Covariance Structure Regularization via Entropy Loss Function

Lin, Lijing and Higham, Nicholas J. and Pan, Jianxin

2012

MIMS EPrint: **2012.61**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Covariance Structure Regularization via Entropy Loss Function

Lijing Lin, Nicholas J. Higham and Jianxin Pan*

School of Mathematics, University of Manchester,

Manchester M13 9PL, UK

June 10, 2012

Abstract

The need to estimate structured covariance matrices arises in a variety of applications and the problem is widely studied in statistics. We propose a new method for regularizing the covariance structure of a given covariance matrix, in which the underlying structure is usually blurred due to random noises particularly when the dimension of the covariance matrix is high. The regularization is made by choosing an optimal structure from an available class of covariance structures in terms of minimizing the discrepancy, defined via the entropy loss function, between the given matrix and the class. A range of potential candidate structures such as tridiagonal, compound symmetry, AR(1), and Toeplitz are considered. Simulation studies are conducted, showing that the proposed new approach is reliable in regularization of covariance structures. The approach is also applied to real data analysis, demonstrating the usefulness of the proposed approach in practice.

Keywords: Covariance estimation; Covariance structure; Entropy loss function; Kullback-Leibler divergence; Regularization.

1 Introduction

The need to estimate structured covariance matrices arises in a variety of application fields including signal processing (Pascal et al., 2008) and control problems (Lin and Jovanović, 2009) and the problem is widely studied in statistics; see, e.g.,

*Corresponding author. Tel: +44 161 2755864; Fax: +44 161 275 5819.
Email Addresses: Lijing.Lin@manchester.ac.uk (L. Lin); Nicholas.J.Higham@manchester.ac.uk (N. J. Higham); Jianxin.Pan@manchester.ac.uk (J. Pan)

Pourahmadi (1999) and Pan and Mackenzie (2003). A conventional way, known as the "Burg technique", is to find the maximum likelihood estimation for a covariance matrix that has a specific/regularized structure using random samples drawn from a stochastic process (Burg et al., 1982). However, this method has some drawbacks, including that (a) it is based on the presumption that the stochastic process is multivariate normal, (b) the structure of the covariance must be prespecified, and (c) the sample covariance matrix must be available. This method fails when the dimension of the covariance matrix is large or close to the sample size, because the sample covariance matrix is then ill conditioned. Furthermore, it is difficult to deduce the underlying covariance structure from the sample covariance matrix, because of random noise or large dimension of the matrix, for example.

To overcome these difficulties, in this paper we propose a new method for regularizing the underlying structure of a given covariance matrix. Our method is based on the entropy loss function (Dey and Srinivasan, 1985; James and Stein, 1961)

$$L(A, B) = \mathrm{tr}(A^{-1}B) - \log(\det(A^{-1}B)) - m, \tag{1.1}$$

where A and B are $m \times m$ matrices and, to ensure that $\log\det(A^{-1}B)$ is real, we assume that A and B are symmetric positive definite. The entropy loss function, also known as the Kullback-Leibler divergence, is a well-accepted nonsymmetric measure of the discrepancy between two probability distributions (Pan and Fang, 2002). It is a special case of the Bregman divergence (Dhillon and Tropp, 2007) and has been widely used in statistics (Pan and Fang, 2002). The problem of interest here is, given a covariance matrix A whose underlying structure is blurred due to random noises particularly when the dimension m is high, to identify the underlying structure of A from a class of candidate covariance structures. To demonstrate our idea, we introduce the following notation. Let $\mathcal{S}$ be the set of all positive definite covariance matrices with structure s. We define the discrepancy between a given positive definite covariance matrix A and the set $\mathcal{S}$ by

$$D(A, \mathcal{S}) = \min_{B \in \mathcal{S}} L(A, B), \tag{1.2}$$

where $L(A, B)$ is the entropy loss function in (1.1). Our idea is that, among a given class of k candidate covariance structures $\{s_1, s_2, \ldots, s_k\}$, the structure with which A has the smallest discrepancy can be viewed as the possible underlying structure of A. We refer to the replacement of A by a matrix B achieving the minimum in (1.2) as the process as regularizing A. It is worth pointing out that the matrix A is not necessarily a sample covariance matrix. It can be any estimator of a covariance matrix, obtained by statistical methods such as those based on modified Cholesky decomposition methods (Pan and Mackenzie, 2003; Ye and Pan, 2006). Regularization of the given covariance matrix helps the understanding of the underlying correlation/covariance process and simplifies complicated and high-dimensional data problems.

In this paper we consider the following four candidate covariance structures that are commonly used in practice, for example, in longitudinal and spatial studies.

(1) The order-1 moving average structure, MA(1), has a tridiagonal and Toeplitz covariance matrix

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & 0 \\ c & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & c & 1 \end{bmatrix}, \tag{1.3}$$

where $\sigma^2 > 0$ and $-1/\cos(\pi/(m+1)) < c < 1/\cos(\pi/(m+1))$.

(2) The covariance of compound symmetry (CS) structure assumes that the correlation coefficients of any two observations are the same, i.e.,

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & c \\ c & 1 & \cdots & c \\ \vdots & \ddots & \ddots & \vdots \\ c & \cdots & c & 1 \end{bmatrix}, \tag{1.4}$$

where $\sigma^2 > 0$ and $-1/(m-1) < c < 1$.

(3) The covariance of autoregression of order 1, AR(1), has the property that the correlation between any pair of observations decays exponentially towards zero as the distance between two observations increases. It is of the form

$$B = \sigma^2 \begin{bmatrix} 1 & c & c^2 & \cdots & c^{m-1} \\ c & 1 & c & \cdots & c^{m-2} \\ c^2 & c & 1 & \cdots & c^{m-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{bmatrix}, \tag{1.5}$$

where $\sigma^2 > 0$ and $-1 < c < 1$.

(4) More generally, banded Toeplitz covariance matrices have constant subdiagonal entries, i.e., constants at lag 1, lag 2, ..., and lag p:

$$B = \sigma^2 \begin{bmatrix} 1 & c_1 & \cdots & c_p & \cdots & 0 \\ c_1 & 1 & c_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & c_p \\ c_p & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & c_1 \\ 0 & \cdots & c_p & \cdots & c_1 & 1 \end{bmatrix}, \tag{1.6}$$

where $\sigma^2 > 0$ and $c_1, c_2, \ldots, c_p$ are nonzero.

The main task now is to calculate the discrepancy $D(A, \mathcal{S})$ for each of the candidate covariance structures listed in (1)–(4) above, where the covariance matrix A is given. Equivalently, it is to find for each covariance structure a positive definite matrix B that minimizes the discrepancy $L(A, B)$ within the set of matrices with that structure. Accordingly, structure s in the candidate class $\{s_1, ..., s_k\}$ that has the smallest discrepancy is the most likely covariance structure, among the candidate class, for the matrix A.

The rest of this paper is organized as follows. In section 2, we consider first the case without a structure constraint. Our result shows that, given a symmetric positive definite covariance matrix A, $L(A, B) = 0$ if and only if $B = A$. We then show in section 3 that the problem of finding B with structure (1.3)–(1.5) that minimizes $L(A, B)$ reduces to computing the zeros of a nonlinear function in one variable. In section 4 we discuss solution of the problem for Toeplitz covariance structure. In section 5 we carry out simulation studies, illustrating how our techniques of computing the structured covariance matrix that minimizes the entropy loss function can be used in regularizing the underlying covariance structure. We also apply the proposed approach to some real data experiments. Some further remarks and discussion are given in section 6.

2 Unconstrained case

We first show that the problem of finding a general covariance matrix B that minimizes the entropy loss function $L(A, B)$ for a given A is trivial. Define $f : \mathbb{R}_+^{m\times m} \to \mathbb{R}$ where $\mathbb{R}_+^{m\times m}$ is the set of all $m \times m$ symmetric positive definite matrices and $f(B) := L(A, B)$. We denote by $\nabla_B f = (\partial f/\partial b_{ij})$ the gradient of f, where b_{ij} is the (i, j) entry of B. Ignoring the symmetry of A and B and using results from Magnus and Neudecker (1999) we have

$$\nabla_B \operatorname{tr}(A^{-1}B) = A^{-T} = A^{-1},$$
$$\nabla_B \log\det(B) = B^{-T} = B^{-1},$$

and then

$$\nabla_B f = A^{-1} - B^{-1}. \tag{2.7}$$

Write $b = \operatorname{vec}(B) \in \mathbb{R}^{m^2}$ where vec denotes the vector obtained by stacking the columns of its matrix argument on top of each other from first to last. Taking f as a function from $\mathbb{R}^{m^2}$ to $\mathbb{R}$, the Hessian of f is then given by

$$\nabla_b^2 f := \left(\frac{\partial^2 f}{\partial b_i \partial b_j}\right) = B^{-T} \otimes B^{-1} = B^{-1} \otimes B^{-1} \tag{2.8}$$

(Magnus and Neudecker, 1999). Since B is positive definite, $B^{-1} \otimes B^{-1}$ is positive definite, and so $f(B)$ is a strictly convex function of B. Therefore, $f(B)$ has a global minimum which is attained at $\nabla_B f = 0$, i.e., $B = A$.

We mention in passing that provided A is symmetric positive definite, the entropy loss function $L(A,B)$ at the boundary of the set of symmetric positive definite matrices is $+\infty$, that is, $L(A,B) \to +\infty$ as $\det(B) \to 0$.

3 Two-parameter problems

We begin by considering the two-parameter matrices (1.3)–(1.5), for which the problems reduces to minimizing a function of two variables. Define the map $B(c,\sigma) : \mathbb{R}^2 \to \mathbb{R}_+^{m\times m}$ where $B(c,\sigma)$ are as in (1.3)–(1.5) and $g(c,\sigma) := f(B(c,\sigma)) : \mathbb{R}^2 \to \mathbb{R}$, where $f(B) := L(A,B)$. Note that here $\sigma > 0$. We now take $B(c,\sigma)$ as a map from $\mathbb{R}^2$ to $\mathbb{R}^{m^2}$ and denote the Jacobian matrix of $B(c,\sigma)$ by $\nabla_{c,\sigma}B$, i.e.,

$$\nabla_{c,\sigma}B = \begin{bmatrix} \partial b_1/\partial c & \partial b_1/\partial \sigma \\ \partial b_2/\partial c & \partial b_2/\partial \sigma \\ \vdots & \vdots \\ \partial b_{m^2}/\partial c & \partial b_{m^2}/\partial \sigma \end{bmatrix} \in \mathbb{R}^{m^2\times 2}, \tag{3.9}$$

where $b = \operatorname{vec}(B(c,\sigma))$. We denote the second order Jacobian matrix $B(c,\sigma)$ by $\nabla^2_{c,\sigma}B \in \mathbb{R}^{m^2\times 2\times 2}$. By applying the chain rule, we obtain the gradient of g

$$\nabla_{c,\sigma}g = (\nabla_{c,\sigma}B)^T\nabla_B f \ \in \mathbb{R}^2, \tag{3.10}$$

where $\nabla_B f := (\partial f/\partial b_i) \in \mathbb{R}^{m^2}$, and the Hessian

$$\nabla^2_{c,\sigma}g = (\nabla_{c,\sigma}B)^T\nabla^2_B f\nabla_{c,\sigma}B + (\nabla_B f)^T\nabla^2_{c,\sigma}B \ \in \mathbb{R}^{2\times 2}. \tag{3.11}$$

3.1 Tridiagonal matrices

Recall that the tridiagonal matrix (1.3)

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & 0 \\ c & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & c & 1 \end{bmatrix}$$

can be rewritten

$$B(c,\sigma) = \sigma^2(I + cT_1), \tag{3.12}$$

where T_1 is a symmetric matrix with the first superdiagonal and subdiagonal equal to 1 and all other elements equal to 0. Note that the eigenvalues of $B(c,\sigma)$ are (Higham, 2002, Sec. 28.5)

$$\lambda_j = \sigma^2(1 + 2cs_j), \quad j = 1:m,$$

where $s_j = \cos(\pi j/(m+1))$, so that $\det(B(c,\sigma)) = \sigma^{2m}\prod_{j=1}^{m}(1+2cs_j)$ and hence $B(c,\sigma)$ is positive definite if and only if

$$-\frac{1}{2s_1} < c < \frac{1}{2s_1}. \tag{3.13}$$

Given a positive definite covariance matrix A, the loss function is now

$$f(c,\sigma) := \sigma^2\mathrm{tr}(A^{-1}) + c\sigma^2\mathrm{tr}(A^{-1}T_1) + \log(\det(A)) - m\log\sigma^2 - \sum_{j=1}^{m}\log(1+2cs_j) - m. \tag{3.14}$$

It follows that

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial c} \\ \dfrac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \sigma^2\mathrm{tr}(A^{-1}T_1) - \sum_{j=1}^{m}\dfrac{2s_j}{1+2cs_j} \\ 2\sigma\mathrm{tr}(A^{-1}) - 2m/\sigma + 2c\sigma\mathrm{tr}(A^{-1}T_1) \end{bmatrix} \tag{3.15}$$

and

$$\nabla^2 f := \begin{bmatrix} \dfrac{\partial^2 f}{\partial c^2} & \dfrac{\partial^2 f}{\partial c\partial\sigma} \\ \dfrac{\partial^2 f}{\partial c\partial\sigma} & \dfrac{\partial^2 f}{\partial\sigma^2} \end{bmatrix} = \begin{bmatrix} \displaystyle\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} & 2\sigma\mathrm{tr}(A^{-1}T_1) \\ 2\sigma\mathrm{tr}(A^{-1}T_1) & 2\mathrm{tr}(A^{-1}) + \dfrac{2m}{\sigma^2} + 2c\mathrm{tr}(A^{-1}T_1) \end{bmatrix}. \tag{3.16}$$

The stationary points (c,σ) satisfy the equations

$$\begin{cases} \sigma^2 = \displaystyle\sum_{j=1}^{m}\frac{2s_j}{1+2cs_j}/\mathrm{tr}(A^{-1}T_1) \\ h(c) := \displaystyle\sum_{j=1}^{m}\frac{2s_j}{1+2cs_j} - \frac{m\mathrm{tr}(A^{-1}T_1)}{\mathrm{tr}(A^{-1}) + c\mathrm{tr}(A^{-1}T_1)} = 0. \end{cases} \tag{3.17}$$

Recalling (3.13), as $c \to -1/(2s_1)$ we have $2s_1/(1+2cs_1) \to +\infty$, so $h(c) \to +\infty$, while as $c \to 1/(2s_1) = -1/(2s_m)$ we have $2s_m/(1+2cs_m) \to -\infty$ and so $h(c) \to -\infty$. Therefore, there exists at least one zero of $h(c)$ on $[-1/(2s_1), 1/(2s_1)]$. Since for every c satisfying $h(c) = 0$,

$$\begin{aligned} h'(c) &= -\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} + \frac{m(\mathrm{tr}(A^{-1}T_1))^2}{(\mathrm{tr}(A^{-1}) + c\mathrm{tr}(A^{-1}T_1))^2} \\ &= -\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} + \frac{1}{m}\left(\sum_{j=1}^{m}\frac{2s_j}{1+2cs_j}\right)^2 < 0, \end{aligned}$$

where the last inequality is from the Cauchy–Schwarz inequality $|x^T e| \le \|x\|_2\|e\|_2$, with $e = [1, 1\ldots, 1]^T$, there exists only one zero of $h(c)$ on $[-1/(2s_1), 1/(2s_1)]$ and thus a unique stationary point of $f(c,\sigma)$ satisfying (3.17).

It is clear that the $(1,1)$ element of the Hessian matrix $\nabla^2 f$ in (3.16) is positive and it can be easily verified that for any (c,σ) satisfying (3.17), we have

$$\begin{aligned}
\det(\nabla^2 f(c,\sigma)) &= \frac{4m}{\sigma^2}\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} - 4\sigma^2(\mathrm{tr}(A^{-1}T_1))^2 \\
&= \frac{4}{\sigma^2}\left(m\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} - \sigma^4(\mathrm{tr}(A^{-1}T_1))^2\right) \\
&= \frac{4}{\sigma^2}\left(m\sum_{j=1}^{m}\frac{(2s_j)^2}{(1+2cs_j)^2} - \left(\sum_{j=1}^{m}\frac{2s_j}{1+2cs_j}\right)^2\right) \\
&> 0,
\end{aligned}$$

where the last inequality is from the Cauchy–Schwarz inequality. Therefore the Hessian matrix $\nabla^2 f$ is positive definite and so the stationary point is a minimum point. Note that $f(c,\sigma)$ is defined on the open set $\{(c,\sigma) : \sigma > 0 \text{ and } (3.13) \text{ holds}\}$ on which $B(c,\sigma)$ is positive definite. As (c,σ) approaches the boundary of the set or as $\sigma \to \infty$, we have $f(c,\sigma) \to +\infty$. It follows immediately that the global minimum of $f(c,\sigma)$ is obtained at the unique stationary point.

We summarize the discussion above in the following theorem.

Theorem 3.1 *Given a positive definite covariance matrix A, there exists a unique tridiagonal positive definite matrix $B(c,\sigma)$ of the form* (3.12) *that minimizes the loss function $f(c,\sigma) := L(A, B(c,\sigma))$ given by* (3.14)*. Furthermore, the minimum is attained at (c,σ) satisfying* (3.17)*.*

3.2 Compound symmetry

The matrix in (1.4) can be rewritten as

$$B(c,\sigma) = \sigma^2\begin{bmatrix} 1 & c & \cdots & c \\ c & 1 & \cdots & c \\ \vdots & \ddots & \ddots & \vdots \\ c & \cdots & c & 1 \end{bmatrix} = \sigma^2(I + c(ee^T - I)). \tag{3.18}$$

The eigenvalues of $B(c,\sigma)$ are $\sigma^2(1+(m-1)c)$ and $\sigma^2(1-c)$ of multiplicities 1 and $m-1$, respectively, so $B(c,\sigma)$ is a positive definite matrix if and only if (Borsdorf et al., 2010, Lem. 2.1)

$$-\frac{1}{m-1} < c < 1.$$

Given A, we define $f(c,\sigma) := L(A, B(c,\sigma))$, where $L(A,\cdot)$ is the entropy loss function in (1.1). We want to find an explicit solution to the corresponding optimization problem

$$\min_{\substack{\sigma>0 \\ -1/(m-1)<c<1}} f(c,\sigma). \tag{3.19}$$

First, it is clear that $\det(B(c,\sigma)) = \sigma^{2m}(1-c)^{m-1}(1+(m-1)c)$. Denoting $t := \mathrm{tr}(A^{-1}(ee^T - I))$, we have

$$\begin{aligned} f(c,\sigma) &= \sigma^2 \mathrm{tr}(A^{-1}) + c\sigma^2 t + \log(\det(A)) - m\log(\sigma^2) \\ &\quad - (m-1)\log(1-c) - \log(1+(m-1)c) - m. \end{aligned}$$

Second, we have

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial c} \\ \dfrac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \sigma^2 t + \dfrac{m-1}{1-c} - \dfrac{m-1}{1+(m-1)c} \\ 2\sigma \mathrm{tr}(A^{-1}) + 2\sigma c t - 2m/\sigma \end{bmatrix} \tag{3.20}$$

and

$$\nabla^2 f := \begin{bmatrix} \dfrac{\partial^2 f}{\partial c^2} & \dfrac{\partial^2 f}{\partial c \partial \sigma} \\ \dfrac{\partial^2 f}{\partial c \partial \sigma} & \dfrac{\partial^2 f}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{m-1}{(1-c)^2} + \dfrac{(m-1)^2}{(1+(m-1)c)^2} & 2\sigma t \\ 2\sigma t & 2(\mathrm{tr}(A^{-1}) + ct) + \dfrac{2m}{\sigma^2} \end{bmatrix}. \tag{3.21}$$

Therefore, the stationary points (c,σ) of $f(c,\sigma)$ must satisfy

$$\begin{cases} h(c) := \sigma^2 t + \dfrac{m-1}{1-c} - \dfrac{m-1}{1+(m-1)c} = 0 \\ \dfrac{m}{\sigma^2} = \mathrm{tr}(A^{-1}) + ct. \end{cases} \tag{3.22}$$

Third, it is clear that $h(c)$ is continuous in the interval $(-1/(m-1), 1)$. Since $h(c) \to +\infty$ as $c \to 1$ and $h(c) \to -\infty$ as $c \to -1/(m-1)$, there exists at least one solution to $h(c) = 0$. The stationary points are obtained immediately by solving $h(c) = 0$, which gives $c = -t/((m-1)\mathrm{tr}(A^{-1}) + (m-2)t)$.

Finally, since

$$(\nabla^2 f)_{11} = \frac{m-1}{(1-c)^2} + \frac{(m-1)^2}{(1+(m-1)c)^2} > 0$$

at the stationary points (c,σ) satisfying (3.22), and we also have

$$\begin{aligned} \det(\nabla^2 f) &= \frac{4m}{\sigma^2}\left(\frac{m-1}{(1-c)^2} + \frac{(m-1)^2}{(1+(m-1)c)^2}\right) - 4\sigma^2 t^2 \\ &= (m-1)\left(\frac{m-1}{1+(m-1)c} + \frac{1}{1-c}\right)^2 \\ &> 0, \end{aligned}$$

it follows that $\nabla^2 f$ is positive definite. Thus every stationary point is a minimum point. Since $f(c,\sigma)$ is defined on an open set such that $B(c,\sigma) > 0$ and as (c,σ) approaches the boundary or as $\sigma \to \infty$, we have $f(c,\sigma) \to +\infty$, the global minimum of $f(c,\sigma)$ is obtained at the unique stationary point.

We summarize the above discussion in the following theorem.

Theorem 3.2 *Given a positive definite covariance matrix $A \in \mathbb{R}^{m\times m}$, define $f(c,\sigma) := L(A, B(c,\sigma))$ where $B(c,\sigma)$ is a positive definite covariance matrix of compound symmetry in (3.18). Then the global minimum of $f(c,\sigma)$ over $\sigma > 0$ and $c \in (-1/(m-1), 1)$ is attained at*

$$\begin{cases} c = -\dfrac{t}{(m-1)\mathrm{tr}(A^{-1}) + (m-2)t}, \\ \dfrac{m}{\sigma^2} = \mathrm{tr}(A^{-1}) + ct, \end{cases} \tag{3.23}$$

where $t = \mathrm{tr}(A^{-1}(ee^T - I))$.

3.3 AR(1)

We rewrite B in (1.5) as

$$B(c,\sigma) = \sigma^2 \begin{bmatrix} 1 & c & c^2 & \cdots & c^{m-1} \\ c & 1 & c & \cdots & c^{m-2} \\ c^2 & c & 1 & \cdots & c^{m-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{bmatrix} = \sigma^2 \sum_{i=0}^{m-1} c^i T_i, \tag{3.24}$$

where $T_0 = I$ and T_i is a symmetric matrix with ones on the ith superdiagonal and subdiagonal and zeros elsewhere. It can be shown that the $k \times k$ leading principal minor of $B(c,\sigma)$ is $\sigma^{2k}(1-c^2)^{k-1}$, $k = 2 : m$ (Horn and Johnson, 1985, Prob. 7.12). Therefore, $B(c,\sigma)$ is a positive definite covariance matrix if and only if

$$-1 < c < 1. \tag{3.25}$$

The entropy loss function is now

$$f(c,\sigma) := \sigma^2 \sum_{i=0}^{m-1} c^i \mathrm{tr}(A^{-1}T_i) + \log\det(A) - m\log\sigma^2 - (m-1)\log(1-c^2) - m.$$

We find that

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial c} \\ \dfrac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \sigma^2 \sum_{i=1}^{m-1} i c^{i-1} \mathrm{tr}(A^{-1}T_i) + \dfrac{2(m-1)c}{1-c^2} \\ 2\sigma \sum_{i=0}^{m-1} c^i \mathrm{tr}(A^{-1}T_i) - 2m/\sigma \end{bmatrix}. \tag{3.26}$$

So the stationary points (c,σ) of $f(c,\sigma)$ satisfy

$$\begin{cases} \dfrac{m\sum_{i=1}^{m-1} i c^{i-1}\mathrm{tr}(A^{-1}T_i)}{\sum_{i=0}^{m-1} c^i \mathrm{tr}(A^{-1}T_i)} + \dfrac{2(m-1)c}{1-c^2} = 0 \\ \dfrac{m}{\sigma^2} = \displaystyle\sum_{i=0}^{m-1} c^i \mathrm{tr}(A^{-1}T_i). \end{cases} \tag{3.27}$$

Since $\sum_{i=0}^{m-1} c^i \mathrm{tr}(A^{-1}T_i) = \mathrm{tr}(A^{-1}B)/\sigma^2 > 0$ and $1 - c^2 \neq 0$ for $c \in (-1, 1)$, by rearranging the first equality in (3.27) we have

$$h(c) := m\sum_{i=1}^{m-1} ic^{i-1}\mathrm{tr}(A^{-1}T_i) - m\sum_{i=1}^{m-1} ic^{i+1}\mathrm{tr}(A^{-1}T_i) + 2(m-1)\sum_{i=0}^{m-1} c^{i+1}\mathrm{tr}(A^{-1}T_i) = 0.$$

Since $h(c)$ is continuous in $(-1, 1)$, $h(-1) = -2(m-1)e^T A^{-1}e < 0$ and $h(1) = 2(m-1)e^T A^{-1}e > 0$, there exists at least one root of $h(c)$ in $(-1, 1)$. Numerical experiments show that in some cases there exists more than one solution to $h(c) = 0$. We then can only expect to find a local minimum in general.

We summarize the discussion above in the following theorem.

Theorem 3.3 *Given a positive definite covariance matrix $A \in \mathbb{R}^{m\times m}$, define $f(c, \sigma) := L(A, B(c, \sigma))$ where $B(c, \sigma)$ is a positive definite covariance matrix of the AR(1) model as in* (3.24)*. Then the local minima of $f(c, \sigma)$ are attained at the points (c, σ) satisfying* (3.27)*.*

4 Toeplitz problems

The banded Toeplitz matrix (1.6) can be rewritten as

$$B(c, \sigma) = \sigma^2 \begin{bmatrix} 1 & c_1 & \cdots & c_p & \cdots & 0 \\ c_1 & 1 & c_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & c_p \\ c_p & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & c_1 \\ 0 & \cdots & c_p & \cdots & c_1 & 1 \end{bmatrix} = \sigma^2(I + \sum_{i=1}^{p} c_i T_i), \tag{4.28}$$

where $c = [c_1, c_2, \ldots, c_p]^T \in \mathbb{R}^p$ and T_i is a symmetric matrix with the ith super-diagonal and subdiagonal equal to 1 and zeros elsewhere. Define

$$p(t) = 1 + 2\sum_{k=1}^{p} c_k \cos(kt). \tag{4.29}$$

Then $B(c, \sigma)$ is positive-definite if and only if $p(t) \geq 0$, $p(t) \not\equiv 0$, for all $t \in \mathbb{R}$ (Parter, 1962, Remark II).

Recall that our problem is

$$\begin{aligned} \min \quad & L(A, B) = \mathrm{tr}(A^{-1}B) - \log(\det(A^{-1}B)) - m, && \text{(4.30a)} \\ \text{subject to} \quad & B \in \Omega := \big\{ B \in \mathbb{R}^{m\times m} : B = \sigma^2(I + \sum_{i=1}^{p} c_i T_i) \\ & \qquad\qquad \text{is positive definite} \big\}. && \text{(4.30b)} \end{aligned}$$

We note that this problem was mentioned in a recent work by Ning et al. (2011), but no solution method was developed there. Setting $f(c,\sigma) := L(A, B(c,\sigma))$ and applying the chain rule, we have the gradient of f,

$$\begin{aligned} \nabla_{c_i} f &= \sigma^2 \mathrm{tr}(T_i(A^{-1} - B^{-1})), \quad i = 1:p, \\ \nabla_{\sigma} f &= 2(\mathrm{tr}(A^{-1}B) - m)/\sigma, \end{aligned}$$

for which a necessary condition for a local minimum of (4.30) can be obtained immediately.

Davis et al. (1998) proposed an approach based on the method of *outer approximation* where they consider the problem of computing the structured covariance estimator through maximum likelihood estimation. Suppose that we are now to minimize a general function $f(B)$ subject to B being a symmetric positive definite Toeplitz matrix. The idea is to expand the semi-infinite problem with constraint $y^T B y > 0$ for all $y \in \mathbb{R}^m$ into an infinite sequence of ordinary nonlinear programming problems $\{P^k\}_{k=0}^{\infty}$ with P^k having the form

$$P^k: \ \min f(B) \text{ subject to } B \in \Omega_k := \Big\{ B \in \mathbb{R}^{m\times m} : B = \sigma^2 (I + \sum_{i=1}^{p} c_i T_i), \quad (4.31a)$$

$$y^T B y > 0, y \in Y_k \Big\}, \quad (4.31b)$$

where Y_k is a finite subset of the ball $\big\{ y \in \mathbb{R}^m : \|y\| = 1 \big\}$. This is called an outer approximation because the constraint $\Omega \subset \Omega_k$. It is proved that any limit point of the sequence generated by outer approximation is a solution of the original problem. Davis, Evans, and Polak (1998) use the linearly convergent Polak–He algorithm to solve P^k. However, this idea cannot be applied to our case because $B \in \Omega_k$ may not imply $B > 0$ and our objective function is defined only when $B > 0$.

A related problem to (4.30) is to find a symmetric positive semidefinite Toeplitz matrix that is the nearest to a given matrix in the Frobenius norm. A method using alternating projections is proposed by Suffridge and Hayden (1993) and a sequential quadratic programming (SQP) algorithm is given by Al-Homidan (2002). This ability to project onto the constraint set in (4.30) motivates us to use a spectral projected gradient method (SPGM) introduced by Birgin et al. (2000, 2001). The method aims to minimize a continuously differentiable function $f(x)$ on a closed convex set by generating a sequence of vectors that is guaranteed to converge r-linearly to a stationary point of f. It generates vectors of the form $x_{k+1} = x_k + \alpha_k d_k$ with the spectral projected gradient direction $d_k = P(x_k - \lambda_k \nabla f(x_k)) - x_k$, where P denotes projection onto the convex set of constraints, $\lambda_k > 0$ is a precomputed scalar, and α_k is chosen by a nonmonotone line search strategy. The direction d_k is guaranteed to be a descent direction (Birgin et al., 2000, Lem. 2.1). The method is intended particularly for problems where it is computationally efficient to project onto the feasible set. The computational cost for finding the nearest symmetric positive semidefinite Toeplitz matrix to a given matrix in the Frobenius norm is

$O(m^3)$ flops for the alternating projection method and $O(m^2)$ flops for each iteration of SQP, so it may be quite expensive to apply SPGM to our problem when m is large.

Another closely related approach is semidefinite programming (Vandenberghe and Boyd, 1996) which deals with optimization problems over symmetric positive semidefinite matrix variables with linear cost function and linear constraints; some techniques therein can be extended to our problem.

Comparing the various different approaches to solving the optimization problem (4.30) is beyond the scope of this paper, so we leave it to future work.

5 Numerical Experiments

In this section, we illustrate numerically how the techniques discussed above can be used in regularizing the underlying covariance structure. We first carry out simulation studies and then apply our techniques to real data analysis. All computations were performed with MATLAB 2011a.

5.1 Simulation studies

Recall that our idea is, given a covariance matrix and a class of possible candidate covariance structures, to find for each structure a covariance matrix that minimizes the entropy loss function. The structure of the minimizer that has the smallest entropy loss function value among the class is considered to be the most likely underlying covariance structure for the given covariance matrix. To examine the idea, our simulation experiments were carried out as follows. We first generate an $m \times n$ data matrix R with columns randomly drawn from the multivariate normal distribution with a common mean vector $\mu = \sigma^2 e \in \mathbb{R}^m$ (recall that e is the vector of ones) and a common covariance matrix $\Sigma = \sigma^2 V(c) \in \mathbb{R}^{m\times m}$, where $\sigma^2 > 0$ and the matrix V is fully determined by the correlation coefficients c. We then compute the sample covariance matrix A with the generated data R: $A = n^{-1}\sum_{i=1}^n (r_i - \bar{r})(r_i - \bar{r})^T$, where r_i is the ith column of R and $r = n^{-1}\sum_{i=1}^n r_i$ is the sample mean. We test with the true covariance matrix Σ of various dimensions m and structures as discussed in the previous sections, where for each structure we consider several different values for σ^2 and c. The sample size is chosen as $n = 1000$. We summarize the experimental results in Tables 1–3, which are for the experiments with covariance matrix size $m = 100$, and Tables 4–6, which are for $m = 200$. We choose $c \in \{0.2, 0.5, 0.75\}$ and $\sigma^2 \in \{2, 4, 8\}$. In Tables 1–6 each row stands for one experiment and for each experiment we report the results averaged over 1000 repeated simulations. The first column gives the true underlying covariance structure and the second column presents the discrepancy between the true covariance matrix Σ and the sample covariance matrix A under the measure of entropy loss function. The rest of the columns report the results from the computed matrix B with different

structures. Note we do not include a row for tridiagonal Σ with $c = 0.75$ because there does not exist such a positive definite covariance matrix in this case. The notation and abbreviation for the results reported in the tables are summarized

- Σ: true covariance matrix.
- A: sample covariance matrix.
- B: the computed covariance matrix that has a certain structure and minimizes the entropy loss function $L(A, B)$ in (1.1).
- T_m: the averaged time (in second) used to find the optimal matrix B in one simulation study.
- $L_{\Sigma,A}$, $L_{A,B}$ and $L_{\Sigma,B}$: the entropy loss function $L(\Sigma, A)$, $L(A, B)$ and $L(\Sigma, B)$, respectively.

[Tables 1-6 are about here.]

In Tables 1–6, we have the following observations. First, we have $L_{\Sigma,B} < L_{\Sigma,A}$ for the all cases as long as the estimated covariance matrix B has the same structure as the true covariance matrix Σ. In other words, the regularized estimator B that has the same structure as Σ is much better than the sample covariance matrix A in terms of the entropy loss function. It shows that regularization of the sample covariance matrix, or any other available estimators of the covariance matrix, is really necessary not only for the convenient use of known structure but also for the accuracy of covariance estimation. Second, for each Σ with an assigned structure, among different minimizers B the one with the smallest value $L_{\Sigma,B}$ turns out to have the same structure as the Σ. The same observation can be made from the discrepancy $L_{A,B}$. The latter is extremely important because in practice the true covariance is usually unknown and so is $L_{\Sigma,B}$. Thus, the discrepancy $L_{A,B}$ can be used to identify the correct covariance structure as long as the class of the candidate structures is broad enough.

The observations above are common to all choices of the structure of Σ in the class we considered, the various values of c and σ^2 and the dimension m of the covariance matrix. Therefore, the findings are reliable in this sense.

5.2 Real data analysis

We also did experiments with some real data. Kenward's (1987) cattle data was analyzed by various statistical methods for longitudinal data in the literature (Pourahmadi, 1999; Pan and Mackenzie, 2003). In the experiment, 60 cattle were assigned randomly to two treatment groups 1 and 2, each of which consists of 30 cattle, and received a certain treatment. The cattle in each group were weighed 11 times over a

nineteen-week period. The weighing times for all cattle were the same, so that the cattle data is a balanced longitudinal data set. The aim of Kenward's study was to investigate treatment effects on intestinal parasites of the cattle. Our experiments are carried out with the cattle data in a similar way as in section 5.1 and the results are reported in Table 7.

[Table 7 is about here.]

Note that in this real data analysis the true covariance matrix Σ is not available, so $L_{\Sigma,A}$ and $L_{\Sigma,B}$ are not available, where A is the sample covariance matrix. Instead, we use the discrepancy $L_{A,B}$ to identify the most likely covariance structure among the possible candidate structures, tridiagonal, CS and AR(1).

From Table 7, it is clear that the underlying covariance structures for both groups are very likely to be AR(1) among the three possible candidate structures, as the discrepancy $L_{A,B}$ has a smaller value than others. This agrees with the finding by Pourahmadi (1999) and Pan and Mackenzie (2003).

6 Discussion

We have proposed a method to regularize the underlying covariance structure with a given covariance matrix A and a class of candidate covariance structures, based on minimizing the entropy loss function between the given covariance matrix and the matrix that has a certain structure. Our simulation studies demonstrate the reliability of the proposed method. Our simulation experiments were carried out with the given matrix A being the sample covariance matrix. In principle, any available estimated covariance matrix using a statistical method can be chosen as the given matrix A. In general, the structure behind the matrix A is not obvious due to noise in the matrix A, in particular, when the dimension m of the matrix A is high. Our aim is to regularize the matrix A, so as to filter the noise in A and to have a standard structure to characterize the covariance/correlation process of the data studied.

Our proposed method can overcome the difficulties that are met by alternative approaches and can produce a reliable estimator of the covariance matrix even if the dimension of the matrix is very high. For example, it is not an easy task to directly calculate the maximum likelihood estimator of a covariance matrix that has a certain structure. In contrast, the proposed method does not require any distribution assumption of the data, and can provide a regularized covariance structure estimator as long as an estimator of the covariance matrix is given.

A restriction of our studies here is that the class of candidate covariance structures we have considered in this paper may not be very broad. We have considered four possible structures: tridiagonal, CS, AR(1), and banded Toeplitz structures. In principle, the ideas and the proposed approach are applicable to any structured

covariance matrices. We are currently studying some other covariance structures, including linearly structured covariance, Hankel matrices, and Toeplitz covariance matrices, all of which are very useful in statistics. However, with more complicated covariance structures more challenging work is inevitably involved in computing the structured covariance matrix that minimizes the entropy loss function. For example, we have carried out some similar simulation experiments for banded Toeplitz matrices but are unable to draw similar conclusions as we have had from Tables 1–6. A crucial reason is the lack of an accurate and efficient algorithm to solve the optimization problem (4.30), as discussed in Section 4. Note that for the three structures considered in our experiments, the dimension m of the covariance matrix does not affect the cost of computing the optimal covariance matrix B, because the optimization problem reduces to computing the zeros of a nonlinear function of a single variable. For more complicated covariance structures, the optimization problem is liable to have more variables and the number of variables may increase with the dimension m of the covariance matrix, making the solution of the optimization problem much more challenging.

Acknowledgement

This work was supported by the EPSRC grant EP/I01912X/1 and European Research Council Advanced Grant MATFUN (267526).

References

S. S. Al-Homidan. SQP algorithms for solving Toeplitz matrix approximation problem. *Numerical Linear Algebra with Applications*, 9(8):619–627, 2002.

E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10(4):1196–1211, 2000.

E. G. Birgin, J. M. Martínez, and M. Raydan. Algorithm 813: SPG—Software for convex-constrained optimization. *ACM Trans. Math. Software*, 27(3):340–349, 2001.

R. Borsdorf, N. J. Higham, and M. Raydan. Computing a nearest correlation matrix with factor structure. *SIAM J. Matrix Anal. Appl.*, 31(5):2603–2622, 2010.

J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9):963–974, 1982.

L. Davis, R. Evans, and E. Polak. Maximum likelihood estimation of positive definite Hermitian Toeplitz matrices using outer approximations. In *Statistical Signal and Array Processing, 1998. Proceedings., Ninth IEEE SP Workshop on*, pages 49–52. IEEE, 1998.

D. K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein's loss. *The Annals of Statistics*, 13(4):1581–1591, 1985.

I. S. Dhillon and J. A. Tropp. Matrix nearness problems with Bregman divergences. *SIAM J. Matrix Anal. Appl.*, 29(4):1120–1146, 2007.

N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2002. ISBN 0-89871-521-0.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985. ISBN 0-521-30586-1.

W. James and C. Stein. Estimation with quadratic loss. In J. Neyman, editor, *Proceedings of the Fourth Berkeley Symposium*, volume 1 of *Mathematical Statistics and Probability*, pages 361–379. University of California Press, 1961. The Statistical Laboratory, University of California, June 30-July 30, 1960.

M. G. Kenward. A method for comparing profiles of repeated measurements. *Appl. Statist.*, pages 296–308, 1987.

F. Lin and M. R. Jovanović. Least-squares approximation of structured covariances. *IEEE Trans. Automat. Control*, 54(7):1643–1648, 2009.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, UK, revised edition, 1999. ISBN 0-471-98633-X.

L. Ning, X. Jiang, and T. Georgiou. Geometric methods for estimation of structured covariances. Technical Report arXiv:1110.3695v1, 2011.

J. Pan and K. Fang. *Growth curve models and statistical diagnostics*. Springer-Verlag, New York, 2002.

J. Pan and G. Mackenzie. On modelling mean-covariance structures in longitudinal studies. *Biometrika*, 90(1):239–244, 2003.

S. V. Parter. An observation on the numerical solution of difference equations and a theorem of Szegö. *Numerische Mathematik*, 4:293–295, 1962.

F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis. *IEEE Trans. Signal Processing*, 56(1):34–48, 2008.

M. Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999. ISSN 0006-3444.

T. J. Suffridge and T. L. Hayden. Approximation by a Hermitian positive semidefinite Toeplitz matrix. *SIAM J. Matrix Anal. Appl.*, 14:721–734, 1993.

L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, pages 49–95, 1996.

H. Ye and J. Pan. Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika*, 93(4):927–941, 2006. ISSN 0006-3444.

Table 1: Simulation results with $m = 100$; $\sigma^2 = 2$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.23	5.42	0.55	2.25e-3	9.75	4.88	1.39e-3	5.60	0.73	3.45e-2
CS	5.23	7.70	2.83	2.15e-3	5.42	0.55	1.35e-3	7.70	2.83	3.37e-2
AR(1)	5.23	5.57	0.70	2.17e-3	9.21	4.34	1.32e-3	5.42	0.55	3.35e-2
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.24	5.43	0.55	2.41e-3	290.10	285.29	1.39e-3	185.73	180.90	3.45e-2
CS	5.23	9.03	4.16	2.13e-3	5.42	0.55	1.30e-3	9.03	4.16	3.35e-2
AR(1)	5.23	10.25	5.37	2.23e-3	26.84	21.96	1.37e-3	5.42	0.55	3.38e-2
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	5.23	10.11	5.25	2.16e-3	5.42	0.55	1.38e-3	10.11	5.25	3.40e-2
AR(1)	5.23	23.87	19.00	2.25e-3	47.94	43.08	1.35e-3	5.42	0.55	3.42e-2

Table 2: Simulation results with $m = 100$; $\sigma^2 = 4$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.23	5.42	0.55	2.22e-3	9.75	4.88	1.36e-3	5.60	0.73	3.40e-2
CS	5.23	7.70	2.83	2.15e-3	5.42	0.55	1.35e-3	7.70	2.83	3.37e-2
AR(1)	5.23	5.57	0.70	2.15e-3	9.21	4.34	1.31e-3	5.42	0.55	3.32e-2
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.23	5.42	0.55	2.37e-3	290.21	285.31	1.37e-3	185.81	180.91	3.42e-2
CS	5.23	9.03	4.16	2.12e-3	5.42	0.55	1.30e-3	9.03	4.16	3.35e-2
AR(1)	5.23	10.24	5.37	2.21e-3	26.82	21.96	1.35e-3	5.42	0.55	3.37e-2
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	5.23	10.12	5.24	2.10e-3	5.42	0.55	1.33e-3	10.12	5.24	3.34e-2
AR(1)	5.23	23.87	19.00	2.22e-3	47.96	43.08	1.33e-3	5.42	0.55	3.40e-2

Table 3: Simulation results with $m = 100$; $\sigma^2 = 8$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.23	5.42	0.55	2.21e-3	9.75	4.88	1.35e-3	5.60	0.73	3.35e-2
CS	5.23	7.71	2.83	2.16e-3	5.42	0.55	1.36e-3	7.71	2.83	3.36e-2
AR(1)	5.23	5.57	0.70	2.19e-3	9.20	4.33	1.35e-3	5.42	0.55	3.35e-2
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	5.23	5.42	0.55	2.37e-3	290.03	285.29	1.36e-3	185.66	180.89	3.39e-2
CS	5.23	9.03	4.16	2.19e-3	5.42	0.55	1.34e-3	9.03	4.16	3.37e-2
AR(1)	5.23	10.24	5.37	2.29e-3	26.83	21.96	1.40e-3	5.42	0.55	3.43e-2
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	5.23	10.12	5.25	2.12e-3	5.42	0.55	1.35e-3	10.12	5.25	3.31e-2
AR(1)	5.23	23.87	19.00	2.27e-3	47.93	43.08	1.36e-3	5.42	0.55	3.40e-2

Table 4: Simulation results with $m = 200$; $\sigma^2 = 2$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.62	23.31	4.69	7.23e-3	32.08	13.47	3.48e-3	23.67	5.05	2.96e-1
CS	21.61	26.24	7.63	7.24e-3	23.30	4.68	3.55e-3	26.24	7.63	2.98e-1
AR(1)	21.63	23.62	5.01	7.26e-3	31.04	12.41	3.54e-3	23.32	4.70	2.99e-1
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.61	23.30	4.69	7.54e-3	727.81	709.43	3.75e-3	499.72	481.20	2.97e-1
CS	21.61	27.60	8.99	7.22e-3	23.30	4.69	3.54e-3	27.60	8.99	2.98e-1
AR(1)	21.62	33.00	14.39	7.28e-3	67.04	48.41	3.60e-3	23.31	4.69	2.98e-1
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	21.62	28.70	10.08	7.21e-3	23.31	4.69	3.56e-3	28.70	10.08	2.98e-1
AR(1)	21.63	60.21	41.59	7.25e-3	110.44	91.83	3.49e-3	23.32	4.70	2.97e-1

Table 5: Simulation results with $m = 200$; $\sigma^2 = 4$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.62	23.30	4.70	7.32e-3	32.09	13.48	3.54e-3	23.67	5.06	2.97e-1
CS	21.61	26.24	7.64	7.07e-3	23.30	4.69	3.46e-3	26.24	7.64	2.95e-1
AR(1)	21.63	23.62	5.00	7.27e-3	31.02	12.40	3.56e-3	23.32	4.69	2.97e-1
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.61	23.29	4.69	7.57e-3	727.61	709.39	3.58e-3	499.51	481.16	2.98e-1
CS	21.63	27.62	8.99	7.20e-3	23.32	4.69	3.53e-3	27.62	8.99	2.97e-1
AR(1)	21.62	33.00	14.39	7.25e-3	67.00	48.40	3.54e-3	23.31	4.69	2.97e-1
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	21.62	28.70	10.08	7.16e-3	23.31	4.69	3.52e-3	28.70	10.08	2.97e-1
AR(1)	21.61	60.18	41.58	7.31e-3	110.43	91.82	3.56e-3	23.30	4.69	2.98e-1

Table 6: Simulation results with $m = 200$; $\sigma^2 = 8$.

		B								
$c = 0.20$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.63	23.32	4.70	7.27e-3	32.11	13.48	3.54e-3	23.68	5.06	2.99e-1
CS	21.62	26.24	7.63	7.22e-3	23.30	4.69	3.57e-3	26.24	7.63	2.98e-1
AR(1)	21.63	23.62	5.00	7.21e-3	31.02	12.40	3.50e-3	23.31	4.69	2.96e-1
		B								
$c = 0.50$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
Tri-diag	21.62	23.31	4.70	7.51e-3	728.40	709.55	3.65e-3	499.52	481.16	2.98e-1
CS	21.60	27.59	8.99	7.14e-3	23.29	4.69	3.50e-3	27.59	8.99	2.97e-1
AR(1)	21.63	33.02	14.40	7.25e-3	67.05	48.42	3.55e-3	23.32	4.70	2.98e-1
		B								
$c = 0.75$		Tri-diag			CS			AR(1)		
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m	$L_{A,B}$	$L_{\Sigma,B}$	T_m
CS	21.62	28.70	10.08	7.16e-3	23.31	4.69	3.56e-3	28.70	10.08	2.99e-1
AR(1)	21.62	60.21	41.59	7.31e-3	110.46	91.83	3.56e-3	23.31	4.70	2.99e-1

Table 7: Results of experiments on Kenward's cattle data.

	Tri-diag		CS		AR (1)	
	$L_{A,B}$	T_m	$L_{A,B}$	T_m	$L_{A,B}$	T_m
Group 1	9.86	1.37e-003	8.55	1.06e-003	5.22	3.16e-003
Group 2	8.05	7.41e-004	5.92	5.97e-004	3.15	2.57e-003