int_0^\infty \xi^{-1/2}p(\xi)d\xi}$$

With this modified scale density, the density of x evaluated at $\sqrt{x}$ becomes,

$$p(\sqrt{x}) = \frac{\mathcal{Z}}{(2\pi)^{1/2}} \int_0^\infty e^{-\frac{1}{2}\xi^{-1}x} \tilde{p}(\xi)d\xi \tag{3}$$

where,

$$\mathcal{Z} = \int_0^\infty \xi^{-1/2}p(\xi)d\xi$$

Proceeding as we did for odd d, taking the kth derivative of both sides of (3), with $k = d/2$, we get,

$$d \text{ even}: \qquad p(\mathbf{x}) = \mathcal{Z}^{-1}\sqrt{2\pi}^{-(d-1)/2}(-D)^{d/2}p(\sqrt{x})\big|_{x=\|\mathbf{x}\|^2} \tag{4}$$

2.3   Posterior Moments of Gaussian Scale Mixtures

To use scale mixtures in the EM context, it is necessary to calculate posterior moments of the scaling random variable. This section indicates how this is accomplished [5]. Differentiating under the (absolutely convergent) integral we get,

$$p'(x) = \frac{d}{dx}\int_0^\infty p(x|\xi)p(\xi)d\xi = -\int_0^\infty \xi^{-1}xp(x,\xi)\,d\xi$$

$$= -xp(x)\int_0^\infty \xi^{-1}p(\xi|x)\,d\xi$$

Thus, with $p(x) = \exp(-f(x))$, we see that,

$$E(\xi_i^{-1}|x_i) = \int_0^\infty \xi_i^{-1}p(\xi_i|x_i)\,d\xi_i = -\frac{p'(x_i)}{x_ip(x_i)} = \frac{f'(x_i)}{x_i} \tag{5}$$

Similar formulae can be derived for higher order posterior moments, and moments of multivariate scale parameters. These results are used in deriving EM algorithms for fitting univariate and multivariate Gaussian scale mixtures.

2.4   Example: 3D Dependent Logistic

Suppose we wish to formulate a dependent Logistic type density on $\mathbb{R}^3$. The scale mixing density in the Gaussian scale mixture representation for the Logistic density is related to the Kolmogorov-Smirnov distance statistic [2,3,7], which is only expressible in series form. However, we may determine the multivariate density produced from the product,

$$\mathbf{x} = \xi^{1/2}\mathbf{z}$$

where $\mathbf{x}, \mathbf{z} \in \mathbb{R}^3$, and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Using the formula (2) with $d = 3$, we get,

$$p(\mathbf{x}) = \frac{1}{8\pi}\frac{\sinh\left(\frac{1}{2}\|\mathbf{x}\|\right)}{\|\mathbf{x}\|\cosh^3\left(\frac{1}{2}\|\mathbf{x}\|\right)}$$

3   Non-radially Symmetric Dependency Models

A possible limitation of the Gaussian scale mixture dependent subspace model is the implied radial symmetry of vectors in the subspace, which leads to non-identifiability of features within the subspace—only the subspace itself can be identified. However, a similar approach using multivariate Generalized Gaussian scale mixtures can be developed, in which the multivariate density becomes a function of the p-norm of the subspace vector rather than the radially symmetric 2-norm, maintaining the directionality and identifiability of the within-subspace features, while preserving their (non-affine) dependence.

The mixing density of the generalized hyperbolic distribution is the generalized inverse Gaussian, which has the form,

$$\mathcal{N}^\dagger(\delta, \kappa, \lambda) = \frac{(\kappa/\delta)^\lambda}{2K_\lambda(\delta\kappa)}\,\xi^{\lambda-1}\exp\left(-\tfrac{1}{2}\left(\delta^2\xi^{-1} + \kappa^2\xi\right)\right) \tag{6}$$

where K_λ is the Bessel K function, or modified Bessel function of the second kind. The moments of the generalized inverse Gaussian [6] are given by,

$$E(\xi^r) = \left(\frac{\delta}{\kappa}\right)^r \frac{K_{\lambda+r}(\delta\kappa)}{K_\lambda(\delta\kappa)} \tag{7}$$

The isotropic generalized hyperbolic distribution [3] in dimension d,

$$\mathcal{GH}(\delta,\kappa,\lambda) = \frac{1}{(2\pi)^{d/2}} \frac{\kappa^{d/2}}{\delta^\lambda K_\lambda(\delta\kappa)} \frac{K_{\lambda-d/2}\left(\kappa\sqrt{\delta^2+\|\mathbf{x}\|^2}\right)}{\left(\delta^2+\|\mathbf{x}\|^2\right)^{d/4-\lambda/2}} \tag{8}$$

is derived as a Gaussian scale mixture with $\mathcal{N}^\dagger(\delta,\kappa,\lambda)$ mixing density. Now, for a generalized Gaussian scale mixture,

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}(\mathbf{p})} \int_0^\infty \xi^{-\sum_i p_i^{-1}} \exp\left(-\xi^{-1}\sum_i |x_i|^{p_i}\right) p(\xi)\, d\xi \tag{9}$$

where,

$$\mathcal{Z}(\mathbf{p}) = 2^d \prod_{i=1}^d \Gamma(1+1/p_i)$$

with $\mathcal{N}^\dagger$ mixing density $p(\xi)$, the posterior density of ξ given $\mathbf{x}$ is also $\mathcal{N}^\dagger$,

$$p(\xi|\mathbf{x}) = \mathcal{N}^\dagger\left(\sqrt{\delta^2 + 2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}}, \kappa, \lambda - d/\bar{p}\right) \tag{10}$$

where $\bar{p}$ is the harmonic mean $d/\sum_i p_i^{-1}$, and

$$\|\mathbf{x}\|_{\mathbf{p}} \triangleq \left(\sum_{i=1}^d |x_i|^{p_i}\right)^{1/\bar{p}}$$

For $\mathbf{x}$ we then get the anisotropic distribution,

$$p(\mathbf{x};\delta,\kappa,\lambda,\mathbf{p}) = \frac{1}{\mathcal{Z}(\mathbf{p})} \frac{\kappa^{d/\bar{p}}}{\delta^\lambda K_\lambda(\delta\kappa)} \frac{K_{\lambda-d/\bar{p}}\left(\kappa\sqrt{\delta^2+2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}}\right)}{\left(\delta^2+2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}\right)^{(d/\bar{p}-\lambda)/2}} \tag{11}$$

Using (7) and (10), we have,

$$E(\xi^{-1}|\mathbf{x}) = \frac{\kappa}{\sqrt{\delta^2+2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}}} \frac{K_{\lambda-d/\bar{p}-1}\left(\kappa\sqrt{\delta^2+2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}}\right)}{K_{\lambda-d/\bar{p}}\left(\kappa\sqrt{\delta^2+2\|\mathbf{x}\|_{\mathbf{p}}^{\bar{p}}}\right)} \tag{12}$$

The EM algorithm does not require that the complete log likelihood be maximized at each step, but only that it be increased, yielding the generalized EM (GEM) algorithm [4,13]. We employ this method here to increase the complete likelihood in (9) (see [13,12]).

4   Skew Models

4.1   Construction of Multivariate Skew Densities from Gaussian Scale Mixtures

Given a Gaussian scale mixture $\mathbf{x} = \xi^{1/2}\mathbf{z}$,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\xi^{-1}\mathbf{x}^T \Sigma^{-1}\mathbf{x}\right) p(\xi)\, d\xi$$

we have, trivially, for arbitrary β,

$$\frac{p(\mathbf{x})\exp(\beta^T \Sigma^{-1}\mathbf{x})}{\varphi\left(\tfrac{1}{2}\beta^T \Sigma^{-1}\beta\right)} = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \times$$

$$\int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\xi^{-1}\mathbf{x}^T \Sigma^{-1}\mathbf{x} + \beta^T \Sigma^{-1}\mathbf{x} - \tfrac{1}{2}\xi\beta^T \Sigma^{-1}\beta\right) \frac{p(\xi)\exp\left(\tfrac{1}{2}\xi\beta^T \Sigma^{-1}\beta\right)}{\varphi\left(\tfrac{1}{2}\beta^T \Sigma^{-1}\beta\right)}\, d\xi \tag{13}$$

where $\varphi(t) = E \exp t\xi$ is the moment generating function of ξ. Now (13) can be written,

$$\tilde{p}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\xi^{-1}\|\mathbf{x} - \xi\beta\|_{\Sigma^{-1}}^2\right) \tilde{p}(\xi;\beta)\, d\xi \tag{14}$$

where,

$$\tilde{p}(\mathbf{x}) = \frac{p(\mathbf{x})\exp(\beta^T \Sigma^{-1}\mathbf{x})}{\varphi\left(\tfrac{1}{2}\|\beta\|_{\Sigma^{-1}}^2\right)}, \quad \tilde{p}(\xi;\beta) = \frac{p(\xi)\exp\left(\tfrac{1}{2}\xi\|\beta\|_{\Sigma^{-1}}^2\right)}{\varphi\left(\tfrac{1}{2}\|\beta\|_{\Sigma^{-1}}^2\right)}$$

We have thus constructed a skewed density $\tilde{p}(\mathbf{x})$ in terms of the isotropic density $p(\mathbf{x})$ and the moment generating function φ of the scale mixing density $p(\xi)$. The skewed density is a that of a location-scale mixture [3] of the Gaussian $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$,

$$\mathbf{x} = \xi^{1/2}\mathbf{z} + \xi\beta.$$

4.2   EM Algorithm Posterior Updates

We now assume arbitrary location parameter μ, along with drift β, and structure matrix Σ. To use the EM algorithm with the Gaussian complete log likelihood in (14), we need to calculate posterior expectation of ξ^{-1}.

We do this using the method of §2.3. If we take the derivative of $-\log p(\mathbf{x} - \mu)$ with respect to $\tfrac{1}{2}\|\mathbf{x} - \mu\|_{\Sigma^{-1}}^2$, then we get,

$$\frac{\partial}{\partial \tfrac{1}{2}\|\mathbf{x} - \mu\|_{\Sigma^{-1}}^2}\left(-\log \int p(\mathbf{x} - \mu, \xi) d\xi\right)$$

$$= \frac{\int \xi^{-1} p(\mathbf{x} - \mu, \xi)\, d\xi}{\int p(\mathbf{x} - \mu, \xi)\, d\xi} = \frac{\int \xi^{-1}\tilde{p}(\mathbf{x} - \mu, \xi)\, d\xi}{\int \tilde{p}(\mathbf{x} - \mu, \xi)\, d\xi} = E(\xi^{-1}|\mathbf{x})$$

Thus, from (2) and (4), with $k \triangleq \lfloor d/2 \rfloor$ (the greatest integer less than $d/2$) we have,

$$E(\xi^{-1}|\mathbf{x}) = \frac{-1}{\|\mathbf{x} - \mu\|_{\Sigma^{-1}}} \frac{p^{(k+1)}\left(\|\mathbf{x} - \mu\|_{\Sigma^{-1}}\right)}{p^{(k)}\left(\|\mathbf{x} - \mu\|_{\Sigma^{-1}}\right)}$$

where $p^{(k)}$ is kth derivative of the univariate scale mixture $p(x)$.

4.3   Closed Form Parameter Updates

Given N observations $\{\mathbf{x}_k\}_{k=1}^N$, the μ that maximizes the complete log likelihood is found to be,

$$\mu = \frac{\frac{1}{N}\sum_k \gamma_k \mathbf{x}_k - \beta}{\frac{1}{N}\sum_k \gamma_k} \tag{15}$$

where $\gamma_k = E(\xi^{-1}|\mathbf{x}_k)$.

The estimation equation to be solved for β, which does not involve the posterior estimate of ξ_k, is,

$$\frac{\varphi'\left(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\right)}{\varphi\left(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\right)}\beta = \mathbf{c} - \mu \tag{16}$$

where $\mathbf{c} = \frac{1}{N}\sum_k \mathbf{x}_k$. This gives β in terms of μ up to a scale factor. Given μ, the optimal β, denoted β^*, may be found by first determining $\zeta \triangleq \frac{1}{2}\|\beta^*\|^2_{\Sigma^{-1}}$ from,

$$h(\zeta) \triangleq \left(\frac{\varphi'(\zeta)}{\varphi(\zeta)}\right)^2 \zeta = \frac{1}{2}\|\mathbf{c} - \mu\|^2_{\Sigma^{-1}}$$

assuming that the univariate function h is invertible. Then β^* is given as,

$$\beta^* = \frac{\varphi(\zeta)}{\varphi'(\zeta)}(\mathbf{c} - \mu)$$

Given β^*, we may determine the optimal μ^* by substituting β^* into (15). Repeated iteration constitutes a coordinate ascent EM algorithm for μ and β.

An alternative method suggests itself: if we fix the norm of β in the mixing density, then we can solve for new estimates of μ and β simultaneously. Let,

$$a \triangleq \frac{1}{N}\sum_k \gamma_k, \quad \mathbf{b} \triangleq \frac{1}{N}\sum_k \gamma_k \mathbf{x}_k, \quad \tau \triangleq \frac{\varphi'\left(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\right)}{\varphi\left(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\right)}$$

Then from (15) and (16), we have,

$$a\mu^* + \beta^* = \mathbf{b}$$
$$\mu^* + \tau\beta^* = \mathbf{c}$$

Solving for the components μ_i, β_i, $i = 1, \dots, d$, we get,

$$\begin{bmatrix} a & 1 \\ 1 & \tau \end{bmatrix}\begin{bmatrix} \mu_i^* \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} b_i \\ c_i \end{bmatrix} \quad \Rightarrow \quad \mu_i^* = \frac{\tau b_i - c_i}{a\tau - 1}, \quad \beta_i^* = \frac{ac_i - b_i}{a\tau - 1}$$

For the structure matrix, Σ, setting the complete log likelihood gradient to zero, we get,

$$\Sigma = \frac{1}{N}\sum_k \gamma_k(\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T - \frac{2}{N}\sum_k(\mathbf{x}_k - \mu)\beta^T$$
$$= \frac{1}{N}\sum_k \gamma_k(\mathbf{x}_k - \mu - \gamma_k^{-1}\beta)(\mathbf{x}_k - \mu - \gamma_k^{-1}\beta)^T - \left(\sum_k \gamma_k^{-1}\right)\beta\beta^T.$$

5 Conclusion

We have shown how to derive general multivariate Gaussian scale mixtures in terms of scalar Gaussian scale mixtures, and how to optimize them using an EM algorithm. We generalized the spherically (or ellipsoidally) symmetric Gaussian scale mixture by introducing a generalization of Barndorff-Nielsen's generalized hyperbolic density using Generalized Gaussian scale mixtures, yielding a multivariate dependent anisotropic model. We also introduced the modeling of skew in ICA sources, deriving a general form of skewed multivariate Gaussian scale mixture, and an EM algorithm to update the location, drift, and structure parameters.

References

1. Amari, S.-I., Cichocki, A.: Adaptive blind signal processing—neural network approaches. Proceedings of the IEEE 86(10), 2026–2047 (1998)
2. Andrews, D.F., Mallows, C.L.: Scale mixtures of normal distributions. J. Roy. Statist. Soc. Ser. B 36, 99–102 (1974)
3. Barndorff-Nielsen, O., Kent, J., Sørensen, M.: Normal variance-mean mixtures and z distributions. International Statistical Review 50, 145–159 (1982)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38 (1977)
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Iteratively reweighted least squares for linear regression when errors are Normal/Independent distributed. In: Krishnaiah, P.R. (ed.) Multivariate Analysis V, pp. 35–57. North Holland Publishing Company, Amsterdam (1980)
6. Eltoft, T., Kim, T., Lee, T.-W.: Multivariate scale mixture of Gaussians modeling. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 799–806. Springer, Heidelberg (2006)
7. Gneiting, T.: Normal scale mixtures and dual probability densities. J. Statist. Comput. Simul. 59, 375–384 (1997)
8. Hyvärinen, A., Hoyer, P.O.: Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. Neural Computation 12, 1705–1720 (2000)
9. Hyvärinen, A., Hoyer, P.O., Inki, M.: Topographic independent component analysis. Neural Computation 13(7), 1527–1558 (2001)
10. 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)
11. Kim, T., Eltoft, T., Lee, T.-W.: Independent vector analysis: An extension of ICA to multivariate components. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 165–172. Springer, Heidelberg (2006)
12. Palmer, J.A., Kreutz-Delgado, K., Makeig, S.: Super-Gaussian mixture source model for ICA. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 854–861. Springer, Heidelberg (2006)
13. Palmer, J.A., Kreutz-Delgado, K., Wipf, D.P., Rao, B.D.: Variational EM algorithms for non-gaussian latent variable models. In: Advances in Neural Information Processing Systems, NIPS 2005, MIT Press, Cambridge (2006)