Covariance Decompositions for Accurate Computation in Bayesian Scale-Usage Models

Online Supplemental Materials

Chris Hans[1], **Greg M. Allenby**[2], **Peter F. Craigmile**[1],
Ju Hee Lee[1], **Steven N. MacEachern**[1], **and Xinyi Xu**[1]

[1]Department of Statistics, The Ohio State University, Columbus, OH 43210.

[2]Department of Marketing, The Ohio State University, Columbus, OH 43210.

C Optimal Decomposition in Special Cases

C.1 Independence structure

Suppose that the distribution of $\mathbf{Y}_i \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ is normal with diagonal covariance matrix $\boldsymbol{\Sigma}$. Then the optimal $\mathbf{D}$ is $\mathbf{D} = \boldsymbol{\Sigma}$, and all the eigenvalues of $\mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$ are 1. Thus, $\lambda_1(\mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}) = 1$ and $\mathbf{R} = \boldsymbol{\Sigma} - \mathbf{D} = \mathbf{0}$. Convergence is immediate, and the Markov chain yields independent draws from the limiting distribution.

Suppl. Example 1 *(One-way ANOVA, independence prior) The case of* $\boldsymbol{\Sigma} = (\sigma_a^2 + \sigma_e^2)\mathbf{I}$ *with* σ_a^2 *and* σ_e^2 *known corresponds to a model with an i.i.d.* $N(0, \sigma_a^2)$ *treatment effect plus an i.i.d.* $N(0, \sigma_e^2)$ *error. The optimal* $\mathbf{D}$ *maximizing the convergence rate is* $\boldsymbol{\Sigma}$.

C.2 Exchangeable correlation structure

Suppose that, among the coordinates of $\mathbf{Y}_i \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$, a set of variables is exchangeable in the sense that the correlation matrix $\mathbf{C}$ remains unchanged under any permutation of these variables. In this case, the following proposition shows that an optimal $\mathbf{D}$ based on the correlation matrix $\mathbf{C} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2}$ has equal diagonal elements at the exchangeable coordinates.

Suppl. Proposition 1 *Suppose that $\mathbf{Y}_i \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ follows a multivariate normal distribution, and there exist two variables (Y_{ij}, Y_{ik}) that are exchangeable. Then there exists an optimal matrix $\mathbf{D} = diag(d_1, \cdots, d_M)$ based on the correlation matrix $\mathbf{C}$ with $d_j = d_k$.*

Proof of Suppl. Proposition 1 Assume that $\mathbf{D}_0 = \mathrm{diag}(d_0^1, \cdots, d_0^M) \in \mathrm{DS}(\mathbf{C})$ is an optimal matrix that minimizes the largest eigenvalue value of $\mathbf{D}_0^{-1/2}\mathbf{C}\mathbf{D}_0^{-1/2}$ and $d_0^i \neq d_0^j$. We can construct a new diagonal matrix $\mathbf{D}_1 = \mathrm{diag}(d_1^1, \cdots, d_1^M)$ where $d_1^k = d_0^k$ for $k \neq i$ or j, $d_1^i = d_0^j$ and $d_i^j = d_0^i$, i.e., $\mathbf{D}_1$ swaps the positions of d_0^i and d_0^j. This corresponds to a relabeling of coordinates i and j. Thus, $\mathbf{D}_1 \in \mathrm{DS}(\mathbf{C})$, and the eigenvalues of $\mathbf{D}_1^{-1/2}\mathbf{C}\mathbf{D}_1^{-1/2}$ are the same as those of $\mathbf{D}_0^{-1/2}\mathbf{C}\mathbf{D}_0^{-1/2}$, although the eigenvectors may differ. Now let $\mathbf{D}^* = \frac{1}{2}(\mathbf{D}_0 + \mathbf{D}_1)$, then $\mathbf{D}^*$ is still in $\mathrm{DS}(\mathbf{C})$, and by a well-known result in linear algebra (see, e.g., Bhatia, 1996),

$$\begin{aligned} \lambda_1((\mathbf{D}^*)^{-1/2}\mathbf{C}(\mathbf{D}^*)^{-1/2}) &\leq \frac{1}{2}\left[\lambda_1(\mathbf{D}_0^{-1/2}\mathbf{C}\mathbf{D}_0^{-1/2}) + \lambda_1(\mathbf{D}_1^{-1/2}\mathbf{C}\mathbf{D}_1^{-1/2})\right] \\ &= \lambda_1(\mathbf{D}_0^{-1/2}\mathbf{C}\mathbf{D}_0^{-1/2}). \end{aligned}$$

Therefore, $\lambda_1((\mathbf{D}^*)^{-1/2}\mathbf{C}(\mathbf{D}^*)^{-1/2})$ is at most as large as $\lambda_1(\mathbf{D}_0^{-1/2}\mathbf{C}\mathbf{D}_0^{-1/2})$, and so $\mathbf{D}^*$ is an optimal matrix based on the correlation matrix. ‡

The key idea in Suppl. Proposition 1 is that, for a given i, the Y_{ij}'s can be reordered without changing the correlation matrix, then the corresponding diagonal elements of a matrix $\mathbf{D}$ can be reordered in the same way without affecting the convergence rate of the Markov chain. Convexity suggests that averaging $\mathbf{D}$ and its reordered version can only hasten convergence. This argument can be applied or extended in various cases to obtain optimal decompositions. We illustrate this procedure for the following well known models.

Case 1: Exchangeable correlation structure Consider the exchangeable correlation structure, where all coordinates of $\mathbf{Y}_i \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ are exchangeable, i.e., the correlation matrix $\mathbf{C}$ is of the form $\mathbf{C} = a\mathbf{I} + b\mathbf{J}$ with $a = 1 - b$. The following corollary shows that in this situation, an optimal $\mathbf{D}$ for $\boldsymbol{\Sigma}$ is proportional to $\mathrm{diag}(\boldsymbol{\Sigma})$.

Suppl. Corollary 1 *If the correlation matrix $\mathbf{C}$ is of the form $a\mathbf{I} + b\mathbf{J}$ where $a = 1 - b$, then the matrix $\mathbf{D} = d\ diag(\boldsymbol{\Sigma})$ is optimal, where $d = 1 - b$ if $b \geq 0$ and $d = 1 + (M-1)b$ if $b < 0$.*

Proof of Suppl. Corollary 1 By Proposition 2 in Section 4.1 of the main text of the paper, it suffices to show that an optimal $\mathbf{D}$ based on the correlation matrix is $d\mathbf{I}$. Since, when the correlation matrix is $a\mathbf{I}+b\mathbf{J}$, any pair within $Y_{i1}, \cdots, Y_{iM} \mid \mu, \tau_i, \sigma_i^2$ is exchangeable, by Suppl. Proposition 1 an optimal $\mathbf{D}$ is proportional to the identity matrix. Moreover, it is easy to see that for any fixed t, the largest eigenvalue of $(t\mathbf{I})^{-1/2}\mathbf{\Sigma}(t\mathbf{I})^{-1/2}$ is $\lambda_1(\mathbf{C})/t$. Thus, to retain $\mathbf{R}$ as a non-negative definite matrix, the largest t that can be used is $t_{\max} = \lambda_M(\mathbf{C})$, which is $1-b$ if $b \geq 0$ and is $1+(M-1)b$ if $b<0$. ‡

Suppl. Example 2 *(One-way ANOVA, hierarchical prior) With the one-way ANOVA model now assume that the prior distribution for the treatment is hierarchical. The center of the distribution of the treatment effects, μ, follows the $N(0, \sigma_\mu^2)$ distribution. The M treatment effects are jointly $N(\mu\mathbf{1}, \mathbf{I})$, conditional on their center. This implies that the treatment effects are jointly $N(\mathbf{0}, \mathbf{I}+\sigma_\mu^2\mathbf{J})$, and thus the correlation matrix is $\mathbf{C} = a\mathbf{I}+b\mathbf{J}$, where $a = 1-b$ and $b = \sigma_\mu^2/(1+\sigma_e^2+\sigma_\mu^2)$. Appealing to Corollary 1, the optimal $\mathbf{D}$ is $(1+\sigma_e^2)/(1+\sigma_e^2+\sigma_\mu^2)\mathbf{I}$.*

Case 2: Circular correlation structure Consider the circular correlation structure, where $Y_{ik}, \cdots, Y_{iM}, Y_{i1}, \cdots, Y_{i,k-1} \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ is the same as the distribution of $Y_{i1}, \cdots, Y_{iM} \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ for any $k = 1, \ldots, M$. Thus, the covariance matrix remains the same under a circular transformation of the coordinates. In this case, we can easily extend the symmetric argument in Suppl. Proposition 1 to show that an optimal $\mathbf{D}$ is a multiplier of the identity matrix.

Suppl. Corollary 2 *Suppose that $Y_{i1}, \cdots, Y_{iM} \mid \boldsymbol{\mu}, \tau_i, \sigma_i^2$ follows a multivariate normal distribution and its covariance matrix is invariant under circular transformations. Then an optimal $\mathbf{D}$ in the decomposition $\mathbf{\Sigma} = \mathbf{D}+\mathbf{R}$ is $\mathbf{D} = t\mathbf{I}$, where $t = \lambda_M(\mathbf{\Sigma})$.*

Proof of Suppl. Corollary 2 Following similar steps as in the proof of Proposition 1, we can see that there exists an optimal $\mathbf{D}$ that satisfies $(d_k, \cdots, d_M, d_1, \cdots, d_{k-1}) = (d_1, \cdots, d_M)$ for each $k = 1, \ldots, M$. That is, $d_1 = \cdots = d_M = t$ for some $t > 0$. To retain $\mathbf{R}$ as a nonnegative definite matrix, the largest t that we can choose is $t_{\max} = \lambda_M(\mathbf{C})$. Therefore, this optimal $\mathbf{D}$ is equal to $\lambda_M(\mathbf{\Sigma})\mathbf{I}$. ‡

We demonstrate this theory in the case of the circular AR(1) process. Further details of the process, eigenvalue decomposition, and another example (the circular MA(1) process) are provided in Appendix D below.

Suppl. Example 3 *(Circular AR(1) process) For $M > 2$, the stationary circular autoregressive process of order 1 has a covariance matrix $\mathbf{\Sigma}$ that is circular, with $\Sigma_{jk} = \gamma_{i-j \bmod M}$ and $\gamma_h = \sigma^2(\phi^{|h|} + \phi^{M-|h|})/((1-\phi^2)(1-\phi^M))$, $h = 0, \dots, M-1$, for autocorrelation parameter $-1 < \phi < 1$ and innovation variance $\sigma^2 > 0$. The smallest eigenvalue of $\mathbf{\Sigma}$ when $\phi \geq 0$ is $\lambda_M(\mathbf{\Sigma}) = \sigma^2/(1 - 2\phi\cos(2\pi\lfloor M/2\rfloor/M) + \phi^2)$ (which simplifies to $\lambda_M(\mathbf{\Sigma}) = \sigma^2/(1+\phi)^2$ when M is even). When $\phi < 0$, the smallest eigenvalue is $\lambda_M(\mathbf{\Sigma}) = \sigma^2/(1-\phi)^2$. Therefore, by Corollary 2 an optimal $\mathbf{D}$ based on $\mathbf{\Sigma}$ is $\lambda_M(\mathbf{\Sigma})\mathbf{I}$.*

Case 3: Reversible correlation structure Consider the reversible correlation structure, where the correlation matrix remains unchanged when the order of the variables is reversed. In this case, we characterize a property of the optimal $\mathbf{D}$. The proof follows the argument in Suppl. Proposition 1.

Suppl. Corollary 3 *Suppose that the correlation matrix has the form $C_{ij} = C_{M+1-i\, M+1-j}$, for $j = 1, \dots, M$. Then the optimal $\mathbf{D}$ matrix based on $\mathbf{C}$ has, with diag$(\mathbf{D}) = (d_1, \dots, d_M)$, $d_i = d_{M+1-i}$ for $i = 1, \dots, M$.*

Suppl. Example 4 *(AR(1) process). An autoregressive process of order 1 has a correlation matrix which is reversible. Direct application of Suppl. Corollary 3 implies that the optimal $\mathbf{D}$ is symmetric, with $d_j = d_{M+1-j}$ for $j = 1, \dots, M$.*

C.3 Block diagonal structure

Suppose that the observations $Y_{i1}, \cdots, Y_{iM}$ $(i = 1, \dots, N)$ can be divided into several parts where variables in different parts are independent conditional on $\boldsymbol{\mu}, \tau_i, \sigma_i^2$, i.e., the covariance matrix $\mathbf{\Sigma}$ is a block diagonal matrix. Then we can divide the matrix $\mathbf{D}$ into several corresponding blocks. The optimization problems in different blocks are independent, and the overall convergence rate is determined by the "worst" block. The next result shows that one can optimize each block separately and then paste the pieces together to get a big $\mathbf{D}$ matrix.

Suppl. Proposition 2 *Suppose the covariance matrix* $\mathbf{\Sigma}$ *is block diagonal with* k *blocks, and* $\mathbf{D} \in DS(\mathbf{\Sigma})$ *is diagonal. Then the matrix* $\mathbf{D}^{-1/2}\mathbf{\Sigma}\mathbf{D}^{-1/2}$ *is also block diagonal with* k *blocks, and its eigenvalues are of the form* $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k$*, where* $\boldsymbol{\lambda}_i$ *is the vector of eigenvalues from the* i^{th} *block. Solving each block and collecting the results together guarantees an optimal solution.*

Suppl. Example 5 *(Two-way ANOVA) The prior distribution for the* J *treatment effects is* $N(\mu\mathbf{1}, \sigma_\alpha^2\mathbf{I})$*, conditional on a known value* μ*. The* n_j *replicate measurements on treatment* j *are conditionally independent, with mean equal to the treatment mean and variance* σ_e^2*. The covariance matrix of* $\mathbf{Y}_i$ *is* $\mathbf{\Sigma} = diag(\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_J)$*, with* $\mathbf{\Sigma}_i = \sigma_e^2\mathbf{I} + \sigma_\alpha^2\mathbf{J}$*. By Suppl. Proposition 2 and Suppl. Corollary 1, an optimal* $\mathbf{D} = diag(\mathbf{D}_1, \ldots, \mathbf{D}_J)$ *where* $\mathbf{D}_j = (1+\sigma_e^2)/(1+\sigma_e^2+\sigma_\alpha^2)\mathbf{I}$.

D Eigenvalues of a circulant matrix

For a positive integer M, a circulant $M \times M$ covariance matrix $\mathbf{\Sigma}$ is defined by the relation $\Sigma_{jk} = \gamma_{j-k \bmod M}$, for M constants $\gamma_0, \ldots, \gamma_{M-1}$, such that $\mathbf{\Sigma}$ is positive definite. In this case there is a closed from expression for the eigenvalues of $\mathbf{\Sigma}$. The unordered eigenvalues are calculated using the discrete Fourier transform (DFT) of $\{\gamma_k : k = 0, \ldots, M-1\}$ (Gray, 2006), where $i = \sqrt{-1}$:

$$\psi_j = \sum_{k=0}^{M-1} \gamma_k e^{-i2\pi(j-1)k/M}, \qquad j = 1, \ldots, M.$$

An example of a process with a circular covariance matrix is the circular autoregressive process of order one. For an integer $M \geq 2$, let $\{U_t : t = 0, \ldots, M-1\}$ be a set of uncorrelated mean zero random variables with variance σ^2, such that $0 < \sigma^2 < \infty$. Then the circular AR(1) process is defined by the recursion, $\eta_t = \phi\eta_{t-1 \bmod M} + U_t$, $t = 0, \ldots, M-1$. For $|\phi| < 1$ this process is stationary and for each t we can express η_t as $\eta_t = (1-\phi^M)^{-1}\sum_{k=0}^{M-1}\phi^k U_{t-k \bmod M}$, which leads to that fact that $E(\eta_t) = 0$ for all t and for $|h| < M$,

$$\gamma_h = \mathrm{cov}(\eta_t, \eta_{t+h}) = \frac{\sigma^2(\phi^{|h|} + \phi^{M-|h|})}{(1-\phi^2)(1-\phi^M)}.$$

The DFT of $\{\gamma_k\}$ is

$$\psi_j \;=\; \sum_{k=0}^{M-1} \gamma_k e^{-i2\pi(j-1)k/M} \;=\; \frac{\sigma^2}{1-2\phi\cos(2\pi(j-1)/M)+\phi^2}, \quad j=1,\ldots,M.$$

For $\phi \geq 0$, the smallest eigenvalue occurs at $j = \lfloor M/2 \rfloor + 1$, with value $\sigma^2/(1-2\phi\cos(2\pi\lfloor M/2\rfloor/M)+\phi^2)$, which simplifies to $\sigma^2/(1+\phi^2)$ when M is even. For $\phi < 0$, the smallest eigenvalue occurs at $j = 1$ with value $\sigma^2/(1-\phi^2)$.

Another example is the circular moving average (MA) process of order one. For some integer $M \geq 3$ and $\theta \neq 0$, the circular MA(1) process is defined by $\eta_t = U_t + \theta U_{t-1 \bmod M}$, $t = 0, \ldots, M-1$, where $\{U_t\}$ was defined as for the circular AR(1) process. We restrict to the case that the process is invertible; i.e., when $|\theta| < 1$ (e.g., Brockwell and Davis, 2002). Then this process has mean zero with a covariance structure described, for $|h| < M$ by,

$$\gamma_h = \operatorname{cov}(\eta_t, \eta_{t+h}) = \begin{cases} \sigma^2(1+\theta^2), & h = 0 \\ \sigma^2\theta, & h = \pm 1, \pm(M-1) \\ 0, & \text{otherwise.} \end{cases}$$

The $M \times M$ covariance matrix again is circular with $\Sigma_{jk} = \gamma_{j-k \bmod M}$, and the unordered eigenvalues are $\psi_j = \sigma^2\left(1 + 2\theta\cos(2\pi(j-1)/M) + \theta^2\right)$ for $j = 1, \ldots, M$. When θ is positive, the minimum eigenvalue occurs at $j = \lfloor M/2 \rfloor + 1$, with value $\sigma^2\left(1 + 2\theta\cos(2\pi\lfloor M/2\rfloor/M) + \theta^2\right)$, which simplifies to a value of $\sigma^2\left(1-\theta\right)^2$ when M is even. When θ is negative, the minimum eigenvalue occurs at $j = 1$, with value $\sigma^2 = \sigma^2(1 + 2\theta + \theta^2) = \sigma^2(1+\theta)^2$.

References

Bhatia, R. (1996). *Matrix Analysis.* Springer, New York.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting.* Springer, New York.

Gray, R. M. (2006). *Toeplitz and Circulant Matrices: A review.* Now Publishers, Norwell, Massachusetts.