Statistica Sinica Preprint No: SS-2023-0051	
Title	Optimal Designs for Functional Principal and Empirical Component Scores
Manuscript ID	SS-2023-0051
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0051
Complete List of Authors	Ming-Hung Kao and Ping-Han Huang
Corresponding Authors	Ming-Hung Kao
E-mails	mkao3@asu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Statistica Sinica

OPTIMAL DESIGNS FOR FUNCTIONAL PRINCIPAL AND EMPIRICAL COMPONENT SCORES

MING-HUNG KAO AND PING-HAN HUANG

Arizona State University

Abstract: Sparse functional data analysis (FDA) is powerful for making inference on the underlying random function when noisy observations are collected at sparse time points. To have a precise inference, knowledge on optimal designs that allow the experimenters to collect informative functional data is crucial. Here, we propose a framework for selecting optimal designs to precisely predict functional principal and empirical component scores. Our work gives a relevant generalization of previous results on the design for predicting individual response curves. We obtain optimal designs, and evaluate the performance of commonly used designs. We demonstrate that without a judiciously selected design, there can be a great loss in statistical efficiency.

Key words and phrases: Design efficiency, Exact designs, Mixed model equations, Random function, Sparse orthonormal approximation.

1 Introduction

We are concerned with optimal experimental designs for sparse functional data analysis (FDA). With limited resources and practical constraints, it is common to have only a small number of observations from the underlying random function. Many sparse FDA methods are developed and shown powerful in analyzing such sparse functional data (e.g., Shi et al., 1996; Yao et al., 2005a,b; Hall et al., 2006, 2008; Müller, 2008; Nie et al., 2022; Zhong et al.,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

2022). However, knowledge on optimal designs for collecting informative sparse functional data remains scarce. For many sparse FDA methods, there is no guidance on the evaluation and selection of designs, and one may (randomly) select a design without having a good knowledge of its performance. As demonstrated in Ji and Müller (2017), and in this work, imprudently selected designs can result in a great loss in statistical efficiency, making the experiment inefficient, or even impossible, to achieve the study objectives of interest. The development of knowledge on optimal designs is crucial.

Due to its importance, the research on the design for sparse FDA has recently drawn some attention. As a pioneering work, Ji and Müller (2017) proposed optimal designs for two study objectives, including (i) trajectory recovery of the underlying random function, and (ii) prediction of the scalar response under a scalar-on-function regression. Park et al. (2018) studied bi-objective optimal designs by considering compound criteria of both study objectives. Rha et al. (2020) developed a probabilistic subset selection (PSS) algorithm for finding optimal designs, and extended the previous design works to function-on-function regression. These works provide methods for evaluating designs and obtaining optimal designs. They also highlighted the practical usefulness of optimal designs in real applications. For example, Rha et al. (2020) discussed a study of Alzheimer's disease (AD), where the optimal design provides a guidance on scheduling the patient's visits for taking a clinical rating, called the AD Assessment Scale-Cognitive Subscale (ADAS-Cog), to evaluate the cognitive impairment in AD. Park et al. (2018) demonstrated the use of optimal designs in determining the time points for performing abdominal ultrasound scans to give precise predictions of fetal growth trajectories. Ji and Müller (2017) illustrated the advantages

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

of optimal designs through Mediterranean fruit fly application, and the Baltimore longitudinal study of aging. It is noteworthy that, in these previous works, the optimal designs are obtained by taking advantages of prior knowledge (e.g., from pilot studies) on some characteristics of the underlying random function. In the sense of Chernoff (1953), the methods proposed in these previous studies are locally optimal design approaches; see also discussions by, e.g., Ford et al. (1992) and Stufken and Yang (2012) on the theoretical importance and practical relevance of locally optimal designs. For cases where the prior knowledge is vague, Rha et al. (2021) proposed an approach to enhance the quality of designs by borrowing the strengths of bagging (bootstrap aggregating). They demonstrated the usefulness of their proposed method, and applied it in finding optimal designs for functional quadratic regression. In addition, Li and Xiao (2020) considered another design problem on the classification of functional data.

In this work, we study two related design issues for collecting sparse functional data. Note that the designs that we consider here are collections of sampling time points (or locations) for making observations from underlying random functions $X_i(t)$'s. Following the previous works in the same research line, the sampling time points for each $X_i(t)$ are selected from a specified grid on the compact domain of $X_i(t)$. We first discuss optimal designs for the sparse functional principal component analysis (FPCA; e.g., Yao et al., 2005a), and then work on a design issue for a recently proposed alternative method of FPCA, namely the functional empirical component approach (FECA) of Nie et al. (2022). The FPCA is popular, and is a main focus of the previously discussed works. In contrast to these works, our focus here is on designs that render a precise prediction of the FPC scores,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

which is essential for further inferences, including objectives (i) and (ii) mentioned above (see also Yao et al., 2005a,b). We note that our problem formulation for the design of FPC scores allows us to consider the commonly used optimality criteria for evaluating optimal designs. It can then be shown (see Theorem 1) that the A-optimal design for predicting FPC scores gives an optimal design for the previously described objective (i) of Ji and Müller (2017) for recovering the trajectory of $X_i(t)$. Similarly, the compound criterion studied in Park et al. (2018) for objectives (i) and (ii) is essentially the L-optimality criterion for the FPC scores. Thus, the design issues considered in these previous studies can be viewed as special cases of the design problem formulated in Section 2 for predicting FPC scores.

In addition, we study optimal designs for the FECA that is recently proposed by Nie et al. (2022) as an improved alternative to the FPCA. To our knowledge, there is no published optimal design research for this sparse FDA method, and guidance on selecting good designs is currently unavailable. Building upon the design works for FPCA, we propose here a framework for evaluating and selecting designs to give a precise prediction of FEC scores. Our proposed framework not only allows experimenters to evaluate and identify good designs for FECA, but also provides a generalization of a previous work of Prus and Schwabe (2016) on predicting individual random effects under mixed effects models. This generalization, as also indicated by Prus and Schwabe (2016), is practically relevant. Specifically, we obtain designs for situations where the method of Prus and Schwabe (2016) cannot be applied. These situations are very common for sparse FDA. For cases where their method is applicable, our proposed method allows us to identify designs that outperform the single-support designs found by their method, as well as those obtained

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

for the FPCA (e.g., Ji and Müller, 2017). We also provide discussions and comparisons of these competing designs, and the commonly used random designs.

The designs that we consider here are exact designs, in the sense that they are obtained for a prespecified number, n, of subjects. Exact designs can directly be used in practice. However, obtaining these designs for FECA becomes rather involved, due to the enormous space of candidate designs. We adapt here a wide used exchange algorithm for finding optimal designs. To allow a feasible computation, we derive alternative representations of the optimality criterion to give a great reduction in the computational effort. In our numerical study, we are able to combine the exchange algorithm with a step that involves an exhaustive search (Section 3). For problems where exhaustive search is infeasible, we also present a possible alternative by considering the PSS approach of Rha et al. (2020). We further note that our formulation of the optimality criterion also makes it possible to consider combining the exchange algorithm with a second-order cone programming (SOCP) solver (as another alternative for the exhaustive search). A discussion, along with a needed result, for this is provided in Appendix A.

In the next section, we present the sparse FPCA setting, and discuss the optimal issue for predicting FPC scores, along with its link with the design problems in existing works. Section 2 also gives some background knowledge for our work on FECA. We then describe our proposed optimal design approach for predicting FEC scores in Section 3. Section 4 provides a numerical study to give insights into the performance of competing designs, including our obtained designs for FEC scores , randomly generated designs, single-support designs from the method of Prus and Schwabe (2016), and designs for FPC scores. This is

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

followed by a discussion in Section 5. The use of the PSS algorithm, and SOCP solver can be found in Appendix A. Some proofs of our results are provided in Appendix B.

2 Optimal Designs for Functional Principal Components

Let $\mathcal{T}$ be a compact (time) interval such as $\mathcal{T} = [0, 1]$, and $X = \{X(t) : t \in \mathcal{T}\}$ be a mean-square continuous, second-order stochastic process with its realizations in the separable Hilbert space $\mathbb{L}_2(\mathcal{T})$ of square-integrable functions on $\mathcal{T}$. The random function X has continuous mean and covariance functions, which are denoted as $\mu_x(t) = E\{X(t)\}$, and $C_x(t_1, t_2) = Cov\{X(t_1), X(t_2)\}$, respectively. With Mercer theorem and Karhunen-Lóeve expansion, we may write

$$\begin{aligned} C_x(t_1, t_2) &= \sum_{j=1}^{\infty} \lambda_j \psi_j(t_1)\psi_j(t_2), \text{ and} \\ X(t) &= \mu_x(t) + \sum_{j=1}^{\infty} \langle X - \mu_x, \psi_j \rangle \psi_j(t). \end{aligned} \tag{2.1}$$

Here, $\{\psi_j(t) : j = 1, 2, ...\}$ is a set of orthonormal eigenfunctions of (the covariance operator of) X, λ_j is the j^{th} greatest eigenvalue with $\sum_{j=1}^{\infty} \lambda_j < \infty$, $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t)\, dt$ (Lebesgue integration) is the inner product, and $\zeta_j = \langle X - \mu_x, \psi_j \rangle$ gives the j^{th} leading FPC score having $E(\zeta_j) = E(\zeta_j\zeta_k) = 0$ for $j \neq k$, and $E(\zeta_j^2) = \lambda_j$. See Hsing and Eubank (2015) for further mathematical details on such random functions.

In most cases, the entire sample path of X is unavailable, but the functional data, $\{u_{ik} : k = 1, ..., K\}$, of subject i form a noisy subsample of the underlying X_i at time points $\{t_{ik} : k = 1, ..., K\} \subset \mathcal{T}$; $i = 1, ..., n$ and $t_{ik} < t_{i,k+1}$. That is, $u_{ik} = X_i(t_{ik}) + \varepsilon_{ik}$, where ε_{ik}'s are iid noise with mean 0 and variance σ^2, and are independent of X_i. Our focus

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

is on cases where limited resources and practical constraints permit only a relatively small K for each subject. The sampling times, $\boldsymbol{t}_i = \{t_{i1}, ..., t_{iK}\}$, for each X_i may be irregularly spaced over $\mathcal{T}$, and possibly, $\boldsymbol{t}_i \neq \boldsymbol{t}_j$ for some $i \neq j$. Under this sparse functional data setting, we are concerned with the selection of an optimal design, which is a multiset $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_n\}$, for collecting informative functional data u_{ik}'s to render precise statistical inference. For clarity, the set $\boldsymbol{t}_i$ of sampling time points for each subject will be refer to as an elemental design (of the design $\boldsymbol{d}$ for a study of size n).

Under the above setting, Ji and Müller (2017), Park et al. (2018), and Rha et al. (2020, 2021) studied some optimal design issues. With a focus on the FPCA, they considered the following approximation for a given (sufficiently large) J.

$$X_i(t) \approx \mu_x(t) + \sum_{j=1}^{J} \zeta_{ij}\psi_j(t), \tag{2.2}$$

where $\zeta_{ij} = \langle X_i - \mu_x, \psi_j \rangle$ is the j^{th} FPC score of X_i, and the remaining terms are defined as in (2.1). With sparse functional data $\boldsymbol{u}_i = (u_{i1}, ..., u_{iK})^\top$, a prediction of the FPC scores $\boldsymbol{\zeta}_i = (\zeta_{i1}, ..., \zeta_{iJ})^\top$ is (e.g., Yao et al., 2005a):

$$\tilde{\boldsymbol{\zeta}}_i = \boldsymbol{\Lambda}\boldsymbol{\Psi}_i^\top \left(\sigma^2 \boldsymbol{I}_K + \boldsymbol{\Psi}_i \boldsymbol{\Lambda} \boldsymbol{\Psi}_i^\top\right)^{-1} (\boldsymbol{u}_i - \boldsymbol{\mu}_{x,i}) = \left(\sigma^2 \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Psi}_i^\top \boldsymbol{\Psi}_i\right)^{-1} \boldsymbol{\Psi}_i^\top (\boldsymbol{u}_i - \boldsymbol{\mu}_{x,i}),$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_J)$ with $\lambda_j > 0$ for all $j = 1, ..., J$, $\boldsymbol{\Psi}_i = [\psi_1(\boldsymbol{t}_i), ..., \psi_J(\boldsymbol{t}_i)]$, $\psi_j(\boldsymbol{t}_i) = (\psi_j(t_{i1}), ..., \psi_j(t_{iK}))^\top$, $\boldsymbol{I}_K$ is the identity matrix of order K, and $\boldsymbol{\mu}_{x,i} = (\mu_x(t_{i1}), ..., \mu_x(t_{iK}))^\top$. This $\tilde{\boldsymbol{\zeta}}_i$ is used in (2.2) for predicting X_i. The quality of the prediction will depend on $\boldsymbol{t}_i$. One of the design goals of the previous studies is thus on finding optimal $\boldsymbol{t}_i^*$ that minimizes the corresponding mean integrated squared error (MISE); see, e.g., Park et al. (2018).

$$MISE(\boldsymbol{t}_i) = tr\{\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Psi}_i^\top (\sigma^2 \boldsymbol{I}_K + \boldsymbol{\Psi}_i \boldsymbol{\Lambda} \boldsymbol{\Psi}_i^\top)^{-1} \boldsymbol{\Psi}_i \boldsymbol{\Lambda}\}. \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Here, $tr\{\cdot\}$ is the matrix trace. Ji and Müller (2017), and Park et al. (2018) considered a greedy search, and Rha et al. (2020) proposed a probabilistic subset selection (PSS) algorithm for finding optimal $\boldsymbol{t}_i^*$ that minimizes the MISE. It is noteworthy that the MISE depends not only on $\boldsymbol{t}_i$, but also on other quantities, including σ^2, J, and (λ_i, ψ_j) for $j = 1, ..., J$. As suggested by Ji and Müller (2017), the values of these quantities can be estimated from a pilot study with the methods discussed in, e.g., Yao et al. (2005a). Consequently, $\boldsymbol{t}_i^*$ is obtained by minimizing the estimated MISE, and can thus be viewed as a locally optimal design in the sense of Chernoff (1953). A design approach that partly takes the uncertainty of the estimated MISE into can be found in Rha et al. (2021). We note that the approach of Rha et al. (2021) can be extended to our current study. But for clarity, we follow the previous works to consider locally optimal designs, and for convenience, the term 'locally' is omitted in this paper.

The inference on X_i requires a prediction of the FPC scores $\boldsymbol{\zeta}_i$, whose prediction precision will again depend on the selected elemental design $\boldsymbol{t}_i$. Let $\mathbb{I}(\boldsymbol{\zeta}_i; \boldsymbol{t}_i) = \sigma^2 Cov^{-1}(\tilde{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)$ be the (standardized) precision matrix for $\boldsymbol{\zeta}_i$ under the elemental design $\boldsymbol{t}_i$. With design $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_n\}$, the precision matrix for $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\top, ..., \boldsymbol{\zeta}_n^\top)^\top$ is then $\mathbb{I}(\boldsymbol{\zeta}; \boldsymbol{d}) = \oplus_{i=1}^n \mathbb{I}(\boldsymbol{\zeta}_i; \boldsymbol{t}_i)$, where $\oplus$ is the matrix direct sum. To evaluate the goodness of design, we follow Kiefer (1959, 1974) to consider specific scalar function $\Phi(\mathbb{I})$ of the precision matrix $\mathbb{I}$ of interest as the optimality criterion. For example, the A-optimality criterion is $\Phi_A(\mathbb{I}) = tr(\mathbb{I}^{-1})$, and the L-criterion has $\Phi_L(\mathbb{I}) = tr(\boldsymbol{L}\mathbb{I}^{-1})$ with given constant matrix $\boldsymbol{L}$. One may also consider the D-criterion to aim at designs minimizing $det(\mathbb{I}^{-1})$. We note that $\mathbb{I}^{-1}$ might not exists for some designs. These designs are considered the worst, and their criterion values are set

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

to $\Phi = +\infty$.

With the above discussions, we immediately observe the following result, which can be derived with similar arguments as in Sec. 5.3 of Fedorov and Hackl (1997). The proof is omitted.

Theorem 1. *The A-optimal design for predicting the FPC scores* $\boldsymbol{\zeta}$ *of* n *subjects can be formed by* $\boldsymbol{d}^* = \{\boldsymbol{t}_1^*, ..., \boldsymbol{t}_n^*\}$, *where* $\boldsymbol{t}_i^*$ *is an elemental design minimizing* $\Phi_A\{\mathbb{I}(\boldsymbol{\zeta}_i; \boldsymbol{t}_i)\} = tr\{(\sigma^2\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Psi}_i^\top\boldsymbol{\Psi}_i)^{-1}\}$, *or equivalently, minimizing* $MISE(\boldsymbol{t}_i)$ *of* (2.3).

Theorem 1 suggests that finding the optimal $\boldsymbol{t}_i^*$ for predicting X_i is equivalent to finding A-optimal elemental design for $\boldsymbol{\zeta}_i$. By applying the same $\boldsymbol{t}_i^*$ to all the n subject, we have an A-optimal design for predicting the n sets of FPC scores $\boldsymbol{\zeta}$. It can also be seen that with specific constant matrix $\boldsymbol{L}$, the L-optimal design for $\boldsymbol{\zeta}_i$ will minimize the mean square prediction error (MSPE) for predicting the response in a functional regression studied in, e.g., Rha et al. (2020). The compound criterion of MISE and MSPE considered in the previously mentioned works also corresponds to an L-optimality criterion for $\boldsymbol{\zeta}_i$. As demonstrated in Rha et al. (2020), the PSS algorithm gives an efficient tool for finding optimal elemental designs for these study objectives. Another possible approach is by recasting this optimization problem to a second-order cone programming (SOCP) one. With the Φ_A in Theorem 1, this can be easily done by the method of Harman and Prus (2018). The application of PSS and SOCP approaches to the present work is discussed in Appendix A.

From Theorem 1, it also can be seen that there are m^n optimal designs for $\boldsymbol{\zeta}$ of n subjects for cases having m distinct optimal designs for each $\boldsymbol{\zeta}_i$. The previous studies mainly focused

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

on designs having only one elemental designs; i.e., $\boldsymbol{d}^* = \{\boldsymbol{t}_i^*, ...\boldsymbol{t}_i^*\}$. This is partly because that the proposed algorithms do not aim at finding multiple $\boldsymbol{t}_i^*$'s, and practically, we only need one $\boldsymbol{t}_i^*$ for achieving the study objectives discussed in this section. Hereinafter, a design formed by a single elemental design will be referred to as a single-support design. It turns out that these single-support designs can be suboptimal for predicting FEC scores to be discussed in the next section. In some cases commonly encountered in sparse FDA, these designs do not even allow the prediction of FEC scores, and thus have zero statistical efficiency.

3 Optimal Designs for Functional Empirical Components

As an improved alternative to the FPCA, Nie et al. (2022) proposed the use of FECA by considering the following approximation.

$$X_i(t) \approx \sum_{j=1}^{J} \alpha_{ij}\eta_j(t), \tag{3.1}$$

where $\eta_j(t)$ is the eigenfunction of the nonnegative-definite, self-adjoint, Hilbert–Schmidt integral operator $\mathcal{H}_x$ having $[\mathcal{H}_x\varphi](s) = \int_{\mathcal{T}} H_x(t,s)\varphi(t)\,dt$, $H_x(t_1,t_2) = E\{X(t_1)X(t_2)\}$ is the (uncentered) second-moment function of X, $\langle\eta_j,\eta_\ell\rangle = \delta_{j\ell}$, $\delta_{j\ell}$ is the Kronecker delta, and $\alpha_{ij} = \langle X_i,\eta_j\rangle$ is referred to as the FEC score. We also write $\theta_j = E(\alpha_{ij}) = \langle\mu_x,\eta_j\rangle$, and $\tau_j = E(\alpha_{ij}^2) > 0$ for $j = 1, ..., J$, and note that the integer J of (3.1) might not equal that of (2.2). With (3.1), we consider the following model for the functional data u_{ik}.

$$u_{ik} = \sum_{j=1}^{J} \alpha_{ij}\eta_j(t_{ik}) + \varepsilon_{ik};\ i = 1, ..., n,\ k = 1, ..., K. \tag{3.2}$$

Following the discussion in Section 2, our focus here is on optimal designs $\boldsymbol{d}^*$ for precise prediction of the FEC scores $\boldsymbol{\alpha}_i = (\alpha_{i1}, ..., \alpha_{iJ})^\top$; $i = 1, ..., n$. With specified η_j (e.g.,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

from pilot study), (3.2) presents a mixed-effects model, and some results on designs for predicting individual random effects can be found in Prus and Schwabe (2016). However, the application of their results requires assumptions that may not hold in the sparse FDA setting. In particular, Prus and Schwabe (2016) allowed a sampling time point t_{ij} to appear more than once in an elemental design $\boldsymbol{t}_i$. But in almost all realistic FDA settings, we do not make more than one observation from $X_i(t)$ at a given time point t. Moreover, the design $\boldsymbol{d}$ considered by Prus and Schwabe (2016) is formed by the same elemental design $\boldsymbol{t}$; i.e., it is a single-support design. A necessary condition for these designs to be useful for predicting FEC scores is $K \geq J$. However, cases with $K < J$ are also very common. In addition, even with $K \geq J$, we may still require more than one elemental design for generating an optimal $\boldsymbol{d}$, as demonstrated in our numerical study in Section 4. New developments are thus called for.

Here, we develop new results on optimal designs for predicting FEC scores by relaxing the constraint on K and J. To derive the optimality criterion for predicting the FEC scores $\boldsymbol{\alpha}$, we consider the following working model for (3.2):

$$\boldsymbol{u} = \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon};\ Cov\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{\Delta} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2 \boldsymbol{I}_{nK} \end{pmatrix}, \tag{3.3}$$

where $\boldsymbol{u} = (\boldsymbol{u}_1^\top, ..., \boldsymbol{u}_n^\top)^\top$ represents the functional data of the n subjects, $\boldsymbol{\theta} = (\theta_1, ..., \theta_J)^\top$, $\boldsymbol{F} = [\boldsymbol{F}_1^\top, ..., \boldsymbol{F}_n^\top]^\top$, $\boldsymbol{F}_i = [\eta_1(\boldsymbol{t}_i), ..., \eta_J(\boldsymbol{t}_i)]$, $\eta_j(\boldsymbol{t}_i) = (\eta_j(t_{i1}), ..., \eta_j(t_{iK}))^\top$, $\boldsymbol{\gamma} = \boldsymbol{\alpha} - \boldsymbol{j}_n \otimes \boldsymbol{\theta}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, ..., \boldsymbol{\alpha}_n^\top)^\top$ is the vector of the n sets of FEC scores, $\boldsymbol{j}_n$ is the vector of n ones, $\otimes$ is the Kronecker product, $\boldsymbol{R} = \oplus_{i=1}^n \boldsymbol{F}_i$, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, ..., \varepsilon_{nK})^\top$ is noise, $\boldsymbol{\Delta} = diag(\tau_1, ..., \tau_J)$, and $\boldsymbol{O}$ represents a zero matrix of appropriate dimensions. In line with Prus and Schwabe

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

(2016) and Nie et al. (2022), we do not impose Gaussian assumption on $\boldsymbol{\alpha}$, and in contrast to the FPCA, prior knowledge of $\mu_x(t)$ is not required; see also Nie et al. (2022) for a discussion on this. For predicting $\boldsymbol{\alpha}$, we consider the mixed model equations for (3.3) that can be reduced to the following with working $\boldsymbol{W}_i = \sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top\boldsymbol{F}_i$ (cf. Appendix A of Prus and Schwabe, 2016, and references therein).

$$\begin{pmatrix} \boldsymbol{F}^\top\boldsymbol{F} & \boldsymbol{F}^\top\boldsymbol{R} \\ \boldsymbol{R}^\top\boldsymbol{F} & \oplus_{i=1}^n\boldsymbol{W}_i \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ \tilde{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{F}^\top\boldsymbol{u} \\ \boldsymbol{R}^\top\boldsymbol{u} \end{pmatrix}. \tag{3.4}$$

The unique solution for $(\tilde{\boldsymbol{\theta}}^\top, \tilde{\boldsymbol{\gamma}}^\top)^\top$ of (3.4) is presented in Theorem 2 below. Assumptions for this and subsequent results are as follows. First, $\boldsymbol{\Delta}$ is nonsingular, and thus, $\boldsymbol{W}_i = \sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top\boldsymbol{F}_i$ is positive definite. This is not a strong assumption, and holds if $\tau_j > 0$ for $j = 1, ..., J$. Extending our derivations in Appendix B to cases with a singular $\boldsymbol{\Delta}$ is possible, but is beyond the scope of our study. In addition, the predictability of $\boldsymbol{\alpha}$ will hinge on the estimability of its mean $\boldsymbol{\theta}$. Thus, we will focus only on designs $\boldsymbol{d}$ allowing an estimable $\boldsymbol{\theta}$ (and hence, a predictable $\boldsymbol{\alpha}$). As suggested by Theorem 2, this means that $\boldsymbol{d}$ yields a full-column-rank $\boldsymbol{F}$ (i.e., $rank(\boldsymbol{F}) = J$). The collection of such designs will be denoted as Ξ_d hereinafter. We note that any $\boldsymbol{d} \in \Xi_d$ has at least J distinct t_{ij}'s (even when $K < J$). Moreover, for typical FDA settings, the K sampling time points $t_{i1}, ..., t_{iK}$ for an $X_i(t)$ are all distinct. We thus require each t_{ij} to appear no more than once in a $\boldsymbol{t}_i$. We now present Theorem 2.

Theorem 2. *Eq.* (3.4) *has the following unique solution if and only if* $\boldsymbol{F}$ *has full column*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

rank.

$$\begin{aligned}\tilde{\boldsymbol{\gamma}} &= (\tilde{\boldsymbol{\gamma}}_1^\top, ..., \tilde{\boldsymbol{\gamma}}_n^\top)^\top;\ \tilde{\boldsymbol{\gamma}}_i = \boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top(\boldsymbol{u}_i - \boldsymbol{F}_i\tilde{\boldsymbol{\theta}});\\ \tilde{\boldsymbol{\theta}} &= (\sum_{j=1}^{n}\boldsymbol{W}_j^{-1}\boldsymbol{F}_j^\top\boldsymbol{F}_j)^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{u}_i. \qquad (3.5)\end{aligned}$$

It is noteworthy that $\tilde{\boldsymbol{\gamma}}$ is the best linear unbiased prediction for $\boldsymbol{\gamma}$, and $\tilde{\boldsymbol{\theta}}$ gives the best linear unbiased estimate satisfying $\boldsymbol{F}^\top\boldsymbol{\Sigma}_u^{-1}\boldsymbol{F}\tilde{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}$, where $\boldsymbol{\Sigma}_u = \oplus_{i=1}^{n}(\sigma^2\boldsymbol{I}_K + \boldsymbol{F}_i\boldsymbol{\Delta}\boldsymbol{F}_i^\top)$ is the covariance matrix of $\boldsymbol{u}$; see also, Henderson et al. (1959), and Theorem 12.3.1 of Christensen (2002). We also note that if $\boldsymbol{F}_i = \boldsymbol{F}_1$ has a full column rank for all i, the solution is reduced to the one presented in Prus and Schwabe (2016). But here, we also consider cases where $\boldsymbol{F}_i \neq \boldsymbol{F}_j$ for some $i \neq j$, and allow $rank(\boldsymbol{F}_i) < J$. In other words, we do not restrict ourselves to single-support designs, and allow different subjects to have different elemental designs $\boldsymbol{t}_i$. With this relaxation of assumptions, our results provide a generalization of Prus and Schwabe (2016).

Theorem 2 also gives a prediction of the FEC scores $\boldsymbol{\alpha}$, which in turn allows the formulation of an optimality criterion for evaluating competing designs. With $\boldsymbol{\alpha}_i = \boldsymbol{\gamma}_i + \boldsymbol{\theta}$, a prediction of $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, ..., \boldsymbol{\alpha}_n^\top)^\top$ is

$$\tilde{\boldsymbol{\alpha}} = \boldsymbol{G}\begin{pmatrix}\tilde{\boldsymbol{\theta}}\\ \tilde{\boldsymbol{\gamma}}\end{pmatrix},\ \text{where } \boldsymbol{G} = [\boldsymbol{j}_n, \boldsymbol{I}_n] \otimes \boldsymbol{I}_J. \qquad (3.6)$$

Some discussions on the designs for the fixed effects $\boldsymbol{\theta}$ can be found in Sec. 5.3 of Fedorov and Hackl (1997) and Prus (2022). But, the estimation precision of $\tilde{\boldsymbol{\theta}}$ is not the main concern here. We instead search over the previously defined Ξ_d for an optimal design yielding the highest prediction precision for $\boldsymbol{\alpha}$. To this end, we again consider specific real-function $\Phi\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}$ of the precision matrix $\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})$ as the optimality criterion

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

for evaluating competing designs. To derive such an optimality criterion, we present below $Cov(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \sigma^2 \mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d})$, i.e., the MSE of $\tilde{\boldsymbol{\alpha}}$; this also gives a generalization of Theorem 2 of Prus and Schwabe (2016).

Theorem 3. *Consider a design $\boldsymbol{d} \in \Xi_d$ for model (3.3), and the prediction $\tilde{\boldsymbol{\alpha}}$ in (3.6). The covariance matrix of $(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ can be written as a partitioned matrix $Cov(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = ((\sigma^2 \boldsymbol{C}_{ij}))_{i,j=1,\ldots,n}$, where $\boldsymbol{C}_{ij}$ is an $J \times J$ matrix with*

$$\boldsymbol{C}_{ij} = \delta_{ij} \boldsymbol{W}_i^{-1} + (\boldsymbol{W}_j \boldsymbol{A} \boldsymbol{W}_i)^{-1}, \ \boldsymbol{A} = n\sigma^{-2}\boldsymbol{\Delta} - \sum_{i=1}^{n} \boldsymbol{W}_i^{-1}, \tag{3.7}$$

δ_{ij} is the Kronecker delta, and $\boldsymbol{W}_i$ is as in (3.4). $\boldsymbol{A}^{-1}$ exists and is positive definite if and only if $\boldsymbol{F}$ has full column rank.

With Theorem 3, the A-optimality criterion (Section 2) for predicting the FEC scores $\boldsymbol{\alpha}$ can be written as the Φ_A below. Similarly, the L-optimality criterion can be easily derived. We note that the unknown quantities involved in $\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}$ can be estimated from, e.g., a pilot study with the method of Nie et al. (2022).

Corollary 1. *The A-optimality criterion for evaluating a design $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_n\} \in \Xi_d$ in predicting the FEC scores $\boldsymbol{\alpha}$ can be written as*

$$\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\} = \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-1}) + \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-2} \boldsymbol{A}^{-1}). \tag{3.8}$$

For convenience, we will refer to the first term of Φ_A as $\Phi_{A1} = \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-1})$, and set $\Phi_{A2} = \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-2} \boldsymbol{A}^{-1})$. We note that with unknown $\boldsymbol{\theta}$, the precision matrix $\mathbb{I}(\boldsymbol{\alpha})$ for the FEC scores is no longer a block-diagonal matrix, and has a more complex structure than its counterpart $\mathbb{I}(\boldsymbol{\zeta})$ in Section 2. We also note that the A-criterion for $\boldsymbol{\alpha}$ is not

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

reduced to that of $\boldsymbol{\zeta}$, even when $\mu_x = 0$ (although in this case, $\boldsymbol{\Delta} = \boldsymbol{\Lambda}$, and $\boldsymbol{F}_i = \boldsymbol{\Psi}_i$). Specifically, $\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}$ includes an additional term $\sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-2}\boldsymbol{A}^{-1})$, as can be seen by comparing Corollary 1 and Theorem 1. But with $\mu_x = 0$, $\Phi_A\{\mathbb{I}(\boldsymbol{\zeta}; \boldsymbol{d})\}$ gives a lower bound of $\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}$. This fact leads to a lower bound for the A-efficiency as presented later in Theorem 5, where a sharper lower bound is also provided.

Finding an A-optimal design for $\boldsymbol{\alpha}$ will require a search of the best combination of n elemental designs $\boldsymbol{t}_i$. But in contrast to Theorem 1, each of these $\boldsymbol{t}_i$'s is not necessarily optimal for studying the individual $\boldsymbol{\alpha}_i$. When used alone, $\boldsymbol{t}_i$ might not even allow the prediction of $\boldsymbol{\alpha}_i$, especially when $K < J$. Another noteworthy feature of the current design issue is that the dimension of $\mathbb{I}(\boldsymbol{\alpha})$ grows with n, and the Φ_A-value, or in a sense, the amount of information, is not additive by replicating a design for a larger n. Corollary 2 provides some insights into this, and suggests that Φ_{A1} is increasingly important than Φ_{A2} with an increased number of replicates. The same argument gives Corollary 3 for single-support designs, which are in Ξ_d only if $K \geq J$. This result can also be derived from Corollary 2 of Prus and Schwabe (2016) by applying, e.g., the Woodbury matrix identity or Theorem 18.2.8 of Harville (1997).

Corollary 2. *As in Corollary 1, let* $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_n\} \in \Xi_d$ *be a design for* $\boldsymbol{\alpha}$*, and* $\boldsymbol{d}_{rn}$ *be the* $r(> 1)$ *copies of* $\boldsymbol{d}$ *for the FEC scores of* rn *subjects. The A-optimality criterion of* $\boldsymbol{d}_{rn}$ *has:*

$$\Phi_A(\boldsymbol{d}_{rn}) = r\sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-1}) + \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-2}\boldsymbol{A}^{-1}).$$

Corollary 3. *Suppose* $K \geq J$*, and* $\boldsymbol{d} \in \Xi_d$ *is a single-support design formed by the* n *copies of an elemental design* $\boldsymbol{t}_1$*; i.e.,* $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_1\}$*. The A-optimality criterion of* $\boldsymbol{d}$ *on*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

predicting $\boldsymbol{\alpha}$ *is reduced to the following.*

$$\Phi_A(\boldsymbol{d}) = ntr(\boldsymbol{W}_1^{-1}) + tr\{\boldsymbol{W}_1^{-2}(\sigma^{-2}\boldsymbol{\Delta} - \boldsymbol{W}_1^{-1})^{-1}\}.$$

Our aim is at an exact optimal design $\boldsymbol{d}$ for n future subjects. To our knowledge, no approach is currently available for finding such designs under the FECA setting. A naive approach is perhaps by comparing candidate designs with the trace of inverse precision matrix of $(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ obtained directly from (3.4), and (3.6). For each design, one then needs to calculate the inverse of an $(n+1)J$-by-$(n+1)J$ matrix, which requires much computational effort and computer memory. With Theorem 3, and the results below (Eq. (3.9) and Theorem 4), we are allowed to consider an approach that involves inverting matrices of reduced dimensions (J-by-J). To obtain an optimal design, we utilize exchange-type algorithms which are widely considered in tackling various optimal design problems (see also Atkinson et al., 2007, Ch. 12). The algorithm that we consider begins with an initial design $\boldsymbol{d}^{(0)} = \{\boldsymbol{t}_1^{(0)}, ..., \boldsymbol{t}_n^{(0)}\}$. At the t^{th} iteration of the algorithm, a substitution of $\boldsymbol{t}_j^{(t-1)}$ of the current design $\boldsymbol{d}^{(t-1)}$ with another elemental design $\boldsymbol{t}_p$ is proposed. The proposed substitution is accepted to give the next design $\boldsymbol{d}^{(t)}$ when it reduces the Φ_A-value; otherwise, the design remains intact, and $\boldsymbol{d}^{(t)} = \boldsymbol{d}^{(t-1)}$. This procedure is repeated until no further improvement is expected. To identify the pair $(\boldsymbol{t}_j^{(t-1)}, \boldsymbol{t}_p)$ at each iteration, we derive (3.9) below, which gives an updating formula for the Φ_A-value after the exchange. Without loss of generality, (3.9) is presented with $j = n$. We also use $\boldsymbol{d}_p$ to denote the resulting design after the substitution. We require both designs to allow an estimable $\boldsymbol{\theta}$ (i.e., $\boldsymbol{d}, \boldsymbol{d}_p \in \Xi_d$). We now provide the updating formula for the Φ_A-criterion when moving

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

from $\boldsymbol{d}$ to $\boldsymbol{d}_p$.

$$\begin{aligned}\Phi_A\{\mathbb{I}(\boldsymbol{\alpha};\boldsymbol{d}_p)\} &= \Phi_A\{\mathbb{I}(\boldsymbol{\alpha};\boldsymbol{d})\} + \Phi_-(\boldsymbol{d}) + \Phi_+(\boldsymbol{d}_p); \\ \Phi_-(\boldsymbol{d}) &= -tr(\boldsymbol{W}_n^{-1} + \sum_{i=1}^{n} \boldsymbol{W}_i^{-2}\boldsymbol{A}^{-1}); \text{ and} \\ \Phi_+(\boldsymbol{d}_p) &= tr(\boldsymbol{W}_p^{-1} + \sum_{i=1}^{n-1} \boldsymbol{W}_i^{-2}\boldsymbol{A}_p^{-1} + \boldsymbol{W}_p^{-2}\boldsymbol{A}_p^{-1}).\end{aligned} \tag{3.9}$$

Here, $\boldsymbol{W}_p$ is as $\boldsymbol{W}_i$ defined under (3.4), but $\boldsymbol{W}_p$ is obtained with $\boldsymbol{t}_p$. Similarly, $\boldsymbol{A}_p$ is from $\boldsymbol{d}_p$, but otherwise is the same as $\boldsymbol{A}$ (for $\boldsymbol{d}$) in (3.7). Ideally, we would like to simultaneously identify $\boldsymbol{t}_j^{(t-1)}$ and $\boldsymbol{t}_p$ that give a large reduction in the Φ_A-value. To have a feasible computing time, we instead separate this into two steps, namely the deletion of a $\boldsymbol{t}_j^{(t-1)}$, followed by the addition of a $\boldsymbol{t}_p$. Each of the two steps will change the dimensions of the precision matrix. However, deriving an updating formula for Φ_A by taking this change into account does not seem advantageous. We thus implement the two steps by fixing the order of the precision matrix to nJ. With (3.9), we first identify a $\boldsymbol{t}_j \in \boldsymbol{d}$ that has the smallest Φ_-, and then replace it with a $\boldsymbol{t}_p$ that minimizes Φ_+. Note that the last term of Φ_- is constant across all elemental designs in $\boldsymbol{d}$, and can thus be omitted when selecting $\boldsymbol{t}_j$. To facilitate the calculation of Φ_+ for identifying the best $\boldsymbol{t}_p$, we provide below an alternative expression.

Theorem 4. *With the same notation as in* (3.9), *let* $\boldsymbol{A}_{n-1} = n\sigma^{-2}\boldsymbol{\Delta} - \sum_{i=1}^{n-1} \boldsymbol{W}_i^{-1}$. *Suppose that* $\boldsymbol{d}, \boldsymbol{d}_p \in \Xi_d$. *Then,*

$$\Phi_+(\boldsymbol{d}_p) = tr\{(\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1})^{-1}(\boldsymbol{I}_J + \boldsymbol{A}_{n-1}^{-1}\sum_{i=1}^{n-1} \boldsymbol{W}_i^{-2}\boldsymbol{A}_{n-1}^{-1})\} + tr(\boldsymbol{A}_{n-1}^{-1}\sum_{i=1}^{n-1} \boldsymbol{W}_i^{-2}).$$

We note that $\boldsymbol{A}_p^{-1}$ in (3.9) is a function of $\boldsymbol{t}_p$, and needs to be calculated for every candidate design in every iteration of the algorithm. By contrast, $\boldsymbol{A}_{n-1}^{-1}$ in Theorem 4

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

depends only on the $(n-1)$ elemental designs of $\boldsymbol{d}$ that remain intact in the same iteration, and is free of $\boldsymbol{t}_p$. It only needs to be calculated once per iteration. The same is true for $\sum_{i=1}^{n-1} \boldsymbol{W}_i^{-2}$. Consequently, the only term in $\Phi_+(\boldsymbol{d}_p)$ of Theorem 4 that needs to be calculated for each candidate design is $(\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1})^{-1}$, which can also be used to check the validity of $\boldsymbol{t}_p$ as suggested in the proof of Theorem 4 in Appendix B. Moreover, the last term of that $\Phi_+(\boldsymbol{d}_p)$ can be ignored in the search of the best $\boldsymbol{t}_p$. In our numerical study, we find the best $\boldsymbol{t}_p$ in each iteration with an exhaustive search over all candidate elemental designs. For such an exchange algorithm, it often is recommended to implement it multiple times with different initials. We follow this recommendation in our numerical study with parallel computing, but further 'refine' the obtained design (i.e., the best one among the multiple runs) by substituting each of its elemental design with the one minimizing the Φ_+ in Theorem 4. We note that when the candidate set grows, the exhaustive search can be infeasible, and we propose to adapt the PSS algorithm of Rha et al. (2020) as described in Appendix A. Systematic comparisons of the algorithms will be reported elsewhere (Kao and Huang, 2023), but combining the exchange algorithm with the PSS algorithm tends to give an efficient tool for finding optimal designs. We note that Theorem 4 also allows us to recast the search of optimal $\boldsymbol{t}_p$ as an SOCP problem. This can be done by applying the method of Harman and Prus (2018). A discussion of this, along with a needed result, can be found in Appendix A.

With the above discussions, we now provide lower bounds for the design efficiency of $\boldsymbol{d} \in \Xi_d$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Theorem 5. *For a design* $\boldsymbol{d}$*, define the relative A-efficiency as*

$$RE_A(\boldsymbol{d}) = \frac{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d}^*)\}}{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}},$$

where $\boldsymbol{d}^*$ *is the A-optimal design for predicting* $\boldsymbol{\alpha}$*. Suppose that* $\tilde{\boldsymbol{d}} \in \Xi_d$ *minimizes* $\widetilde{\Phi}(\boldsymbol{d}) = \sum_{i=1}^n tr(\widetilde{\boldsymbol{W}}_i^{-1})$ *with* $\widetilde{\boldsymbol{W}}_i = \frac{n-1}{n}\sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top \boldsymbol{F}_i$*. We have the following lower bounds for* $RE_A(\boldsymbol{d})$*.*

$$RE_A(\boldsymbol{d}) = \frac{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d}^*)\}}{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}} > \frac{\widetilde{\Phi}(\boldsymbol{d}^*)}{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}} \geq \frac{\widetilde{\Phi}(\tilde{\boldsymbol{d}})}{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}} > \frac{n \min_{t_i} tr(\boldsymbol{W}_i^{-1})}{\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}}.$$

With FPCA, Park et al. (2018) showed that the prediction precision for $\boldsymbol{\zeta}$ is no worse after including an additional sampling time point to the current design. A similar result is provided in Theorem 6 for the FECA. The same result holds for any smaller-the-better criterion Φ that is nonincreasing in the Löwner ordering; i.e., $\Phi(\mathbb{I}_1) \leq \Phi(\mathbb{I}_2)$ if $\mathbb{I}_1 \geq_L \mathbb{I}_2$, where $\geq_L$ represents the Löwner ordering. Most commonly used criteria possess this property (Pukelsheim, 1993).

Theorem 6. *Let* $\boldsymbol{d} = \{\boldsymbol{t}_1, ..., \boldsymbol{t}_n\} \in \Xi_d$*, and* $\boldsymbol{d}_a$ *be obtained by augmenting a sampling time point to* $\boldsymbol{d}$*. We have* $\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d}_a) \geq_L \mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})$*, and* $\Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d}_a)\} \leq \Phi_A\{\mathbb{I}(\boldsymbol{\alpha}; \boldsymbol{d})\}$*.*

4 A Numerical Study

To provide some additional insights, we conduct a numerical study to obtain and study designs for FPC and FEC scores, with a focus on the latter. For the eigenfunctions η_j, we consider Fourier bases of the following form (see also Park et al., 2018): $\eta_j(t) = \sqrt{2}\sin\{(j+1)\pi t\}$ for odd j, and $\eta_j(t) = \sqrt{2}\cos\{(j+1)\pi t\}$ for even j, with $t \in \mathcal{T} = [0, 1]$. The corresponding eigenvalue has $\tau_j = 10/2^j$, and the number of effective FECs is $J = 3, 5$, or

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

7. The number of observations per subject is $K \in \{3, 5, 7\}$. We also assume that we are designing for studies having $n = 10, 50$ or 70 subjects, and set $\sigma^2 = 1$. The numerical study thus has $3^3 = 27$ full-factorial settings. The sampling time points in each elemental design $\boldsymbol{t}_i$ are selected from a regular grid on $[0, 1]$ with 21 grid points. We thus have s^n candidate designs for $\boldsymbol{d}$ (or $(n + s - 1)!/\{n!(s - 1)!\}$ designs if the search algorithm can identify designs that are equivalent under subject permutations); here, $s = 21!/\{K!(21 - K)!\}$. We obtain each design by implementing the previously mentioned exchange algorithm 1,000 times with different random initials. This is followed by the refinement of the obtained design, as described in Section 3. The numerical study is conducted on a 3.7GHz Intel Core i7-8700k 6-core processor with 32GB RAM; there are two hyperthreads per physical core. The R codes, which are available upon request to the first author, are written to allow a parallel computing by assigning ten hyperthreads for executing the exchange algorithm with multiple random initials.

With no guidance on design selection, random designs are commonly adopted in practice. In Figure 1, we present the box plot of the relative A-efficiency of 1,000 randomly generated designs to our obtained design in each of the 27 scenarios. There, Scenarios $3q+1$, $3q+2$, and $3q+3$ have the same (J, K), but are for $n = 10, 50$ and 70, respectively, and here, $q = \{3(J-3)+(K-3)\}/2$ for $J, K = 3, 5, 7$. These randomly generated designs do not perform well and can have at least 17% to 72% efficiency loss in predicting the FEC scores. Our designs, which are referred to as the FECA designs, and are presented in Table S1 in the supplementary document, significantly outperform randomly generated designs.

For $K \geq J$, we also obtain optimal single-support designs $\boldsymbol{d}_s = \{\boldsymbol{t}_s, ..., \boldsymbol{t}_s\} \in \Xi_d$ by

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

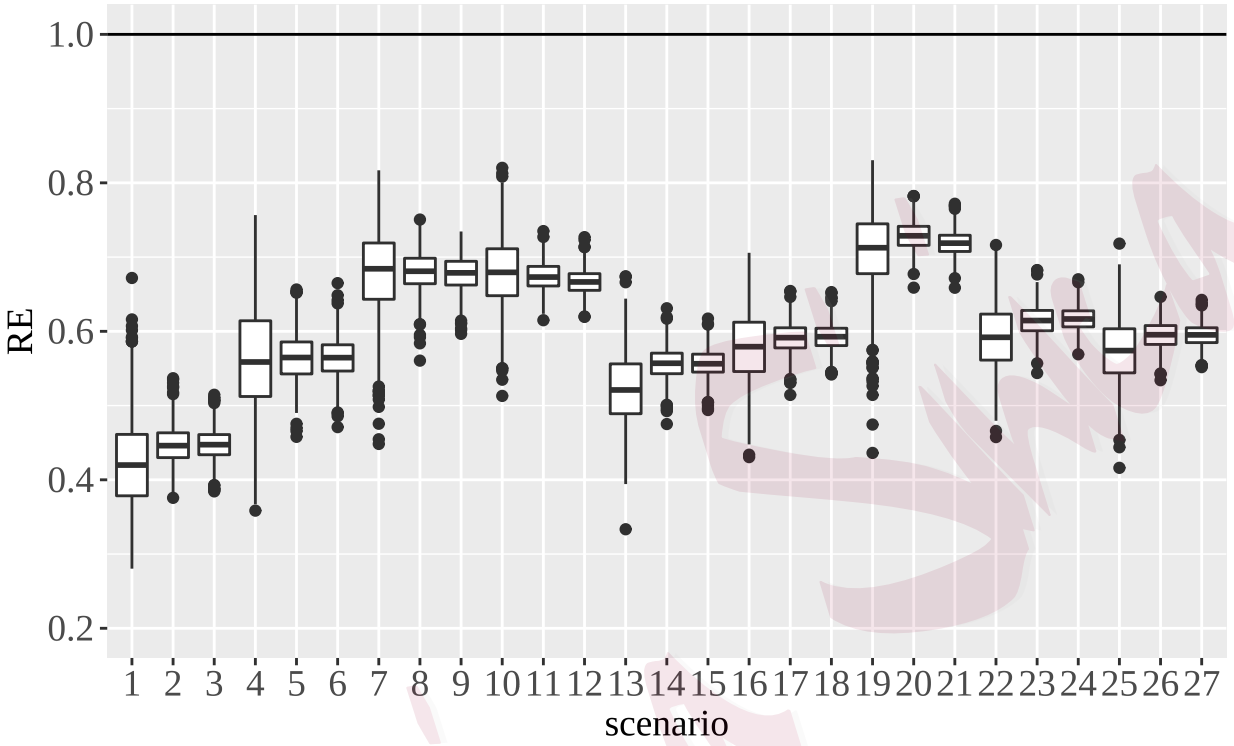

Figure 1: Relative efficiency (%) of 1,000 random designs to the obtained optimal FECA designs in predicting FEC scores.

minimizing Φ_A of Corollary 3 with an exhaustive search over all the $21!/\{K!(21-K)!\}$ elemental designs. In addition, we use the same method to find single-support design $\boldsymbol{d}_{fpc}$ that minimize $\Phi_{A1} = \sum_{i=1}^{n} tr(\boldsymbol{W}_i^{-1}) = ntr(\boldsymbol{W}_1^{-1})$. When $\mu_x = 0$, these $\boldsymbol{d}_{fpc}$ are A-optimal for predicting the FPC scores. There exist multiple $\boldsymbol{d}_s$ and $\boldsymbol{d}_{fpc}$ (see Table S2 in the supplementary document), and these two designs agree in all the cases that we consider, except for $K = J = 7$. The elemental designs of $\boldsymbol{d}_{fpc}$ and $\boldsymbol{d}_s$ for $K = J = 7$ are provided in Table 1, where we present the indices of the $K = 7$ time points from the 21-point regular grid on $\mathcal{T} = [0, 1]$. The optimality of $\boldsymbol{d}_{fpc}$ does not depend on n as discussed in Section 2. Nevertheless, $\boldsymbol{d}_s$ can be different for different n, as a result of Corollary 2. We provide

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Table 1: Elemental designs for single-support designs with $K = J = 7$

$\boldsymbol{d}_{fpc}$	$\boldsymbol{d}_s$ for n=10	$\boldsymbol{d}_s$ for n=50 & 70
	$(2, 4, 7, 9, 12, 16, 18)$	$(2, 4, 7, 10, 13, 16, 18)$
$(2, 5, 7, 10, 13, 16, 19)$	$(2, 6, 8, 12, 14, 17, 19)$	$(2, 5, 8, 10, 14, 16, 19)$
$(3, 6, 9, 12, 15, 17, 20)$	$(3, 5, 8, 10, 14, 16, 20)$	$(3, 6, 8, 12, 14, 17, 20)$
	$(4, 6, 10, 13, 15, 18, 20)$	$(4, 6, 9, 12, 15, 18, 20)$

below further discussions on the single-support designs, and their comparisons with our obtained A-optimal designs $\boldsymbol{d}^*$ for predicting FEC scores.

A further analysis of the above single-support designs and our obtained optimal $\boldsymbol{d}^*$ reveals that for most cases with $K \geq J$, each elemental design of $\boldsymbol{d}^*$ minimizes Φ_{A1} defined below Corollary 1. Following Theorem 1, such $\boldsymbol{d}^*$ also give an optimal design for FPC scores (if $\mu_x = 0$). But combining these elemental designs helps to further reduce Φ_{A2}, and gives the obtained $\boldsymbol{d}^*$. However, Φ_{A2} is elevated when each of these elemental designs is used alone in generating a single-support design. This can be seen from the relative Φ_A-value of $\boldsymbol{d}_s = \boldsymbol{d}_{fpc}$ to $\boldsymbol{d}^*$ for $(J, K) \neq (7, 7)$, as in Table 2. In many cases with $K \geq J$, $\boldsymbol{d}_s$ is highly efficient, but remains suboptimal. For $J = K = 7$, the elemental designs of the obtained $\boldsymbol{d}^*$ are those of the two $\boldsymbol{d}_{fpc}$'s in Table 1. While both elemental designs minimize Φ_{A1}, none of them gives a full-column-rank $\boldsymbol{F}_i$. The resulting single-support design $\boldsymbol{d}_{fpc}$ thus have zero efficiency in predicting the FEC scores. It also is noteworthy that the elemental design for each $\boldsymbol{d}_s$ in Table 1 has a higher Φ_{A1} than that of $\boldsymbol{d}_{fpc}$. As reported in the last column of Table 2, these $\boldsymbol{d}_s$ are suboptimal with $< 93\%$ design efficiency for predicting

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Table 2: Relative efficiency (%) of single-support designs $\boldsymbol{d}_s \in \Xi_d$ to the obtained optimal FECA designs in predicting FEC scores with $K \geq J$. Note. Single-support designs have 0 efficiency for $K < J$.

	J=3			J=5		J=7
n	K=3	K=5	K=7	K=5	K=7	K=7
10	99.9628	99.9987	99.9971	99.8364	99.9955	91.4515
50	99.9924	99.9997	99.9994	99.9666	99.9991	92.5612
70	99.9946	99.9998	99.9996	99.9761	99.9993	92.8818

FEC scores. We note that the design comparisons here suggest the possible suboptimality of imprudently selected designs. But as for most numerical studies, the achieved relative efficiencies reported in Table 2 (and Table 3 below) may vary across experimental settings; see also Kao and Huang (2023). We also note that the single-support designs with $K < J$ are outside of Ξ_d, and will always have 0 efficiency for predicting $\boldsymbol{\alpha}$.

In Table 3, we present the performance of our obtained $\boldsymbol{d}^*$ for FEC scores in predicting FPC scores of n subjects when $\mu_x = 0$. With Φ_A in Theorem 1, the relative efficiency reported there is

$$\frac{\Phi_A\{\mathbb{I}(\boldsymbol{\zeta}; \boldsymbol{d}_{fpc})\}}{\Phi_A\{\mathbb{I}(\boldsymbol{\zeta}; \boldsymbol{d}^*)\}} \times 100\%, \text{ where } \boldsymbol{d}_{fpc} = \{\boldsymbol{t}_{fpc}, ..., \boldsymbol{t}_{fpc}\}$$

As discussed above, our obtained $\boldsymbol{d}^*$ are also A-optimal for FPC scores when $K \geq J$. Unlike the $\boldsymbol{d}_{fpc}$ that has 0 efficiency in predicting FEC scores when $K < J$, $\boldsymbol{d}^*$ attains a relatively

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Table 3: Relative efficiency (%) of $\boldsymbol{d}^*$ to the single-support FPCA designs $\boldsymbol{d}_{fpc}$ in predicting FPC scores when $\mu_x(t) = 0$.

	J=3			J=5			J=7		
n	K=3	K=5	K=7	K=3	K=5	K=7	K=3	K=5	K=7
10	100	100	100	91.1277	100	100	86.1183	98.4277	100
50	100	100	100	95.2612	100	100	95.2581	99.2092	100
70	100	100	100	96.0966	100	100	96.5176	99.2092	100

high efficiency in predicting $\boldsymbol{\zeta}$ for most of these cases. But moving away from $\mu_x = 0$, the eigenpairs used in the two methods can be different, and we recommend finding optimal designs under the corresponding setting with prior knowledge from, e.g., pilot studies.

In addition, we use $\widetilde{\Phi}$ to obtain a lower bound for the relative A-efficiency of the obtained $\boldsymbol{d}^*$ as in Theorem 5. Many of the obtained $\boldsymbol{d}^*$ has $> 99\%$ efficiency (Table 4). For cases where the lower bound is below 95%, additional searches were conducted, but a slightly improved designs was found only for the case with $(J, K, n) = (5, 3, 50)$. This was the design considered in our previous discussions. But the difference in Φ_A is small (with a 0.000294% improvement), which does not change our conclusion. We further note that for some cases, $\boldsymbol{d}^*$ for an n can be obtained by re-weighting (i.e., adjusting the number of replicates of) the elemental designs of the $\boldsymbol{d}^*$ for another n. This seems to be the case for $K \geq J$, but there is an exception for $(J, K) = (7, 5)$. For this latter case, the design for $n = 10$ is slightly different from that for $n = 50$ and 70. As suggested by Corollary 2, having a small Φ_{A2} tends to be more important for smaller n, which provides a partial

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Table 4: A lower bound (%) of relative A-efficiency

	J=3			J=5			J=7		
n	K=3	K=5	K=7	K=3	K=5	K=7	K=3	K=5	K=7
10	99.8212	99.9515	99.9747	77.3935	99.4299	99.7848	66.4967	87.7406	95.6981
50	99.9606	99.9894	99.9945	89.9769	99.8748	99.9533	85.2123	96.5283	99.0570
70	99.9717	99.9924	99.9960	91.5263	99.9100	99.9665	88.0291	97.2737	99.3221

explanation of the different elemental designs needed for $n = 10$ in this case. For this and other scenarios with $K < J$, we also studied the relative efficiency of the designs obtained from re-weighting the designs of a different n. The re-weighted designs that we studied (omitted here) did not perform as well as the $\boldsymbol{d}^*$ obtained from correctly specified n, although these re-weighted designs are expected to be efficient for an n that is nearby. Having some knowledge on the number of future subjects that we would like to study is thus helpful for selecting a good design.

To further demonstrate the usefulness of optimal designs, we compare the empirical MSEs for predicting $X_i(t)$ of these designs with 1,000 random designs. Specifically, for each of the 27 combinations of (J, K, n) of this numerical study, we simulate 100 sets of 'true' $X_1(t), ..., X_n(t)$ from (3.1), where $\eta_j(t)$ is the previously mentioned Fourier basis functions, and as in Nie et al. (2022), α_{ij} are generated from the centered gamma distribution, $Gamma(1, \sqrt{\tau_j}) - \sqrt{\tau_j}$, with shape parameter 1, and scale parameter $\sqrt{\tau_j} = \sqrt{10/2^j}$; i.e. $\theta_j = 0$. For each true $X_i(t)$, 100 sets of K observations u_{ik}'s are then generated from (3.2), where t_{ik}'s are determined by the design being evaluated, and $\varepsilon_{ik} \overset{iid}{\sim} N(0, \sigma^2)$. Each set of $u_{i1}, ..., u_{iK}$ is used to give $\hat{X}_i(t) = \sum_{j=1}^{J} \tilde{\alpha}_{ij}\eta_j(t)$ as described in Section 3. Thus, we have

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

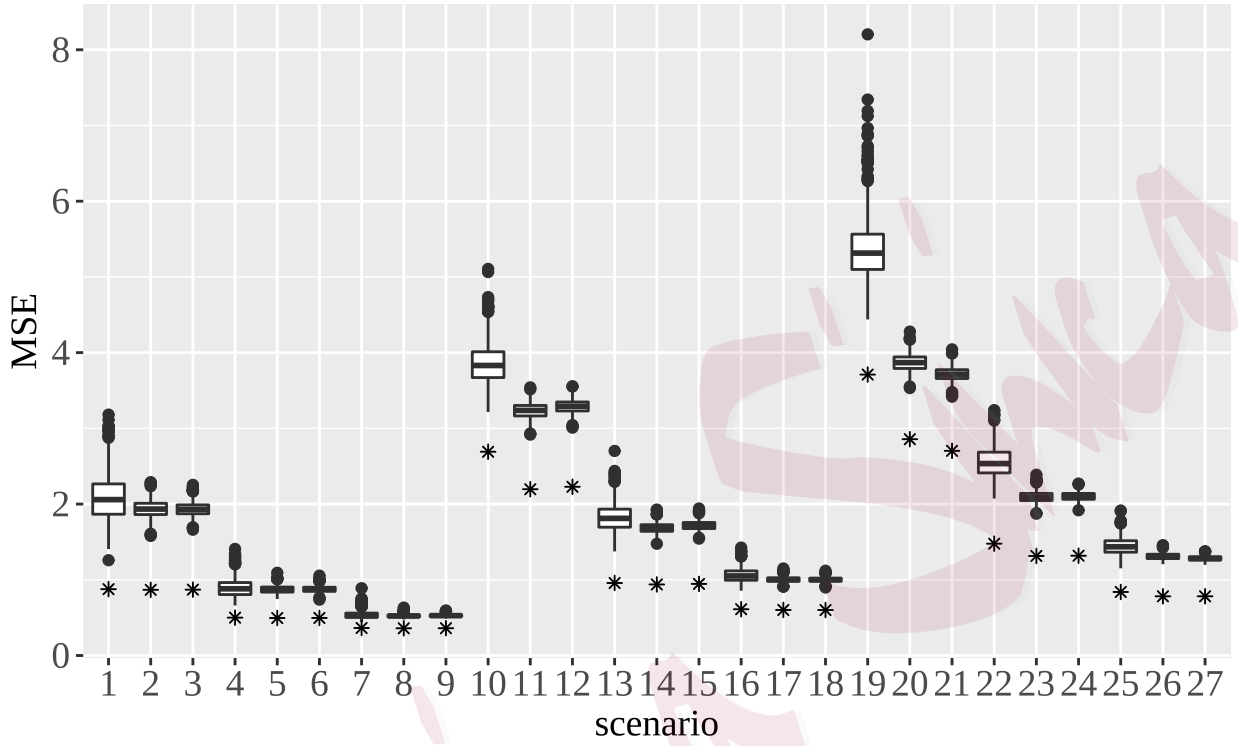

Figure 2: Empirical MSE in predicting X for FECA designs ($*$) vs. 1,000 random designs (box plot) with $\sigma^2 = 1$

100 $\hat{X}_i(t)$ for each of the 100 simulated $X_i(t)$ at all the 21 grid points of the time domain $[0, 1]$; $i = 1, ..., n$. The empirical MSE is set to the average of $\{X_i(t) - \hat{X}_i(t)\}^2$ taken over the n curves, the grid points, and the 100×100 simulations. Figure 2 presents the box plots of this empirical MSE of the 1,000 random designs. The empirical MSEs for our obtained optimal designs are marked as '$*$', and they are consistently smaller than that of the random designs across all the 27 scenarios; the scenarios are defined as in Figure 1.

In addition to $\sigma^2 = 1$, we also consider the σ^2 that is ten times greater (i.e., $\sigma^2 = 10$). Our approach is applied again to obtain optimal FECA designs with this σ^2, and the empirical MSEs of our obtained designs and 1,000 random designs are compared in Figure

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

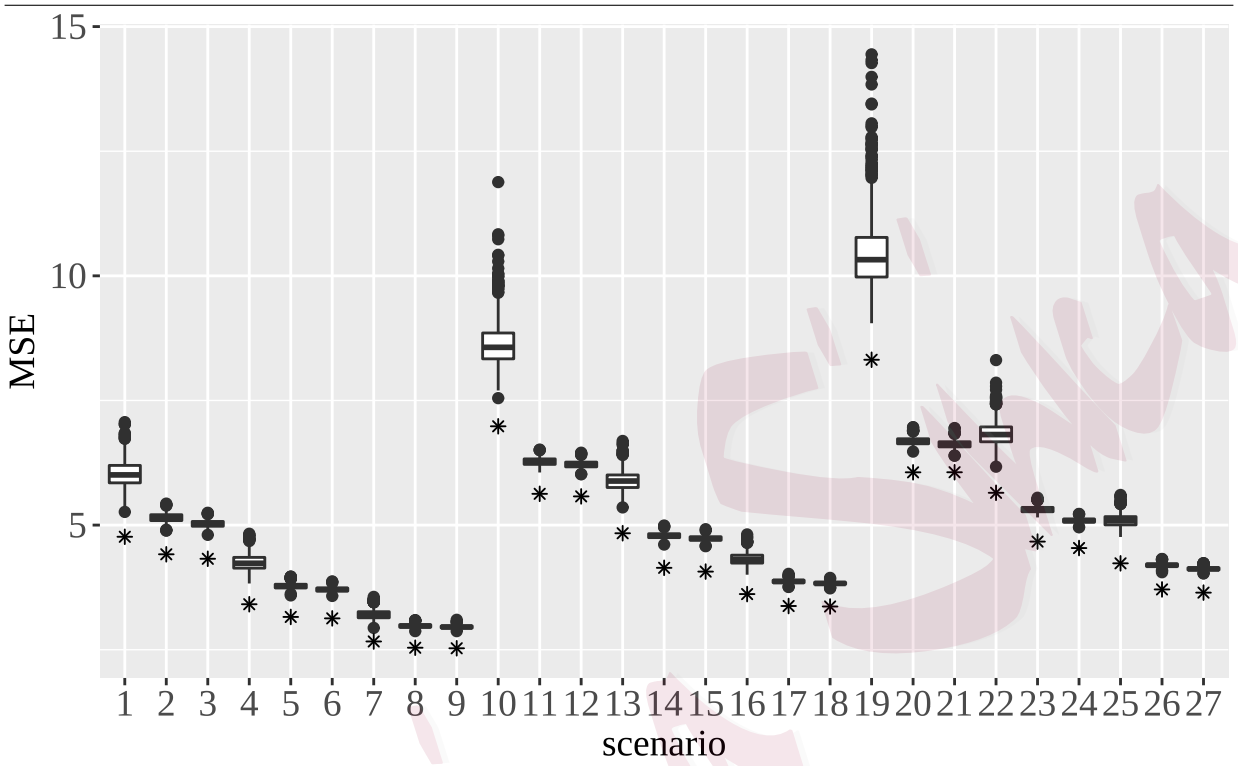

Figure 3: Empirical MSE in predicting X for FECA designs ($*$) vs. 1,000 random designs (box plot) with $\sigma^2 = 10$

3. Both Figures 2 and 3 again suggest that the optimal designs outperform random designs.

In Table 5, we present the computing time spent in obtaining optimal FECA designs for $\sigma^2 = 1$ by implementing the exchange algorithm 1,000 times with random initials. The computing time for $\sigma^2 = 10$ is similar, and is omitted. As described in Section 3, each iteration of the exchange algorithm involves the identification of the 'worst' $\boldsymbol{t}_j$ in the current design, followed by an exhaustive search of the 'best' $\boldsymbol{t}_p$ among the $21!/\{K!(21-K)!\}$ elemental designs. The exhaustive search is used here to render high-quality designs, but it contributes the most to the consumed computational resources. For cases with large n and K, one may consider replacing the exhaustive search with a 'shortcut' method, e.g.,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Table 5: Computing time (hr) of the exchange algorithm with 1,000 initials.

	J=3			J=5			J=7		
n	K=3	K=5	K=7	K=3	K=5	K=7	K=3	K=5	K=7
10	0.02	0.27	1.40	0.02	0.27	1.53	0.02	0.32	1.60
50	0.08	1.16	6.49	0.07	1.23	6.99	0.08	1.29	7.40
70	0.11	1.60	8.96	0.10	1.73	9.70	0.11	1.80	10.24

the PSS algorithm, as also mentioned Section 3 and Appendix A.

5 A discussion

In this work, we study optimal designs for predicting FPC and FEC scores. We first demonstrate that the design issues considered in some previous design works for sparse FDA (e.g., Ji and Müller, 2017) can be formulated as A- or L-optimal design problems for FPC scores. We then present a framework for finding designs that give precise prediction of FEC scores. We note that the FPCA is popular and has a wide application. On the other hand, the FECA is recently proposed by Nie et al. (2022), who demonstrated the advantages of this new method over the FPCA. As also described in Section 3, the latter method does not require precise information on $\mu_x(t)$, and can be considered even when the FEC scores do not follow a Gaussian distribution; see also our numerical study in the previous section, where $\boldsymbol{\alpha}$ has a centered Gamma distribution.

As a byproduct, our work also gives a generalization of Prus and Schwabe (2016). This

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

and existing works on the designs for sparse FDA only considered single-support designs, which do now allow the FECA with $K < J$ as explained in this paper. In sparse FDA, cases with $K < J$ are common. For example, an analysis of the Mediterranean fruit flies data set (i.e., the *medfly* data set of the `FDAPACE` R package) reveals that we need $J = 5$ eigenpairs to explain $\geq 95\%$ of the total variation in the numbers of eggs laid per day from Day 1 to Day 25. Explaining $\geq 99\%$ of the total variation requires at least $J = 8$ eigenpairs. We thus can easily have $K < J$, such as the case in Ji and Müller (2017), where $K = 3$. For these situations, our proposed framework allows to generate high-quality designs. We also note that even with $K \geq J$, single-support design can be suboptimal, and some single-support designs for FPCA can still have zero design efficiency for predicting FEC scores.

The proposed framework not only gives high-quality designs, but also provides a method for evaluating other (e.g., random) designs to avoid wasting time and money on conducting experiments with inefficient designs. It thus provides an important tool for practitioners and researchers for making informed decisions when selecting a design for their studies. An important future research is the development of efficient algorithms for finding optimal designs. The exchange algorithm with an exhaustive search described in Section 3 allowed us to identify good designs for our numerical study. There are cases where the exhaustive search can become infeasible. A possible alternative is by considering the PSS algorithm as described in Appendix A. Our experience suggests that the PSS algorithm requires much less computing time than the exhaustive search, and finds very efficient designs.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Supplementary Material

Some obtained designs are provided in the online supplementary document.

Acknowledgments

We thank the anonymous referees for raising questions that resulted in an improvement of this paper.

References

Atkinson, A. C., A. N. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics 24*(4), 586–602.

Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models, 3rd Ed.* New York: Springer.

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments*. New York: Springer.

Ford, I., B. Torsney, and C. F. J. Wu (1992). The use of a canonical form in the construction of locally optimal designs for nonlinear problems. *Journal of the Royal Statistical Society Series B-Methodological 54*(2), 569–583.

Hall, P., H.-G. Müller, and J.-L. Wang (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics 34*(3), 1493–

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

1517.

Hall, P., H.-G. Müller, and F. Yao (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society. Series B, Statistical methodology 70*(4), 703–723.

Harman, R. and M. Prus (2018). Computing optimal experimental designs with respect to a compound bayes risk criterion. *Statistics & Probability Letters 137*, 135–141.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics 31*(2), 423–447.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*(2), 192–218.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operations*. Chichester, England: John Wiley & Sons.

Ji, H. and H.-G. Müller (2017). Optimal designs for longitudinal and functional data. *Journal of the Royal Statistical Society: Series B 79*(3), 859–876.

Kao, M.-H. and P.-H. Huang (2023). Hybrid exact-approximate design approach for sparse functional data. *Computational Statistics & Data Analysis 190*, 107850.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B 21*(2), 272–319.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Annals of Statistics 2(5), 849–879.

Li, C. and L. Xiao (2020). Optimal design for classification of functional data. *Canadian Journal of Statistics 48*(2), 285–307.

Müller, H.-G. (2008). *Functional Modeling of Longitudinal Data*, Book section 10. Boca Raton, FL: Chapman and Hall/CRC.

Nie, Y., Y. Yang, L. Wang, and J. Cao (2022). Recovering the underlying trajectory from sparse and irregular longitudinal data. *Canadian journal of statistics 50*(1), 122–141.

Park, S. Y., L. Xiao, J. D. Willbur, A.-M. Staicu, and N. L. Jumbe (2018). A joint design for functional data with application to scheduling ultrasound scans. *Computational Statistics and Data Analysis 122*, 101–114.

Prus, M. (2022). Equivalence theorems for multiple-design problems with application in mixed models. *Journal of Statistical Planning and Inference 217*, 153–164.

Prus, M. and R. Schwabe (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B 78*(1), 175–191.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: John Wiley & Sons.

Rha, H., M.-H. Kao, and R. Pan (2020). Design optimal sampling plans for functional regression models. *Computational Statistics and Data Analysis 146*, 106925.

Rha, H., M.-H. Kao, and R. Pan (2021). Bagging-enhanced sampling schedule for functional quadratic regression. *Journal of Statistical Theory and Practice 15*(4), 91, 1–20.

Shi, M., R. E. Weiss, and J. M. G. Taylor (1996). An analysis of paediatric cd4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Royal Statistical Society. Series C (Applied Statistics) 45(2), 151–163.

Stufken, J. and M. Yang (2012). *Optimal Designs for Generalized Linear Models.* In *Design and Analysis of Experiments, Volume 3: Special Designs and Applications,* edited by Klaus Hinkelmann and Oscar Kempthorne, pp. 137–164. John Wiley & Sons.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*(6), 2873–2903.

Zhong, R., S. Liu, H. Li, and J. Zhang (2022). Robust functional principal component analysis for non-gaussian longitudinal data. *Journal of Multivariate Analysis 189*, 104864.

Appendix

A The PSS algorithm and SOCP

We present the probabilistic subset selection (PSS) algorithm of Rha et al. (2020), and its application to our setting. This algorithm takes advantages of prior knowledge on the moments of the underlying random function $X(t)$, and allows a higher probability in exploring time points where $X(t)$ has a greater uncertainty. This has been demonstrated efficient in finding some high quality designs for sparse FDA. With the notation in this paper, let Φ be the objective function of interest, $\mathcal{T}_g$ be the grid of size N_g on the domain $\mathcal{T}$. The PSS algorithm uses the following steps to search for an optimal K-point elemental

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

design $\boldsymbol{t} = \{t_1, ..., t_K\}$ minimizing Φ.

Step 1. Generate an initial K-point design $\boldsymbol{t}_K^{(0)}$. Specify an integer $\ell < N_g - K$; e.g., $\ell = 3$. Set $m = 1$.

Step 2. For the m^{th} iteration, let $\boldsymbol{t}_K^{(m-1)}$ be the best design obtained in the previous iteration. With probability proportional to $Var\{X(t)\}$, randomly draw a $\boldsymbol{t}_\ell$ of ℓ distinct points from the $N_g - K$ points in the set $(\mathcal{T}_g - \boldsymbol{t}_K^{(m-1)})$. Set $\boldsymbol{t}_{K+\ell} = \boldsymbol{t}_K^{(m-1)} \cup \boldsymbol{t}_\ell$.

Step 3. Select a Φ-optimal K-point design among the $(K+\ell)!/K!\ell!$ designs from $\boldsymbol{t}_{K+\ell}$. The obtained design is used as $\boldsymbol{t}_K^{(m)}$ for the next iteration. Set $m = m + 1$.

Step 4. Repeat Steps 2-3 until a stopping rule is met; e.g., when the same design is obtained for $M = 30$ consecutive iterations.

To find designs for FPC scores, Φ is the Φ_A in Theorem 1, $\boldsymbol{t}_K^{(0)}$ in Step 1 is randomly generated, and $Var\{X(t)\}$ in Step 2 is approximated by $\sum_{j=1}^{J} \lambda_j \psi_j^2(t)$. The PSS algorithm can also be considered for obtaining $\boldsymbol{t}_p$ in the exchange algorithm to find designs for FEC scores. In this case, Φ is set to Φ_+ in Theorem 4, and $\boldsymbol{t}_K^{(0)}$ is the $\boldsymbol{t}_j^{(t-1)}$ selected in each iteration of the exchange algorithm as described in Section 3. As for $Var\{X(t)\}$ in Step 2, we propose to replace it by $E\{X^2(t)\}$, which is approximated by $\sum_{j=1}^{J} \tau_j \eta_j^2(t)$.

Harman and Prus (2018) proposed the use of second-order cone programming (SOCP) solvers for minimizing a criterion of the following form:

$$\Phi\{\boldsymbol{d}\} = \sum_{j=1}^{s} tr\left\{(\boldsymbol{M}(\boldsymbol{d}) + \boldsymbol{B}_j)^{-1}\boldsymbol{H}_j\right\},$$

where $\boldsymbol{B}_j$'s are nonnegative definite matrices, and $\boldsymbol{H}_j$'s are positive definite. The Φ_A in our Theorem 1 for FPC scores is a special case with $s = 1$, $\boldsymbol{M}(\boldsymbol{d}) = \boldsymbol{\Psi}_i^\top \boldsymbol{\Psi}_i$, $\boldsymbol{B}_1 = \sigma^2 \boldsymbol{\Lambda}^{-1}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

and $\boldsymbol{H}_1 = \boldsymbol{I}_J$. Their method can thus be directly applied, with the constraint that a time point cannot appear more than once in an elemental design.

We also note that the same method can be adapted to search for $\boldsymbol{t}_p$ in each iteration of the exchange algorithm for obtaining FECA designs. This is done by focusing on minimizing the first term of Φ_+ in Theorem 4 with $s = 1$, and $\boldsymbol{H}_1 = \boldsymbol{I}_J + \boldsymbol{A}_{n-1}^{-1} \sum_{i=1}^{n-1} \boldsymbol{W}_i^{-2} \boldsymbol{A}_{n-1}^{-1}$. $\boldsymbol{H}_1$ is positive definite, but we also need the following result.

Lemma A.1. *$\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1} = \boldsymbol{F}_p^\top \boldsymbol{F}_p + \sigma^2 \boldsymbol{\Delta}^{-1} - \boldsymbol{A}_{n-1}^{-1}$, and $\boldsymbol{B}_1 = (\sigma^2 \boldsymbol{\Delta}^{-1} - \boldsymbol{A}_{n-1}^{-1})$ is nonnegative definite.*

Proof. Let $\leq_L$ represent the Löwner ordering. We would like to show that $\sigma^{-2}\boldsymbol{\Delta} \leq_L \boldsymbol{A}_{n-1}$. But

$$\boldsymbol{A}_{n-1} = n\sigma^{-2}\boldsymbol{\Delta} - \sum_{i=1}^{n-1} \boldsymbol{W}_i^{-1} = \sigma^{-2}\boldsymbol{\Delta} + \sum_{i=1}^{n-1} (\sigma^{-2}\boldsymbol{\Delta} - \boldsymbol{W}_i^{-1}).$$

Our claim then follows from the following observation.

$$\boldsymbol{W}_i^{-1} = (\sigma^2 \boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top \boldsymbol{F}_i)^{-1} \leq_L (\sigma^2 \boldsymbol{\Delta}^{-1})^{-1} = \sigma^{-2}\boldsymbol{\Delta}.$$

B Proofs of theorems

We first present a useful lemma, which extends from Lemma 18.2.3 of Harville (1997).

Lemma B.1. *Suppose $c > 0$, $\boldsymbol{V}$ is a v-by-v positive definite matrix, and $\boldsymbol{S}$ is s-by-v. We have*

$$\boldsymbol{S}^\top (c\boldsymbol{I}_s + \boldsymbol{S}\boldsymbol{V}\boldsymbol{S}^\top)^{-1} = \boldsymbol{V}^{-1}(c\boldsymbol{V}^{-1} + \boldsymbol{S}^\top \boldsymbol{S})^{-1}\boldsymbol{S}^\top.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Proof. With $c > 0$, $(c\boldsymbol{I}_s + \boldsymbol{S}\boldsymbol{V}\boldsymbol{S}^\top)$ and $(c\boldsymbol{V}^{-1} + \boldsymbol{S}^\top\boldsymbol{S})$ are positive definite, and are thus nonsingular. We then have the following.

$$\begin{aligned}\boldsymbol{S}^\top(c\boldsymbol{I}_s + \boldsymbol{S}\boldsymbol{V}\boldsymbol{S}^\top)^{-1} &= c^{-1}\boldsymbol{S}^\top(\boldsymbol{I}_s + c^{-1}\boldsymbol{S}\boldsymbol{V}\boldsymbol{S}^\top)^{-1} = c^{-1}(\boldsymbol{I}_v + c^{-1}\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{V})^{-1}\boldsymbol{S}^\top \\ &= c^{-1}\boldsymbol{V}^{-1}(\boldsymbol{V}^{-1} + c^{-1}\boldsymbol{S}^\top\boldsymbol{S})^{-1} = \boldsymbol{V}^{-1}(c\boldsymbol{V}^{-1} + \boldsymbol{S}^\top\boldsymbol{S})^{-1}\boldsymbol{S}^\top.\end{aligned}$$

Proof of Theorem 2. Following Theorem 8.5.11 of Harville (1997), Eq. (3.4) have a unique solution requiring that $(\boldsymbol{I}_n \otimes \sigma^2\boldsymbol{\Delta}^{-1}) + \boldsymbol{R}^\top\boldsymbol{R}$ is full rank, and $\boldsymbol{Q} = \boldsymbol{F}^\top\boldsymbol{F} - \boldsymbol{F}^\top\boldsymbol{R}(\oplus_{i=1}^n \boldsymbol{W}_i)^{-1}\boldsymbol{R}^\top\boldsymbol{F}$ is nonsingular. With $\sigma^2 > 0$, the former matrix is positive definite, and is thus nonsingular. $\boldsymbol{Q}$ is nonsingular if and only if $\boldsymbol{F}$ has full column rank. This can be seen from the following equivalent expressions of $\boldsymbol{Q}$, and the positive-definiteness of $\boldsymbol{\Sigma}_u = \oplus_{i=1}^n(\sigma^2\boldsymbol{I}_K + \boldsymbol{F}_i\boldsymbol{\Delta}\boldsymbol{F}_i^\top)$, even when $\boldsymbol{F}_i$ does not have full column rank.

$$\begin{aligned}\boldsymbol{Q} &= \sum_{i=1}^n \boldsymbol{F}_i^\top\{\boldsymbol{I}_K - \boldsymbol{F}_i(\sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top\boldsymbol{F}_i)^{-1}\boldsymbol{F}_i^\top\}\boldsymbol{F}_i \\ &= \sigma^2\sum_{i=1}^n \boldsymbol{F}_i^\top(\sigma^2\boldsymbol{I}_K + \boldsymbol{F}_i\boldsymbol{\Delta}\boldsymbol{F}_i^\top)^{-1}\boldsymbol{F}_i = \sigma^2\boldsymbol{F}^\top\boldsymbol{\Sigma}_u^{-1}\boldsymbol{F} \\ &= \sigma^2\boldsymbol{\Delta}^{-1}\sum_{i=1}^n(\sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top\boldsymbol{F}_i)^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i \qquad \text{(B.1)}\end{aligned}$$

The second equality is from the Woodbury matrix identity (Theorem 18.2.8 of Harville, 1997), and the last equality is from Lemma B.1. With some algebra (omitted), the solution $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\gamma}})$ can then be found from Lemma B.1, and (B.1) with Theorem 8.5.11 of Harville (1997).

Proof of Theorem 3. Suppose $\boldsymbol{F}$ has full column rank, and $\boldsymbol{W}_i = \sigma^2\boldsymbol{\Delta}^{-1} + \boldsymbol{F}_i^\top\boldsymbol{F}_i$ with $(\boldsymbol{I}_n \otimes \sigma^2\boldsymbol{\Delta}^{-1}) + \boldsymbol{R}^\top\boldsymbol{R} = \oplus_{i=1}^n \boldsymbol{W}_i$. Following Henderson (1975), Theorem 8.5.11 of Harville

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

(1997), and (B.1), we then have

$$Cov\left\{\boldsymbol{G}\begin{pmatrix}\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\\ \tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\end{pmatrix}\right\}=\sigma^2\boldsymbol{G}\begin{pmatrix}\boldsymbol{F}^\top\boldsymbol{F} & \boldsymbol{F}^\top\boldsymbol{R}\\ \boldsymbol{R}^\top\boldsymbol{F} & \oplus_{i=1}^n\boldsymbol{W}_i\end{pmatrix}^{-1}\boldsymbol{G}^\top=\sigma^2\boldsymbol{G}\begin{pmatrix}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12}\\ \boldsymbol{Q}_{12}^\top & \boldsymbol{Q}_{22}\end{pmatrix}\boldsymbol{G}^\top;$$

$$\begin{aligned}\boldsymbol{Q}_{11} &= \boldsymbol{Q}^{-1}=\sigma^{-2}\left(\sum_{i=1}^n\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i\right)^{-1}\boldsymbol{\Delta};\\ \boldsymbol{Q}_{12} &= -\boldsymbol{Q}^{-1}\boldsymbol{F}^\top\boldsymbol{R}(\oplus_{i=1}^n\boldsymbol{W}_i^{-1})=-\boldsymbol{Q}^{-1}\left[\boldsymbol{F}_1^\top\boldsymbol{F}_1\boldsymbol{W}_1^{-1},...,\boldsymbol{F}_n^\top\boldsymbol{F}_n\boldsymbol{W}_n^{-1}\right];\end{aligned}$$

$$\begin{aligned}\boldsymbol{Q}_{22} &= (\oplus_{i=1}^n\boldsymbol{W}_i^{-1})+(\oplus_{i=1}^n\boldsymbol{W}_i^{-1})\boldsymbol{R}^\top\boldsymbol{F}\boldsymbol{Q}^{-1}\boldsymbol{F}^\top\boldsymbol{R}(\oplus_{i=1}^n\boldsymbol{W}_i^{-1});\\ &= (\oplus_{i=1}^n\boldsymbol{W}_i^{-1})+\begin{pmatrix}\boldsymbol{W}_1^{-1}\boldsymbol{F}_1^\top\boldsymbol{F}_1\\ \vdots\\ \boldsymbol{W}_n^{-1}\boldsymbol{F}_n^\top\boldsymbol{F}_n\end{pmatrix}\boldsymbol{Q}^{-1}\left[\boldsymbol{F}_1^\top\boldsymbol{F}_1\boldsymbol{W}_1^{-1},...,\boldsymbol{F}_n^\top\boldsymbol{F}_n\boldsymbol{W}_n^{-1}\right].\end{aligned}$$

Here, $\boldsymbol{G}=[\boldsymbol{j}_n,\boldsymbol{I}_n]\otimes\boldsymbol{I}_J$, and we observe that $\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i=\boldsymbol{I}_J-\sigma^2\boldsymbol{W}_i^{-1}\boldsymbol{\Delta}^{-1}$. The $\boldsymbol{C}_{ij}$ in Theorem 3 can then be derived as follows.

$$\begin{aligned}\boldsymbol{C}_{ij} &= \boldsymbol{Q}^{-1}-\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i\boldsymbol{Q}^{-1}-\boldsymbol{Q}^{-1}\boldsymbol{F}_j^\top\boldsymbol{F}_j\boldsymbol{W}_j^{-1}+\delta_{ij}\boldsymbol{W}_i^{-1}+\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i\boldsymbol{Q}^{-1}\boldsymbol{F}_j^\top\boldsymbol{F}_j\boldsymbol{W}_j^{-1}.\\ &= \delta_{ij}\boldsymbol{W}_i^{-1}+\boldsymbol{Q}^{-1}-(\boldsymbol{I}_J-\sigma^2\boldsymbol{W}_i^{-1}\boldsymbol{\Delta}^{-1})\boldsymbol{Q}^{-1}-\boldsymbol{Q}^{-1}(\boldsymbol{I}_J-\sigma^2\boldsymbol{\Delta}^{-1}\boldsymbol{W}_j^{-1})\\ &\quad+(\boldsymbol{I}_J-\sigma^2\boldsymbol{W}_i^{-1}\boldsymbol{\Delta}^{-1})\boldsymbol{Q}^{-1}(\boldsymbol{I}_J-\sigma^2\boldsymbol{\Delta}^{-1}\boldsymbol{W}_j^{-1})\\ &= \delta_{ij}\boldsymbol{W}_i^{-1}+\sigma^4\boldsymbol{W}_i^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{W}_j^{-1}\\ &= \delta_{ij}\boldsymbol{W}_i^{-1}+\boldsymbol{W}_i^{-1}\left(\sum_{i=1}^n\sigma^{-2}\boldsymbol{\Delta}-\boldsymbol{W}_i^{-1}\right)^{-1}\boldsymbol{W}_j^{-1}.\end{aligned}$$

The last equality is due to the following.

$$\begin{aligned}\boldsymbol{Q} &= \sigma^2\boldsymbol{\Delta}^{-1}\sum_{i=1}^n\boldsymbol{W}_i^{-1}\boldsymbol{F}_i^\top\boldsymbol{F}_i=\sigma^2\boldsymbol{\Delta}^{-1}\sum_{i=1}^n(\boldsymbol{I}_J-\sigma^2\boldsymbol{W}_i^{-1}\boldsymbol{\Delta}^{-1})\\ &= \sigma^4\boldsymbol{\Delta}^{-1}\left(\sum_{i=1}^n\sigma^{-2}\boldsymbol{\Delta}-\boldsymbol{W}_i^{-1}\right)\boldsymbol{\Delta}^{-1}.\end{aligned}\tag{B.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

With (B.2), and the above proof for Theorem 2, we see that $\boldsymbol{A}^{-1} = (n\sigma^{-2}\boldsymbol{\Delta} - \sum_{i=1}^{n} \boldsymbol{W}_i^{-1})^{-1}$ exists and is positive definite if and only if $\boldsymbol{F}$ has full column rank.

Proof of Theorem 4. As in the proof for Theorem 3, $\boldsymbol{W}_1, ..., \boldsymbol{W}_n$, and $\boldsymbol{A} = n\sigma^{-2}\boldsymbol{\Delta} - \sum_{i=1}^{n} \boldsymbol{W}_i^{-1}$ are positive definite under the design $\boldsymbol{d}$, and so is $\boldsymbol{A}_{n-1} = \boldsymbol{A} + \boldsymbol{W}_n^{-1}$. In addition, the two matrices $\boldsymbol{A}_p = \boldsymbol{A}_{n-1} - \boldsymbol{W}_p^{-1}$ and $\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1}$ have the same rank (Theorem 18.2.4 of Harville, 1997). They are nonsingular if and only if $\boldsymbol{d}_p$ yields a full-column rank $\boldsymbol{F}$. We then have

$$\begin{aligned}
\boldsymbol{A}_p^{-1} &= (\boldsymbol{A}_{n-1} - \boldsymbol{W}_p^{-1})^{-1} = -\boldsymbol{W}_p + \boldsymbol{W}_p(\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1})^{-1}\boldsymbol{W}_p; \\
\boldsymbol{W}_p^{-1} &+ \boldsymbol{W}_p^{-1}\boldsymbol{A}_p^{-1}\boldsymbol{W}_p^{-1} = (\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1})^{-1}; \text{ and} \\
\boldsymbol{W}_i^{-2}\boldsymbol{A}_p^{-1} &= \boldsymbol{W}_i^{-2}(\boldsymbol{A}_{n-1} - \boldsymbol{W}_p^{-1})^{-1} = \boldsymbol{W}_i^{-2}\boldsymbol{A}_{n-1}^{-1} + \boldsymbol{W}_i^{-2}\boldsymbol{A}_{n-1}^{-1}(\boldsymbol{W}_p - \boldsymbol{A}_{n-1}^{-1})^{-1}\boldsymbol{A}_{n-1}^{-1}.
\end{aligned}$$

The result in Theorem 4 then follows with simple algebra.

Proof of Theorem 5. For any design $\boldsymbol{d} \in \Xi_d$, and given i, we observe that

$$\boldsymbol{W}_i^{-1} + \boldsymbol{W}_i^{-1}\boldsymbol{A}^{-1}\boldsymbol{W}_i^{-1} >_L \boldsymbol{W}_i^{-1} + \boldsymbol{W}_i^{-1}(n\sigma^{-2}\boldsymbol{\Delta} - \boldsymbol{W}_i^{-1})^{-1}\boldsymbol{W}_i^{-1} = \widetilde{\boldsymbol{W}}_i^{-1} >_L \boldsymbol{W}_i^{-1},$$

where $>_L$ is Löwner ordering, and $\boldsymbol{M}_1 >_L \boldsymbol{M}_2$ if $(\boldsymbol{M}_1 - \boldsymbol{M}_2)$ is positive definite. Our claim then follows from the Löwner isotonicity of trace.

Proof of Theorem 6. With the same $\boldsymbol{G}$ as in the proof for Theorem 3, we have

$$\mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d}) = \boldsymbol{G}[\boldsymbol{G}^\top(\oplus_{i=1}^{n}\boldsymbol{F}_i^\top\boldsymbol{F}_i)\boldsymbol{G} + \{\boldsymbol{O}_J \oplus (\boldsymbol{I}_n \otimes \sigma^2\boldsymbol{\Delta}^{-1})\}]^{-1}\boldsymbol{G}^\top.$$

Without loss of generality, let $\boldsymbol{d}_a$ be obtained by adding a point, t_a, to $\boldsymbol{t}_1$ of $\boldsymbol{d}$. Then $\mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d}_a)$ can be obtained from $\mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d})$ by replacing $\boldsymbol{F}_1^\top\boldsymbol{F}_1$ with $\boldsymbol{F}_1^\top\boldsymbol{F}_1 + \boldsymbol{f}\boldsymbol{f}^\top (\geq_L \boldsymbol{F}_1^\top\boldsymbol{F}_1)$ with $\boldsymbol{f} = (\eta_1(t_a), ..., \eta_J(t_a))^\top$; $\mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d}_a) = \mathbb{I}^{-1}(\boldsymbol{\alpha}; \boldsymbol{d})$ if $\eta_j(t_a) = 0$ for $j = 1, ..., J$. Our claim then follows from Theorems 14.2.9 & 18.3.4 of Harville (1997).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0051

Ming-Hung Kao

School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona 85287, U.S.A.

E-mail: mkao3@asu.edu

Ping-Han Huang

School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona 85287, U.S.A.

E-mail: phuang37@asu.edu