This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: http://orca.cf.ac.uk/98928/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:


Publishers page: http://dx.doi.org/10.1016/j.jmva.2017.03.001
<http://dx.doi.org/10.1016/j.jmva.2017.03.001>

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
A calibration method for non-positive definite covariance matrix in multivariate data analysis

Chao Huang 1, Daniel Farewell 2 and Jianxin Pan 3*

1 South East Wales Trials Unit, Cardiff University
Heath Park, Cardiff, CF14 4YS, UK
2 Division of Population Medicine, School of Medicine, Cardiff University
Heath Park, Cardiff, CF14 4YS, UK
3 School of Mathematics, University of Manchester,
Manchester M13 9PL, UK

Abstract

Covariance matrices that fail to be positive definite arise often in covariance estimation. Approaches addressing this problem exist, but are not well supported theoretically. In this paper, we propose a unified statistical and numerical matrix calibration, finding the optimal positive definite surrogate in the sense of Frobenius norm. The proposed algorithm can be directly applied to any estimated covariance matrix. Numerical results show that the calibrated matrix is typically closer to the true covariance, while making only limited changes to the original covariance structure.

Keywords:
Covariance matrix calibration, Nearness problem, Non-positive definiteness, Spectral decomposition

1. Introduction

The estimation of covariance matrices plays an essential role in multivariate data analysis. Covariances are required by many statistical modelling
approaches, including multivariate regression and the analysis of spatial data. Often, well-estimated covariance matrices improve efficiency in estimating parameters in a mean function [22]. In some circumstances, the covariance matrix may itself be of direct scientific interest: for instance, in spatial variation analysis for geographical data, and in volatility analysis for financial data.

However, it is not uncommon that estimators of covariance matrices fail to be positive definite. A typical example is the sample covariance matrix, which is often singular when the sample size is close to, or less than, the dimension of the random samples [3]. If singularity is caused by collinearity, conventional ridge regression [18] or modern variable selection [6, 21] approaches may solve the problem by excluding redundant variables. Dimension reduction approaches such as Principle Component Analysis [19] can also help to exclude eigenvalues with ignorable contributions.

However, these resolutions only apply in cases where such redundancy truly exists. More often, non-positive definiteness may be put down to the generic difficulty of maintaining positive definiteness in covariance estimation; resulting estimators may not even be positive semidefinite. Even for elaborately designed statistical approaches, the estimators of covariance matrices can be ill-conditioned [5, 14]. A number of approaches have been proposed to resolve this issue. However, these are either limited to special circumstances or lack theoretical support. For instance, one alternative is to use the Moore-Penrose inverse of a non-positive definite matrix to replace the regular inverse typically used in statistical inferences [20]. However, this does not directly resolve the non-positive definiteness, and its lack of statistical interpretation. Alternatively, a smoothing approach exists [23] in which non-positive eigenvalues of the covariance matrix estimator are replaced by certain positive values. However, justification for the selection of these positive values was scant.

Based on the fundamental work of Halmos [7], Higham [9] proposed a solution for finding the nearest (in the sense of Frobenius norm) positive semidefinite matrix to an arbitrary input matrix. However, this surrogate positive semidefinite matrix is still singular [9, 10], so difficulty persists in using the surrogate matrix in statistical practice. Rebonto and Jäckel [17] considered a correlation matrix calibration using the hyperspherical decomposition and eigenvalue correction, which again leads to positive semidefinite correlation matrices. Hendrikse et al. [8] proposed an eigenvalue correction method using bootstrap resampling in order to reduce the bias arising in
sample eigenvalues. Their work focused on the correction of the sample covariance, where the performance of the correction method relies on the assumed distribution of the covariance matrix eigenvalues in the population.

In this paper, we propose a unified approach to calibrate a non-positive definite covariance matrix to ensure positive definiteness. The calibrated covariance matrix is usually closer to the true covariance matrix than the original covariance matrix estimator. Our proposed approach is implemented through a straightforward screening algorithm. In Section 2, we briefly review the matrix nearness problem, before proposing our novel calibration method together with its integrated criterion and algorithm. In Section 3 we conduct two simulation studies, and in Section 4 we discuss two case studies, including a calibration of the non-positive definite covariance matrix obtained by nonparametric regression in Diggle and Verbyla [5]. Conclusions are presented in Section 5.

2. Calibration method

2.1. The matrix nearness problem

In numerical analysis, a nearness problem involves finding, for a given matrix and a particular matrix norm, the nearest matrix that has certain important properties. Examples include finding the nearest covariance matrix [9] or correlation matrix [2, 16] in the sense of the Frobenius norm (or 2-norm).

Given an arbitrary square matrix $X$ of order $n$, we denote its Frobenius norm by $\|X\| = \text{trace}(X^\top X)^{1/2}$. The nearness problem involves finding the nearest symmetric positive semidefinite matrix $P_0(X)$:

$$P_0(X) = \arg \min_{A \geq 0} \|X - A\|$$

(1)

Throughout, we shall assume that $A \geq 0$ denotes both non-negative definiteness and symmetry $A = A^\top$. Higham [9] used a polar decomposition to show that the solution to (1) has the explicit form $P_0(X) = (B + H)/2$, where $B = S(X) = (X + X^\top)/2$ is the symmetric matrix version of $X$, and $H$ is the symmetric polar factor of $B$, satisfying $B = UH$ with $U$ a unitary matrix and $H \geq 0$. This solution has been compiled in a MATLAB file named poldex.m, which can be found in the Matrix Computation Toolbox [11]. Clearly, if $X$ is symmetric then the solution becomes $P_0(X) = (X + H)/2$. If, further, we are given the spectral decomposition of a symmetric $X = X^\top$
(that is, $X = QΛQ^\top$ for $Q^\top Q = I$ and $Λ = \text{diag}(\lambda_1, \ldots, \lambda_n)$), we have $P_0(X) = Q\text{diag}\{\max(\lambda_1, 0), \ldots, \max(\lambda_n, 0)\}Q^\top$. In words, the nearest positive semidefinite matrix $P_0(X)$ can be obtained by replacing by zero any negative eigenvalues of a symmetric $X$ [10], eliminating the corresponding columns of $Q$ (and causing some information loss). A immediate alternative is to instead replace negative eigenvalues by positive values, so that a positive definite correction of $X$ is formed without this loss of information about $Q$.

However, the theory of this idea need to be justified, particularly on how to choose appropriate replacement positive values, for which we will address in this paper.

2.2. A new calibration approach

We now aim to find a positive definite matrix surrogate for a generic $X$. First, we formulate this question as a nearness problem. For $c \geq 0$, let $\mathcal{D}_c = \{A : A - cI \geq 0\}$ be the set of positive definite matrices with no eigenvalue smaller than $c$. Given $X$, finding the nearest matrix $P_c(X) \in \mathcal{D}_c$ to $X$ in terms of the Frobenius norm amounts to defining

$$P_c(X) = \underset{A \in \mathcal{D}_c}{\text{arg min}} \|X - A\|. \quad (2)$$

An explicit expression for $P_c(X)$ is given in Theorem 1.

**Theorem 1.** Given $X$ and a constant $c \geq 0$, the nearest (in the sense of Frobenius norm) matrix $P_c(X) \in \mathcal{D}_c$ to $X$ is of the form

$$P_c(X) = P_0(X - cI) + cI \quad (3)$$

where (as before) $P_0(X - cI) = (B + H)/2$ for $B = S(X - cI)$ and $H$ the polar factor of $B$. Furthermore, if $X$ is symmetric with spectral decomposition $X = Q\text{diag}(\lambda_1, \ldots, \lambda_n)Q^\top$ then $P_c(X)$ has the simplified form

$$P_c(X) = Q\text{diag}\{\max(\lambda_1, c), \ldots, \max(\lambda_n, c)\}Q^\top. \quad (4)$$

**Proof:** The details of the proof are deferred to the Appendix.

Maintaining symmetry in covariance estimation is typically not difficult, so direct use of (4) will often be sufficient in practice. For non-symmetric $X$, one may directly symmetrize $X$ before calibration. Note that $P_c(X)$ only depends on $X$ via its symmetric version $S(X)$, so (4) can equivalently be applied to $S(X)$.  

4
2.3. Selection criterion

Clearly \( P_c(X) \) varies with \( c \) and, as \( c \) decreases, the domain \( D_c \) of \( A \) expands. At \( c = 0 \), \( P_c(X) = P_0(X) \) becomes positive semidefinite (unless, of course, all eigenvalues of \( X \) are already positive). Consequently, we require a criterion for selecting an appropriate positive value, \( c = c_* \), say. Let \( \lambda_{\min}^+ \) be the smallest positive eigenvalue of an estimated covariance matrix \( X \). In order to maintain, as far as possible, the covariance structure of \( X \), it is reasonable to constrain \( 0 \leq c_* \leq \lambda_{\min}^+ \). Rather than make simple choices such as \( c_* = \frac{\lambda_{\min}^+}{2} \), here we propose a tuning approach, balancing proximity to \( X \) with proximity to singularity. Writing \( c_\alpha = 10^{-\alpha} \lambda_{\min}^+ \), where \( \alpha \geq 0 \) is a tuning parameter, with \( c_0 = \lambda_{\min}^+ \) and \( c_\infty = 0 \) (i.e., \( c \to 0 \) as \( \alpha \to \infty \)), we choose \( c_* \) as follows:

**Definition 1.** Define \( c_* = c_\alpha \) via

\[
\alpha_* = \arg \min_{\alpha} \| X - P_{c_\alpha}(X) \| + \alpha, \tag{5}
\]

where (as before) \( c_\alpha = 10^{-\alpha} \lambda_{\min}^+ \).

Rather than simply minimize the quantity \( \| X - A \| \), in (5) we also add a penalty (namely, \( \alpha \)) that penalises small values of \( c \). Such penalty terms are widely used in a variety of statistical contexts, such as the AIC/BIC and penalty functions [1, 6, 21]. Reassuringly, positive definite covariance matrices remain unchanged after calibration. To see this, note that \( P_{\lambda_{\min}^+}(X) = X \) if \( X \) is itself positive definite. In this case, choosing \( \alpha = 0 \) (so \( c = \lambda_{\min}^+ \)) thus makes both \( \| X - P_{c_\alpha}(X) \| \) and \( \alpha \) vanish, so \( c_* = \lambda_{\min}^+ \) and the solution \( P_*(X) = P_{c_*}(X) \) completely reduces to \( X \).

The tuning parameter \( \alpha \) can also be interpreted in terms of the condition number of the matrix \( P_{c_\alpha}(X) \) [24, p146]. For a positive semi-definite matrix, the condition number is the ratio of its biggest to smallest eigenvalues. The condition number can warn us the numerical inaccuracy in calculating the inverse of a given matrix. In our case, let \( \lambda_{\max}^+ \) be the biggest positive eigenvalues of \( X \) and \( d = \frac{\lambda_{\max}^+}{\lambda_{\min}^+} \). Then the condition number of the calibrated matrix \( P_{c_\alpha}(X) \) is \( \kappa(P_{c_\alpha}(X)) = 10^{\alpha d} \). Therefore, the penalty \( \alpha \) approximates the number of digits of accuracy we are prepared to sacrifice in the inversion of \( P_{c_\alpha}(X) \) in order to reduce \( \| X - P_{c_\alpha}(X) \| \).
2.4. Algorithm

In practice, we implement a screening-search strategy for the tuning parameter $\alpha$. Rather than let $\alpha \in [0, +\infty)$, we constrain the screening to a feasible region. This strategy is employed in the following algorithm.

Step 1. Given a feasible region of $\alpha$, say $[0, \alpha_N]$, create a partition $0 = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_N$. For $\alpha \in \{\alpha_0, \ldots, \alpha_N\}$, compute the corresponding $c_\alpha$ and use (4) to calculate the resulting solution matrix $P_{c_\alpha}(X)$. We choose

$$\alpha_\star = \arg\min_{\alpha \in \{\alpha_0, \ldots, \alpha_N\}} \|X - P_{c_\alpha}(X)\| + \alpha. \quad (6)$$

Step 2. Set $c_\star = c_{\alpha_\star}$, and return $P_\star(X) = P_{c_\star}(X)$ as the final calibrated covariance matrix.

In terms of the screening region $[0, \alpha_N]$ and its partition, we make the following recommendations. In most applications, $10^{-\alpha}$ become negligible when $\alpha > 10$, so we take our default option to be $\alpha_N = 10$. Options of larger $\alpha_N$ are possible when the original $\lambda_{\text{min}}$ is in large scale. However, we would not recommend a too large $\alpha_N$, as it corresponds to a large condition number of $P_\star(X)$. When screening $\alpha \in [0, \alpha_N]$, we suggest a uniform partition of the region: for example, given $\alpha_N = 10$, we could use the partition $0, 1, \ldots, 10$ or $0, 0.5, \ldots, 10$. More refined partition would be preferable when extra accuracy in calibration is demanded.

3. Simulation studies

In this section, we carry out two simulation studies to assess the performance of our proposed calibration method. In Simulation 1, we consider three commonly used covariance structures: compound symmetry, first-order autoregressive ($\text{AR}(1)$) and tri-diagonal. In Simulation 2, a more general covariance structure formed by the modified Cholesky decomposition [15] is investigated. Covariance matrices are fitted via the nonparametric covariance estimation approach of Diggle and Verbyla [5]. The advantage of that approach lies in that the variogram considered therein has a clear statistical interpretation, being useful in describing spatial correlation in geo-statistics [4], and in judging if the covariance structure is stationary. Here we focus on the non-positive definite covariance matrices obtained by that approach to assess the performance of the proposed calibration method.
The longitudinal data we consider here are described by \((y_{ij}, t_{ij}), i = 1, \ldots, n, j = 1, \ldots, n_i\), where \(y_{ij}\) represents the measurement \(j\) (out of \(n_i\)) on subject \(i\) and \(t_{ij}\) is its measurement time. Let \(\mu_{ij} = \mu_i(t_{ij})\) be the mean of \(y_{ij}\) and \(\mu_i = (\mu_{i1}, \ldots, \mu_{in_i})^\top\) be the vector of the means of the responses \(y_i = (y_{i1}, \ldots, y_{in_i})^\top\). Assume \(\Sigma_i\) is the covariance matrix of the responses \(y_i\), where the elements of \(\Sigma_i\) are defined by a generic covariance function \(\Sigma_i)_{j,k} = \Sigma(t_{ij}, t_{ik})\). Following [5], a multivariate normal distribution is assumed, i.e., \(y_i \sim \mathcal{N}(\mu_i, \Sigma_i)\).

The main covariance fitting process of [5] is briefly summarized as follows. Firstly, a local polynomial smoothing technique is used to estimate the variances in \(\hat{\Sigma}_i\), \((\Sigma_i)_{j,j}, j = 1, \ldots, n_i\), where \(\hat{\mu}_{ij}\) are the fitted means through certain nonparametric regression estimation methods, such as [25]. Secondly, a bivariate local polynomial smoothing method is used to model the variograms \(\hat{v}_{ijk}\) through the sample variograms \(\{(y_{ij} - \hat{\mu}_{ij}) - (y_{ik} - \hat{\mu}_{ik})\}^2/2\). Finally, the off-diagonal elements of \(\Sigma_i\), \((\Sigma_i)_{j,k}, j \neq k\), are estimated through \((\Sigma_i)_{j,k} = \{(\Sigma_i)_{j,j} + (\Sigma_i)_{k,k}\}/2 - \hat{v}_{ijk}\).

3.1. Simulation 1

We generate 100 datasets based on the Gaussian process mechanism described above. In each dataset there are \(n = 50\) subjects and \(n_i = m = 10\) or 20 repeated measurements for each subject. The means \(\mu_{ij}\) are formulated as \(\mu_{ij} = t_{ij} + \sin(t_{ij})\) for all \(i, j\), with measurement times \(t_{ij} = j\) for all \(j\). Given a common variance \(\sigma^2\) and a correlation parameter \(\rho\), the covariance structure \(\Sigma_i\) of subject \(i\) is assumed to have the particular structure described below:

1. Compound symmetry. Within-subject correlation is assumed equal for any disjoint pair of observations. In other words, \(\Sigma_i = \sigma^2\{(1-\rho)I + \rho J\}\), where \(\rho \in (-1/(m-1), 1)\), \(I\) is an identity matrix and \(J\) is a matrix of ones with order \(m\).

2. AR(1). Within-subject correlation decreases with the time separation as \(\Sigma_i = \sigma^2(\rho^{j-k})\) \((j, k = 1, \ldots, m, \rho \in (-1, 1))\).

3. Tri-diagonal. Within-subject correlation vanishes except for adjacent observations, i.e., the \((j, k)\)-th element of \(\Sigma_i\) is given by

\[
(\Sigma_i)_{j,k} = \begin{cases} 
\sigma^2, & j = k \\
\sigma^2\rho, & |j - k| = 1 \\
0, & |j - k| \geq 2 
\end{cases}
\]
where $\rho \in (-\frac{1}{2}\{\cos(\frac{\pi}{m+1})\}^{-1}, \frac{1}{2}\{\cos(\frac{\pi}{m+1})\}^{-1})$.

Table 1: Calibration for conventional covariance structures with $m = 10$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>0</th>
<th>0.2</th>
<th>0.5</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound non-P.D. frequency (%)</td>
<td>13.8075</td>
<td>13.8161</td>
<td>13.8507</td>
<td>13.8815</td>
</tr>
<tr>
<td>$|\Sigma - \Sigma|$</td>
<td>94</td>
<td>96</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$|P_*(\Sigma) - \Sigma|$</td>
<td>13.7601</td>
<td>13.7694</td>
<td>13.7956</td>
<td>13.8280</td>
</tr>
<tr>
<td>$|\hat{\Sigma} - P_*(\hat{\Sigma})|$</td>
<td>0.6648</td>
<td>0.7194</td>
<td>1.0337</td>
<td>1.3536</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>0</th>
<th>0.2</th>
<th>0.5</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-diagonal non-P.D. frequency (%)</td>
<td>13.8075</td>
<td>13.8161</td>
<td>13.8507</td>
<td>13.8815</td>
</tr>
<tr>
<td>$|\Sigma - \Sigma|$</td>
<td>94</td>
<td>67</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>$|P_*(\Sigma) - \Sigma|$</td>
<td>13.7601</td>
<td>13.8371</td>
<td>13.9826</td>
<td></td>
</tr>
<tr>
<td>$|\hat{\Sigma} - P_*(\hat{\Sigma})|$</td>
<td>0.6648</td>
<td>0.4311</td>
<td>0.4305</td>
<td></td>
</tr>
<tr>
<td>$|\hat{\Sigma} - \Sigma|$</td>
<td>13.8075</td>
<td>13.8687</td>
<td>14.0129</td>
<td></td>
</tr>
</tbody>
</table>

We explore with four different values of $\rho$ (0, 0.2, 0.5, 0.8), representing different degrees of within-subject correlation. An exception exists in the tri-diagonal case, where $\rho$ can only take three different values $\rho = 0, 0.2, 0.5$, as $\rho = 0.8$ leads to a non-positive definite matrix $\Sigma_i$. Table 1 and Table 2 display the simulation results with $m = 10$ and $m = 20$. The labels $\|\Sigma - \Sigma\|$, $\|\hat{\Sigma} - \Sigma\|$, and $\|\hat{\Sigma} - P_*(\hat{\Sigma})\|$ denote the average Frobenius norm between the true covariance matrix $\Sigma$, the estimated covariance matrix $\hat{\Sigma}$ and the calibrated covariance matrix $P_*(\hat{\Sigma})$. The row headed “non-P.D.frequency” records the proportion of non-positive definite $\hat{\Sigma}$ arising in the covariance estimation procedure. The very high rate of non-positive definite $\hat{\Sigma}$, particularly when $m = 20$, demonstrates the need for calibration. Also, the $\|P_*(\Sigma) - \Sigma\|$ are smaller than $\|\hat{\Sigma} - \Sigma\|$, indicating that, on average, $P_*(\Sigma)$ is closer to the true covariance matrix than $\hat{\Sigma}$. Furthermore, the reported $\|\hat{\Sigma} - P_*(\hat{\Sigma})\|$ are relatively small, meaning that the calibration method makes no essential changes to the original covariance matrix, other than in correcting its non-positive definiteness. Comparing Table 2 with Table 1, the averaged Frobenius norms $\|P_*(\Sigma) - \Sigma\|$, $\|\hat{\Sigma} - \Sigma\|$ and $\|\hat{\Sigma} - P_*(\hat{\Sigma})\|$ increase with $m$, indicating the dimension dependency of the Frobenious norm. Another potential reason is that $\Sigma$ with higher dimension were more frequently to be non-positive definiteness, so that more actual calibrations were manipulated. When $\rho = 0$ all three structures reduce to the identity structure and the results become identical.
the calibrated covariance matrix $P$ of the Frobenius norm ($\| \cdot \|$).

Comparing the case of $m$ average, a closer-to-true surrogate covariance matrix. In the case of presented in Table 3. Again, we see that our proposed method provides, on population 1, 100 simulated datasets are generated. The numerical results are here we assume $\ln \sigma_i = q(t_{ij})$, where $g(\ldots)$ and $q(\cdot)$ are two- and one-dimensional smoothing functions, respectively. With different specifications for $g(\ldots)$ and $q(\cdot)$, the covariance matrix $\Sigma_i$ encompasses a wide range of covariance structures. Here we assume $g(t_{ij}, t_{ik}) = m^{-2}(t_{ij}^2 + t_{ik}^2) \exp\{- (t_{ij} - t_{ik})/4\}$ and $q(t_{ik}) = 2 \ln[\ln(t_{ik}/(m + 2))], with m = 10, 20. With the same mean function of Simulation 1, 100 simulated datasets are generated. The numerical results are presented in Table 3. Again, we see that our proposed method provides, on average, a closer-to-true surrogate covariance matrix. In the case of $m = 20$, the calibrated covariance matrix $P_i(\Sigma)$ substantially improves $\Sigma$ in the sense of the Frobenius norm ($\| P_i(\Sigma) - \Sigma \|$ = 36.5162 while $\| \Sigma - \Sigma \|$ = 45.5746). Comparing the case of $m = 10$ to $m = 20$, $\| \Sigma - P_i(\Sigma) \|$ substantially increases.

Table 2: Calibration for conventional covariance structures with $m = 20$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$| \Sigma - \Sigma |$</th>
<th>$| \Sigma - P_i(\Sigma) |$</th>
<th>$| \Sigma - \Sigma |$</th>
<th>$| \Sigma - P_i(\Sigma) |$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>29.0470</td>
<td>28.7750</td>
<td>3.2369</td>
<td>3.2369</td>
</tr>
<tr>
<td>0.5</td>
<td>29.2198</td>
<td>29.8966</td>
<td>2.9150</td>
<td>2.5349</td>
</tr>
<tr>
<td>0.8</td>
<td>29.2109</td>
<td>29.1032</td>
<td>2.1133</td>
<td>1.7128</td>
</tr>
<tr>
<td>1</td>
<td>29.3859</td>
<td>29.3403</td>
<td>0.7128</td>
<td></td>
</tr>
</tbody>
</table>

3.2. Simulation 2

We now consider a more general covariance structure via the modified Cholesky decomposition [15]. With a covariance matrix $\Sigma_i$ of order $m$, the modified Cholesky decomposition of $\Sigma_i$ is specified by $T_i \Sigma_i T_i^\top = D_i$, where

$$T_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{i21} & 1 & 0 & \cdots & 0 \\ -\phi_{i31} & -\phi_{i32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{im1} & -\phi_{im2} & -\phi_{im3} & \cdots & 1 \end{pmatrix}, D_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{im}^2 \end{pmatrix},$$

and where $\phi_{ij}$ and $\sigma_{ij}^2$ are the generalized autoregressive parameters and innovation variances, respectively. We then parameterize $\phi_{ij}$ and $\sigma_{ij}^2$ as functions of their corresponding measurement times, $\phi_{ij} = g(t_{ij}, t_{ik})$ and $\ln \sigma_{ij}^2 = q(t_{ij})$, where $g(\ldots)$ and $q(\cdot)$ are two- and one-dimensional smoothing functions, respectively. With different specifications for $g(\ldots)$ and $q(\cdot)$, the covariance matrix $\Sigma_i$ encompasses a wide range of covariance structures.
from 1.1041 to 26.5426. This is partly because of the dimension dependency of the F-norm. It may also imply that the covariance estimate $\hat{\Sigma}$ with $m = 20$ were more ill-conditioned and deeper calibrations were made.

Table 3: Calibration of a more general covariance structure with $m = 10$ and $m = 20$

<table>
<thead>
<tr>
<th></th>
<th>$m = 10$</th>
<th>$m = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\hat{\Sigma} - \Sigma|$</td>
<td>14.4988</td>
<td>45.5746</td>
</tr>
<tr>
<td>non-P.D. frequency (%)</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>$|P_*(\hat{\Sigma}) - \Sigma|$</td>
<td>14.4718</td>
<td>36.5162</td>
</tr>
<tr>
<td>$|\hat{\Sigma} - P_*(\hat{\Sigma})|$</td>
<td>1.1041</td>
<td>26.5426</td>
</tr>
</tbody>
</table>

4. Example analyses

In this section, we revisit two substantive analyses to demonstrate the properties of the proposed calibration method for covariance matrix calibration.

4.1. Cattle data

Kenward’s cattle data [13] involves 60 cattle assigned randomly into two treatment groups: half of the cattle received treatment A, and the other half received treatment B. The cattle were weighed 11 times in total over a nineteen-week period. The data are balanced in the sense that the weighing times were the same for every cow. Using Diggle and Verbyla’s nonparametric regression approach [5] to model these two treatment groups separately, we calculate the associated covariance matrix estimates for the two groups. In this instance, both groups’ estimated covariance matrices are positive definite. Here, we illustrate the need for calibration in missing data settings, by artificially removing different portions of the cattle data. For each subject, the first six repeated measurements are kept observed, but from the seventh repeated measurement onwards the measurement is set to be missing with probability $\theta$. Once a measurement is missing, all subsequent measurements for this subject are also dropped. With this manipulation, non-positive definite covariance estimates emerge as $\theta$ increases. The missingness rate is set to be $\theta = 0\%$ (full data), 10%, 20%, 30%, respectively. These data sets with missingness are then analyzed and calibrated where necessary. Because the true covariance $\Sigma$ is unknown, the sample covariance based on the full data, $\Sigma_s$, is used as a benchmark for $\Sigma$. 

10
Table 4: Calibration results of cattle data with/without missingness

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Missing Rate</th>
<th>Positive Definiteness of $\hat{\Sigma}$</th>
<th>$|\hat{\Sigma} - \Sigma_s|$</th>
<th>$|P_*(\hat{\Sigma}) - \Sigma_s|$</th>
<th>$|\hat{\Sigma} - P_*(\hat{\Sigma})|$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0% Yes</td>
<td>116.8018</td>
<td>116.8018</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>10% Yes</td>
<td>197.2300</td>
<td>197.2300</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20% No</td>
<td>215.1713</td>
<td>213.5823</td>
<td>12.7821</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>30% No</td>
<td>347.9251</td>
<td>344.2363</td>
<td>4.3888</td>
<td>12.7821</td>
</tr>
<tr>
<td>B</td>
<td>0% Yes</td>
<td>109.7911</td>
<td>109.7911</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>10% Yes</td>
<td>307.9060</td>
<td>307.9060</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20% Yes</td>
<td>348.7275</td>
<td>348.7275</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>30% No</td>
<td>511.1006</td>
<td>498.6707</td>
<td>30.1888</td>
<td>12.7821</td>
</tr>
</tbody>
</table>

Table 4 shows that for the full data set or cases with relatively low missingness rates (treatment A with missing rate up to 10%, treatment B with missing rate up to 20%), the $\hat{\Sigma}$ are positive definite. In these cases, the calibrated matrices $P_*(\hat{\Sigma})$ are identical to $\hat{\Sigma}$, the calibration keeping $\hat{\Sigma}$ unchanged. When the missing rate increases to 20% for treatment A and 30% for treatment B, $\hat{\Sigma}$ become non-positive definite. In these circumstances, the proposed calibration method yields surrogate matrices $P_*(\hat{\Sigma})$ that are positive-definite and whose Frobenius distances to $\Sigma_s$ are shorter than those from $\hat{\Sigma}$.

4.2. CD4+ data

Figure 1: Variograms of CD4+ data
The CD4+ data comes from an AIDS cohort study [12] comprising 369 infected patients. In total, 2376 repeated measurements of CD4+ cell counts were taken over a period of eight and half years. The data are highly unbalanced, with measurement times varying from subject to subject. Diggle and Verbyla [5] analyzed the CD4+ data using their proposed nonparametric covariance structure estimation method. Their estimated covariance matrix turns to be non-positive definite, however. We reanalyze the CD4+ data and then use our proposed calibration method to calibrate the original covariance matrix estimate.

![Figure 2: Original covariance estimate \( \hat{\Sigma} \) (left) and its calibration matrix \( P_r(\hat{\Sigma}) \) (right) in the CD4+ data](image)

The estimated variogram surface is presented in Figure 1, corresponding to Figure 8 of [5]. The variogram varies for time pairs with equal lags, implying that the underlying longitudinal process for the CD4+ cell counts may be non-stationary. In Figure 2, we plot the original covariance matrix estimate \( \hat{\Sigma} \) and its calibrated covariance matrix \( P_r(\hat{\Sigma}) \), where \( \hat{\Sigma} \) is found to be non-positive definite, as mentioned by [5]. From Figure 2 we can see that \( \hat{\Sigma} \) and \( P_r(\hat{\Sigma}) \) are very similar in terms of shape and structure, indicating that the calibration approach maintains the major characteristics of the original covariance matrix.
5. Discussion

We have proposed a calibration approach that provides a positive definite surrogate for any given non-positive definite matrix. The calibrated covariance matrix preserves the major characteristics of the original matrix, while being closer to the true covariance than the original matrix, in the sense of the Frobenius norm. Figure 3 illustrates the idea behind our approach, where the dashed circle represents the domain $\mathcal{D}_0$ of all positive semidefinite matrices whilst the solid circle is the domain $\mathcal{D}_c$ of all positive definite matrices with eigenvalues no smaller than $c$. Given a positive constant $c$, a non-positive definite matrix $\hat{\Sigma}$’s nearest positive definite matrix $P_c(\hat{\Sigma})$ will be closer to the true covariance matrix, provided $\Sigma \in \mathcal{D}_c$. We might therefore expect that using our positive definite surrogate will improve efficiency and accuracy in mean estimation.

One potential extension is to replace the domain $\mathcal{D}_c$ by a more general set $\{A : A - \text{diag}(c_1, \ldots, c_n) \succeq 0\}$ where the $c_i$ are all positive. However, this extension implies different restrictions imposed on the eigenvalues at different positions, which may be difficult to justify. It also uses the same number of parameters as the dimension of the covariance matrix, requiring
intensive computational efforts, particularly for high-dimensional data. We feel that this particular extension is unlikely to offer substantial benefit.

Our proposed method is not constrained by model assumptions, and hence can be used in both likelihood-based methods (such as generalized linear models) and moment-based approaches (such as generalized estimating equations). Neither is it limited by data structures, indicating it can be applied into any multivariate data setting. In principle, it is applicable to any field of multivariate data analysis where non-positive definiteness of a covariance matrix estimator is a concern. Since the proposed approach is a calibration approach, rather than a covariance estimation approach itself, it can be directly incorporated in any existing covariance estimation process, and offers a routine check and calibration of covariance matrix estimators.

Appendix: Proof of Theorem 1

We seek

$$P_c(X) = \arg\min_{A \in D_c} \|X - A\|.$$ 

Rewrite this as

$$\arg\min_{A: (A - cI) \in D_0} \|(X - cI) - (A - cI)\| = \arg\min_{A' \in D_0} \|(X - cI) - A'\| + cI.$$ 

From [9], this latter is just $P_0(X - cI) + cI$, as required. If, further, $X$ is symmetric, it has spectral decomposition $X = Q\Lambda Q^\top$ (say) for orthogonal $Q$ and diagonal $\Lambda$. Therefore $X - cI = Q(\Lambda - cI)Q^\top$, and

$$P_c(X) = Q\text{diag}\{\max(\lambda_1 - c, 0), \ldots, \max(\lambda_n - c, 0)\}Q^\top + cI.$$ 

But $cI = QcQ^\top$, so $P_c(X) = Q\text{diag}\{\max(\lambda_1, c), \ldots, \max(\lambda_n, c)\}Q^\top$, as required.

□

References


