Received: 6 May 2020 | Revised: 25 September 2020 | Accepted: 25 September 2020

DOI: 10.1002/bimj.202000129

Biometrical Journal

RESEARCH PAPER

D-optimal designs of mean-covariance models for longitudinal data

Siyu Yi[1,2] | Yongdao Zhou[1] | Jianxin Pan[3]

[1] School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, P. R. China

[2] College of Mathematics, Sichuan University, Chengdu, P. R. China

[3] Department of Mathematics, University of Manchester, Manchester, UK

Correspondence
Yongdao Zhou, School of Statistics and Data Science, LPMC, and KLMDASR, Nankai University, Tianjin 300071, P. R. China.
Email: ydzhou@nankai.edu.cn

Funding information
National Key R&D Programme of China, Grant/Award Number: 2018YFC0831900

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to their computational complexity.

Abstract

Longitudinal data analysis has been very common in various fields. It is important in longitudinal studies to choose appropriate numbers of subjects and repeated measurements and allocation of time points as well. Therefore, existing studies proposed many criteria to select the optimal designs. However, most of them focused on the precision of the mean estimation based on some specific models and certain structures of the covariance matrix. In this paper, we focus on both the mean and the marginal covariance matrix. Based on the mean–covariance models, it is shown that the trick of symmetrization can generate better designs under a Bayesian D-optimality criterion over a given prior parameter space. Then, we propose a novel criterion to select the optimal designs. The goal of the proposed criterion is to make the estimates of both the mean vector and the covariance matrix more accurate, and the total cost is as low as possible. Further, we develop an algorithm to solve the corresponding optimization problem. Based on the algorithm, the criterion is illustrated by an application to a real dataset and some simulation studies. We show the superiority of the symmetric optimal design and the symmetrized optimal design in terms of the relative efficiency and parameter estimation. Moreover, we also demonstrate that the proposed criterion is more effective than the previous criteria, and it is suitable for both maximum likelihood estimation and restricted maximum likelihood estimation procedures.

KEYWORDS

Bayesian, cost function, D-optimality criterion, sequential number-theoretic optimization (SNTO)

1 | INTRODUCTION

Longitudinal data are very common in practice, for example, the randomized controlled trials in health and medical sciences, the quality control in industry and the growth curve analysis in biological and agriculture. In longitudinal data analysis, measurements are taken from the same subject repeatedly over time. The responses between subjects may be independent, but the repeated measurements within subjects are very likely to be correlated.

It is well known that the misspecification of the covariance structure may lead to a great loss of efficiency of the mean parameter estimators. Also, if the longitudinal data contain certain missing values and/or are not normally distributed, the mean parameter estimators may be biased when the covariance structure is misspecified (Daniels & Zhao, 2003).

Thus, a good covariance modelling approach improves the statistical inference of the mean of interest. Furthermore, the covariance matrix itself may be of scientific interest. Actually, the impact of the covariance misspecification is of particular concern to researchers conducting longitudinal randomized controlled trials where optimistic standard errors may lead to over-reporting of beneficial intervention effects and vice versa (MacKenzie & Reeves, 2002). However, the issue is more general, and these difficulties may occur whenever the covariance structure is misspecified. Thus many authors studied how to model the mean and the covariance matrix for longitudinal data. The main difficulty in modelling the covariance matrix is the constraint that it must be positive definite. Pourahmadi (2000) provided a convenient reparameterization of the marginal covariance matrix and proposed the joint mean–covariance models for balanced longitudinal data, in which the parameters were estimated by maximum likelihood estimation (MLE). This modelling method overcomes this difficulty and changes the constrained problem to unconstrained and statistically meaningful parameter estimation. MacKenzie and Pan (2001) extended the work to model the dependence of this covariance structure on baseline covariates, time and their interaction. Pan and MacKenzie (2003) used a BIC-based (Bayesian information criterion (BIC)) model selection criterion to identify the optimum degrees of the models. Ye and Pan (2006) considered the modelling of covariance structures in generalized estimating equations (GEEs). To alleviate the problem of inefficient estimation and downward bias in the variance estimates, Papageorgiou (2012) developed restricted maximum likelihood estimation (RMLE) for the parameters in mean–covariance models.

From the perspective of experimental designs, it is well known that optimal designs of experiments are more efficient when more knowledge about the underlying model is available. Based on those modelling methods, a natural problem is to select the optimal designs, that is optimal numbers of repeated measurements and independent subjects and optimal allocation of time points as well, to estimate the mean and the covariance matrix well. The same allocation of time points for each subject is suitable in the selection process for practical application. Then the collected longitudinal data based on the selected optimal designs will be balanced if there is no missing observation. Many studies discussed the modelling methods for the balanced longitudinal data besides Pourahmadi (2000). Forcina (1992) discussed the MLEs of the marginal covariance parameters. Gosho et al. (2011) proposed a criterion for selecting an appropriate correlation structure used in the GEE approach for the balanced longitudinal data. Westgate (2014) proposed a novel criterion that utilized the trace of the empirical covariance matrix to improve the selection approach.

However, most of the existing studies focused on the estimation precision of the mean and few on the covariance matrix. Ouwens et al. (2002) discussed the maximin D-optimal designs. They maximized the minimal relative efficiency, which was defined on the mean parameter estimator with a given covariance parameter space. Because of practical restrictions, we may hope that the cost can be controlled or be as low as possible. Tekle et al. (2008) compared different types of D-optimal cohort designs and took the cost function into account. However, both of them were based on specific linear mixed model (LMM; Laird and Ware, 1982) and restricted the structure of the covariance matrix, such as independence or exponential correlation, which may be unreasonable in practice. Therefore, in this paper, we focus on the estimation effects of both the mean and the marginal covariance matrix and select optimal designs based on the mean–covariance models.

This paper is organized as follows. Section 2 gives a brief introduction to the mean–covariance models. Then we introduce the trick of symmetrization for the designs under the Bayesian D-optimality criterion. Section 3 shows the proposed criterion which takes both the parameter estimation and the cost for recruiting subjects and repeated measurements into account. The numerical algorithm MSNTO (modified sequential number theoretic optimization) for solving the problem is given in Section 4. Section 5 presents an application to a real dataset, and some simulation studies are also conducted. Section 6 gives some conclusions.

2 | D-OPTIMALITY CRITERION BASED ON THE MEAN–COVARIANCE MODELS

In this section, we first briefly describe the concept of the mean–covariance models. The contents are from the statements in Pourahmadi (1999, 2000). Then based on the mean–covariance models, we develop some theoretical results about the goodness of the symmetrized designs under the D-optimality criterion.

Like Ouwens et al. (2002), we consider designs where all subjects are measured at the same time points for practical application. The collected data should be balanced if there is no missing data, and we assume the within-subject correlation is the same for each subject. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})'$ be an $n \times 1$ stacked vector of n responses made at times $t_1, t_2, \dots, t_n$ for the ith subject, $i = 1, \dots, m$. It is assumed that $\boldsymbol{y}_i \sim N_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{in})'$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$ are the $n \times 1$ mean vector and the $n \times n$ covariance matrix of $\boldsymbol{y}_i$, respectively. $\boldsymbol{\Sigma}$ is assumed to be positive definite. Pourahmadi (1999) showed that the modified Choleskey decomposition Newton (1988) of $\boldsymbol{\Sigma}^{-1}$ offers a simple unconstrained and statistically

meaningful reparameterization of the covariance matrix. In fact, there exists a unique lower triangular matrix $\boldsymbol{T}$ with 1s as the diagonal entries and a unique diagonal matrix $\boldsymbol{D}$ with positive diagonals such that

$$\boldsymbol{T\Sigma T}' = \boldsymbol{D} \quad \text{or} \quad \boldsymbol{\Sigma}^{-1} = \boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}.$$

This decomposition has a simple statistical interpretation: the below-diagonal entries of $\boldsymbol{T}$ are the negatives of the autoregressive coefficients, ϕ_{jg}, in the model $\hat{y}_{ij} = \mu_{ij} + \sum_{g=1}^{j-1} \phi_{jg}(y_{ig} - \mu_{ig})$, and the diagonal entries of $\boldsymbol{D}$ are the prediction error (innovation) variances $\sigma_j^2 = var(y_{ij} - \hat{y}_{ij})$, for $1 \le j \le n, 1 \le i \le m$. Since ϕ_{jg} and $\log \sigma_j^2$ are unconstrained, they may be modelled in terms of covariates. Thus, for $1 \le g \le j-1$, $1 \le j \le n$ and $1 \le i \le m$, Pourahmadi (1999) proposed the joint mean–covariance models which could estimate the mean and the marginal covariance matrix simultaneously, and there is no restriction on the structure of the covariance,

$$\mu_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}, \quad \log \sigma_j^2 = \boldsymbol{z}_j'\boldsymbol{\lambda}, \quad \phi_{jg} = \boldsymbol{z}_{jg}'\boldsymbol{\gamma}, \tag{1}$$

where $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of covariates which may contain baseline covariates, polynomials of time and their interactions as well, $\boldsymbol{z}_j$ and $\boldsymbol{z}_{jg}$ are $d \times 1$ and $q \times 1$ vectors of polynomials of time, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are the associated parameters. The three submodels in (1) are called mean model, innovation variance model and autoregressive model, respectively. For example, when we use polynomials in time to model the mean, autoregressive parameters and innovation variances, the covariates may take the forms $\boldsymbol{x}_{ij} = (1, t_j, \ldots, t_j^{p-1})'$, $\boldsymbol{z}_j = (1, t_j, \ldots, t_j^{d-1})'$,$\boldsymbol{z}_{jg} = (1, (t_j - t_g), \ldots, (t_j - t_g)^{q-1})'$, where the within-subject correlation only depends on the elapsed time.

Following Pourahmadi (2000), for estimating the parameters in (1), MLE is a suitable choice and the log-likelihood function, up to the additive constant $mn \log 2\pi$, is as follows:

$$\begin{aligned} -2l(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) &= m \log |\boldsymbol{\Sigma}| + \sum_{i=1}^{m} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \\ &= m \sum_{j=1}^{n} \log \sigma_j^2 + \sum_{i=1}^{m} \{\boldsymbol{r}_i - \boldsymbol{Z}(i)\boldsymbol{\gamma}\}'\boldsymbol{D}^{-1}\{\boldsymbol{r}_i - \boldsymbol{Z}(i)\boldsymbol{\gamma}\} \\ &= m \sum_{j=1}^{n} \log \sigma_j^2 + \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(r_{ij} - \hat{r}_{ij})^2}{\sigma_j^2}, \end{aligned}$$

where $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in})'$, $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} = (r_{ij})_{j=1}^n$, $\boldsymbol{Z}(i) = (\boldsymbol{z}(i,1), \ldots, \boldsymbol{z}(i,n))'$, $\boldsymbol{z}(i,j) = \sum_{g=1}^{j-1} r_{ig}\boldsymbol{z}_{jg}$ and $\hat{r}_{ij} = \sum_{g=1}^{j-1} r_{ig}\phi_{jg} = \boldsymbol{z}'(i,j)\boldsymbol{\gamma}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}')'$. The corresponding score function $U(\boldsymbol{\theta}) = (U_1'(\boldsymbol{\beta}), U_2'(\boldsymbol{\lambda}), U_3'(\boldsymbol{\gamma}))'$ and the Fisher information matrix I_θ are obtained by $U_1(\boldsymbol{\beta}) = \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}_i$, $U_2(\boldsymbol{\lambda}) = \frac{1}{2}\sum_{i=1}^m \boldsymbol{Z}'(\boldsymbol{D}^{-1}\boldsymbol{R}_i - \boldsymbol{1}_n)$, $U_3(\boldsymbol{\gamma}) = \sum_{i=1}^m \boldsymbol{Z}'(i)\boldsymbol{D}^{-1}(\boldsymbol{r}_i - \boldsymbol{Z}(i)\boldsymbol{\gamma})$ and

$$I_\theta = I_\theta(n, t_1, \ldots, t_n, m) = \begin{pmatrix} \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{m}{2}\boldsymbol{Z}'\boldsymbol{Z} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & m\boldsymbol{W} \end{pmatrix}, \tag{2}$$

where $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, $\boldsymbol{R}_i = (e_{i1}, \ldots, e_{in})'$, $e_{ij} = (r_{ij} - \hat{r}_{ij})^2$, and $\boldsymbol{W} = \sum_{j=1}^n \sigma_j^{-2}(\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}\sigma_{kl}\,\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}')$. Then the MLEs $\hat{\boldsymbol{\beta}} = (\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i)^{-1}\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_i$ and $\hat{\boldsymbol{\gamma}} = (\sum_{i=1}^m \boldsymbol{Z}'(i)\boldsymbol{D}^{-1}\boldsymbol{Z}(i))^{-1}\sum_{i=1}^m \boldsymbol{Z}'(i)\boldsymbol{D}^{-1}\boldsymbol{r}_i$. For $\boldsymbol{\lambda}$, the Fisher scoring algorithm is applied to update current value $\tilde{\boldsymbol{\lambda}}$ to $\hat{\boldsymbol{\lambda}}$ where the iterative formula is $\hat{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}} + I_{22}^{-1}(\tilde{\boldsymbol{\lambda}})U_2(\tilde{\boldsymbol{\lambda}})$ and we regard the convergence solution as the MLE of it. Furthermore, under some regularity conditions, the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\lambda}}', \hat{\boldsymbol{\gamma}}')'$ were also obtained in Pourahmadi (2000). Given $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$ can be regarded as asymptotically orthogonal and their asymptotic covariance matrices are the block diagonal elements of $I_{\theta_0}^{-1}$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Detailed discussion refers to Pourahmadi (2000) and Ye and Pan (2006).

In addition, it is well known that in the LMMs, the variance parameter estimators have less biases with RMLE than with MLE. The RMLE for $\boldsymbol{\theta}$ in the mean–covariance models, which adjusts for the degrees of freedom lost due to the estimation of the mean parameter $\boldsymbol{\beta}$, was also discussed in Papageorgiou (2012). Following Papageorgiou (2012), since $\boldsymbol{\beta}$

is orthogonal to $(\boldsymbol{\lambda}, \boldsymbol{\gamma})$, the estimate of $\boldsymbol{\beta}$ varies only slowly with $(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ and almost identical estimates of $\boldsymbol{\beta}$ are obtained by MLE and RMLE. It also turned out that the correlation parameter $\boldsymbol{\gamma}$ is not misestimated by MLE, while RMLE indeed brings better estimates for the variance parameter $\boldsymbol{\lambda}$. However, both Jiang (1996) and Papageorgiou (2012) stated that when the number of the parameters is fixed or bounded, MLE and RMLE are asymptotically equivalent, that is the asymptotic covariance matrix of RMLEs is the same as that of MLEs. Moreover, utilizing the asymptotic property to construct the optimality criterion is common in optimal designs. Thus if the criterion is based on the asymptotic covariance matrix, optimal designs can be suitable for both MLE and RMLE procedures.

D-optimal allocation of the time points for longitudinal models has been studied in the literature. In this paper, we also consider the D-optimality. A design is called D-optimal if it minimizes the determinant of the covariance matrix of the parameter estimators. A D-optimal design also minimizes the volume of the confidence ellipsoid for the estimators, and it is invariant to non-degenerate linear transformations of the variables in the model. However, it can be seen that I_θ in (2) still depends on the unknown covariance matrix $\boldsymbol{\Sigma}$. The prior information about $\boldsymbol{\Sigma}$, equivalently, that about $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$, is needed. To solve this problem, one mainstream approach is to derive the optimal design for a particular guessed value of $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$. Here, we would adopt the Bayesian optimal design approach, which is a good compromise over a range of the parameter values as a prior information. Let the time points design space be Δ and $\tau = (t_1, \dots, t_n) \in \Delta$ with $t_i \in [-T, T]$, $i = 1, \dots, n$. We define the Bayesian D-optimality criterion as

$$\max_{\Delta} \frac{1}{|\Phi|} \sum_{\Phi} |I_\theta(n, \tau, m)|^{\frac{1}{p+d+q}},$$

where Φ is the prior parameter space, $|\Phi|$ is the cardinality of Φ and we take the $(p + d + q)$th root of $|I_\theta|$ to eliminate the effect of the order. Thus the selected design shall be suitable for both MLE and RMLE.

In the following part, we discuss some novel results about the designs based on the Bayesian D-optimality criterion. The proofs of the following proposition and theorem are given in the Appendix. Assume that the contained baseline covariates are not changed with th e considered time points. That is, for each subject and any $\tau_1, \tau_2 \in \Delta$, the values of these baseline covariates under τ_1 are equal to those under τ_2. Define the symmetric design of any design τ by multiplying each time point by -1. Inspired by the Property 1 of Ouwens et al. (2002), we can obtain the similar proposition as follows.

Proposition 2.1. *Given two joint mean–covariance models M_1 and M_2. Let the baseline covariates in X_i, p, d and q be the same for both models, and let the associated parameters $\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^1$ for M_1 and $\boldsymbol{\lambda}^2, \boldsymbol{\gamma}^2$ for M_2 satisfy $\boldsymbol{\lambda}^2 = \boldsymbol{A}\boldsymbol{\lambda}^1$ and $\boldsymbol{\gamma}^2 = \boldsymbol{E}\boldsymbol{\gamma}^1$, where $\boldsymbol{A}$ and $\boldsymbol{E}$, respectively, are $d \times d$ and $q \times q$ diagonal matrices with odd diagonal entries being 1 and even entries being -1. Then,*

$$det\{I_\theta(n, \tau, m)|M_1\} = det\{I_\theta(n, \tilde{\tau}, m)|M_2\}, \tag{3}$$

where $\tilde{\tau}$ is the symmetric design of τ in Δ.

Based on Proposition 2.1, with a specific prior $(\boldsymbol{\lambda}_0^1, \boldsymbol{\gamma}_0^1)$, if τ^* is the locally optimal design for M_1, then with the specific prior $(\boldsymbol{A}\boldsymbol{\lambda}_0^1, \boldsymbol{E}\boldsymbol{\gamma}_0^1)$, $\tilde{\tau}^*$ is also the locally optimal design for M_2. If τ^* is symmetric, that is $\tau^* = -\tau^*$, it is the locally optimal design for both M_1 and M_2. Define the prior parameter space as $\Phi = \{(\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^1), (\boldsymbol{\lambda}^2, \boldsymbol{\gamma}^2) : \boldsymbol{\lambda}^2 = \boldsymbol{A}\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^2 = \boldsymbol{E}\boldsymbol{\gamma}^1\}$ where the ranges of $(\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^1)$ are decided by the practical prior information. Then we can obtain the following theorem.

Theorem 2.2. *For any design $\tau \in \Delta$, we have*

$$\sum_{\Phi} |I_\theta(n, \xi, m)|^{\frac{1}{p+d+q}} \geq \sum_{\Phi} |I_\theta(n, \tau, m)|^{\frac{1}{p+d+q}}, \tag{4}$$

where $\xi = \{(\tau, \frac{1}{2}), (\tilde{\tau}, \frac{1}{2})\}$ and $\tilde{\tau}$ is the symmetric design of τ in Δ.

Proposition 2.1 and Theorem 2.2 are in the spirit of Ouwens et al. (2002) in which they only discussed the maximin designs for the mean estimation in the polynomial random slope model. However, we consider the Bayesian designs for both the mean and the covariance matrix estimations in the mean–covariance models and we extend it to any order of the models. The maximin case can be proved similarly. According to Theorem 2.2, it is known that for any design $\tau \in \Delta$,

the derived symmetrized design ξ is better than τ under the Bayesian D-optimality criterion. It implies that we can first select a general optimal design τ^*, and then for better estimates, half of the subjects are measured according to τ^* and half of the subjects are measured according to $\tilde{\tau}^*$. On the other hand, if τ is a symmetric design, we have $\xi = \tau$. It provides a possible search direction to symmetric designs, which reduce the search space and may lead to good designs.

3 | OPTIMALITY CRITERION FOR PARAMETER ESTIMATIONS AND SAVING COST

In Section 2, based on the mean–covariance models, we discuss the properties of the designs under Bayesian D-optimality criterion. Bayesian D-optimal design can make the estimation of $\boldsymbol{\theta}$ as accurate as possible in the mean–covariance models. However, in practical application we also hope that the total cost is as low as possible. Thus we modify the Bayesian D-optimality criterion and propose an optimality criterion $C_\theta(n, \tau, m)$ to select the optimal design $d_{\text{opt}} = (n^*, \tau^*, m^*)$, that is,

$$\begin{aligned} \max\ & C_\theta(n,\tau,m) \triangleq \log\Big(\frac{1}{|\Phi|}\sum_{\Phi}|I_\theta(n,\tau,m)|^{\frac{1}{p+d+q}}\Big) - (c_1 m + c_2 mn) \\ s.t.\ & m_{\min} \le m \le M_{\max}, \\ & \max\{p,d,q\} \le n \le N_{\max}, \\ & -T_{\max} < t_1 < \cdots < t_n \le T_{\max}, \\ & \min_{i=1:(n-1)}\{t_{i+1} - t_i\} \ge \ell. \end{aligned} \tag{5}$$

Here we use the *log* function to reduce the order of magnitude. Since the numbers of subjects and repeated measurements for each subject usually need to be restricted because of monetary cost, subject fatigue or other logistical reasons, we consider the cost function $c = c_1 m + c_2 mn$, where c_1 and c_2, respectively, are the cost of each subject and repeated measurement. They can adjust the weight between the generalized variance $|I_\theta|$ and the cost function. In practice, the cost of recruiting a new subject for a study is often larger than that of obtaining a measurement for each subject, that is $c_1 \ge c_2$. Comparing with the cost part of the objective function in Tekle et al. (2008), the difference is that we do not fix the cost but want to minimize it instead. The cost function can be treated as a penalty to make a trade-off between the Fisher information matrix and m, m and n. From (2), it is known that larger m indicates more information because of the positive definiteness of the Fisher information matrix, while larger m also increases the cost. Larger n often makes smaller m when the cost function is fixed. The optimization result is also influenced substantially by c_1 and c_2, and they can be set according to the actual situation. Additionally, we make some necessary constraints. The maximal values of m, n and the elements of τ are limited by $M_{\max}, N_{\max}$ and $T_{\max}$, the minimal value of m should also be limited by $m_{\min}$ to guarantee certain estimation accuracy of those parameters. In order to make the Fisher information matrix non-singular, n should be the maximum of $\{p, d, q\}$ at least. The minimal interval of the time points is required to be not less than ℓ for practical application. Additionally, in the computer implementation, because the number of the optimization variables in the optimization problem is different for different n, we sequentially maximize the criterion value for each n and then choose the (n, τ, m) corresponding to the maximal criterion value as the d_{opt}.

Compared with the previous studies that discussed the optimal designs for LMMs, such as Ouwens et al. (2002) and Tekle et al. (2008), the proposed optimality criterion $C_\theta(n, \tau, m)$ takes the estimation precision of both the mean vector and the covariance matrix into account and there is no any restriction on the structure of the covariance matrix, and the literature work only considered the mean alone and made the covariance matrices of the random effect and the random error known. If we just focus on the mean and ignore the cost, the optimality criterion will degenerate into the existing situations.

Furthermore, in order to compare the performance of different designs, we define the relative efficiency (RE). Actually, the values of n and m are not replaceable, for example $n_1 = 2n_2$ cannot be offset by $m_1 = m_2/2$, because of the within-subject correlation. Doubling up the number of time points for each subject does not double up the information, while doubling up the number of subjects indeed doubles up the amount of information. Thus for any two designs $d_1 = (n_1, \tau_1, m_1)$

and $d_2 = (n_2, \tau_2, m_2)$, the RE is defined as

$$RE(d_1, d_2; \Phi) = RE(d_1, d_2) = \frac{m_2}{m_1} \cdot \frac{\sum_{\Phi} |I_\theta(d_1)|^{\frac{1}{p+d+q}}}{\sum_{\Phi} |I_\theta(d_2)|^{\frac{1}{p+d+q}}}, \tag{6}$$

where the term m_2/m_1 is used to eliminate the impact of number of subjects. Given the prior space Φ, if d_2 is an optimal design under C_θ-criterion, the RE is also the efficiency of d_1 relative to the optimal design. Further, if $RE(d_1, d_2) > 1 (< 1)$, the performance of d_1 is better (worse) than d_2, and $RE(d_1, d_2) = 1$ means that the two designs have the same performance.

4 | NUMERICAL IMPLEMENTATION

Many powerful algorithms have been discussed in the literature to solve the optimization problems, such as the gradient descent method, the Gauss–Newton method and so on. However, those methods only provide a local maximal solution and require the derivatives of the objective function. It is not suitable to solve the optimization problem in (5) by those methods since the objective function may have multiple local maxima, and it is difficult to calculate the derivatives of the function. Therefore, we need a new method to solve such a problem. It is known that the SNTO (sequential number-theoretic optimization) algorithm (Fang & Wang, 1994) can find the global maximum and the maximum point of a continuous function over a closed and bounded domain. The SNTO is a sequential algorithm, in which the feasible region is set as a hypercube. It uses the uniform designs to arrange points in the domain and contracts the domain according to some criterion until the required accuracy is achieved. Thus it can serve for the optimization of a multiple modal function without requirements of continuous derivatives and save computing time. The feasible region in (5) is not a hypercube, and thus we consider a modified SNTO, denoted by MSNTO, to solve the optimization problem in (5). Here we define the standard feasible region as

$$N_{n+\upsilon} = \{(t_1, \dots, t_{n+\upsilon}) : 0 < t_1 < \cdots < t_n \le 1, 0 < t_{n+i} \le 1, i = 1, \dots, \upsilon, \min_i\{t_{i+1} - t_i\} \ge \kappa\},$$

where κ should be less than $1/(n-1)$. The key to the MSNTO is to construct the uniform designs on $N_{n+\upsilon}$. Theorem 1.6 of Fang and Wang (1994) shows that applying the inverse transform method can find a set of points uniformly scattered on $N_{n+\upsilon}$ based on $[0,1]^{n+\upsilon}$. This theorem is also provided in the Appendix. Similarly, if $\{\boldsymbol{c}_k\}$ are the points uniformly scattered on $[0,1]^{n+\upsilon}$, we can obtain $\{\boldsymbol{t}_k\}$ which are uniformly scattered on $N_{n+\upsilon}$ by a specific transformation. We subsequently call it $\boldsymbol{h}$-transformation. To derive the $\boldsymbol{h}$-transformation, we first consider the mapping from $[0,1]^{n+\upsilon}$ to $N_{n+\upsilon}$. Let $\boldsymbol{t} = (t_1, \dots, t_{n+\upsilon})$ be a random vector which is uniformly distributed on $N_{n+\upsilon}$, then we have the stochastic representation

$$t_i = \begin{cases} (i-1)\kappa + B\phi_i \cdots \phi_n & i = 1, \dots, n, \\ \phi_i & i = n+1, \dots, n+\upsilon, \end{cases} \tag{7}$$

where $B = 1 - (n-1)\kappa$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{n+\upsilon}) \in [0,1]^{n+\upsilon}$.

Based on the stochastic representation, we can prove that

(1) $\phi_1, \dots, \phi_{n+\upsilon}$ are mutually independent;

(2) ϕ_j has the cumulative distribution function (c.d.f.)$F_j(\phi) = \begin{cases} \phi^j & 1 \le j \le n \\ \phi & n+1 \le j \le n+\upsilon \end{cases}$ and the inverse function of $F_j(\phi)$ is $F_j^{-1}(\phi) = \begin{cases} \phi^{\frac{1}{j}} & 1 \le j \le n \\ \phi & n+1 \le j \le n+\upsilon \end{cases}$, where $j = 1, \dots, n+\upsilon$, $0 \le \phi \le 1$.

Then, we obtain the $\boldsymbol{h}$-transformation $\boldsymbol{t}_k = \boldsymbol{h}(\boldsymbol{c}_k)$ as follows.

Proposition 4.1. *Suppose that* $\{\boldsymbol{c}_k = (c_{k1}, \dots, c_{k(n+\upsilon)})\}$ *are points uniformly scattered on* $[0,1]^{n+\upsilon}$. *Let*

$$t_{ki} = \begin{cases} (i-1)\kappa + B c_{ki}^{\frac{1}{i}} \cdots c_{kn}^{\frac{1}{n}} & i = 1, \dots, n \\ c_{ki} & i = n+1, \dots, n+\upsilon, \end{cases} \tag{8}$$

where $B = 1 - (n-1)\kappa$. *Then* $\{\boldsymbol{t}_k = (t_{k1}, \dots, t_{k(n+\upsilon)})\}$ *are points uniformly scattered on* $N_{n+\upsilon}$.

The proof of Proposition 4.1 is given in the Appendix. Our purpose is to convert the uniform points on $[0,1]^{n+\upsilon}$ to $N_{n+\upsilon}$ and retain the uniformity. Both (7) and (8) can convert $[0,1]^{n+\upsilon}$ to $N_{n+\upsilon}$, but they have different uniformities. We give an illustration for two-dimensional case to verify the validity of Proposition 4.1 in Supplementary Material A.1. Further, after obtaining the $\boldsymbol{h}$-transformation in Proposition 4.1, the MSNTO algorithm can be summarized as the following steps for each fixed n and υ,

Step 1. **Initialization**. Set $w = 0, N_{n+\upsilon}^{(0)} = N_{n+\upsilon}, E^{(0)} = [0,1]^{n+\upsilon}, \boldsymbol{a}^{(0)} = \boldsymbol{0}_{(n+\upsilon)\times 1}, \boldsymbol{b}^{(0)} = \boldsymbol{1}_{(n+\upsilon)\times 1}$;

Step 2. **Generate points uniformly scattered on** $N_{n+\upsilon}^{(w)}$. Generate points $\{\boldsymbol{c}_k^{(w)}, k = 1, \dots, n_w\}$ uniformly scattered on $E^{(w)}$. Then we can obtain points $\mathcal{P}^{(w)} = \{\boldsymbol{t}_k^{(w)} \mid \boldsymbol{t}_k^{(w)} = \boldsymbol{h}(\boldsymbol{c}_k^{(w)}), k = 1, \dots, n_w\}$ uniformly scattered on $N_{n+\upsilon}^{(w)}$;

Step 3. **Compute new approximation**. Find $\boldsymbol{t}^{(w)} \in \mathcal{P}^{(w)} \bigcup \{\boldsymbol{t}^{(w-1)}\}$ and the value of the objective function $C^{(w)}$ such that $C^{(w)} = C(\boldsymbol{t}^{(w)}) \geq C(\boldsymbol{y}), \forall \boldsymbol{y} \in \mathcal{P}^{(w)} \bigcup \{\boldsymbol{t}^{(w-1)}\}$;

Step 4. **Termination criterion**. Let $\boldsymbol{c}^{(w)} = (\boldsymbol{b}^{(w)} - \boldsymbol{a}^{(w)})/2$. If $\max \boldsymbol{c}^{(w)} < \delta$, a pre-determined small number, then $\boldsymbol{t}^{(w)}$ and $C^{(w)}$ are acceptable; terminate algorithm. Otherwise, go to the next step;

Step 5. **Contract domain**. Let $\boldsymbol{o}^{(w)}$ be the point of $[0,1]^{n+\upsilon}$ corresponding to $\boldsymbol{t}^{(w)}$ by $\boldsymbol{t}^{(w)} = \boldsymbol{h}(\boldsymbol{o}^{(w)})$. Form the new domain $E^{(w+1)} = [\boldsymbol{a}^{(w+1)}, \boldsymbol{b}^{(w+1)}]$, where $a_i^{(w+1)} = \max(o_i^{(w)} - \gamma_0 c_i^{(w)}, 0)$, $b_i^{(w+1)} = \min(o_i^{(w)} + \gamma_0 c_i^{(w)}, 1)$, $i = 1, \dots, n+\upsilon$, and γ_0 is a predefined contraction ratio. Let $N_{n+\upsilon}^{(w+1)}$ be the subdomain of $N_{n+\upsilon}$, which corresponds to $E^{(w+1)}$ by the relation $N_{n+\upsilon}^{(w+1)} = \{\boldsymbol{t} \mid \boldsymbol{t} = \boldsymbol{h}(\boldsymbol{u}), \boldsymbol{u} \in E^{(w+1)}\}$. Set $w = w + 1$. Go to Step 2.

Let $D_{n+1} = \{(t_1, \dots, t_n, m) : -T_{\max} < t_1 < \cdots < t_n \leq T_{\max}, m_{\min} \leq m \leq M_{\max}, \min_i\{t_{i+1} - t_i\} \geq \ell\}$. We use the MSNTO algorithm to solve the optimization problem in (5) through simple transformation from N_{n+1} to D_{n+1}. In Step 2, we use the good lattice point method with a power generator (Korobov, 1959) to construct some designs where each factor has n_w levels and calculate their mixture discrepancy (MD) values (Zhou et al., 2013). The MD is a reasonable measure of uniformity for uniform designs. For a design $\boldsymbol{D} = (d_{ij})$ with s runs and $n+1$ factors, the square value of MD is

$$MD^2(\boldsymbol{D}) = \left(\frac{19}{12}\right)^{n+1} - \frac{2}{s}\sum_{i=1}^{s}\prod_{j=1}^{n+1}\left(\frac{5}{3} - \frac{1}{4}|d_{ij} - \frac{1}{2}| - \frac{1}{4}|d_{ij} - \frac{1}{2}|^2\right)$$

$$+ \frac{1}{s^2}\sum_{i=1}^{s}\sum_{k=1}^{n+1}\prod_{j=1}^{n+1}\left(\frac{15}{8} - \frac{1}{4}|d_{ij} - \frac{1}{2}| - \frac{1}{4}|d_{kj} - \frac{1}{2}| - \frac{3}{4}|d_{ij} - d_{kj}| + \frac{1}{2}|d_{ij} - d_{kj}|^2\right).$$

We choose the design with the minimal MD value as the uniform design for each iteration and each n and transform the levels of the last column into $[m_{\min}, M_{\max}]$ through pseudo-level transformation (Fang et al., 2006). If n_w is even, we generally construct designs with run size $n_w + 1$ first to obtain more alternative designs. Then we delete the last row and choose the design with best uniformity as the uniform design. We transform the uniform designs onto $[0,1]^{n+1}$ by $d_{ki} = (2u_{ki} - 1)/(2q_i)$, $k = 1, \dots, n_w$, $i = 1, \dots, n+1$, where $q_i = n_w$ for $1 \leq i \leq n$ and $q_{n+1} = M_{\max}$. Set $c_{ki} = a_i + (b_i - a_i)d_{ki}$, $k = 1, \dots, n_w, i = 1, \dots, n+1$. Then $\{\boldsymbol{c}_k, k = 1, \dots, n_w\}$ is a uniform design on $[\boldsymbol{a}, \boldsymbol{b}]$. The parameter κ should be $\ell/(2T_{\max})$ to ensure that the minimal time interval of the transformed time points on $[-T_{\max}, T_{\max}]^n$ is ℓ. In general, the first iteration step needs more runs to increase the possibility for approaching the global maximum. Moreover, we can use the same uniform design on $[0,1]^{n+1}$ for the iteration steps $w \geq 2$. It can decrease the computational complexity and the precision for finding global maximum will be higher on smaller domain with the same uniform design. In Step 3, we compare the current maximum point with the previous one to guarantee that the objective function is monotonically non-decreasing. In Step 4, the termination criterion is that the experimental domain is small enough rather than $|C^{(w)} - C^{(w-1)}| < \delta$. The

reason is that it may occur $C^{(w)} = C^{(w-1)}$ in the iterative process, while the value may increase when the experimental domain continues contracting. And the contraction ratio γ_0 is often taken to be 0.5.

Based on the discussion in Section 2, we can also select the symmetric optimal designs based on the class of symmetric time points for each n. The steps for constructing symmetric time points in the standard feasible region N_n can be found in Supplementary Material A.2. In fact, the run sizes of the uniform designs in MSNTO should be large to guarantee a certain precision, especially for large n. However, if we just consider the symmetric time points, the numbers of the columns that need to be constructed for the uniform designs are only half of the original numbers. Thus, this approach not only reduces the search space but also reduces the computational complexity of the MSNTO algorithm. These statements will be demonstrated in the next section.

5 | DATA ILLUSTRATION: CATTLE DATA

In this section, we apply the proposed C_θ-criterion and the MSNTO algorithm to a real application—the cattle data (Kenward, 1987). In this balanced longitudinal study, 60 cattle were assigned randomly and equally to two treatment groups A and B. The animals were weighted 11 times over the 133-day period at 0, 14, 28, 42, 56, 70, 84, 98, 112, 126 and 133 days on schedule, which closes to a uniform design. Their weights were recorded to study the effect of treatment on the intestinal parasites. Thus it is necessary to estimate the mean and the covariance matrix of the weights for each animal. Focusing on group A, the mean–covariance models are appropriate for analysing these data (Pan & MacKenzie, 2003; Pourahmadi, 1999, 2000). Following Pan and MacKenzie (2003), for the data in group A, the best fitting models for μ_{ij}, $\log\sigma_j^2$ and ϕ_{jg} in the mean–covariance models are the eighth, third and fourth degree polynomials in week, respectively, as shown below (1) with (p, d, q) =(9, 4, 5). The optimal designs only focusing on the mean do not take the estimation of the covariance into account. It can lead to larger estimation bias and variance for the covariance matrix, which will be illustrated in the following simulation. Hence, to improve the estimations of both mean and covariance matrix, it is meaningful to select optimal designs under the mean–covariance models for the follow-up study. Here we use the same models to re-analysed the data with the original times rescaled to $t_{\text{ori}} = -10, -8, \ldots, 8, 9$. The MLEs are $\hat{\boldsymbol{\theta}}_o = (\hat{\boldsymbol{\beta}}_o', \hat{\boldsymbol{\lambda}}_o', \hat{\boldsymbol{\gamma}}_o')'$, where $\hat{\boldsymbol{\beta}}_o$ =(294, 6.27, −0.19, −4.91e-2, −1.42e-2, 1.3e-3, 3.21e-04, −1.06e-05, −1.87e-06)$'$, $\hat{\boldsymbol{\lambda}}_o$ =(3.48, 0.01, 4.44e-3, −7.96e-4)$'$ and $\hat{\boldsymbol{\gamma}}_o$ =(0.30, 0.31, −0.10, 8.64e-3, −2.31e-4)$'$. For locally optimal design, we regard $\hat{\boldsymbol{\alpha}}_o = (\hat{\boldsymbol{\lambda}}_o', \hat{\boldsymbol{\gamma}}_o')$ as the specific priors. For the Bayesian case, we choose the prior parameter space Φ with $\boldsymbol{\lambda}^1 \in [-3.6, -3.2] \bigcup [3.2, 3.6] \times [-0.02, 0.02] \times \{4.44e-3\} \times \{-7.96e-4\}$, $\boldsymbol{\gamma}^2 \in [-0.5, 0.5] \times [-0.4, 0.4] \times [-0.2, 0.2] \times \{8.64e-3\} \times \{-2.31e-4\}$ and any randomly sampled prior is independently and uniformly distributed. Here we fix the small parameters, and one can also choose other priors. We consider the cases for the cost: $(c_1, c_2)/\rho_0 \in \{(0,0), (50,10), (100,10), (500,10)\}$, where ρ_0 = 1e-4, and it balances the magnitude of the generalized variance and the cost. Let the bounds of the numbers of the follow-up subjects $M_{\max}$ and $m_{\min}$ be 100 and 30, the maximal values of the repeated measurements $N_{\max}$ and the specific time points $T_{\max}$ be 11 and 10, the minimal time interval $\ell = 1$. Three types of optimal designs are considered, that is the general optimal design (d_o), the symmetric optimal design (d_{sym}) and the symmetrized optimal design (d_{sym2}). The design d_o is selected from all possible allocation of time points on $[-T_{\max}, T_{\max}]$ to maximize C_θ-criterion. For the design d_{sym}, we restrict the searching to the symmetric time points. The design d_{sym2} is derived from d_o by symmetrization as in Theorem 2.2. These optimal designs are selected using MSNTO. We compare the performance of these optimal designs and the original design (d_{ori} = (11, t_{ori}, 30)). The locally and Bayesian optimal designs are also compared. Further, we study the effect of different costs, the sensitivity of optimal designs to the prior parameter intervals, the comparison of designs selected based on C_θ-criterion and the criterion which only focuses on the mean, and the effects on the selected designs for cases when the mean model and the covariance models are misspecified. Finally, we also compare MLE with RMLE based on the selected optimal designs.

For both local and Bayesian cases, we obtain d_o and d_{sym} under different cost parameter values. For each case, the optimal design is attained at n = 11. In Figure 1, we show the optimal time points, the C-values and the optimal numbers of subjects m^*. When the cost parameters become larger, the number of subjects becomes smaller to reduce the total cost and hence maximize the C-value. Thus the selected optimal design balances the D-optimality and the collection cost. For both local and Bayesian cases, it occurs that d_{sym} is better than d_o. The reason may be that the searching precision in the MSNTO without any restriction is lower than that restricted to the symmetric time points. For the local case, the condition about the prior space Φ in (4) cannot be satisfied, while the constructed prior space of the Bayesian case accords with it. Hence it is more likely to select a better design for the Bayesian case in the symmetric time points, that is except

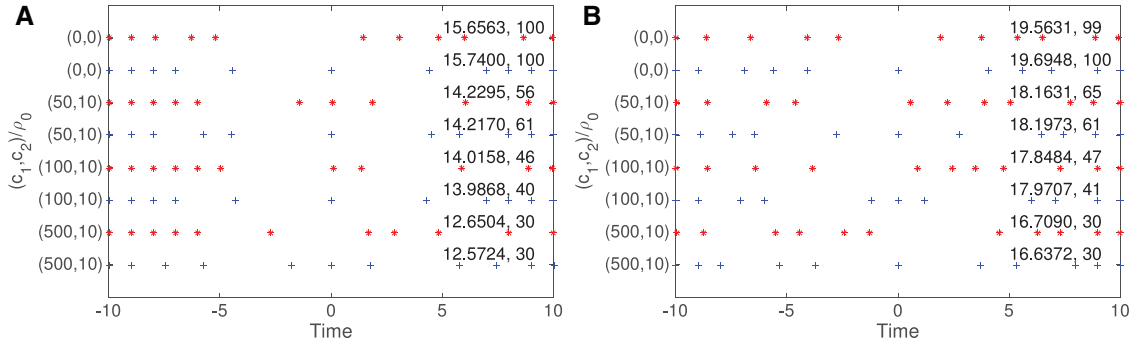

FIGURE 1 (A) The locally optimal designs; (B) the Bayesian optimal designs. The symbols '$*$' and '+' denote for the general and symmetric optimal time points, respectively. Each pair of the values in the subplots, respectively, are the corresponding C-value and m^* for each case

TABLE 1 The RE's among d_{ori}, d_o, d_{sym} and d_{sym2} for local and Bayesian cases

	Locally					Bayesian			
$(c_1,c_2)/\rho_0$	(0,0)	(50,10)	(100,10)	(500,10)	$(c_1,c_2)/\rho_0$	(0,0)	(50,10)	(100,10)	(500,10)
$RE(d_o, d_{\text{ori}}; \hat{\boldsymbol{\alpha}}_o)$	1.67	1.75	1.85	1.71	$RE(d_o, d_{\text{ori}}; \Phi)$	2.85	3.03	2.90	3.38
$RE(d_{\text{sym}}, d_o; \hat{\boldsymbol{\alpha}}_o)$	1.09	0.98	0.98	0.92	$RE(d_{\text{sym}}, d_o; \Phi)$	1.13	1.03	1.14	0.93
$RE(d_{\text{sym2}}, d_o; \hat{\boldsymbol{\alpha}}_o)$	1.27	1.07	1.19	1.29	$RE(d_{\text{sym2}}, d_o; \Phi)$	2.50	3.19	3.18	3.18
$RE(d_{o,n=9}, d_{\text{ori}}; \hat{\boldsymbol{\alpha}}_o)$	1.26	1.21	1.18	1.25	$RE(d_{o,n=9}, d_{\text{ori}}; \Phi)$	1.27	1.25	1.30	1.17
$RE(d_{o,n=10}, d_{\text{ori}}; \hat{\boldsymbol{\alpha}}_o)$	1.43	1.62	1.54	1.34	$RE(d_{o,n=10}, d_{\text{ori}}; \Phi)$	2.36	2.54	2.52	2.36

for the case $(c_1,c_2)/\rho_0$ =(500, 10), the Bayesian d_{sym}s are slightly better than d_os. Thus, under the same cost parameters, the performance of d_{sym} is similar to d_o, even better.

Further, we compare the performance of d_{ori}, d_o, d_{sym} and d_{sym2} through RE for both local (under single prior $\hat{\boldsymbol{\alpha}}_o$) and Bayesian cases (under prior space Φ). From Table 1, d_o shows better performance than d_{ori}, and d_{sym2} is better than d_o. $RE(d_{\text{sym}}, d_o)$ is always around 1. Thus d_o, d_{sym} and d_{sym2} are all better than d_{ori}, and d_{sym} shows similar performance to d_o. This coincides with the C-values in Figure 1 exactly. d_{sym2} is better than both d_o and d_{sym}. It demonstrates the validity of Theorem 2.2. In addition, we also obtain $RE(d_{o,n=9}, d_{\text{ori}})$ and $RE(d_{o,n=10}, d_{\text{ori}})$ for the optimal designs with n = 9, 10. As shown in the last two rows in Table 1, even though the number of the repeated measurements n is smaller, the selected optimal designs are still better than d_{ori}. And as n becomes larger, the optimal designs show better performance. In addition, in the above discussion, we fix the size of the prior parameter interval. To study the sensitivity of the Bayesian optimal designs for the prior parameter intervals, we select optimal designs under various interval sizes and compare them using RE in Supplementary Material A.3. The results show that the selected designs have good robustness to the interval sizes.

For distinction and simplicity, denote the locally and Bayesian optimal designs by d_*^L and d_*^B, respectively, where symbol '$*$' can be $o, sym, sym2, \beta$. In order to investigate the properties of the selected optimal designs above for the parameter estimations of the mean–covariance models, we assume that the true parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t', \boldsymbol{\lambda}_t', \boldsymbol{\gamma}_t')' = (\hat{\boldsymbol{\beta}}_o', \hat{\boldsymbol{\lambda}}_o', \hat{\boldsymbol{\gamma}}_o')'$. The true mean and true covariance matrix are obtained by $\boldsymbol{\mu}_t = \boldsymbol{X}\boldsymbol{\beta}_t$ and $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_t) = \boldsymbol{T}_t^{-1}\boldsymbol{D}_t(\boldsymbol{T}_t')^{-1}$, which relies on the structure of the model. Thus, once the structure and design are given, the true mean and covariance matrix are also determined. Then, we carry out 100 repetitions to sample $\boldsymbol{y}_i, i = 1, \ldots, m^*$ where $\boldsymbol{y}_i \sim N_n(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and estimate the parameters. Here we consider the selected optimal designs under (c_1,c_2) =(0, 0), and the corresponding results of MLEs are given in Table 2, where (i) se_bta, se_lamb and se_gam denote for the products of the standard errors for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$, respectively; (ii) se_prodt denotes for the product of se_bta, se_lamb and se_gam (the total standard errors); (iii) det_OC denotes for the determinant of the observed covariance matrix of the estimators (the generalized variance); (iv) bia_bta, bia_lamb and bia_gam denote for the summations of the elements of the differences between the true values and the averages for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$, respectively; (v) bia_sumt denotes for the summation of bia_bta, bia_lamb and bia_gam (the total biases). There is no limit to the collection cost, so the optimal numbers of subjects for all the optimal designs are nearly 100, the maximal value. Thus we reset the number of subjects for the original design to 100. Additionally, we also discuss the optimal designs under the case $(c_1,c_2)/\rho_0$ =(100, 10) in Supplementary Material A.4.

TABLE 2 The (a) standard errors and (b) biases of the estimators for different types of designs

(a) Design	se_bta	se_lamb	se_gam	se_prodt	det_OC
d_{ori}	3.0426e-26	4.6213e-10	2.3270e-12	3.2720e-47	3.7582e-119
d_o^L	5.9774e-25	4.3687e-10	8.3970e-14	2.1927e-47	1.8828e-123
d_{sym}^L	8.6906e-27	5.7133e-10	3.4044e-14	1.6904e-49	3.9169e-124
d_{sym2}^L	8.4711e-27	4.4610e-10	1.2206e-13	4.6127e-49	3.4555e-125
d_o^B	1.5069e-26	4.4183e-10	1.9045e-13	1.2679e-48	7.7013e-123
d_{sym}^B	1.2534e-26	3.0116e-10	1.1510e-13	4.3447e-49	1.9851e-123
d_{sym2}^B	8.4847e-27	2.6976e-10	2.6213e-13	5.9997e-49	1.8631e-123
d_β^L	4.1155e-27	3.6106e-10	4.4293e-13	6.5816e-49	1.0115e-122
d_β^B	1.7811e-27	4.1255e-10	6.5402e-13	4.8056e-49	1.2156e-122

(b) Design	bia_bta	bia_lamb	bia_gam	bia_sumt
d_{ori}	0.1647	0.0135	0.0306	0.2088
d_o^L	0.0859	0.0119	0.0079	0.1056
d_{sym}^L	0.1387	0.0140	0.0052	0.1579
d_{sym2}^L	0.0567	0.0024	0.0069	0.0660
d_o^B	0.1169	0.0104	0.0067	0.1341
d_{sym}^B	0.1296	0.0172	0.0089	0.1556
d_{sym2}^B	0.0694	0.0063	0.0186	0.0943
d_β^L	0.1402	0.0176	0.0063	0.1641
d_β^B	0.1837	0.0136	0.0088	0.2061

For all optimal designs based on C_θ-criterion, that is the rows 3–8 in Table 2(a) and (b), the total standard errors and total biases are smaller than those for d_{ori}. The optimality of the generalized variance of $\hat{\boldsymbol{\theta}}$ is also guaranteed by the proposed C_θ-criterion. Thus it implies that the optimal designs are better than the original design. Due to the precise prior, the generalized variances of $\hat{\boldsymbol{\theta}}$ for locally optimal designs are smaller than those for Bayesian optimal designs. However, the latter can still ensure that the standard errors of the estimators are small enough and the total biases of the two kinds of optimal designs are comparable. For both cases, the generalized variances and total biases of $\hat{\boldsymbol{\theta}}$ for d_{sym2}s are the smallest compared with d_{sym}'s and d_o's. Thus, the improvement based on symmetrization is significant. Besides, we select locally and Bayesian optimal designs d_β (i.e. d_β^L and d_β^B) based on the previous criteria which only focus on the mean. The results are shown in the last two rows of Table 2(a) and (b). For both local and Bayesian cases, the standard errors of $\hat{\boldsymbol{\beta}}$ are smaller for d_β. But in an overall sense, most of the results for d_β are larger than those for d_o, for example the generalized variances. Hence, d_o performs better than d_β and C_θ-criterion is more effective than the criteria in the literature when the covariance matrix also needs accurate estimation.

Next, we discuss the effects on the selected optimal designs for cases when the mean and/or the covariance model is misspecified. To avoid the confusion of the effects of misspecification and cost on the selected design, we consider the case (c_1, c_2) =(0, 0). Three cases of misspecified covariance parameter are considered, that is $\boldsymbol{\lambda}_m^i = \hat{\boldsymbol{\lambda}}_o + \boldsymbol{\delta}_1^i; \boldsymbol{\gamma}_m^i = \hat{\boldsymbol{\gamma}}_o + \boldsymbol{\delta}_2^i, i$ =1, 2, 3, where $mis_1 : \boldsymbol{\delta}_1^1 = [-0.1, -0.01, 0, 0], \boldsymbol{\delta}_2^1 = [-0.1, -0.1, -0.01, 0, 0]$; $mis_2 : \boldsymbol{\delta}_1^2 = [-0.5, -0.05, 0, 0], \boldsymbol{\delta}_2^2 = [-0.5, -0.5, -0.05, 0, 0]$; $mis_3 : \boldsymbol{\delta}_1^3 = [-1, -0.1, 0, 0], \boldsymbol{\delta}_2^3 = [-1, -1, -0.1, 0, 0]$. We select the locally general optimal design d_{o,mis_i}^L under the single prior $(\boldsymbol{\lambda}_m^i, \boldsymbol{\gamma}_m^i)$ and Bayesian general optimal design d_{o,mis_i}^B under prior space $\Phi_m^i = \{(\boldsymbol{\lambda}^1 + \boldsymbol{\delta}_1^i, \boldsymbol{\gamma}^2 + \boldsymbol{\delta}_2^i) : (\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^2) \in \Phi\}, i$ =1, 2, 3. Eight cases of misspecified model are considered, that is (p, d, q) =(8, 4, 5), (7, 4, 5), (9, 3, 5), (9, 2, 5), (9, 4, 4), (9, 4, 3), (9, 3, 4), (9, 2, 3). We use the original observations and rescaled time in group A to re-estimate the parameters based on the misspecified models. Then for the misspecified models, we regard the estimators as the single prior to select locally general optimal design d_{o,mis_j}^L, or construct the prior space based on the estimators and select the Bayesian general optimal design $d_{o,mis_j}^B, j = 4, \ldots, 11$. After obtaining these designs, similar to Singh and Mukhopadhyay (2016), we compare the *RE*s of them relative to d_o^L under $\hat{\boldsymbol{\alpha}}_o$ which is selected without misspecification. The results are shown in Table 3. From Table 3(a), the performance of locally optimal design worsens as the degree of parameter misspecification

TABLE 3 The *RE*s (efficiencies) of optimal designs under misspecification relative to d_o under $\hat{\boldsymbol{\alpha}}_o$

(a) Parameter misspecification		
Case	**Design (*RE*)**	
mis_1	d^L_{o,mis_1} (0.91)	d^B_{o,mis_1} (0.87)
mis_2	d^L_{o,mis_2} (0.88)	d^B_{o,mis_2} (0.91)
mis_3	d^L_{o,mis_3} (0.65)	d^B_{o,mis_3} (0.92)

(b) Model misspecification			
Case	$(\boldsymbol{p},\boldsymbol{d},\boldsymbol{q})$	**Design (*RE*)**	
mean	(8, 4, 5)	d^L_{o,mis_4} (0.89)	d^B_{o,mis_4} (0.91)
	(7, 4, 5)	d^L_{o,mis_5} (0.79)	d^B_{o,mis_5} (0.96)
var	(9, 3, 5)	d^L_{o,mis_6} (0.99)	d^B_{o,mis_6} (0.86)
	(9, 2, 5)	d^L_{o,mis_7} (0.98)	d^B_{o,mis_7} (0.91)
auto	(9, 4, 4)	d^L_{o,mis_8} (0.89)	d^B_{o,mis_8} (0.89)
	(9, 4, 3)	d^L_{o,mis_9} (0.89)	d^B_{o,mis_9} (0.90)
cov	(9, 3, 4)	$d^L_{o,mis_{10}}$ (0.83)	$d^B_{o,mis_{10}}$ (0.96)
	(9, 2, 3)	$d^L_{o,mis_{11}}$ (0.86)	$d^B_{o,mis_{11}}$ (0.83)

Note: When there is no misspecification, $RE(d^B_o, d^L_o; \hat{\boldsymbol{\alpha}}_o) = 0.92$. Cases *mean*, *var*, *auto* and *cov* represent that the first, second, third, and both second and third submodels in (1) are misspecified, respectively.

TABLE 4 The standard errors of $\hat{\boldsymbol{\lambda}}$ and the generalized variances of $\hat{\boldsymbol{\theta}}$ for MLE and RMLE

	MLE		**RMLE**	
Design	**se_lamb**	**det_OC**	**se_lamb**	**det_OC**
d_{ori}	4.7179e-10	4.4342e-119	3.4713e-10	2.0171e-119
d^B_o	4.1626e-10	7.5473e-123	1.9264e-10	1.3548e-123
d^B_{sym}	4.2327e-10	1.2728e-123	2.2886e-10	4.5237e-124

increases. With slight parameter misspecification, the locally optimal design performs better than the Bayesian. When the misspecification is relatively large, the conclusion is inversed. From the local case in Table 3(b), the selected designs under cases *mean*, *auto* and *cov* become worse. The case *var* has little effect on the performance of selected design. For case *cov*, the performance of the selected design is dominated mainly by the misspecified autoregressive model. When the locally optimal design performs well, the Bayesian optimal design will worsen it. However, if the locally optimal design has a low *RE*, the Bayesian optimal design will improve it in most cases.

Finally, we carry out 100 repetitions again to sample observations with true parameters $\boldsymbol{\theta}_t$ based on d_{ori}, d^B_o and d^B_{sym} under $(c_1, c_2) = (0, 0)$ and estimate $\boldsymbol{\theta}$ using the MLE procedure and RMLE procedure, respectively. Following Papageorgiou (2012), we examine the effect of the underestimation of the innovation variances on the estimation of the diagonal elements of the true covariance matrix in Supplementary Material A.5. Further, we compare the standard errors of $\hat{\boldsymbol{\lambda}}$ and the generalized variances of $\hat{\boldsymbol{\theta}}$ with MLE and RMLE in Table 4. Clearly, RMLE entails smaller standard errors of $\hat{\boldsymbol{\lambda}}$ and generalized variances of $\hat{\boldsymbol{\theta}}$ than MLE. Moreover, with RMLE, the generalized variance of $\hat{\boldsymbol{\theta}}$ for d^B_{sym} is smaller than that for d^B_o, and that for d^B_o is smaller than d_{ori}. It is consistent with those for MLE. Therefore, these results illustrate that the selected optimal designs are also suitable for RMLE. Note that for these simulation results in Tables 1–4 and Figure 1, if one uses different prior space or other sampled observations, the outcomes may be a little bit different, but the main conclusions stated in the paper still hold.

6 | CONCLUSION

For the mean–covariance models, we first show that the trick of symmetrization can generate better designs under the Bayesian D-optimality criterion. Then, we propose a novel criterion to select the optimal designs of the mean–covariance models for longitudinal data. We assume that the optimal allocation of the time points is the same for each subject in the selection process, and there is no missing observation. The proposed criterion is to make the estimations of the mean and

the covariance matrix more accurate, and the total cost is as low as possible. We give the MSNTO algorithm which can be used to solve the optimization problem corresponding to the proposed criterion.

Based on the cattle data, we select the optimal designs using MSNTO for follow-up study and discuss the impacts of different cost parameters, in which larger cost results in a smaller number of subjects and smaller criterion value. We also show the superiority of the symmetric and symmetrized optimal designs for both local and Bayesian cases. The symmetric optimal designs and the general optimal designs are comparable. The symmetrized optimal designs perform better. We compare the locally and Bayesian optimal designs. With precise priors, the locally optimal designs have better performance than the Bayesian optimal designs. However, the latter also performs well. With less precise knowledge, the Bayesian optimal designs are more suitable. This statement is also illustrated in the simulation with parameter misspecificaiton. The sensitivity of the Bayesian optimal designs to the parameter intervals is also considered. In most cases, the designs show good robustness against the prior parameter interval sizes. Moreover, according to the parameter estimation, we compare the proposed criterion with the previous criteria only focusing on the mean and it shows that our proposed criterion is more effective. Finally, based on the Bayesian optimal designs, we also use RMLE procedure to estimate the parameters. It illustrates that the proposed criterion is also suitable for RMLE, which can reduce the bias of the estimate of the covariance matrix.

More generally, we can further consider the case of missing data. Given a design schedule, there is a possibility that the measurements may not be collected at some time points for certain reasons. In this case, the criterion for searching the optimal design should be robust to the measurements that may be missing. It is an important question for further investigations. In addition, instead of giving numerical designs under the cost constraint only, a further interesting question is to relate the number of the subjects and the number of repeated measurements to the power of the hypothesis test and give optimal allocations of them.

ACKNOWLEDGEMENTS

We would like to thank the associate editor and the two referees for their valuable suggestions. This work was supported by the National Natural Science Foundation of China (11871288 and 11871357), the Natural Science Foundation of Tianjin (19JCZDJC31100) and the National Key R&D Programme of China (2018YFC0831900).

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to their computational complexity.

ORCID

Yongdao Zhou https://orcid.org/0000-0003-3805-7021
Jianxin Pan https://orcid.org/0000-0002-7460-6350

REFERENCES

Daniels, M., & Zhao, Y. (2003). Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine*, *22*, 1631–1647.
Fang, K. T., Li, R. Z., & Sudjianto, A. (2006). *Design and modeling for computer experiments*. Chapman and Hall/CRC.
Fang, K. T., & Wang, Y. (1994). *Number-theoretic methods in statistics*. Chapman and Hall.
Forcina, A. (1992). Modelling balanced longitudinal data: Maximum likelihood estimation and analysis of variance. *Biometrics*, *48*, 743–750.
Gosho, M., Hamada, C., & Yoshimura, I. (2011). Criterion for the selection of a working correlation structure in the generalized estimating equation approach for longitudinal balanced data. *Communications in Statistics - Theory and Methods*, *40*, 3839–3856.
Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics*, *24*, 255–286.
Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics*, *36*, 296–308.
Korobov, N. M. (1959). The approximate computation of multiple integrals. *Doklady Akademii Nauk SSSR*, *124*, 1207–1210.
Laird, N. M., & Ware, J. J. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

MacKenzie, G., & Pan, J. X. (2001). On modelling marginal covariance structures in linear mixed model. *Proceedings of the 16th International Workshop of Statistical Modelling*. Odense, Denmark, July, 275–282.

MacKenzie, G., & Reeves, J. (2002) Modelling bivariate longitudinal data with serial correlation. In D. Gregori, et al. (Eds.), *Correlated data modelling* (pp. 19–26). Franco Angeli.

Newton, H. J. (1988). *TIMESLAB: A time series analysis laboratory*. Wadsworth & Brooks/Cole.

Ouwens, M. J. N. M., Tan, F. E. S., & Berger, M. P. F. (2002). Maximin D-optimal designs for longitudinal mixed effects models. *Biometrics*, *58*, 735–741.

Pan, J. X., & MacKenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika*, *90*, 239–44.

Papageorgiou, G. (2012). Restricted maximum likelihood estimation of joint mean-covariance models. *The Canadian Journal of Statistics*, *40*, 225–242.

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

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, *87*, 425–435.

Singh, S. P., & Mukhopadhyay, S. (2016). Bayesian crossover designs for generalized linear models. *Computational Statistics and Data Analysis*, *104*, 35–50

Tekle, F. B., Tan, F. E. S., & Berger, M. P. F. (2008). D-optimal cohort designs for linear mixed-effects models. *Statistics in Medicine*, *27*, 2586–2600.

Westgate, P. (2014). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine*, *33*, 2222–2237.

Ye, H. J., & Pan, J. X. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika*, *93*, 927–941.

Zhou, Y. D., Fang, K. T., & Ning, J. H. (2013). Mixture discrepancy for quasi-random point sets. *Journal of Complexity*, *29*, 283–301.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Yi S, Zhou Y, Pan J. D-optimal designs of mean-covariance models for longitudinal data. *Biometrical Journal*. 2021;*63*:1072–1085. https://doi.org/10.1002/bimj.202000129

APPENDIX

Proof of Proposition 2.1. Without loss of generality, we first consider the case that the model matrices for the mean are all the same which only contain polynomial of time, that is $\boldsymbol{X}_i = \boldsymbol{X}$. Denote the marginal covariance matrices corresponding to M_1 and M_2 by $\boldsymbol{\Sigma} = \boldsymbol{T}^{-1}\boldsymbol{D}(\boldsymbol{T}')^{-1}$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{T}}^{-1}\tilde{\boldsymbol{D}}(\tilde{\boldsymbol{T}}')^{-1}$. The elements of $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are denoted by σ_j^2 and $\tilde{\sigma}_j^2$, respectively, which satisfy $\log\sigma_j^2 = \boldsymbol{z}_j'\boldsymbol{\lambda}$ and $\log\tilde{\sigma}_j^2 = \tilde{\boldsymbol{z}}_j'\tilde{\boldsymbol{\lambda}}$. The elements of T and $\tilde{T}$ are denoted by ϕ_{jg} and $\tilde{\phi}_{jg}$, respectively, which satisfy $\phi_{jg} = \boldsymbol{z}_{jg}'\boldsymbol{\gamma}$ and $\tilde{\phi}_{jg} = \tilde{\boldsymbol{z}}_{jg}'\tilde{\boldsymbol{\gamma}}$. If the time points in the design matrices for M_1 and M_2 are mutually symmetric, that is, $\tilde{\tau} = -\tau$, we have $\tilde{\boldsymbol{z}}_j = \boldsymbol{A}\boldsymbol{z}_j$ and $\tilde{\boldsymbol{z}}_{jg} = \boldsymbol{E}\boldsymbol{z}_{jg}$. Then $\log\tilde{\sigma}_j^2 = \log\sigma_j^2$, $\tilde{\phi}_{jg} = \phi_{jg}$ and thus $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$.

Let $\boldsymbol{A}_1$ is a $p \times p$ diagonal matrix with odd diagonal entries being 1 and even being -1. Due to the fact that $\tilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{A}_1$, $\tilde{\boldsymbol{Z}} = \boldsymbol{Z}\boldsymbol{A}$ where $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n)'$, $\tilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n)'$, it is easy to show that

$$|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{X}}| = |\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|, \quad |\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}| = |\boldsymbol{Z}'\boldsymbol{Z}|, \quad |\tilde{\boldsymbol{W}}| = |\boldsymbol{W}|.$$

Thus $det\{I_\theta(n,\tau,m)|M_1\} = det\{I_\theta(n,\tilde{\tau},m)|M_2\}$. Further, if the model matrices for the mean also contain other baseline covariates and the interaction terms with time. Through the assumption that the values of these baseline covariates under τ equal to those under $\tilde{\tau}$. Then the corresponding A_1 is still a diagonal matrix with entries 1 or -1. It is easy to show that Equation (3) also holds and we omit it here. □

Proof of Theorem 2.2. Let $\tau \in \Delta$. It follows from Proposition 2.1 that $|I_\theta(n,\tau,m)|$ under $\boldsymbol{\lambda}^1, \boldsymbol{\gamma}^1$ is equal to $|I_\theta(n,\tilde{\tau},m)|$ under $\boldsymbol{\lambda}^2, \boldsymbol{\gamma}^2$. Thus, $\sum_\Phi |I_\theta(n,\tau,m)|^{\frac{1}{p+d+q}} = \sum_\Phi |I_\theta(n,\tilde{\tau},m)|^{\frac{1}{p+d+q}}$. Consider the design $\xi = \{(\tau, \frac{1}{2}), (\tilde{\tau}, \frac{1}{2})\}$ for which half of the subjects are measured according to τ and half of the subjects are measured according to $\tilde{\tau}$. Then ξ is symmetric and

for the Bayesian case we have

$$\begin{aligned}\sum_{\Phi}|I_\theta(n,\xi,m)|^{\frac{1}{p+d+q}} &= \sum_{\Phi}\left|\frac{1}{2}I_\theta(n,\tau,m)+\frac{1}{2}I_\theta(n,\tilde{\tau},m)\right|^{\frac{1}{p+d+q}}\\ &\geq \frac{1}{2}\sum_{\Phi}|I_\theta(n,\tau,m)|^{\frac{1}{p+d+q}}+\frac{1}{2}\sum_{\Phi}|I_\theta(n,\tilde{\tau},m)|^{\frac{1}{p+d+q}}\\ &= \sum_{\Phi}|I_\theta(n,\tau,m)|^{\frac{1}{p+d+q}}.\end{aligned}$$

The inequality follows from the fact that $|\boldsymbol{G}+\boldsymbol{K}|^{\frac{1}{s}} \geq |\boldsymbol{G}|^{\frac{1}{s}}+|\boldsymbol{K}|^{\frac{1}{s}}$, where $\boldsymbol{G}$ and $\boldsymbol{K}$ are positive definite matrices of order s. □

Theorem 1.6 in Fang and Wang (1994) *Let D be a closed and bounded domain in R^s and $\boldsymbol{t} \sim U(D)$. Suppose $\boldsymbol{t}$ have a stochastic representation $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is a $\dot{n}$-dimensional random vector with independent marginal p.d.f. $p_i(\phi_i)$ and c.d.f. $F_i(\phi_i)$. Let $\mathcal{P} = \{\boldsymbol{c}_k, k = 1, \dots, s\}$ be a set uniformly scattered on $[0,1]^{\dot{n}}$. Then the set $\mathcal{P}_F = \{\boldsymbol{t}_k, k = 1, \dots, s\}$, where $\boldsymbol{t}_k = \boldsymbol{t}(F_1^{-1}(c_{k1}), \dots, F_t^{-1}(c_{kt}))$ is also uniformly scattered on D.*

Proof of Proposition 4.1. Based on the stochastic representation in (7), we can obtain the Jacobian of the transformation from $\boldsymbol{t}$ to $\boldsymbol{\phi}$,

$$J(\boldsymbol{t}\to\boldsymbol{\phi}) = \begin{vmatrix} B\phi_2\cdots\phi_n & B\phi_1\phi_3\cdots\phi_n & \cdots & B\phi_1\cdots\phi_{n-1} & 0 & \cdots & 0\\ & B\phi_3\cdots\phi_n & \cdots & B\phi_2\cdots\phi_{n-1} & 0 & \cdots & 0\\ & & \ddots & \vdots & \vdots & & \vdots\\ & & & B & 0 & & 0\\ & & & & 1 & & 0\\ & & & & & \ddots & \vdots\\ & & & & & & 1 \end{vmatrix} = B^n\prod_{i=2}^{n}\phi_i^{i-1}.$$

Because $\boldsymbol{x}$ is uniformly distributed on $N_{n+\upsilon}$, we can write the p.d.f. of $\boldsymbol{\phi}$ by

$$p_\phi(\boldsymbol{\phi}) = p_t(\boldsymbol{t})|J(\boldsymbol{t}\to\boldsymbol{\phi})| = \frac{B^n}{\upsilon(N_{n+\upsilon})}\prod_{i=2}^{n}\phi_i^{i-1},$$

where $p_{\boldsymbol{\phi}}$ and $p_{\boldsymbol{t}}$ represent the p.d.f. of $\boldsymbol{\phi}$ and $\boldsymbol{t}$, respectively, and $\upsilon(N_{n+\upsilon})$ is the volume of $N_{n+\upsilon}$ with

$$\upsilon(N_{n+\upsilon}) = \int_{N_{n+\upsilon}} d\boldsymbol{t} = \int_{[0,1]^{n+\upsilon}} |J(\boldsymbol{t}\to\boldsymbol{\phi})|d\boldsymbol{\phi} = B^n\frac{1}{n!}.$$

Then, we have $p_\phi(\boldsymbol{\phi}) = \prod_{i=1}^{n}(i\phi_i^{i-1})$. This indicates that $\phi_1, \dots, \phi_{n+\upsilon}$ are mutually independent and ϕ_j has p.d.f. $p_j(\phi) = \begin{cases} j\phi^{j-1} & 1 \leq j \leq n \\ 1 & n+1 \leq j \leq n+\upsilon \end{cases}$ and c.d.f. $F_j(\phi) = \begin{cases} \phi^j & 1 \leq j \leq n \\ \phi & n+1 \leq j \leq n+\upsilon \end{cases}$, where $0 \leq \phi \leq 1$. And the inverse function of $F_j(\phi)$ is $F_j^{-1}(\phi) = \begin{cases} \phi^{\frac{1}{j}} & 1 \leq j \leq n \\ \phi & n+1 \leq j \leq n+\upsilon \end{cases}$, where $0 \leq \phi \leq 1, j = 1, \dots, n+\upsilon$. Through Theorem 1.6 in Fang and Wang (1994), we can obtain the $\boldsymbol{h}$-transformation in (8). Further, we can also get the inverse $\boldsymbol{h}$-transformation from $\boldsymbol{t}_k$ to $\boldsymbol{c}_k$,

$$c_{ki} = \begin{cases} \left[\frac{t_{ki}-(i-1)\kappa}{x_{k(i+1)}-i\kappa}\right]^i & i = 1, \dots, n-1, \\ \left[\frac{t_{ki}-(i-1)\kappa}{B}\right]^i & i = n, \\ t_{ki} & i = n+1, \dots, n+\upsilon. \end{cases}$$

□