*Monte Carlo Methods Appl.*, Vol. 6, No. 3 (2000), pp. 205–210


# Factorization of Seperable and Patterned Covariance Matrices for Gibbs Sampling

Daniel B. Rowe
H & SS, 228 - 77
California Institute of Technology
Pasadena, California 91125

**Abstract –** Recently the Gibbs sampler has become a very popular estimation technique especially in Bayesian Statistics. In order to implement the Gibbs sampler, matrix factorizations must be computed which normally is not problematic. When the dimension of the matrices to be factored is large, computation time increases to an amount to merit special attention. I have found that when the matrices to be factored are separable or patterned, results from matrix theory can assist in computation time reduction.

**Keywords:** latent roots, latent vectors, random sample.

## 1. Introduction

Recently the Gibbs sampler has become one of the favored techniques for parameter estimation especially in Bayesian Statistics. The Gibbs sampler is a sampling based approach to calculating marginal posterior distributions especially useful when the densities are not integrable in closed form or are too high a dimension for other methods. The Gibbs sampler is a stochastic integration technique that draws samples from conditional densities and uses them to approximate the marginal densities.

In order to generate a random sample $x$ from say a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}$, denoted by $N(0, \boldsymbol{\Omega})$, we must first compute the factorization of the matrix $\boldsymbol{\Omega}$. I will denote this factorization by $\boldsymbol{\Omega} = UU'$. We then generate a random $y$ from a $N(0, I)$ and transform it by $Uy$. Then $x = Uy$ is a random sample from $N(0, \boldsymbol{\Omega} = UU')$ (Press 1982, p. 67). We similarly require matrix factorizations for random samples from Wishart or inverted Wishart distributions.

In most instances computing the factorization is not very time consuming, and if it is we allow for longer computation times. Sometimes, we would like to perform the computations in minimal time. I will consider the factorization of matrices that are separable or patterned.

## 2. Seperable Matrices

Quite often, the matrices to be factored are of the form

$$\underset{np \times np}{\mathbf{\Omega}} = \underset{n \times n}{\mathbf{\Phi}} \otimes \underset{p \times p}{\mathbf{\Psi}}$$

where $\otimes$ denotes the Kroneker product. Matrices of this form are called separable. Separable matrices occur frequently in regression and time series. A single observation vector consisting of observation vectors of smaller dimension, represented by

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{2.1}$$

where the variance of observation vector $i$ is given by the $p \times p$ matrix,

$$var(x_i|\mathbf{\Omega}, m, f, \lambda) = \mathbf{\Omega}_{ii} = \phi_{ii}\mathbf{\Psi}$$

and the covariance between observation vectors $i$ and $j$ is given by the $p \times p$ matrix

$$cov(x_i, x_j|\mathbf{\Omega}, m, f, \lambda) = \mathbf{\Omega}_{ij} = \phi_{ij}\mathbf{\Psi}$$

has a separable covariance matrix. If it is assumed that the observation vectors have a common variance, then this is equilavent to weak stationarity.

The factorization of the covariance matrix $\mathbf{\Omega}$ without taking advantage of its separable nature can be very time consuming. We take advantage of its separable form to reduce computation.

The factorization can be found by computing the latent roots and latent vectors (Press 1982). The matrix $\mathbf{\Omega}$ can be expressed as

$$\mathbf{\Omega} = \Gamma D_\lambda \Gamma' = (\Gamma D_\lambda^{-\frac{1}{2}})(\Gamma D_\lambda^{-\frac{1}{2}})' = UU' \tag{2.2}$$

where $\Gamma$ is an orthogonal matrix of latent vectors as columns and

$$D_\lambda = (\lambda_1, \ldots, \lambda_k, \ldots, \lambda_{Np})' \tag{2.3}$$

is a matrix of latent roots of $\mathbf{\Omega}$. This factorization can be performed numerically for a general covarince matrix.

Quite often, computing the factorization takes appreciable amounts of time especially when we have to calculate them many thousands of times for the Gibbs sampler. We can save computation time by using the following theorem.

**Theorem (Anderson 1994)**
Let the $i^{th}$ latent root of $\mathbf{\Phi}$ be $\alpha_i$ and the vector be $u_i = (u_{1i}, \ldots, u_{ni})'$, $i = 1, \ldots, n$ and the $j^{th}$ latent root of $\mathbf{\Psi}$ be $\beta_j$ and the vector be $v_j = (v_{1j}, \ldots, v_{pj})'$, $j = 1, \ldots, p$. Then, the $k^{th}$ or $(ij)^{th}$ latent root of $\mathbf{\Omega} = \mathbf{\Phi} \otimes \mathbf{\Psi}$ is

$$\lambda_k = \alpha_i \beta_j \tag{2.4}$$

and the vector is

$$\gamma_k = u_i \otimes v_j = (u_{1i} y_j', \ldots, u_{ni} y_j') \tag{2.5}$$

for $k = 1, \ldots, np$. Now we can find $\mathbf{\Omega} = (\Gamma D_\lambda^{-\frac{1}{2}})(\Gamma D_\lambda^{-\frac{1}{2}})' = UU'$

The required computation is simplified by computing separate latent roots and vectors for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ and using the above theorem.

Sometimes, one or both of the matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ have structure. When they have structure we can use the following results for latent roots and vectors of patterned matrices.

## 3. Patterned Matrices

We will consider the factorizations when we have structured or patterned matrices. Let the patterned matrix either $\mathbf{\Phi}$ or $\mathbf{\Psi}$ be $A = WD_\delta W'$ with dimensions $q$.

If $A$ were the *intraclass* matrix

$$A = \begin{pmatrix} a & b & b & \cdots & b \\ & a & b & \cdots & b \\ & & \ddots & & \vdots \\ & & & & b \\ & & & & a \end{pmatrix}, \tag{3.1}$$

with $a$ on the diagonal and $b$ off the diagonal, then the latent roots are

$$\delta_1 = a + (N-1)b,\ \delta_2 = \cdots = \delta_q = a - b \tag{3.2}$$

and vectors are any $q$ mutually orthogonal vectors of unit length, the first of which has components that are identical. For example the columns of the following Helmert matrix

$$W = \begin{pmatrix} \frac{1}{\sqrt{q}} & \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \cdots & \frac{1}{\sqrt{(q-1)\cdot q}} \\ \frac{1}{\sqrt{q}} & -\frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \cdots & \\ \vdots & 0 & -\frac{2}{\sqrt{2\cdot 3}} & \cdots & \vdots \\ & \vdots & & \ddots & \\ \frac{1}{\sqrt{q}} & 0 & \cdots & 0 & -\frac{(q-1)}{\sqrt{(q-1)\cdot q}} \end{pmatrix}. \tag{3.3}$$

If $A$ were the following *tri-diagonal* matrix

$$A = \begin{pmatrix} a & b & & & 0 \\ b & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & & a & b \\ 0 & & & b & a \end{pmatrix}, \tag{3.4}$$

corresponding to a distributed lag model with $a$ on the diagonal and $b$ on the super and sub diagonals, then the latent roots are

$$\delta_l = a + 2b\ \cos\left(\frac{l\pi}{q+1}\right),\ l = 1, \ldots, q \tag{3.5}$$

and vectors are

$$w_l = \left(\sin\left(\frac{l\pi}{q+1}\right), \sin\left(\frac{2l\pi}{q+1}\right), \ldots, \sin\left(\frac{ql\pi}{q+1}\right)\right),\ l = 1, \ldots, q. \tag{3.6}$$

If $A$ were the *circular* matrix

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_q \\ a_q & a_1 & \cdots & a_{q-1} \\ \vdots & & & \vdots \\ a_2 & a_3 & \cdots & a_1 \end{pmatrix}, \tag{3.7}$$

with symmetry imposed on it so that it were a covariance matrix with applications to cyclical time series, then the latent roots are

$$\delta_l = \sum_{s=1}^{q} a_{s-1} \cos\left[\frac{2\pi}{q}(s-1)(l-1)\right],\ l = 1, \dots, q \tag{3.8}$$

and the elements of $W$ are

$$w_{lm} = \frac{1}{\sqrt{q}}\left\{\cos\left[\frac{2\pi}{q}(l-1)(m-1)\right] + \sin\left[\frac{2\pi}{q}(l-1)(m-1)\right]\right\}. \tag{3.9}$$

If $A$ were the Toeplitz matrix corresponding to a *first order Markov* correlation scheme

$$A = \begin{pmatrix} 1 & a & a^2 & \cdots & a^{N-1} \\ a & 1 & a & \cdots & a^{N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ a^{N-1} & a^{N-2} & & \cdots & 1 \end{pmatrix} \tag{3.10}$$

with applications to time series, for example an AR(1), then the latent roots are

$$\delta_1 = 1,\ \delta_2 = \cdots = \delta_q = 1 - a^2 \tag{3.11}$$

but the a general form for the latent vectors $w_l$ is not known to the current author.

It has been found by the current author that a matrix with the above structure can be factorized by

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a & \sqrt{1-a^2} & 0 & \cdots & 0 \\ a^2 & a\sqrt{1-a^2} & \sqrt{1-a^2} & \ddots & \vdots \\ a^3 & a^2\sqrt{1-a^2} & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & \end{pmatrix}, \tag{3.12}$$

where $A = TT'$.

But if $\mathbf{\Phi} = A$ where $A$ is the first order Markov correlation matrix in $\mathbf{\Omega} = \mathbf{\Phi} \otimes \mathbf{\Psi}$ we still do not have an expression for $\mathbf{\Omega} = UU'$.

An approximation to

$$\mathbf{\Omega} \equiv \begin{pmatrix} \mathbf{\Psi} & a\mathbf{\Psi} & a^2\mathbf{\Psi} & \cdots & a^{N-1}\mathbf{\Psi} \\ & \mathbf{\Psi} & & & \\ & & \ddots & & \vdots \\ & & & & \\ & & & & \mathbf{\Psi} \end{pmatrix}, \tag{3.13}$$

can be made which is

$$\Omega = \begin{pmatrix} \Psi & \dots & a^b\Psi & & 0 \\ \vdots & & & \ddots & \\ a^b\Psi & \ddots & \ddots & & \\ & & & & \\ 0 & & & & \end{pmatrix}, \tag{3.14}$$

where it is assumed that $a^{b+1}$ is small enough to be neglected and a routine for exact factorization of band symmetric matrices can be used.

## 4. Conclusion

Factorization of separable or patterned matrices can be performed by taking advantage of the abovementioned theorem and results for latent roots and latent vectors. Implementation of these factorization techniques will show that they are much faster than direct factorization when the dimension of the covariance matrix to be factorized becomes large.

## References

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Inc., New York, p. 600.

[2] Gelfand, A. E. and Smith, A. F. M. (1990). “Sampling based approaches to calculating marginal densities.” *J. Amer. Stat. Assoc.* 85, 398-409.

[3] Geman, S. and Geman, D. (1984). “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.” *IEEE Trans. on pattern analysis and machine intelligence* 6, 721-741.

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

[5] Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. John Wiley and Sons, Inc., New York.