Monte Carlo Methods Appl., Vol. 8, No. 4 (2002), pp. 395–403

Jointly Distributed Mean and Mixing Coefficients for Bayesian Source Separation using MCMC and ICM

Daniel B. Rowe
Biophysics Research Institute
Medical College of Wisconsin
Milwaukee, WI 53226

Abstract – Recent source separation work has described a model which assumes a nonzero overall mean and incorporates prior knowledge regarding it. This is significant because source separation models that have previously been presented, have assumed that the overall mean is zero. However, this work specified that the prior distribution which quantifies available prior knowledge regarding the overall mean be independent of the mixing coefficient matrix. The current paper, generalizes this work by quantifying available prior information regarding the overall mean and mixing matrix with the use of joint prior distributions. This prior knowledge in the prior distributions is incorporated into the inferences along with the current data. Conjugate normal, and generalized conjugate normal distributions are used. Algorithms for estimating the parameters of the model from the joint posterior distribution are derived and they are determined statistically from the posterior distribution using both Gibbs sampling a Markov chain Monte Carlo method and the iterated conditional modes algorithm a deterministic optimization technique for marginal mean and maximum a posterior estimates respectively.

1 Introduction and Model

The source separation problem is that of separating unobservable or latent source signals when mixed signals are observed. To take a set of observed mixed signal vectors and unmix or separate them into a set of true unobservable source signal vectors. This paper adopts a multivariate Bayesian [7,8] statistical approach and the linear synthesis model [3,5,9,10] with an overall mean.

For motivation and illustation of the source separation model, the context of the "cocktail party problem" is adopted [3]. At a cocktail party, there are p microphones that record or observe m partygoers or speakers at n time increments. The observed conversations consist of mixtures of true conversations. In other words, p-dimensional mixed signal vectors $x_i = (x_{i1}, \ldots, x_{ip})'$ are observed and the goal is to separate these observed signal vectors into m-dimensional true underlying source signal vectors, $s_i = (s_{i1}, \ldots, s_{im})'$ where $i = 1, \ldots, n$.

The Bayesian source separation model for the observed vector x_i at time i is

$$\underset{(p\times 1)}{(x_i|\mu,\Lambda,s_i)} = \underset{(p\times 1)}{\mu} + \underset{(p\times m)}{\Lambda}\ \underset{(m\times 1)}{s_i} + \underset{(p\times 1)}{\epsilon_i} ,$$

where it has been assumed that the observed signals have a nonzero mean. The variables in the model are denoted as follows,
μ = the p-dimensional overall mean, $\mu = (\mu_1, \ldots, \mu_p)'$;
Λ = a $p \times m$ matrix of unobserved mixing constants, $\Lambda = (\lambda_1', \ldots, \lambda_p')'$;
s_i = the i^{th} m-dimensional unobservable source vector, $s_i = (s_{i1}, \ldots, s_{1m})'$; and
ϵ_i = the p-dimensional vector of errors or noise terms of the i^{th} observed signal vector $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})'$.

In order to incorporate jointly distributed prior knowledge regarding the mean vector and mixing matrix, the model is rewritten as

$$\begin{array}{ccccccc} (x_i|C, s_i) & = & C & & z_i & + & \epsilon_i, \\ (p \times 1) & & p \times (m+1) & & (m+1) \times 1 & & p \times 1 \end{array}$$

where $C = (\mu, \Lambda)$ and $z_i' = (1, s_i')$.

Analogous to regression, the source separation model can be written in terms of matrices as

$$\begin{array}{ccccccc} (X|C, S) & = & Z & & C' & + & E \quad , \\ (n \times p) & & n \times (m+1) & & (m+1) \times p & & n \times p \end{array}$$

where $X' = (x_1, \ldots, x_n)$ contains the observations as rows of X, $Z' = (z_1, \ldots, z_n)$ contains the unobserved true source vectors in the rows of Z, and $E' = (\epsilon_1, \ldots, \epsilon_n)$ contains the error vectors as rows of E. The time series of observations for the j^{th} microphone is the j^{th} column of X and the time series of unobservables for the k^{th} source is the k^{th} column of S.

2 Likelihood

It is specified that the errors of the observation vectors are independent over time but free to be dependent or correlated within each vector. As in the multivariate regression model, the observation vectors x_i are taken to be normally distributed with mean zero and covariance matrix Ψ. Thus, the likelihood of a given observation vector x_i can be written as

$$p(x_i|C, s_i, \Psi) = (2\pi)^{-\frac{p}{2}} |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i - Cz_i)'\Psi^{-1}(x_i - Cz_i)}.$$

If proportionality is denoted by "$\propto$" then the likelihood for (C, S, Ψ) with all the observation vectors collected into a matrix is

$$p(X|C, S, \Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - ZC')'(X - ZC')}$$

where the n p-variate observation vectors are $X' = (x_1, \ldots, x_n)$, the source vectors are contained in $Z' = (z_1, \ldots, z_n)$, and the errors of observation are $E' = (\epsilon_1, \ldots, \epsilon_n)$. The notation $p(\cdot)$ will generically denote a probability distribution which is distinguished by its argument whose proportionality constant does not depend on its argument.

Available knowledge regarding how probable values of the parameters are in the form of prior distributions is now quantified and incorporated into the inferences.

3 Priors And Posteriors

The source vectors s_i are specified to be normally distributed with mean s_{i0} and covariance matrix R. Regarding the other parameters, information is incorporated as an inverted Wishart distribution for the covariance of the source vectors with η degrees of freedom and scale matrix V, an inverted Wishart distribution for the covariance of the observed vectors with ν degrees of freedom and scale matrix Q. The matrix C which contains the overall mean and the mixing matrix rewritten as a vector $c = vec(C')$ is specified to be either conjugate normally distributed with mean $c_0 = vec(C_0')$ and covariance $\Psi \otimes M$ or generalized conjugate normally distributed with mean $c_0 = vec(C_0')$ and covariance Δ. The inverted Wishart distribution is the multivariate generalization of the inverted gamma distribution which is used as a prior distribution for variances.

More formally, prior information regarding the overall mean and the mixing matrix is quantified using the conjugate and generalized conjugate prior distributions

$$\begin{aligned} p(C|\Psi) &\propto |M|^{-\frac{p}{2}}|\Psi|^{-\frac{m+1}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(C-C_0)M^{-1}(C-C_0)'} \\ p(c) &\propto |\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(c-c_0)'\Delta^{-1}(c-c_0)}. \end{aligned}$$

Note that the overall mean and mixing matrix are jointly distributed thus allowing them to be dependent or correlated.

Prior distributions are assessed for the remaining model parameters. It is specified that the prior distributions for the sources S, the source covariance matrix R, the error covariance matrix Ψ, follow normal, inverted Wishart, and inverted Wishart distributions respectively

$$\begin{aligned} p(S|R) &\propto |R|^{-\frac{n}{2}}e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'}, \\ p(R) &\propto |R|^{-\frac{\eta}{2}}e^{-\frac{1}{2}trR^{-1}V}, \\ p(\Psi) &\propto |\Psi|^{-\frac{\nu}{2}}e^{-\frac{1}{2}tr\Psi^{-1}Q}, \end{aligned}$$

where $\nu > 2p$, $\eta > 2m$. The hyperparameters C_0, M, Δ, S_0, η, V, ν, and Q are to be assessed.

Note that both Ψ and R are full covariance matrices allowing the elements of the observed mixed signal and the unobserved source component vectors to be correlated or dependent.

Upon using Bayes' rule, the posterior distributions for the unknown parameters, taking either the conjugate normal or generalized conjugate normal priors for the jointly distributed overall mean and mixing matrix are either

$$\begin{aligned} p(C,S,R,\Psi|X) &\propto |\Psi|^{-\frac{(n+\nu+m+1)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}G} \\ &\quad \times|R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \end{aligned}$$

or

$$\begin{aligned} p(c,S,R,\Psi|X) &\propto |\Psi|^{-\frac{(n+\nu)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}[(X-ZC')'(X-ZC')+Q]} \\ &\quad \times|R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \\ &\quad \times|\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(c-c_0)\Delta^{-1}(c-c_0)'}, \end{aligned}$$

where

$$G = (X-ZC')'(X-ZC') + (C-C_0)M^{-1}(C-C_0)' + Q.$$

These posterior distributions are now evaluated in order to obtain parameter estimates of the sources, the source covariance matrix, the overall population mean, the mixing matrix, and the errors covariance matrix.

4 Conjugate Estimation

With the posterior distribution, it is not possible to obtain marginal distributions and thus marginal estimates for any of the parameters in an analytic closed form. It is also not possible to find analytic closed form solutions for maximum a posteriori estimates. It is however possible to use both Gibbs sampling, a Markov chain Monte Carlo integration technique to obtain marginal parameter estimates [1,2,11] and the deterministic optimization technique iterated conditional modes (ICM) for maximum a posteriori estimates [4,6,10]. For both estimation procedures, the posterior conditional distributions are required.

4.1 Posterior Conditionals

From the joint posterior distribution we can obtain the posterior conditional distributions.

The conditional posterior distribution for the matrix containing the overall mean and mixing matrix is

$$\begin{aligned} p(C|S,R,\Psi,X) &\propto p(C|\Psi)p(X|C,S,\Psi) \\ &\propto |\Psi|^{-\frac{m+1}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(C-C_0)M^{-1}(C-C_0)'} \\ &\times |\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-ZC')'(X-ZC')} \\ &\propto e^{-\frac{1}{2}tr\Psi^{-1}(C-\tilde{C})(M^{-1}+Z'Z)(C-\tilde{C})'} \end{aligned}$$

where the posterior conditional mean and mode is given by

$$\tilde{C} = [X'Z + C_0M^{-1}](M^{-1}+Z'Z)^{-1}.$$

The conditional distribution for the mixing matrix given the other parameters and the data is normally distributed.

The conditional posterior distribution of the observation error matrix is

$$\begin{aligned} p(\Psi|C,S,R,X) &\propto p(\Psi)p(C|\Psi)p(X|C,S,\Psi) \\ &\propto |\Psi|^{-\frac{\nu}{2}}e^{-\frac{1}{2}tr\Psi^{-1}Q}|\Psi|^{-\frac{m+1}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(C-C_0)M^{-1}(C-C_0)'} \\ &\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-ZC')'(X-ZC')} \\ &\propto |\Psi|^{-\frac{(n+\nu+m+1)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}G} \end{aligned}$$

where

$$G = (X-ZC')'(X-ZC') + (C-C_0)M^{-1}(C-C_0)' + Q$$

with a mode given by

$$\tilde{\Psi} = \frac{G}{n+\nu+m+1} .$$

The conditional distribution of the observation error covariance matrix given the other parameters and the data is an inverted Wishart.

The conditional posterior distribution for the sources is

$$\begin{aligned} p(S|\mu,\Lambda,R,\Psi,X) &\propto p(S|R)p(X|\mu,\Lambda,S,\Psi) \\ &\propto |R|^{-\frac{n}{2}}e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\ &\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\ &\propto e^{-\frac{1}{2}tr(S-\tilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\tilde{S})'} \end{aligned}$$

where the posterior conditional mean and mode is given by

$$\tilde{S} = [(X - e_n\mu')\Psi^{-1}\Lambda + S_0R^{-1}](R^{-1} + \Lambda'\Psi^{-1}\Lambda)^{-1}.$$

The conditional posterior distribution for the sources given the other parameters and the data is normally distributed.

The conditional posterior distribution for the source covariance matrix is

$$\begin{aligned} p(R|C,S,\Psi,X) &\propto p(R)p(S|R)p(X|C,S,\Psi) \\ &\propto |R|^{-\frac{\eta}{2}}e^{-\frac{1}{2}trR^{-1}V}|R|^{-\frac{n}{2}}e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\ &\propto |R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \end{aligned}$$

with the posterior conditional mode given by

$$\tilde{R} = \frac{(S-S_0)'(S-S_0)+V}{n+\eta}.$$

The conditional posterior distribution for the source covariance matrix given the other parameters and the data is inverted Wishart distributed.

4.2 Gibbs Sampling

For Gibbs sampling estimation of the posterior, start with initial values for S and Ψ say $\bar{S}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned} \bar{C}_{(l+1)} &= \text{a random variate from } p(C|\bar{S}_{(l)},\bar{R}_{(l)},\bar{\Psi}_{(l)},X) \\ \bar{\Psi}_{(l+1)} &= \text{a random variate from } p(\Psi|\bar{S}_{(l)},\bar{R}_{(l)},\bar{C}_{(l+1)},X) \\ \bar{R}_{(l+1)} &= \text{a random variate from } p(R|\bar{S}_{(l)},\bar{C}_{(l+1)},\bar{\Psi}_{(l+1)},X) \\ \bar{S}_{(l+1)} &= \text{a random variate from } p(S|\bar{C}_{(l+1)},\bar{\Psi}_{(l+1)},\bar{R}_{(l+1)},X) \end{aligned}$$

and the first random variates called the “burn in” are discarded compute from the next L variates

$$\bar{S} = \frac{1}{L}\sum_{l=1}^{L}\bar{S}_{(l)} \quad \bar{R} = \frac{1}{L}\sum_{l=1}^{L}\bar{R}_{(l)} \quad \bar{C} = \frac{1}{L}\sum_{l=1}^{L}\bar{C}_{(l)} \quad \bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Psi}_{(l)}$$

which are the sampling based marginal posterior mean estimates of the parameters.

4.3 Maximum A Posteriori

The ICM estimaton procedure consists of starting with an initial value for S say $\tilde{S}_{(0)}$, forming $\tilde{Z}_{(0)} = (e_n, \tilde{S}_{(0)})$ then iterating through

$$\begin{aligned} \tilde{C}_{(l+1)} &= [X'\tilde{Z}_{(l)} + C_0M^{-1}](M^{-1} + \tilde{Z}'_{(l)}\tilde{Z}_{(l)})^{-1} \\ \tilde{\Psi}_{(l+1)} &= [(X - \tilde{Z}_{(l)}\tilde{C}'_{(l+1)})'(X - \tilde{Z}_{(l)}\tilde{C}'_{(l+1)}) + \\ &\quad (\tilde{C}_{(l+1)} - C_0)M^{-1}(\tilde{C}_{(l+1)} - C_0)' + Q]/(n+\nu+m+1) \\ \tilde{R}_{(l+1)} &= \frac{(\tilde{S}_{(l)} - S_0)'(\tilde{S}_{(l)} - S_0) + V}{n+\eta} \\ \tilde{S}_{(l+1)} &= (X - e_n\tilde{\mu}'_{(l+1)})\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)}(\tilde{R}^{-1}_{(l+1)} + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)})^{-1} \end{aligned}$$

until convergence is reached. The converged values $(\tilde{C}, \tilde{S}, \tilde{R}, \tilde{\Psi})$ are joint posterior modal (maximum a posteriori) estimators of the parameters.

5 Generalized Conjugate Estimation

With the joint generalized conjugate normal distribution for the overall mean and mixing coefficients, it is not possible to obtain all or any of the marginal distributions and thus marginal estimates in closed form. It is also not possible to find analytic closed form solutions for maximum a posteriori estimates. It is possible to use both Gibbs sampling, a Monte Carlo integration technique to obtain marginal parameter estimates [1,2,11] and the deterministic optimization technique iterated conditional modes (ICM) for maximum a posteriori estimates [4,6,10]. For this reason, marginal estimates are found using these two algorithms.

5.1 Posterior Conditionals

Both Gibbs sampling and ICM require the posterior conditionals. Gibbs sampling requires the conditionals for the generation of random variates while ICM requires them for maximization by cycling through their modes.

The conditional posterior distribution of the sources is

$$\begin{aligned} p(S|\mu,\Lambda,R,\Psi,X) &\propto p(S|R)p(X|\mu,\Lambda,S,\Psi) \\ &\propto e^{-\frac{1}{2}tr(S-S_0)'R^{-1}(S-S_0)}e^{-\frac{1}{2}tr(X-e_n\mu'-S\Lambda')\Psi^{-1}(X-e_n\mu'-S\Lambda')'} \end{aligned}$$

which after some algebra can be written as

$$p(S|\mu,\Lambda,R,\Psi,X) \propto e^{-\frac{1}{2}tr(S-\tilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\tilde{S})'}$$

where $\tilde{S} = (X - e_n\mu')\Psi^{-1}\Lambda(R^{-1} + \Lambda'\Psi^{-1}\Lambda)^{-1}$.

That is, the sources given the other parameters and the data is normally distributed.

The conditional posterior distribution of the source covariance matrix is

$$\begin{aligned} p(R|\mu,\Lambda,S,\Psi,X) &\propto p(R)p(S|R) \\ &\propto |R|^{-\frac{\nu}{2}}e^{-\frac{1}{2}trR^{-1}V}|R|^{-\frac{n}{2}}e^{-\frac{1}{2}trR^{-1}(S-S_0)(S-S_0)'} \\ &\quad |R|^{-\frac{(n+\nu)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)(S-S_0)'+V]}. \end{aligned}$$

That is, the conditional distribution of the error covariance matrix given the other parameters and the data is inverted Wishart distributed.

The conditional posterior distribution of the mixing matrix is

$$\begin{aligned} p(c|S,R,\Psi,X) &\propto p(\lambda)p(X|c,S,\Psi) \\ &\propto |\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(c-c_0)'\Delta^{-1}(c-c_0)} \\ &\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-ZC')'(X-ZC')} \end{aligned}$$

which after some algebra becomes

$$p(c|S,R,\Psi,X) \propto e^{-\frac{1}{2}(c-\tilde{c}')[\Delta^{-1}+\Psi^{-1}\otimes Z'Z](c-\tilde{c})}$$

where

$$\tilde{c} = [\Delta^{-1} + \Psi^{-1} \otimes Z'Z]^{-1}[\Delta^{-1}c_0 + (\Psi^{-1} \otimes Z'Z)\hat{c}]$$

and

$$\hat{c} = vec[(Z'Z)^{-1}Z'X].$$

The conditional posterior distribution of the mixing matrix given the other parameters and the data is normally distributed.

The conditional posterior distribution of the error covariance matrix is

$$\begin{aligned} p(\Psi|C,S,R,X) &\propto p(\Psi)p(X|C,S,\Psi) \\ &\propto |\Psi|^{-\frac{(n+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}[(X-ZC')'(X-ZC')+Q]}. \end{aligned}$$

That is, the conditional distribution of the error covariance matrix given the other parameters and the data is inverted Wishart distributed.

The modes of these conditional distributions are $\tilde{S}$, $\tilde{c}$ (as defined above),

$$\tilde{R} = \frac{(S-S_0)(S-S_0)'+V}{n+\eta},$$

and

$$\tilde{\Psi} = \frac{(X-ZC')'(X-ZC')+Q}{n+\nu},$$

respectively.

5.2 Gibbs Sampling

For Gibbs estimation of the posterior, start with initial values for S and Ψ say $\bar{S}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned} \bar{R}_{(l+1)} &= \text{a random variate from } p(R|\bar{S}_{(l)}, \bar{\Psi}_{(l)}, \bar{C}_{(l)}, X) \\ \bar{c}_{(l+1)} &= \text{a random variate from } p(c|\bar{S}_{(l)}, \bar{\Psi}_{(l)}, \bar{R}_{(l+1)}, X) \\ \bar{\Psi}_{(l+1)} &= \text{a random variate from } p(\Psi|\bar{S}_{(l)}, \bar{R}_{(l+1)}, \bar{C}_{(l+1)}, X) \\ \bar{S}_{(l+1)} &= \text{a random variate from } p(S|\bar{R}_{(l+1)}, \bar{C}_{(l+1)}, \bar{\Psi}_{(l+1)}, X) \end{aligned}$$

and the first random variates called the "burn in" are discarded compute from the next L variates

$$\bar{S} = \frac{1}{L}\sum_{l=1}^{L}\bar{S}_{(l)} \quad \bar{R} = \frac{1}{L}\sum_{l=1}^{L}\bar{R}_{(l)} \quad \bar{c} = \frac{1}{L}\sum_{l=1}^{L}\bar{\lambda}_{(l)} \quad \bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Psi}_{(l)}$$

which are the sampling based marginal posterior mean and modal estimates of the parameters.

5.3 Maximum A Posteriori

For the ICM estimation of the parameters start with an initial values for $\tilde{S}$ and Ψ, say $\tilde{S}_{(0)}$ and $\tilde{\Psi}_{(0)}$ then cycle through

$$\begin{aligned}
\hat{c}_{(l+1)} &= vec[(\tilde{Z}'_{(l)}\tilde{Z}_{(l)})^{-1}\tilde{Z}'_{(l)}X] \\
\tilde{c}_{(l+1)} &= [\Delta^{-1} + \tilde{\Psi}^{-1}_{(l)} \otimes \tilde{Z}'_{(l)}\tilde{Z}_{(l)}]^{-1}[\Delta^{-1}c_0 + (\tilde{\Psi}^{-1}_{(l)} \otimes \tilde{Z}'_{(l)}\tilde{Z}_{(l)})\hat{c}_{(l+1)}] \\
\tilde{\Psi}_{(l+1)} &= \frac{(X - \tilde{Z}_{(l)}\tilde{C}'_{(l+1)})'(X - \tilde{Z}_{(l)}\tilde{C}'_{(l+1)}) + Q}{n+\nu} \\
\tilde{R}_{(l+1)} &= \frac{(\tilde{S}_{(l)} - S_0)(\tilde{S}_{(l)} - S_0)' + V}{n+\eta} \\
\tilde{S}_{(l+1)} &= (X - e_n\tilde{\mu}'_{(l+1)})\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)}(\tilde{R}^{-1}_{(l+1)} + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)})^{-1}.
\end{aligned}$$

until convergence is reached with the joint modal (maximum a posteriori) estimator for the unknown parameters $(\tilde{S}, \tilde{C}, \tilde{\Psi})$.

6 Conclusion

It is seen that the overall mean and the mixing coefficients do not need to be constrained to be independent. Taking the overall mean and the mixing coefficients to be independent is analagous to taking the intercept and the dependent variable coefficients to be independent in a regression model. Further, available information about the overall mean and mixing coefficients can be quantified and incorporated into the inferences using conjugate normal and generalized conjugate normal distributions. From the posterior distribution, both marginal mean and maximum a posteriori estimates can be determined.

References

[1] A. E. Gelfand and A. F. M Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741, 1984.

[3] Kevin Knuth. A Bayesian approach to source separation. In C. Jutten J.-F. Cardoso and P. Loubaton, editors, *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99, Aussios, France*, pages 283–288, 1999.

[4] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34(1), 1972.

[5] Ali Mohammad-Djafari. A Bayesian approach to source separation. In *Proceedings of the Nineteenth International Conference on Maximum Entropy and Bayesian Methods, (August 2-6, 1999, Boise, Idaho, USA)*, 1999.

[6] Anthony O'Hagen. *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference*. John Wiley and Sons Inc., New York, 1994.

[7] S. James Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Robert E. Krieger Publishing Company, Malabar, Florida, 1982.

[8] S. James Press. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley and Sons, New York, 1989.

[9] Daniel B. Rowe. Bayesian source separation of FMRI signals. In A. Mohammad-Djafari, editor, *Maximum Entropy and Bayesian Methods*. American Institute of Physics, College Park, Maryland, USA, 2000.

[10] Daniel B. Rowe. Bayesian source separation for reference function determination in FMRI. *Magnetic Resonance in Medicine*, 45(5):374–378, 2001.

[11] Daniel B. Rowe. Factorization of separable and patterned covariance matrices for Gibbs sampling. *Monte Carlo Methods and Applications*, 6(3):205–210, 2000.

[12] Daniel B. Rowe. A Bayesian approach to blind source separation. *Journal of Interdisciplinary Mathematics*, 5(1):49–76, 2002.

[13] Daniel B. Rowe. Bayesian source separation of functional sources. *Journal of Interdisciplinary Mathematics*, 5(2), 2002.