Statistica Sinica Preprint No: SS-2022-0328

Title	Necessary and Sufficient Conditions for Multiple Objective Optimal Regression Designs
Manuscript ID	SS-2022-0328
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0328
Complete List of Authors	Lucy L. Gao, Jane J. Ye, Shangzhi Zeng and Julie Zhou
Corresponding Authors	Lucy L. Gao
E-mails	lucy.gao@stat.ubc.ca

doi:10.5705/ss.202022.0328

Statistica Sinica

Necessary and sufficient conditions for multiple objective optimal regression designs

Lucy L. Gao° , Jane J. Ye[‡], Shangzhi Zeng[‡], Julie Zhou[‡]

∘ *Department of Statistics, University of British Columbia*

† *Department of Mathematics and Statistics, University of Victoria*

Abstract: **We typically construct optimal designs based on a single objective function. To better capture the breadth of an experiment's goals, we could instead construct a multiple objective optimal design based on multiple-objective functions. However, although algorithms have been developed to find such designs (e.g., efficiency-constrained and maximin optimal designs), it is far less clear how to verify the optimality of a solution obtained from these algorithms. In this paper, we provide theoretical results that characterize optimality for efficiency-constrained and maximin optimal designs on a discrete design space. Lastly, we demonstrate how to use our results with linear programming algorithms to verify optimality.**

Key words and phrases: optimality conditions, efficiency, maximin design, linear programming, convex optimization, robustness

Author ordering is alphabetical. Corresponding author: lucy.gao@stat.ubc.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

1. Introduction

Consider modeling the output of a designed experiment as:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where y_i is the response variable observed at design point $\mathbf{x}_i \in S$, for $S \subseteq \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathbb{R}^q$ is a vector of unknown regression parameters, and ϵ_i are independent random errors, with $\mathbb{E}[\epsilon_i] = 0$ and $\text{Var}(\epsilon_i) = \sigma^2$. An optimal design chooses values of $\mathbf{x}_i$ to answer the experimental questions of interest as precisely as possible. This problem is often formulated as a *single-objective* optimal design problem, where optimality is defined with respect to a single summary measure of the information, obtained by fitting a single model to the experimental data. For example, for a particular choice of regression function $f(\cdot, \cdot)$ in (1.1) and estimator $\hat{\boldsymbol{\theta}}$, an A-optimal design minimizes the average variance of $\hat{\theta}_1, \ldots, \hat{\theta}_q$.

However, experimenters sometimes have complex goals that cannot be captured fully by a single-objective optimal design criterion. For example, an experimenter may fit a single model for inference and prediction, but there is little overlap between the single-objective optimality criteria that measure inferential and predictive power. Furthermore, depending on the parameters being examined, an inference answers different research questions, with varying importance to the experimenter (e.g., main effects are more important than interaction terms, or vice versa). Single-objective optimal design criteria do not reflect this variation. Another consideration is that experimenters may be uncertain about the functional form of the relationship between y_i and $\mathbf{x}_i$. Thus, they may want a design with good inferential or predictive power for multiple models, rather than a single model, of the form (1.1).

Multi-objective optimal designs combine several single-objective optimal design criteria. Here, common formulations include the *compound*, efficiency-constrained, and maximin formulations. The compound formulation optimizes the weighted sum of the criteria for a set of user-specified weights. The *efficiency-constrained* formulation optimizes one criterion, while requiring the design efficiency with respect to the other criteria to be higher than user-specified values. The *maximin* formulation maximizes the minimum efficiency across the set of optimality criteria (Wong, 1999; Wong and Zhou, 2023). Here, we focus on the efficiency-constrained and maximin formulations, because it is difficult to interpret the practical significance of the weights in a compound formulation. Numerous algorithms exist for finding efficiency-constrained and maximin optimal designs (Huang and Wong, 1998; Imhof and Wong, 2000; Cheng and Yang, 2019). Wong and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Zhou (2023) provide a particularly flexible algorithm. They formulate many efficiency-constrained and maximin optimal design problems as convex optimization problems, and then apply an off-the-shelf convex optimization solver, such as the `MATLAB` package `CVX` (Grant and Boyd 2014).

We formulate efficiency-constrained and maximin problems as convex optimization problems, following Wong and Zhou (2023). We then consider how to verify the optimality of an efficiency-constrained or a maximin optimal design obtained from `CVX`, providing a complete characterization of optimality for efficiency-constrained and maximin efficiency designs on a discrete design space. Related results appear in the literature for efficiency-constrained optimal designs (see, e.g., Cook and Wong 1994 and Clyde and Chaloner 1996). To the best of our knowledge, our characterization of optimality for minimax efficiency designs is new, although there are related works on minimax and maximin single-objective optimization problems (see, e.g., Müller and Pázman 1998 and Dette et al. 2007).

Characterizations of optimality for many popular single-objective optimal design criteria (e.g., D- and A-) require that the optimal design satisfies a set of easily computable inequalities. In contrast, our characterizations of optimality for efficiency-constrained and maximin designs posit the *existence* of a set of quantities that satisfy a set of equalities and inequalities

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

involving the optimal design. These types of results are thought to be impractical for optimality verification, because it is unclear how to find a suitable set of quantities efficiently. Previous works on efficiency-constrained optimal design problems search for a suitable set of quantities using a grid search and bisection search. However, the computational complexity of such methods grows exponentially in the number of objective functions (Cheng and Yang, 2019). We overcome this challenge of finding a suitable set of quantities by solving linear programming problems (Luenberger and Ye, 2016), which are a cornerstone of mathematical optimization, and can be solved accurately and efficiently using off-the-shelf software.

The rest of the paper is organized as follows. In Section 2, we review concepts related to single-objective optimality criteria, including the necessary and sufficient conditions for optimality. In Sections 3 and 4, we describe how to solve efficiency-constrained and maximin optimal designs, respectively, including verifying the optimality of the obtained solutions. Several theoretical results are derived. We apply our approach to several examples in Section 5. Section 6 concludes the paper. Proofs are provided in the Appendix.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

2. Single-objective optimal designs

We consider a discrete design space $S_N = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\} \subseteq S$ with N points, where $\mathbf{u}_1, \ldots, \mathbf{u}_N$ and S are user-specified. If S is a continuous design space, then S_N approximates S. We denote a design ξ on S_N by $\xi(\mathbf{w}) = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \\ w_1 & w_2 & \cdots & w_N \end{pmatrix}$, where $\mathbf{w}$ is an N-vector, with the ith entry w_i representing the proportion of design points with value $\mathbf{u}_i$, for $i = 1, 2, \ldots, N$. Let $\Omega \equiv \left\{\mathbf{w} \in \mathbb{R}^N : \sum_{i=1}^{N} w_i = 1, w_i \geq 0\right\}$.

2.1 Optimality criteria

Let $\mathbf{z}_f(\mathbf{x})$ be the q-vector with jth entry $\frac{\partial f(\mathbf{x},\boldsymbol{\theta})}{\partial \theta_j}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$, where $\boldsymbol{\theta}^*$ is the true value of $\boldsymbol{\theta}$. The asymptotic covariance matrix of the ordinary least squares estimator of $\boldsymbol{\theta}$ in model (1.1) with regression function $f(\cdot,\cdot)$ at design $\xi(\mathbf{w})$ is proportional to $\mathcal{I}_f^{-1}(\mathbf{w})$, where

$$\mathcal{I}_f(\mathbf{w}) = \sum_{i=1}^{N} w_i \mathbf{z}_f(\mathbf{u}_i)\mathbf{z}_f^T(\mathbf{u}_i) \tag{2.2}$$

is the expected information matrix for model (1.1) with regression function $f(\cdot,\cdot)$, under the assumption of normally distributed errors. If $f(\mathbf{x}, \boldsymbol{\theta})$ is nonlinear in $\boldsymbol{\theta}$, then $\mathcal{I}_f(\mathbf{w})$ may depend on $\boldsymbol{\theta}^*$. If $\mathcal{I}_f(\mathbf{w})$ depends on $\boldsymbol{\theta}^*$, then optimizing the design criteria involving $\mathcal{I}_f(\mathbf{w})$ yields locally optimal designs. In practice, $\boldsymbol{\theta}^*$ is typically unknown, so we must replace it with a "guess" about its value, for example, an estimate of $\boldsymbol{\theta}$ from a small pilot study.

Many single-objective optimal design criteria on S_N can be transformed into convex optimization problems of the form $\min_{\mathbf{w}\in\Omega} \Phi(\mathbf{w})$, where $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ for a convex function ϕ defined on the set of all $q \times q$ positive-definite matrices; see, for example, Table 1. Note that the function $\Phi(\mathbf{w})$ is convex as a composition of a convex function and a linear function. We measure the quality of a design $\mathbf{w}$ using its *efficiency* relative to the optimal design, denoted as Eff($\mathbf{w}$).

Table 1: Single-objective optimality criteria that solve $\min_{\mathbf{w}\in\Omega} \Phi(\mathbf{w})$, where $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ for a convex function ϕ defined on the set of all positive-definite matrices. We use $\lambda_{min}(\mathbf{M})$ to denote the smallest eigenvalue of $\mathbf{M}$.

Criteria	*D*-	*A*-	*c*-, for $\mathbf{c} \in \mathbb{R}^q$	*L*-, for $\mathbf{L} \in \mathbb{R}^{q\times q'}$	*E*-
$\phi(\mathbf{M})$	$-\log\det(\mathbf{M})$	$\mathrm{trace}(\mathbf{M}^{-1})$	$\mathbf{c}^T\mathbf{M}^{-1}\mathbf{c}$	$\mathrm{trace}(\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L})$	$-\lambda_{min}(\mathbf{M})$
Eff($\mathbf{w}$)	$\left(\frac{\exp\left\{\min_{\mathbf{w}'\in\Omega}\Phi(\mathbf{w}')\right\}}{\exp\{\Phi(\mathbf{w})\}}\right)^{1/q}$	$\frac{\min_{\mathbf{w}'\in\Omega}\Phi(\mathbf{w}')}{\Phi(\mathbf{w})}$	$\frac{\min_{\mathbf{w}'\in\Omega}\Phi(\mathbf{w}')}{\Phi(\mathbf{w})}$	$\frac{\min_{\mathbf{w}'\in\Omega}\Phi(\mathbf{w}')}{\Phi(\mathbf{w})}$	$\frac{\Phi(\mathbf{w})}{\min_{\mathbf{w}'\in\Omega}\Phi(\mathbf{w}')}$

The `MATLAB` package `CVX` (Grant and Boyd, 2014) is a user-friendly option for solving a special subclass of convex optimization problems that includes the convex optimization problems described in Table 1; further details on `CVX` are provided in Section 3.1. The `CVX` package has previously been applied to solve many single-objective optimal design problems; see, for example, Gao and Zhou (2017) and Wong and Zhou (2019).

2.2 Necessary and sufficient conditions for optimality

All of the criteria in Table 1 lead to convex objective functions, but these objective functions are not all differentiable everywhere. For example, the E-optimal design criterion leads to a convex objective function that is non-differentiable at designs $\mathbf{w}$, such that the smallest eigenvalue of $\mathcal{I}_f(\mathbf{w})$ has geometric multiplicity greater than one. Thus, the optimality conditions in this setting rely on *subdifferentials*, which generalize derivatives to the class of convex functions. We denote the subdifferential of a convex function $\Phi : \mathbb{R}^N \to \mathbb{R}$ at a point $\mathbf{w}$ as $\partial\Phi(\mathbf{w})$. The following result describes basic properties of subdifferentials (Chapter 2, Mordukhovich and Nam 2013).

Lemma 1. *Suppose that* Φ *and* Φ' *are two finite-valued convex functions defined on* $\mathbb{R}^N$*. Then, for any* $\mathbf{w} \in \mathbb{R}^N$*:*

1. *If* Φ *is differentiable at* $\mathbf{w}$*, then* $\partial\Phi(\mathbf{w}) = \{\nabla\Phi(\mathbf{w})\}$*, where* $\nabla\Phi(\mathbf{v})$ *is the* N*-vector with* i*th entry* $\frac{\partial\Phi}{\partial w_i}\big|_{\mathbf{w}=\mathbf{v}}$.

2. *If* $a \geq 0$*, then* $\partial(a\Phi)(\mathbf{w}) = a\partial\Phi(\mathbf{w}) \equiv \{a\mathbf{g} : \mathbf{g} \in \partial\Phi(\mathbf{w})\}$.

3. $\partial(\Phi+\Phi')(\mathbf{w}) = \partial\Phi(\mathbf{w})+\partial\Phi'(\mathbf{w}) \equiv \{\mathbf{g}+\mathbf{g}' : \mathbf{g} \in \partial\Phi(\mathbf{w}), \mathbf{g}' \in \partial\Phi'(\mathbf{w})\}$.

Let $\mathbf{e}_i$ denote the N-vector with ith entry equal to one and all other entries equal to zero. The following result characterizes optimality for convex single-objective optimal design criteria on a discrete design space.

Theorem 1. *Suppose that* $\Phi : \mathbb{R}^N \to \mathbb{R}$ *is a convex function. Let* $\mathbf{w}^* \in \Omega$*. Then,* $\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\Omega} \Phi(\mathbf{w})$ *if and only if*

$$\exists\ \mathbf{g} \in \partial\Phi(\mathbf{w}^*) \text{ such that } \mathbf{g}^T(\mathbf{w}^* - \mathbf{e}_i) \leq 0, \text{ for all } i = 1, 2, \ldots, N. \quad (2.3)$$

The rest of this subsection is devoted to results that help us evaluate condition (2.3) in the special case where $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ for a convex function ϕ. First, if ϕ is convex and differentiable at $\mathcal{I}_f(\mathbf{w}^*)$, then Lemma 1 states that $\partial\Phi(\mathbf{w}^*) = \{\nabla\Phi(\mathbf{w}^*)\}$, and condition (2.3) simplifies to

$$[\nabla\Phi(\mathbf{w}^*)]^T (\mathbf{w}^* - \mathbf{e}_i) \leq 0 \text{ for all } i = 1, 2, \ldots, N. \quad (2.4)$$

The following result follows from the matrix chain rule (Section 2.8.1 of Petersen and Pedersen 2012), and characterizes the left-hand side of (2.4).

Lemma 2. *Let* $\mathbf{w}^* \in \Omega$*. If* $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ *for a convex function* ϕ *and a regression function* $f(\cdot,\cdot)$*, and* ϕ *is differentiable at* $\mathcal{I}_f(\mathbf{w}^*)$*, then*

$$[\nabla\Phi(\mathbf{w}^*)]^T (\mathbf{w}^* - \mathbf{e}_i) = d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*), \quad \text{for all } i = 1, 2, \ldots, N,$$

where for all $i = 1, 2, \ldots, N$*, we define*

$$d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*) \equiv \mathit{trace}\left([\nabla\phi(\mathcal{I}_f(\mathbf{w}^*))]^T \left[\mathcal{I}_f(\mathbf{w}^*) - \mathbf{z}_f(\mathbf{u}_i)\mathbf{z}_f^T(\mathbf{u}_i)\right]\right), \quad (2.5)$$

and where $\nabla\phi(\mathbf{M}^*)$ *is the* $q \times q$ *matrix with* (j, j')*th entry* $\frac{\partial \phi}{\partial M_{jj'}}\big|_{\mathbf{M}=\mathbf{M}^*}$.

It follows from Theorem 1 and Lemma 2 that characterizing optimality for single-objective optimality criteria with $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ for a differentiable convex function ϕ amounts to checking whether $d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*) \leq 0$, for all $i = 1, 2, \ldots, N$. Furthermore, $d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*)$ is straightforward to compute, given the formula for $\nabla\phi(\mathbf{M}^*)$; see Table 2.

Table 2: Optimality criteria that solve $\min_{\mathbf{w}\in\Omega} \phi(\mathbf{A}_f(\mathbf{w}))$ for all differentiable convex functions ϕ given in Table 1. The formulae for $\nabla\phi(\mathbf{M})$ are from Petersen and Pedersen (2012).

Criteria	D-	A-	c-, for $\mathbf{c} \in \mathbb{R}^q$	L-, for $\mathbf{L} \in \mathbb{R}^{q\times q'}$
$\phi(\mathbf{M})$	$-\log\det(\mathbf{M})$	$\mathrm{trace}(\mathbf{M}^{-1})$	$\mathbf{c}^T\mathbf{M}^{-1}\mathbf{c}$	$\mathrm{trace}(\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L})$
$\nabla\phi(\mathbf{M})$	$-\mathbf{M}^{-1}$	$-\mathbf{M}^{-2}$	$-\mathbf{M}^{-1}\mathbf{c}\mathbf{c}^T\mathbf{M}^{-1}$	$-\mathbf{M}^{-1}\mathbf{L}\mathbf{L}^T\mathbf{M}^{-1}$

Combining Theorem 1, Lemma 2, and Table 1 yields various classical equivalence theorems on a discrete design space; see, for example, Kiefer (1974).

In the case of E-optimality, we have $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ for a *non-differentiable* convex function ϕ (Table 1); thus Lemma 2 does not always apply. We address this with the following result.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Lemma 3. *Suppose that* $\Phi(\mathbf{w}) = -\lambda_{min}(\mathcal{I}_f(\mathbf{w}))$ *for a regression function* $f(\cdot,\cdot)$*. Let* $\mathbf{w}^* \in \Omega$ *and* r^* *be the geometric multiplicity of* $\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*))$*.*

1. *If* $r^* = 1$*, then* $\partial\Phi(\mathbf{w}^*) = \{\nabla\Phi(\mathbf{w}^*)\}$*, and*

   $$[\nabla\Phi(\mathbf{w}^*)]^T(\mathbf{w}^*-\mathbf{e}_i) = [(\mathbf{v}^*)^T\mathbf{z}_f(\mathbf{u}_i)]^2-\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*)),\ \textit{for all } i = 1, 2, \ldots, N,$$

   where $\mathbf{v}^*$ *denotes an arbitrary unit eigenvector associated with* $\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*))$*.*

2. *If* $r^* > 1$*, then for any* $\mathbf{g} \in \partial\Phi(\mathbf{w}^*)$*, there exist* $a_1, \ldots, a_{r^*} \geq 0$ *such that* $\sum_{j=1}^{r^*} a_j = 1$ *and*

   $$\mathbf{g}^T(\mathbf{w}^* - \mathbf{e}_i) = d_{-\lambda_{min},f,\mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*),\ \textit{for all } i = 1, 2, \ldots, N,$$

   where for all $i = 1, 2, \ldots, N$*, we define*

   $$d_{-\lambda_{min},f,\mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*) \equiv \sum_{j=1}^{r^*} a_j[(\mathbf{v}_j^*)^T\mathbf{z}_f(\mathbf{u}_i)]^2 - \lambda_{min}(\mathcal{I}_f(\mathbf{w}^*)), \qquad (2.6)$$

   where $\mathbf{v}_1^*, \ldots, \mathbf{v}_{r^*}^*$ *denotes an arbitrary set of orthonormal eigenvectors associated with* $\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*))$*.*

Combining Theorem 1 with Lemma 3 yields the equivalence theorem for E-optimality on a discrete design space (Kiefer, 1974).

3. Efficiency-constrained optimal designs

Suppose that an experimenter is interested primarily in optimizing one particular single-objective optimality criterion Φ_1, without losing too much efficiency with respect to the other criteria $\Phi_2, \ldots, \Phi_K$, for $K \geq 2$. Let $\text{Eff}_k(\mathbf{w})$ denote the efficiency of the design $\xi(\mathbf{w})$ with respect to criterion Φ_k, for $k = 1, \ldots, K$. Given experimenter-specified constants $m_2, \ldots, m_K \in (0, 1)$, an *efficiency-constrained optimal design* on S_N solves

$$\left\{ \begin{array}{ll} \min\limits_{\mathbf{w} \in \mathbb{R}^\mathbf{N}} & \Phi_1(\mathbf{w}) \\ \text{subject to:} & \text{Eff}_k(\mathbf{w}) \geq m_k,\ k = 2, \ldots, K, \\ & \sum\limits_{i=1}^{N} w_i = 1, w_i \geq 0,\ i = 1, 2, \ldots, N. \end{array} \right\}. \tag{3.7}$$

All designs may fail to satisfy the constraints in (3.7) when the desired minimum efficiencies $m_2, \ldots, m_K$ are large.

3.1 Convex optimization problem

Suppose that $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$ for all $k = 1, \ldots, K$, where $\phi_1, \ldots, \phi_K$ are continuous convex functions chosen from Table 1, and $\mathcal{I}_{f_k}(\mathbf{w})$ in (2.2) is the expected information matrix for a model of the form (1.1) with regression function $f_k(\mathbf{x}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathbb{R}^{q_k}$ at design $\xi(\mathbf{w})$. We use the definitions of $\text{Eff}_k(\mathbf{w})$ in Table 1 to rewrite (3.7) as the following problem:

$$\left\{ \begin{array}{ll} \min\limits_{\mathbf{w} \in \mathbb{R}^\mathbf{N}} & \Phi_1(\mathbf{w}) \\ \text{subject to:} & \Phi_k(\mathbf{w}) \leq h_k(m_k),\ k = 2, \ldots, K, \\ & \sum\limits_{i=1}^{N} w_i = 1,\ w_i \geq 0,\ i = 1, 2, \ldots, N. \end{array} \right\}, \tag{3.8}$$

where we define $h_k(m)$ as

$$h_k(m) = \begin{cases} \left(\min_{\mathbf{w}'\in\Omega} \Phi_k(\mathbf{w}')\right) - q_k \log(m), & \text{if } \Phi_k(\mathbf{w}) = -\log\det(\mathcal{I}_{f_k}(\mathbf{w})), \\ m\left(\min_{\mathbf{w}'\in\Omega} \Phi_k(\mathbf{w}')\right), & \text{if } \Phi_k(\mathbf{w}) = -\lambda_{min}(\mathcal{I}_{f_k}(\mathbf{w})), \\ \frac{1}{m}\left(\min_{\mathbf{w}'\in\Omega} \Phi_k(\mathbf{w}')\right), & \text{otherwise.} \end{cases} \quad (3.9)$$

For all $k = 1, 2, \ldots, K$, $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$ is a convex function, because ϕ_k is a convex function and $\mathcal{I}_{f_k}$ is a linear function. Thus, (3.8) is a convex optimization problem. Note that our formulation differs from that of Wong and Zhou (2022) because we use $\Phi_k(\mathbf{w}) = -\log\det(\mathcal{I}_{f_k}(\mathbf{w}))$ for D-optimality, rather than $\Phi_k(\mathbf{w}) = [\det(\mathcal{I}_{f_k}(\mathbf{w})]^{-1/q_k}$.

In fact, (3.8) is a convex optimization problem that can be solved by using `CVX` (Grant and Boyd, 2014), a `MATLAB` package that works with a special subclass of optimization problems; see Grant and Boyd (2008) for details on this subclass. `CVX` converts (3.8) to a form solvable by a numerical convex optimization solver (e.g., SDPT3 or SeDuMi), and then translates the numerical results back to the original form.

3.2 Necessary and sufficient conditions

The following result characterizes optimality for (3.8), under the assumption that the minimum efficiency inequality constraints can be strictly satisfied.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Theorem 2. *Suppose that there exists* $\mathbf{w} \in \Omega$ *satisfying* $\mathit{Eff}_k(\mathbf{w}) > m_k$*, for all* $k = 2, \ldots, K$*. Let* $\mathbf{w}^*$ *be a feasible solution for problem* (3.8)*. Then,* $\mathbf{w}^*$ *solves problem (3.8) if and only if there exist* $\eta_2, \ldots, \eta_K \geq 0$ *such that*

1. $\eta_k \left(\Phi_k(\mathbf{w}^*) - h_k(m_k)\right) = 0$*, for all* $k = 2, \ldots, K$*, and*
2. $\mathbf{w}^* \in \underset{\mathbf{w}\in\Omega}{\arg\min} \left[\Phi_1(\mathbf{w}) + \sum_{k=2}^{K} \eta_k \Phi_k(\mathbf{w})\right]$.

Theorem 2 is related to the results of Lee (1988), Cook and Wong (1994), and Clyde and Chaloner (1996).

We now discuss how to use the results in Section 2.2 to rewrite Theorem 2. First, suppose that $\phi_1, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Then, it follows from Theorem 1 and Lemma 2 that we can replace Condition 2 in Theorem 2 with

$$d_{\phi_1,f_1}(\mathbf{u}_i, \mathbf{w}^*) + \sum_{k=2}^{K} \eta_k d_{\phi_K,f_K}(\mathbf{u}_i, \mathbf{w}^*) \leq 0, \text{ for all } i = 1, 2, \ldots, N, \quad (3.10)$$

for $d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*)$ defined in (2.5). Table 2 provides formulae for $\nabla\phi_k(\mathbf{M})$.

Otherwise, we must apply Theorem 1 with Lemma 1 and Lemma 3 to rewrite Condition 2 in Theorem 2, as shown in the following example.

Example 1. Suppose that $\phi_1(\mathbf{M}) = -\lambda_{min}(\mathbf{M})$, and $\phi_2, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Then, it follows from Theorem 1 and Lemma 1 that Condition 2 in Theorem 2 holds if and only if there exists

$\mathbf{g} \in \partial\Phi_1(\mathbf{w}^*)$ such that

$$\mathbf{g}^T(\mathbf{w}^* - \mathbf{e}_i) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0, \text{ for all } i = 1, 2, \ldots, N, \quad (3.11)$$

for $d_{\phi,f}(\mathbf{u}_i, \mathbf{w}^*)$ defined in (2.5). Let r^* be the geometric multiplicity of $\lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*))$. It further follows from Lemma 3 that

- Case 1 ($r^* = 1$): Condition 2 in Theorem 2 holds if and only if

$$[(\mathbf{v}^*)^T \mathbf{z}_{f_1}(\mathbf{u}_i, \mathbf{w}^*)]^2 - \lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*)) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0, \forall\, i = 1, \ldots, N, \quad (3.12)$$

where $\mathbf{v}^*$ denotes an arbitrary unit eigenvector of $\lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*))$;

- Case 2 ($r^* > 1$): Condition 2 in Theorem 2 holds if and only if there exist $a_1, \ldots, a_{r^*} \geq 0$ such that $\sum_{j=1}^{r^*} a_j = 1$ and

$$d_{-\lambda_{min}, f_1, \mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0, \ \forall\, i = 1, \ldots, N, \quad (3.13)$$

where $d_{-\lambda_{min}, f, \mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*)$ is defined in (2.6).

3.3 Optimality verification using linear programming

Suppose we have obtained $\mathbf{w}^*$ by solving (3.8) numerically, where $\phi_1, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. We know from Theorem 2 that $\mathbf{w}^*$ is optimal if we can find $\eta_2, \ldots, \eta_K \geq 0$ such that Conditions 1–2 in Theorem 2 are satisfied. However, $\mathbf{w}^*$ is an approximate numerical solution, and is thus unlikely to satisfy Conditions 1–2 exactly. Instead, we check whether $\mathbf{w}^*$ is "close enough" to optimal by searching for $\eta_2, \ldots, \eta_K \geq 0$ such that

$$\eta_k \left(\Phi_k(\mathbf{w}^*) - h_k(m_k)\right) \leq \delta, \quad k = 2, \ldots, K, \tag{3.14}$$

$$-\eta_k \left(\Phi_k(\mathbf{w}^*) - h_k(m_k)\right) \leq \delta, \quad k = 2, \ldots, K, \tag{3.15}$$

$$d_{\phi_1, f_1}(\mathbf{u}_i, \mathbf{w}^*) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq \delta, \quad i = 1, 2, \ldots, N, \tag{3.16}$$

where δ is a small positive constant (e.g., $\delta = 10^{-4}$). Here, (3.14)–(3.15) relax Condition 1 in Theorem 2, (3.16) relaxes Condition 2 in Theorem 2, and δ controls our definition of "close enough" to optimal. Similar ideas appear in single-objective optimal designs (Wong and Zhou, 2019).

We propose solving the following optimization problem:

$$\left\{ \begin{array}{ll} \min\limits_{\boldsymbol{\eta} \in \mathbb{R}^{K-1}} & \mathbf{1}_{K-1}^T \boldsymbol{\eta} \\ \text{subject to: } \boldsymbol{\eta} \geq \mathbf{0}_{K-1}, & \mathbf{B}_1^T \boldsymbol{\eta} \leq \mathbf{b}_1, \quad \mathbf{C}_1 \boldsymbol{\eta} \leq \delta \mathbf{1}_{K-1}, -\mathbf{C}_1 \boldsymbol{\eta} \leq \delta \mathbf{1}_{K-1}. \end{array} \right\} \tag{3.17}$$

where $\leq$ and $\geq$ denote component-wise inequality; $\mathbf{1}_{K-1}$ is a $(K-1)$-vector with every entry equal to one, $\mathbf{B}_1$ is a $(K-1) \times N$ matrix with (k, i)th entry $d_{\phi_{k+1}, f_{k+1}}(\mathbf{u}_i, \mathbf{w}^*)$, where $d_{\phi, f}(\mathbf{u}_i, \mathbf{w}^*)$ is defined in (2.5), for $k = 1, 2, \ldots, K-1$; $\mathbf{b}_1$ is an N-vector with ith entry equal to $\delta - d_{\phi_1, f_1}(\mathbf{u}_i, \mathbf{w}^*)$, and $\mathbf{C}_1 = \text{diag}(\Phi_2(\mathbf{w}^*) - h_2(m_2), \ldots, \Phi_K(\mathbf{w}^*) - h_K(m_K))$.

If we are able to find a solution $\boldsymbol{\eta}^*$ to (3.17), then we know that $\mathbf{w}^*$

and $\boldsymbol{\eta}^*$ jointly satisfy (3.14)–(3.16). This would mean that the conditions in Theorem 2 (approximately) hold, and, thus $\mathbf{w}^*$ is optimal for (3.8). Furthermore, (3.17) is a *linear programming problem* (Luenberger and Ye, 2016): its objective function and constraints are all linear. Thus, we can solve (3.17) by simply applying an off-the-shelf linear programming solver, such as the `linprog` function in the Optimization Toolbox of `MATLAB`.

If one or more of $\phi_1, \ldots, \phi_K$ correspond to E-optimality, then the following example shows that we can still verify the conditions in Theorem 2 using linear programming.

Example 1 (continued). Consider Example 1 in Section 3.2, where $\phi_1(\mathbf{M}) = -\lambda_{min}(\mathbf{M})$ and $\phi_2, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Recall that we defined r^* to be the geometric multiplicity of $\lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*))$. We previously showed that if $r^* = 1$, then Condition 2 in Theorem 3 is equivalent to (3.12), which defines a set of N linear equalities in $\eta_2, \ldots, \eta_K$. Thus, we can minimize $\sum_{k=2}^{K} \eta_k$ subject to the $2(K-1)$ linear inequalities defined in (3.14)–(3.15) and the following relaxed version of (3.12):

$$[(\mathbf{v}^*)^T \mathbf{z}_{f_1}(\mathbf{u}_i, \mathbf{w}^*)]^2 - \lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*)) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq \delta, \forall\, i = 1, \ldots, N.$$

If we can find a solution to this linear programming problem, then we know that $\mathbf{w}^*$ is an optimal solution.

We also showed in Section 3.2 that if $r^* > 1$, then Condition 2 holds if and only if there exist $a_1, \ldots, a_{r^*} \geq 0$ and $\eta_2, \ldots, \eta_K \geq 0$ such that $\sum_{j=1}^{r^*} a_j = 1$ and (3.13) holds. Thus, Theorem 2 says that $\mathbf{w}^*$ is optimal for (3.8) if and only if there exists $a_1, \ldots, a_{r^*} \geq 0$ and $\eta_2, \ldots, \eta_K \geq 0$ such that $\eta_k(\Phi_k(\mathbf{w}^*) - h_k(m_k)) = 0$ for all $k = 2, \ldots, K$ and (3.13) holds. Observe that (3.13) defines N linear inequalities in $\eta_2, \ldots, \eta_K$ and in $a_1, \ldots, a_{r^*}$. Thus, we can minimize $\sum_{k=2}^{K} \eta_k + \sum_{j=1}^{r^*} a_j$ subject to the $2(K-1)$ linear inequalities defined in (3.14)–(3.15) and the following relaxation of (3.13),

$$d_{-\lambda_{min}, f_1, \mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq \delta, \ \forall \ i = 1, \ldots, N.$$

Once again, if we are able to find a solution to this linear programming problem, then we know that $\mathbf{w}^*$ is an optimal solution.

4. Maximin optimal designs

Suppose that an experimenter requires a design that yields reasonable efficiency for *all* K single-objective optimality criteria. We formulate this *maximin* design problem as

$$\left\{ \begin{array}{ll} \max\limits_{\mathbf{w} \in \mathbb{R}^N} & \min\{\text{Eff}_1(\mathbf{w}), \ldots, \text{Eff}_K(\mathbf{w})\} \\ \text{subject to:} & \sum\limits_{i=1}^{N} w_i = 1, \ w_i \geq 0, \ i = 1, 2, \ldots, N. \end{array} \right\}. \tag{4.18}$$

4.1 Convex optimization problem

Problem (4.18) is hard to solve directly, because the objective function involves a minimization. However, we can formulate (4.18) equivalently as

$$\left\{\begin{array}{l} \max\limits_{\mathbf{w}\in\mathbb{R}^N, t\in\mathbb{R}} \quad 1/t \\ \text{subject to: } \mathrm{Eff}_k(\mathbf{w}) \geq 1/t,\ k=1,\dots,K, \\ \qquad\qquad t\geq 0,\ w_i\geq 0, i=1,\dots,N,\ \sum_{i=1}^N w_i = 1 \end{array}\right\}. \tag{4.19}$$

This formulation eliminates the minimization from the objective function by introducing an additional optimization variable (t). Furthermore, when $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$, with $\phi_1,\dots,\phi_K$ chosen from the convex functions in Table 1, (4.19) is equivalent to

$$\left\{\begin{array}{l} \min\limits_{\mathbf{w}\in\mathbb{R}^N, t\in\mathbb{R}} \quad t \\ \text{subject to: } \Phi_k(\mathbf{w}) \leq h_k(1/t),\ k=1,\dots,K, \\ \qquad\qquad t\geq 0, \sum_{i=1}^N w_i = 1,\ w_i\geq 0, i=1,\dots,N. \end{array}\right\}, \tag{4.20}$$

where $h_k(\cdot)$ is defined in (3.9). This is a convex optimization problem, because $h_k(1/t)$ is a concave function of t and $\Phi_k(\mathbf{w})$ is a convex function of $\mathbf{w}$. Furthermore, we can solve (4.19) using `CVX`, because it fits into the `CVX` modeling framework described in Grant and Boyd (2008). Note that our formulation is more general than that of Wong and Zhou (2023), because we allow the user to select any combination of the criteria in Table 1.

4.2 Necessary and sufficient conditions for optimality

The following result characterizes optimality for (4.20).

Theorem 3. *Suppose that* $(\mathbf{w}^*, t^*)$ *are feasible for problem* (4.20)*. Then,* $(\mathbf{w}^*, t^*)$ *solves (4.20) if and only if there exist* $\eta_1, \ldots, \eta_K \geq 0$ *satisfying*

1. $\sum_{k=1}^{K} \eta_k \left(\frac{d}{d_t} h_k(1/t) \Big|_{t=t^*} \right) = 1.$

2. $\eta_k(\Phi_k(\mathbf{w}^*) - h_k(1/t^*)) = 0$*, for all* $k = 1, 2, \ldots, K$*.*

3. $\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\Omega} \left\{ \sum_{k=1}^{K} \eta_k \Phi_k(\mathbf{w}) \right\}.$

We now show how to use the results in Section 2.2 to rewrite Theorem 3. First, suppose that $\phi_1, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Then, it follows from Theorem 1 and Lemma 1 that we can replace Condition 3 in Theorem 3 with

$$\sum_{k=1}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0 \text{ for all } i = 1, 2, \ldots, N, \tag{4.21}$$

where $d_k(\mathbf{u}_i, \mathbf{w}^*)$ is given in (2.5), and the formulae for $\nabla\phi_k(\mathbf{M})$ are given in Table 2. Otherwise, we need to apply Theorem 1 with Lemma 1 and Lemma 3 to rewrite Condition 3, as shown in the following example.

Example 2. Suppose that $\phi_1(\mathbf{M}) = -\lambda_{min}(\mathbf{M})$, and $\phi_2, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Then, it follows from Theorem 1 and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Lemma 1 that Condition 3 in Theorem 3 holds if and only if there exists $\mathbf{g} \in \partial\Phi_1(\mathbf{w}^*)$ such that

$$\eta_1 \mathbf{g}^T (\mathbf{w}^* - \mathbf{e}_i) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0, \ \forall\, i = 1, 2, \ldots, N. \tag{4.22}$$

Let r^* be the geometric multiplicity of $\lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*))$. Then, based on Lemma 3, we can consider two cases:

- Case 1 ($r^* = 1$): Condition 3 in Theorem 3 holds if and only if

  $$\eta_1[(\mathbf{v}^*)^T \mathbf{z}_{f_1}(\mathbf{u}_i, \mathbf{w}^*)]^2 - \eta_1 \lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*)) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq \delta, \forall\, i = 1, \ldots, N, \tag{4.23}$$

  where $\mathbf{v}^*$ denotes an arbitrary unit eigenvector corresponding to $\lambda_{min}(\mathcal{I}_{f_1}(\mathbf{w}^*))$.

- Case 2 ($r^* > 1$): Condition 3 in Theorem 3 holds if and only if there exist $a_1, \ldots, a_{r^*} \geq 0$ such that $\sum_{j=1}^{r^*} a_j = 1$ and

  $$\eta_1 d_{-\lambda_{min}, f_1, \mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*) + \sum_{k=2}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq 0 \text{ for all } i = 1, 2, \ldots, N,$$

  where $d_{-\lambda_{min}, f_1, \mathbf{a}}(\mathbf{u}_i, \mathbf{w}^*)$ is defined in (2.6).

4.3 Optimality verification using linear programming

Suppose we obtain a candidate solution $(\mathbf{w}^*, t^*)$ by solving (4.20) numerically (e.g., using `CVX`), where $\phi_1, \ldots, \phi_K$ all correspond to D-, A-, c-, or L-optimality. Based on the results in Section 4.2, we need to find $\eta_1, \ldots, \eta_K \geq 0$ such that

$$\sum_{k=1}^{K} \eta_k \left(\frac{d}{dt} h_k(1/t)\Big|_{t=t^*} \right) = 1, \tag{4.24}$$

$$\eta_k(\Phi_k(\mathbf{w}^*) - h_k(1/t^*)) \leq \delta, \quad 1 = 2, \ldots, K, \tag{4.25}$$

$$-\eta_k(\Phi_k(\mathbf{w}^*) - h_k(1/t^*)) \leq \delta, \quad 1 = 2, \ldots, K, \tag{4.26}$$

$$\sum_{k=1}^{K} \eta_k d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*) \leq \delta, \quad i = 1, 2, \ldots, N, \tag{4.27}$$

where δ is a small positive constant. Here, we have relaxed Conditions 2 and 3 in Theorem 3, because $\mathbf{w}^*$ is an approximate solution, along the lines of the discussion in Section 3.3. We achieve this goal by solving the following linear programming problem using the `linprog` function in `MATLAB`:

$$\left\{ \begin{array}{l} \min_{\boldsymbol{\eta} \in \mathbb{R}^K} \quad \mathbf{1}_K^T \boldsymbol{\eta} \\ \text{subject to: } \boldsymbol{\eta} \geq \mathbf{0}_K, \ \mathbf{b}_2^T \boldsymbol{\eta} = 1, \ \mathbf{B}_2^T \boldsymbol{\eta} \leq \delta \mathbf{1}_N, \mathbf{C}_2 \boldsymbol{\eta} \leq \delta \mathbf{1}_K, -\mathbf{C}_2 \boldsymbol{\eta} \leq \delta \mathbf{1}_K. \end{array} \right\} \tag{4.28}$$

where $\mathbf{b}_2$ is a K-vector with kth entry equal to $\left(\frac{d}{dt} h_k(1/t)\Big|_{t=t^*} \right)$, for h_k defined in (3.9), $\mathbf{B}_2$ is a $K \times N$ matrix with (k, i)th entry equal to $d_{\phi_k, f_k}(\mathbf{u}_i, \mathbf{w}^*)$, for $d_{\phi, f}(\mathbf{u}_i, \mathbf{w}^*)$ defined in (2.5), and $\mathbf{C}_2 = \text{diag}(\Phi_1(\mathbf{w}^*) - h_1(m_1), \ldots, \Phi_K(\mathbf{w}^*) - h_K(m_K))$.

When one or more of $\phi_1, \ldots, \phi_K$ correspond to E-optimality, we can still rewrite Condition 3 in Theorem 3 as a set of linear inequalities; see Example 2. Thus, we can still verify the conditions in Theorem 3 using

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

linear programming. We omit the details, because the linear programming problem is similar to that in Example 1 in Section 3.3.

5. Applications

In all three of the following applications, we set $\delta = 10^{-4}$ when verifying optimality using linear programming, as described in Sections 3.3 and 4.3. Any choice of δ larger than 10^{-6} yields the same results. All computations are performed on a 2021 M1 Macbook Pro with 10 cores and 16 GB memory. We provide the `MATLAB` code to reproduce all numerical results at `https://github.com/lucylgao/multi-objective-paper-code-2022`.

Application 1. Consider a four-parameter compartment model of the form (1.1), with $p = 1$, $q = 4$, $f(x, \boldsymbol{\theta}) = \theta_1 e^{-\theta_2 x} + \theta_3 e^{-\theta_4 x}$, and $S = [0, 15]$, where the response y_i represents the concentration level of a drug in compartments and x denotes the sampling time. This model has been studied in optimal designs for various optimality criteria, including multi-objective criteria (Huang and Wong, 1998; Cheng and Yang, 2019).

We seek efficiency-constrained optimal designs that solve (3.7) with $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_f(\mathbf{w})]$, for $k = 1, 2, 3$. As in Cheng and Yang (2019), we let ϕ_1 correspond to L-optimality with $\mathbf{L} = \text{diag}\left(\frac{1}{\theta_1}, \frac{1}{\theta_2}, \frac{1}{\theta_3}, \frac{1}{\theta_4}\right)$, ϕ_2 correspond to D-optimality, and ϕ_3 correspond to L-optimality with $\mathbf{L} =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

$\left[\int_2^{10} \mathbf{z}_f(x,\boldsymbol{\theta}^*)\mathbf{z}_f^T(x,\boldsymbol{\theta}^*)dx\right]^{1/2}$, where $\mathbf{z}_f(x,\boldsymbol{\theta}) = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x}, e^{-\theta_4 x}, -\theta_3 x e^{-\theta_4 x})^\top$ and $\boldsymbol{\theta}^* = (5.25, 1.34, 1.75, 0.13)^\top$. We discretize the continuous design space S to form S_N with $u_i = 15(i-1)/(N-1)$, for $i = 1, \ldots, N$.

First, we find the single-objective optimal designs by solving $\min_{\mathbf{w}\in\Omega} \Phi_k(\mathbf{w})$ for each $k = 1, 2, 3$ using `CVX`. Then, we solve (3.7) with $m_2 = 0.9$, $m_3 = 0.8$, and $N = 501$ using `CVX`, and denote the solution as $\mathbf{w}^{*m}$. We report the single-objective optimal designs and $\mathbf{w}^{*m}$ in Table 3. The efficiencies at $\mathbf{w}^{*m}$ are close to those reported in Cheng and Yang (2019).

We then verify the conditions for optimality in Theorem 2 for $\mathbf{w}^{*m}$ by using the `linprog` `MATLAB` function to solve (3.17) with $\delta = 10^{-4}$, as described in Section 3.3. Solving (3.17) yields $\eta_2^* = 36.4870$ and $\eta_3^* = 5.0767$. Because we obtain a solution, we know that $\mathbf{w}^{*m}$ is the efficiency-constrained optimal design (Theorem 2). Figure 1 shows that $d_{\phi_1,f}(u_i,\mathbf{w}^{*m}) + \sum_{k=2}^{3} \eta_k^* d_{\phi_k,f}(u_i,\mathbf{w}^{*m}) \leq \delta$, for all $i = 1, 2, \ldots, N$, showing visually that Condition 2 in Theorem 2 is satisfied for $\mathbf{w}^{*m}$, η_2^*, and η_3^*. Figure 1 also shows that $d_{\phi_k,f}(u_i,\mathbf{w}^{*m})$ is not uniformly non-negative for all $k = 1, 2, 3$. Thus, $\mathbf{w}^{*m}$ is not the single-objective optimal design that minimizes Φ_1, Φ_2, or Φ_3 (Theorem 1 and Lemma 2).

Next, we examine the results of varying m_2 and m_3; see Table 4. For $m_2 = 0.90$ and $m_3 = 0.70$, we find that $\eta_3^* = 0$, because $\Phi_3(\mathbf{w}^{*m}) < h_3(m_3)$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Table 3: For Application 1, single-objective optimal designs and the efficiency-constrained optimal design with $m_2 = 0.9$ and $m_3 = 0.8$.

ϕ_1-optimal points (weights)	ϕ_2-optimal points (weights)	ϕ_3-optimal points (weights)	efficiency-constrained points (weights)
0 (0.0591)	0 (0.2500)	0 (0.1339)	0 (0.1339)
0.6300 (0.1315)	0.6600 (0.2500)	0.9600 (0.0663)	0.6600 (0.1513)
2.9400 (0.3126)	2.8800 (0.2500)	3.300 (0.4502)	3.0300 (0.2481)
13.2900 (0.4968)	11.0100 (0.2441)	9.7500 (0.2231)	3.0600 (0.0942)
	11.0400 (0.0059)	9.7800 (0.2502)	10.8300 (0.0807)
			10.8600 (0.2918)

and η_2^*, η_3^* satisfy Condition 1 in Theorem 2. Similarly, for $m_2 = 0.70$ and $m_3 = 0.70$, we find that $\eta_2^* = \eta_3^* = 0$, because $\Phi_k(\mathbf{w}^{*m}) < h_k(m_k)$ (i.e., $\text{Eff}_k(\mathbf{w}^{*m}) > m_k$) for $k = 2, 3$. This implies that the multi-objective optimal design $\mathbf{w}^{*m}$ is also a single-objective optimal design that maximizes Φ_1 (Theorem 1 and Lemma 2). For $m_2 = 0.90$ and $m_3 = 0.90$, there is no feasible solution.

Computing the optimal designs for $(m_2, m_3) = (0.9, 0.8)$ took 19.5,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

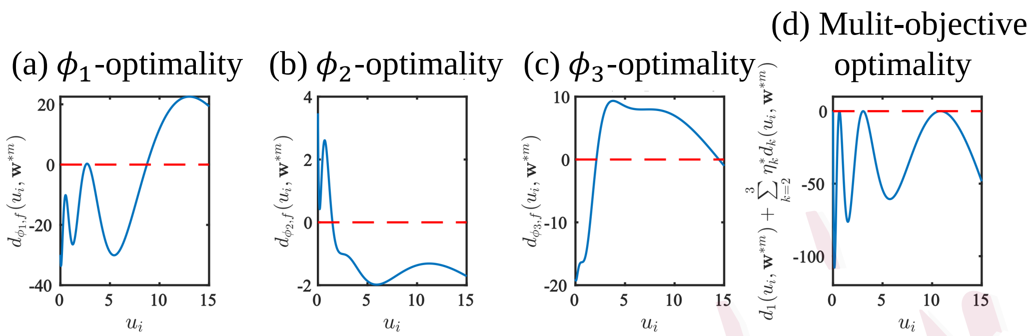

Figure 1: For $(m_2, m_3) = (0.9, 0.8)$ in Application 1, panels (a)–(c) display plots of $d_{\phi_k,f}(u_i, \mathbf{w}^{*m})$ for $k = 1, 2, 3$, and panel (d) displays $d_{\phi_1,f}(u_i, \mathbf{w}^{*m})+ \sum_{k=2}^{3} \eta_k^* d_{\phi_k,f}(u_i, \mathbf{w}^{*m})$. In panel (d), the dashed line represents the horizontal line $y = \delta$, for $\delta = 10^{-4}$.

Table 4: For Application 1, efficiencies and η_2^* and η_3^*, for various (m_2, m_3).

Case (m_2, m_3)	(0.90, 0.80)	(0.90, 0.70)	(0.70, 0.70)	(0.90, 0.90)
$\eta_2^*,\ \eta_3^*$	36.4870, 5.0767	7.2923, 0	0, 0	NA
$\text{Eff}_1(\mathbf{w}^{*m})$	0.8694	0.9360	1.000	NA
$\text{Eff}_2(\mathbf{w}^{*m}),\ \text{Eff}_3(\mathbf{w}^{*m})$	0.9000, 0.8000	0.9000, 0.7035	0.7317, 0.7746	NA

25.6, and 36.2 seconds for $N = 101, 501, 1001$, respectively. Verifying the optimality of the efficiency constrained design took less than a second.

Application 2. Several dose response models are popular in clinical dose finding studies. We consider the four competing regression models of the form (1.1) from Bretz et al. (2010) to construct maximin optimal designs, and to verify the necessary and sufficient conditions in Theorem 3 for the optimal designs. The four models are: (i) linear model: $f_1(x, \boldsymbol{\theta}) = \theta_{11} + \theta_{12}x$; (ii) Emax I model: $f_2(x, \boldsymbol{\theta}) = \theta_{21} + \theta_{22}x/(\theta_{23} + x)$; (iii) Emax II model: $f_3(x, \boldsymbol{\theta}) = \theta_{31} + \theta_{32}x/(\theta_{33} + x)$; and (iv) Logistic model: $f_4(x, \boldsymbol{\theta}) = \theta_{41} + \theta_{42}/\left(1 + \exp[(\theta_{43} - x)/\theta_{44}]\right)$. In all models, $x \in [0, 500]$ (μg) is the dose level. Let S_N contain $N = 501$ equally spaced grid points in $[0, 500]$. As in Bretz et al. (2010), we assume that the true parameter values for the Emax I , Emax II, and logistic models are, respectively, (60, 294, 25), (60, 340, 107.14), and (49.62, 290.51, 150, 45.51). (The information matrix for the linear model does not depend on its true parameter values.)

We let $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$, with ϕ_k corresponding to D-optimality (defined in Table 1), for all $k = 1, \ldots, 4$, and then solve problem (4.20) using `CVX`, denoting the solution as $\mathbf{w}^{*mm}$. The single-objective and maximin D-optimal designs on S_N are given in Table 5. We find that $t^* = 1.1712$, $\text{Eff}_k(\mathbf{w}^{*mm}) = 0.8538$, for $k = 1, 2, 4$, and $\text{Eff}_3(\mathbf{w}^{*mm}) = 0.8547$. Solving problem (4.28) using the `MATLAB` function `linprog` yielded $\eta_1^* = 0.1983, \eta_2^* = 0.1291, \eta_3^* = 0$, and $\eta_4^* = 0.0968$. Because we obtain a solution, we know that $\mathbf{w}^{*mm}$ is the maximin D-optimal design (Theorem 3).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Figure 2 displays plots of $d_{\phi_k,f_k}(u_i,\mathbf{w}^{*mm})$ for $k=1,\ldots,4$ and $\sum_{k=1}^{4}\eta_k^* d_{\phi_k,f_k}(u_j,\mathbf{w}^{*mm})$. Figure 2(e) confirms that Condition 3 in Theorem 3 is satisfied. Figure 2(a)–(d) show that $\mathbf{w}^{*mm}$ is not the single-objective D-optimal design for any of the four models (Theorem 1 and Lemma 2).

Table 5: Optimal designs for Application 2.

linear model	Emax I	Emax II	logistic	maximin
points (weights)	points (weights)	points (weights)	points (weights)	points (weights)
0 (0.5000)	0 (0.3333)	0 (0.3333)	0 (0.2500)	0 (0.2406)
500 (0.5000)	22 (0.3333)	75 (0.3333)	114 (0.2500)	19 (0.1806)
	500 (0.3333)	500 (0.3333)	204 (0.1316)	112 (0.1314)
			205 (0.2500)	204 (0.1070)
			500 (0.2500)	205 (0.0178)
				500 (0.3225)

Computing the optimal designs took 18.2, 31.0, and 48.7 seconds for $N=101,501,1001$, respectively. Verifying the optimality of the maximin design took less than a second.

Application 3: Consider the linear model of the form (1.1) with $p=2$, $q=5$, $f(\mathbf{x};\boldsymbol{\theta})=\theta_1+x_1\theta_2+x_2\theta_3+x_1x_2\theta_4+x_2^2\theta_5$, and $S=\{0,1\}\times[-1,1]$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

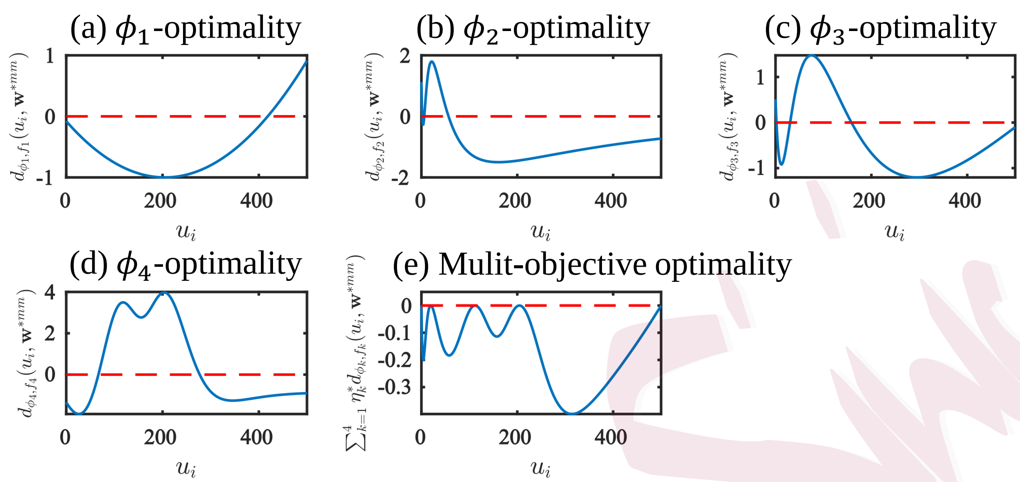

Figure 2: For Application 2, we display plots of (a)–(d) $d_{\phi_k,f_k}(u_i, \mathbf{w}^{*mm})$ for $k = 1, 2, 3, 4$, and (e) $\sum_{k=1}^{4} \eta_k^* d_{\phi_k,f_k}(u_i, \mathbf{w}^{*mm})$. In panel (e), the dashed line represents the horizontal line $y = \delta$, for $\delta = 10^{-4}$.

We let $S_{N/2}^1$ contain 201 equally spaced points on $[-1, 1]$, and discretize the design space S to form $S_N = \{0, 1\} \times S_{N/2}^1$, for $N = 402$. Here, the information matrix does not depend on the true parameter values.

We let $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_f(\mathbf{w})]$, for $k = 1, 2, 3$, with ϕ_1 corresponding to A-optimality, ϕ_2 corresponding to E-optimality, and ϕ_3 corresponding to c-optimality with $\mathbf{c} = (0, 0, 0, 1, 0)^T$. Then we solve problem (4.20) as described in Section 4.1 to obtain the maximin optimal design, and denote the solution as $\mathbf{w}_{lm}^{*mm}$. The single-objective and maximin optimal designs on S_N

doi:10.5705/ss.202022.0328

are given in Table 6. We find that $t^* = 1.2979$, $\text{Eff}_1(\mathbf{w}_{lm}^{*mm}) = 0.9298$, and $\text{Eff}_1(\mathbf{w}_{lm}^{*mm}) = 0.7705$, for $k = 2, 3$. In this case, the geometric multiplicity of $\lambda_{min}(\mathcal{I}_f(\mathbf{w}_{lm}^{*mm}))$ is equal to one. Thus, to verify the optimality of $\mathbf{w}_{lm}^{*mm}$, we use the `linprog` `MATLAB` function to minimize $\sum_{k=1}^{3} \eta_k$, subject to the linear inequality constraints in (4.23) and the linear equality and inequality constraints in (4.24)–(4.26). We find a solution at $\eta_1^* = 0$, $\eta_2^* = 0.2445$, and $\eta_3^* = 0.0151$. Therefore, the maximin design is optimal.

Table 6: Optimal designs for Application 3.

A-optimal	E-optimal	c-optimal	maximin
points [weights]	points [weights]	points [weights]	points [weights]
(0, -1) [0.1859]	(0, -1) [0.2069]	(0, -1) [0.2500]	(0, -1) [0.1926]
(1, -1) [0.1399]	(1, -1) [0.1379]	(1, -1) [0.2500]	(1, -1) [0.1926]
(0, 0) [0.2287]	(0, 0) [0.2414]	(0, 1) [0.2500]	(0, 0) [0.1679]
(1, 0) [0.1197]	(0, 1) [0.0690]	(1, 1) [0.2500]	(1, 0) [0.0616]
(0, 1) [0.1859]	(1, 1) [0.2069]		(0, 1) [0.1926]
(1, 1) [0.1399]			(1, 1) [0.1926]

Computing the optimal designs took 11.2, 14.5, and 23.2 seconds for $N/2 = 101, 201, 401$, respectively. Verifying the optimality of the maximin design took less than a second.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

6. Conclusion

In this paper, we have shown how to solve multi-objective optimal design problems on a discrete design space using convex optimization, and how to verify the optimality of the designs using linear programming. Our approach can be applied to efficiency-constrained or maximin optimal design problems that combine any of the single-objective criteria shown in Table 1.

The multi-objective optimal design setting offers a natural opportunity to gain robustness against parameter and/or model misspecification, because we can include objective functions formulated from a range of guesses for $\boldsymbol{\theta}^*$ and/or from the information matrices of multiple models. A sequential multi-objective optimal design setting may offer opportunities for further robustness, because it would enable us to select design points and weights in stages, and then use the data from each stage to inform the choice of parameters and/or models used in the objective functions for the next stage. This may provide a fruitful avenue for future work.

We were able to achieve the results and algorithms presented here because the inverse of the asymptotic covariance matrix of the ordinary least squares estimator under model (1.1) is a linear function of $\mathbf{w}^*$; see equation (2.2). It would be straightforward to extend the results and algorithms to other models and estimators that have a similar property. For example, we could allow the vector of errors in (1.1) to be heteroskedastic or have a block diagonal covariance structure, and use the generalized least squares estimator. Another example is generalized linear models with a canonical link function, where we estimate $\boldsymbol{\theta}$ using the maximum likelihood estimator.

Necessary and sufficient conditions for optimality that involve a set of unknown parameters appear in contexts outside of the multi-objective design problems we consider here, for example, in any single-objective optimal design problem involving a convex, but non-differentiable objective function (e.g. single-objective E-optimality). When the conditions define linear equalities and inequalities in these unknown parameters, we can verify them using linear programming, as discussed here.

A limitation of our results and algorithms is the assumption of a discrete design space. An important direction for future work is to develop results and algorithms under a continuous design space.

Acknowledgments

This research work was partially supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada.

References

Bretz, F., H. Dette, and J. C. Pinheiro (2010). Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine 29*, 731–742.

Cheng, Q. and M. Yang (2019). On multiple-objective optimal designs. *Journal of Statistical Planning and Inference 200*, 87–101.

Clarke, F. H. (1983). *Optimization and nonsmooth analysis*. New York: Wiley.

Clyde, M. and K. Chaloner (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *Journal of the American Statistical Association 91*, 1236–1244.

Cook, R. D. and W. K. Wong (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association 89*, 687–692.

Dette, H., L. M. Haines, and L. A. Imhof (2007). Maximin and bayesian optimal designs for regression models. *Statistica Sinica 17*, 463–480.

Gao, L. L. and J. Zhou (2017). D-optimal designs based on the second-order least squares estimator. *Statistical Papers 58*, 77–94.

Grant, M. and S. Boyd (2014). CVX: MATLAB software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

Grant, M. C. and S. P. Boyd (2008). Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and M. Kimura (Eds.), *Recent Advances in Learning and Control* *(tribute to M. Vidyasagar)*, pp. 95–110. Springer.

Huang, Y.-C. and W.-K. Wong (1998). Sequential construction of multiple-objective optimal designs. *Biometrics 54*, 1388–1397.

Imhof, L. and W. K. Wong (2000). A graphical method for finding maximin efficiency designs. *Biometrics 56*, 113–117.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics 2*, 849–879.

Lee, C. M.-S. (1988). Constrained optimal designs. *Journal of Statistical Planning and Inference 18*, 377–389.

Lewis, A. (1999). Nonsmooth analysis of eigenvalues. *Mathematical Programming 84*, 1–24.

Luenberger, D. G. and Y. Ye (2016). *Linear and nonlinear programming (4th ed.)*. Cham: Springer.

Mordukhovich, B. S. and N. M. Nam (2013). *An easy path to convex analysis and applications.* Synthesis Lectures on Mathematics & Statistics. Cham: Springer.

Müller, C. H. and A. Pázman (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika 48*, 1–19.

Petersen, K. B. and M. S. Pedersen (2012). The matrix cookbook. Version 20121115.

Wong, W. K. (1999). Recent advances in multiple-objective design strategies. *Statistica Neerlandica 53*, 257–276.

Wong, W. K. and J. Zhou (2019). CVX-based algorithms for constructing various optimal regression designs. *Canadian Journal of Statistics 47*, 374–391.

Wong, W. K. and J. Zhou (2023). Using cvx to construct optimal designs for biomedical studies with multiple objectives. *Journal of Computational and Graphical Statistics 32*, 744–753.

Appendix

A. General convex optimization theory

In this section, we review general results on optimality conditions in convex optimization.

Proposition A1. *Let $\Phi : \mathbb{R}^N \mapsto \mathbb{R}$ be a convex function and C be a closed convex set. Then, $\mathbf{w}^* \in \arg\min_{\mathbf{w}\in C} \Phi(\mathbf{w})$ if and only if there exists $\mathbf{g} \in \partial\Phi(\mathbf{w}^*)$ such that*

$$\mathbf{g}^T(\mathbf{w} - \mathbf{w}^*) \geq 0, \quad \textit{for all } \mathbf{w} \in C, \tag{A.29}$$

where $\partial\Phi(\mathbf{w}^*) \equiv \{\mathbf{g} \in \mathbb{R}^N : \Phi(\mathbf{w}) - \Phi(\mathbf{w}^*) \geq \mathbf{g}^T(\mathbf{w} - \mathbf{w}^*) \ \forall \ \mathbf{w} \in \mathbb{R}^N\}$.

Proposition A1 is a direct consequence of Theorem 4.14 of Mordukhovich and Nam (2013). The following result characterizes optimality for constrained convex optimization problems.

Proposition A2. *Define the following convex optimization problem:*

$$\min_{\mathbf{w}\in C} \quad \Phi_0(\mathbf{w}) \qquad \textit{subject to: } \Phi_l(\mathbf{w}) \leq 0, \ l = 1, \ldots, L, \tag{A.30}$$

where $\Phi_0, \ldots, \Phi_L : \mathbb{R}^N \to \mathbb{R}$ are convex functions and C is a closed convex set. Suppose that Slater's condition holds, i.e. there exists $\mathbf{w}' \in C$ such that $\Phi_l(\mathbf{w}') < 0$ for all $l = 1, \ldots, L$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

Then, a feasible solution $\mathbf{w}^*$ *of* (A.30) *solves* (A.30) *if and only if there exists* $\eta_1, \ldots, \eta_L \geq 0$ *such that* $\eta_l \Phi_l(\mathbf{w}^*) = 0$ *for all* $l = 1, 2, \ldots, L$ *and* $\mathbf{w}^* \in \arg\min_{w \in C} \left\{ \Phi_0(\mathbf{w}) + \sum_{l=1}^{L} \eta_k \Phi_l(\mathbf{w}) \right\}$.

Proof. We assumed that Slater's condition holds. Thus, Theorem 4.18 of Mordukhovich and Nam (2013) says that $\mathbf{w}^*$ is optimal for (A.30) if and only if there exist multipliers $\eta_1, \ldots, \eta_L \geq 0$ such that $\eta_l \Phi_l(\mathbf{w}^*) = 0$ for all $l = 1, 2, \ldots, L$ and

$$0 \in \partial \Phi_0(\mathbf{w}^*) + \sum_{l=1}^{L} \eta_l \partial \Phi_l(\mathbf{w}^*) + \left\{ \mathbf{g} \in \mathbb{R}^N : \mathbf{g}^T \mathbf{w}^* \geq \mathbf{g}^T \mathbf{w} \ \forall \ \mathbf{w} \in C \right\}. \tag{A.31}$$

Thus, by Lemma 1, (A.31) is equivalent to:

$$\exists \mathbf{g} \in \partial \left(\Phi_0 + \sum_{l=1}^{L} \eta_l \Phi_l \right) (\mathbf{w}^*) \ \text{ such that } \ \mathbf{g}^T (\mathbf{w} - \mathbf{w}^*) \geq 0 \quad \text{for all } \mathbf{w} \in C. \tag{A.32}$$

Finally, it follows from Proposition A1 that (A.32) holds if and only if $\mathbf{w}^* \in \arg\min_{\mathbf{w} \in C} \left\{ \Phi_0(\mathbf{w}) + \sum_{l=1}^{L} \eta_l \Phi_l(\mathbf{w}) \right\}$.

□

B. Proof of Theorem 1

Let $\mathbf{w}^* \in \Omega$. Since $\Phi(\mathbf{w}) = \phi[\mathcal{I}_f(\mathbf{w})]$ is convex and Ω is a closed convex set, from Proposition A1, $\mathbf{w}^* \in \arg\min_{\mathbf{w} \in \Omega} \Phi(\mathbf{w})$ if and only if

$$\exists \mathbf{g} \in \partial \Phi(\mathbf{w}^*) \text{ such that } \mathbf{g}^T (\mathbf{w} - \mathbf{w}^*) \geq 0, \text{ for all } \mathbf{w} \in \Omega. \tag{B.33}$$

Suppose that (B.33) holds. Since $\mathbf{e}_i \in \Omega$ for all $i = 1, 2, \ldots, N$, we have that (2.3) holds. Now suppose that (2.3) holds. Then, because all $\mathbf{w} \in \Omega$ have $w_i \geq 0$ for all $i = 1, 2, \ldots, N$, (2.3) implies that there exists $\mathbf{g} \in \partial \Phi(\mathbf{w}^*)$ such that $\sum_{i=1}^{N} w_i \mathbf{g}^T (\mathbf{e}_i - \mathbf{w}^*) \geq 0$, for all $\mathbf{w} \in \Omega$. Observing that $\mathbf{g}^T (\mathbf{w} - \mathbf{w}^*) = \sum_{i=1}^{N} w_i \mathbf{g}^T (\mathbf{e}_i - \mathbf{w}^*)$ completes the proof. □

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

C. Proof of Lemma 3

We will start by establishing properties of $(-\Phi)(\mathbf{w}) = \lambda_{min}(\mathcal{I}_f(\mathbf{w}))$, as this allows us to take advantage of existing theoretical results about the minimum eigenvalue function $\lambda_{min}(\mathbf{M})$.

The function $(-\Phi)(\mathbf{w})$ is not differentiable at every point in $\mathbb{R}^N$. Furthermore, the notion of a subdifferential does not apply to $(-\Phi)(\mathbf{w})$, as $(-\Phi)(\mathbf{w})$ is a *concave* rather than a convex function. However, it has a *Clarke subdifferential* (Clarke, 1983), which generalizes the notion of the gradient to the class of locally Lipschitz continuous functions. The Clarke subdifferential of a locally Lipschitz continuous function $h(\mathbf{w})$ on $\mathbb{R}^N$ is defined as:

$$\partial^C h(\mathbf{w}) = \text{co}\left(\left\{\mathbf{v} \in \mathbb{R}^N : \exists\ \{\mathbf{w}_k\}_{k=1}^{\infty} \text{ s.t.} \lim_{k\to\infty} \mathbf{w}_k \text{ exists}, \nabla h(\mathbf{w}_k) \text{ exists, and } \lim_{k\to\infty} \nabla h(\mathbf{w}_k) = \mathbf{v}\right\}\right),$$

where $\text{co}(S)$ is the convex hull of the set S, i.e. the intersection of all convex sets containing S.

We can characterize the Clarke subdifferential $\partial^C(-\Phi)(\mathbf{w})$ by observing that $(-\Phi)(\mathbf{w})$ is the composition of the non-differentiable concave function λ_{min} with the linear function $\mathcal{I}_f$. Thus, it follows from the Clarke subdifferential chain rule (Theorem 2.3.10 in Clarke 1983) that $\mathbf{g}_c \in \partial^c(-\Phi)(\mathbf{w}^*)$ if and only if there exists $\mathbf{M} \in \partial^c \lambda_{\min}(\mathcal{I}_f(\mathbf{w}^*))$ such that for any $\mathbf{w} \in \mathbb{R}^N$,

$$\mathbf{g}_c^T \mathbf{w} = \text{trace}\left(\mathbf{M}\left(\sum_{i=1}^{N} w_i \mathbf{z}_f(\mathbf{u}_i)\mathbf{z}_f^T(\mathbf{u}_i)\right)\right). \tag{C.34}$$

Furthermore, Corollary 10 of Lewis (1999) says that

$$\partial^c \lambda_{min}(\mathcal{I}_f(\mathbf{w}^*)) = \left\{\sum_{j=1}^{r^*} a_j \mathbf{v}_j^* [\mathbf{v}_j^*]^T : a_j \geq 0, \sum_{j=1}^{r^*} a_j = 1\right\}, \tag{C.35}$$

recalling that in the statement of Lemma 3 we defined $\mathbf{v}_1, \ldots, \mathbf{v}_{r^*}$ to be an arbitrary set of r^* linearly independent unit eigenvectors associated with $\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*))$ where r^* is the geometric

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0328

multiplicity of $\lambda_{min}(\mathcal{I}_f(\mathbf{w}^*))$. It follows from the definition of $\mathcal{I}_f(\mathbf{w}^*)$ in (2.2) and (C.34)–(C.35) that if $\mathbf{g}_c \in \partial^c(-\Phi)(\mathbf{w}^*)$, then there exists $a_1, \dots, a_{r^*} \geq 0$ such that $\sum_{j=1}^{r^*} a_j = 1$ and $\mathbf{g}_c^T(\mathbf{e}_i - \mathbf{w}^*) = \sum_{j=1}^{r^*} a_j([\mathbf{v}_j^*]^T \mathbf{z}_f(\mathbf{u}_i))^2 - \lambda_{\min}(\mathcal{I}_f(\mathbf{w}^*))$, for all $i = 1, 2, \dots, N$. Furthermore, $\partial\Phi(\mathbf{w}^*) = \partial^c\Phi(\mathbf{w}^*) = -\left[\partial^c(-\Phi)(\mathbf{w}^*)\right]$, where the first equality follows from Proposition 2.2.7 of (Clarke, 1983), and the second equality follows from Proposition 2.3.1 of (Clarke, 1983). Therefore, for any $\mathbf{g} \in \partial\Phi(\mathbf{w}^*)$, we know that $-\mathbf{g} \in \partial^c(-\Phi)(\mathbf{w}^*)$. Thus, there exists $a_1, \dots, a_{r^*} \geq 0$ such that $\sum_{j=1}^{r^*} a_j = 1$ and $-\mathbf{g}^T(\mathbf{e}_i - \mathbf{w}^*) = \sum_{j=1}^{r^*} a_j([\mathbf{v}_j^*]^T \mathbf{z}_f(\mathbf{u}_i))^2 - \lambda_{\min}(\mathcal{I}_f(\mathbf{w}^*))$, for all $i = 1, 2, \dots, N$. $\square$

D. Proof of Theorem 2

Since we assumed that there exists $\mathbf{w} \in \Omega$ satisfying $\text{Eff}_k(\mathbf{w}) > m_k$ for all $k = 2, \dots, K$, we have that Slater's condition holds for problem (3.8). Furthermore, for all $k = 1, 2, \dots, K$, $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$ is a convex function and Ω is a convex set. Thus, it follows from Proposition A2 that $\mathbf{w}^* \in \Omega$ solves (3.8) if and only if there exists $\eta_2, \dots, \eta_K \geq 0$ such that $\eta_k(\Phi_k(\mathbf{w}^*) - h_k(m_k)) = 0$ for all $k = 2, 3, \dots, K$ and $\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\Omega} \left\{\Phi_1(\mathbf{w}) + \sum_{k=2}^{K} \eta_k(\Phi_k(\mathbf{w}) - h_k(m_k))\right\}$. Observing that $\sum_{k=2}^{K} \eta_k h_k(m_k)$ does not depend on $\mathbf{w}$ completes the proof. $\square$

E. Proof of Theorem 3

We first confirm that the following restatement of (4.20),

$$\min_{\mathbf{w}\in\Omega, t\geq 0} t \qquad \text{subject to: } \Phi_k(\mathbf{w}) \leq h_k(1/t),\ k = 1, \dots, K, \tag{E.36}$$

satisfies the conditions in Proposition A2. Since Ω and $\mathbb{R}$ are closed convex sets, $\Omega \times \mathbb{R}$ is a closed convex set. Furthermore, for all $k = 1, 2, \ldots, K$, $\Phi_k(\mathbf{w}) = \phi_k[\mathcal{I}_{f_k}(\mathbf{w})]$ is a convex function. We also know that for all $k = 1, 2, \ldots, K$, $h_k(1/t)$ in (3.9) is a concave function of t for all $k = 1, 2, \ldots, K$, as $-\lambda_{\min}(\mathbf{M})/t$ is a concave function of t for any positive definite matrix $\mathbf{M}$, t is a linear function, and $q_k \log(t)$ is a concave function. To show that Slater's condition holds, we need to find $\mathbf{w}' \in \Omega$ and $t' > 0$ with $\Phi_k(\mathbf{w}') < h_k(1/t')$ for all $k = 1, 2, \ldots, K$. It follows from the definition of the efficiency functions $\text{Eff}_k(\mathbf{w}')$ in Table 1 that $\Phi_k(\mathbf{w}') < h_k(1/t')$ if and only if $\text{Eff}_k(\mathbf{w}') > 1/t'$, and that $\text{Eff}_k(\mathbf{w}') > 0$ for all $\mathbf{w}' \in \Omega$. Thus, choosing $\mathbf{w}' = \arg\min_{\mathbf{w}\in\Omega} \Phi_1(\mathbf{w})$ and $t' = 2/\left(\min_{k=1,2,\ldots,K} \text{Eff}_k(\mathbf{w}')\right)$ satisfies Slater's condition.

We can now apply Proposition A2 to (E.36) to yield the following result: a feasible solution $(\mathbf{w}^*, t^*)$ for (E.36) solves problem (E.36) if and only if there exists $\nu, \eta_1, \ldots, \eta_K \geq 0$ satisfying:

$$\nu t^* = 0, \tag{E.37}$$

$$\eta_k(\Phi_k(\mathbf{w}^*) - h_k(1/t^*)) = 0 \text{ for all } k = 1, 2, \ldots, K, \tag{E.38}$$

$$(\mathbf{w}^*, t^*) \in \arg\min_{\mathbf{w}\in\Omega, t\in\mathbb{R}} \left\{ t - \nu t + \sum_{k=1}^{K} \eta_k(\Phi_k(\mathbf{w}) - h_k(1/t)) \right\}. \tag{E.39}$$

The optimization problem in (E.39) is separable. Thus, (E.39) can be rewritten as

$$\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\Omega} \left\{ \sum_{k=1}^{K} \eta_k \Phi_k(\mathbf{w}) \right\}, \tag{E.40}$$

$$t^* \in \arg\min_{t\in\mathbb{R}} \left\{ t - \nu t - \sum_{k=1}^{K} \eta_k h_k(1/t) \right\}. \tag{E.41}$$

Since $g(t) = t - \nu t - \sum_{k=1}^{K} \eta_k h_k(1/t)$ is a convex function, we can rewrite (E.41) as

$$1 - \nu - \sum_{k=1}^{K} \eta_k \left[\frac{d}{dt} h_k(1/t) \Big|_{t=t^*} \right] = 0. \tag{E.42}$$

This means that there exists $\nu, \eta_1, \ldots, \eta_K \geq 0$ satisfying (E.37)–(E.39) if and only if there exists $\eta_1, \ldots, \eta_K \geq 0$ satisfying (E.38), (E.40), and

$$1 - \sum_{k=1}^{K} \eta_k \left[\frac{d}{dt} h_k(1/t) \Big|_{t=t^*} \right] \geq 0, \qquad t^* \left(1 - \sum_{k=1}^{K} \eta_k \left[\frac{d}{dt} h_k(1/t) \Big|_{t=t^*} \right] \right) = 0. \tag{E.43}$$

Since (E.38) is Condition 2 in Theorem 3, and (E.40) is Condition 3 in Theorem 3, it remains to show that (E.43) is equivalent to Condition 1 in Theorem 3. It suffices to show that $t^* > 0$. Recall that (E.36) is equivalent to (4.19), so the optimal value for t in (E.36) is the reciprocal of the maximin efficiency attained by the optimal design. Efficiencies are bounded between 0 and 1, so the optimal value for t must be greater than 1, i.e. $t^* > 1$. □

Lucy L. Gao

Department of Statistics, University of British Columbia, Vancouver, B.C., Canada

E-mail: lucy.gao@stat.ubc.ca

Jane J. Ye, Shangzhi Zeng, Julie Zhou

Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada

E-mails: janeye@uvic.ca, zengshangzhi@uvic.ca, jzhou@uvic.ca