Computing Optimal Experimental Designs via Interior Point Method *

Zhaosong Lu† Ting Kei Pong ‡

October 12, 2012

Abstract

In this paper, we study optimal experimental design problems with a broad class of smooth convex optimality criteria, including the classical A-, D- and pth mean criterion. In particular, we propose an interior point (IP) method for them and establish its global convergence. Further, by exploiting the structure of the Hessian matrix of the aforementioned optimality criteria, we derive an explicit formula for computing its rank. Using this result, we then show that the Newton direction arising in the IP method can be computed efficiently via Sherman-Morrison-Woodbury formula when the size of the moment matrix is small relative to the sample size. Finally, we compare our IP method with the widely used multiplicative algorithm introduced by Silvey et al. [29]. The computational results show that the IP method generally outperforms the multiplicative algorithm both in speed and solution quality.

Key words: Optimal experimental design, A-criterion, c-criterion, D-criterion, pth mean criterion, interior point method

1 Introduction

In this paper, we consider the optimal experimental design problems on a given finite design space $\mathcal{X} = \{x_1, \ldots, x_n\} \subseteq \Re^m$. In this setting, we consider a coefficient matrix $K \in \Re^{m\times k}$ of full column rank and the moment matrix defined as

$$\mathcal{M}(w) = \sum_{i=1}^{n} w_i A_i$$

for $w \in \Omega := \{w : \ w_i \geq 0, \sum_{i=1}^n w_i = 1\}$, where A_i is the expected Fisher information matrix related to x_i, $i = 1, ..., n$. As in [41], throughout this paper we assume that A_i's are $m \times m$ real symmetric positive semidefinite matrices and that there exists an $w \in \Omega$ such that $\mathcal{M}(w)$ is positive definite. This in particular implies that $\mathcal{M}(w)$ is positive definite for all positive $w \in \Omega$. The optimal experimental design problem can then be formulated as the following minimization problem (see [25, Section 7.10]):

$$\begin{array}{rl} f^* := \inf\limits_{w} & \Phi(\mathcal{M}(w)) := \Psi(\mathcal{C}_K(\mathcal{M}(w))) \\ \text{s.t.} & w \in \Omega, \ \text{Range}(K) \subseteq \text{Range}(\mathcal{M}(w)), \end{array} \tag{1}$$

where Ψ is a function defined on the set of positive definite matrices and $\mathcal{C}_K(\mathcal{M}(w))$ is the information matrix defined by $\mathcal{C}_K(\mathcal{M}(w)) := (K^T(\mathcal{M}(w))^\dagger K)^{-1}$. Here $A^\dagger$ denotes the Moore-Penrose

*This work was supported in part by an NSERC Discovery Grant.

†Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada. (email: zhaosong@sfu.ca).

‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. (email: ptingkei@math.uwaterloo.ca).

pseudoinverse of a matrix A. The well-definedness of $\mathcal{C}_K(\mathcal{M}(w))$ is guaranteed by the range inclusion condition in the constraint of (1) and the fact that K has full column rank [25, Chapter 3]. The function Φ in the objective is commonly referred to as an "optimality criterion". Some classical optimality criteria include (see [25, Chapter 6]):

(i) A-criterion $\Phi(X) := \mathrm{tr}(K^T X^\dagger K)$;

(ii) c-criterion $\Phi(X) := c^T X^\dagger c$;

(iii) D-criterion $\Phi(X) := \log\det(K^T X^\dagger K)$;

(iv) pth mean criterion $\Phi(X) := \mathrm{tr}((K^T X^\dagger K)^{-p})$.

for some $p < 0$, $c \in \Re^m$ and $K \in \Re^{m\times k}$ of full column rank.

It is easy to observe that c-criterion is just a special case of A-criterion with $K = c$ and A-criterion is a special case of pth mean criterion with $p = -1$. We shall also mention that pth mean criterion can be defined more generally to include D-criterion as a special case (see [25, Chapter 6] for details). Furthermore, it can be shown that the constraint set of (1) is convex [25, Section 3.3], and the criteria (i)-(iv) are convex functions in the constraint set (by using [25, Theorem 5.14] and [25, Theorem 6.13], or [24, Proposition IV.14] and [24, Proposition IV.15]). Hence, problem (1) with these criteria is a convex optimization problem. Indeed, it is known that (1) with the above criteria can be reformulated as (possibly nonlinear) semidefinite programming (SDP) problems (see, for example, [14, 8, 10, 23]).

The optimal design problems (1) with the aforementioned criteria usually do not have closed form solutions. Numerous procedures have thus been proposed to solve (1) (see, for example, [13, 40, 5, 6, 39, 7, 18, 9, 24, 3, 34, 1, 12, 26, 35, 28]). Among them, the multiplicative algorithm introduced in [29] has been widely explored. For example, Titterington [30], Pázman [24], Dette et al. [12] and Harman and Trnovská [19] studied the multiplicative algorithm for D-criterion. In addition, Fellman [15] and Torsney [33] considered the multiplicative algorithm for A-criterion under the assumption that all A_i's are rank-one. Recently, Yu [41] studied the multiplicative algorithm for a class of convex optimality criteria and proved its global convergence under some assumptions. Nevertheless, for several commonly used optimality criteria, some of those assumptions may not hold and hence there is no theoretical guarantee for its convergence. Indeed, as observed in [41, Section 5], one of the assumptions does not hold for pth mean criterion with $p = -2$. Moreover, for such a criterion, our numerical experiments in Section 5 demonstrate that the multiplicative algorithm appears not to converge when $p < -1$. More details about the multiplicative algorithm for solving (1) are given in Section 2.

In this paper, we consider an alternative approach to solve problem (1). In particular, we propose an interior point (IP) method for (1) and establish its global convergence. The method is a Newton-type method that can be efficiently applied to solve problem (1) with a broad class of convex optimality criteria and moderate-sized matrices A_i's. By exploiting the structure of the Hessian matrix of the classical A-, D- and pth mean criterion, we derive an explicit formula for its rank. Using this result, we further show that the Newton direction arising in the IP method for (1) with the aforementioned classical optimality criteria can be computed efficiently via Sherman-Morrison-Woodbury formula when $n \gg m^2$, i.e., when the size of A_i's is small relative to the sample size. We finally compare the IP method with the multiplicative algorithm. The computational results show that the IP method usually outperforms the multiplicative algorithm in both speed and solution quality.

The rest of this paper is organized as follows. In Subsection 1.1, we introduce the notations that are used throughout the paper. In Section 2, we review the multiplicative algorithm and address its convergence. In Section 3, we propose an IP method for solving problem (1) with a large class of convex optimality criteria and address its convergence. In Section 4, we discuss how the IP method can be applied to solve problem (1) with criteria (i)–(iv) and demonstrate how the Newton direction can be computed efficiently when $n \gg m^2$. In Section 5, we conduct numerical experiments to test the performance of the method and compare it with the multiplicative algorithm. Finally, we present some concluding remarks in Section 6.

1.1 Notations

In this paper, the symbol $\Re_{++}$ denotes the set of all positive real numbers and $\Re^n$ denotes the n-dimensional Euclidean space. For a vector $x \in \Re^n$ and $\mathcal{I} \subseteq \{1,\dots,n\}$, $\|x\|$ denotes the Euclidean norm of x, $x_{\mathcal{I}}$ denotes the subvector of x indexed by $\mathcal{I}$ and $\mathscr{D}(x)$ denotes the diagonal matrix whose ith diagonal entry is x_i for all i. For $\alpha \in \Re$ and a vector $x \in \Re^n$ with positive entries, x^α denotes the vector whose ith entry is x_i^α for all i. For x, $y \in \Re^n$, $x \circ y$ denotes the Hadamard (entry-wise) product of x and y. The letter e denotes the vector of all ones, whose dimension should be clear from the context. The set of all $m \times n$ matrices with real entries is denoted by $\Re^{m\times n}$. For any $A \in \Re^{m\times n}$, $\mathcal{I} \subseteq \{1,\dots,m\}$ and $\mathcal{J} \subseteq \{1,\dots,n\}$, a_{ij} denotes the (i,j)th entry of A, $A_{\mathcal{J}}$ denotes the submatrix of A comprising the columns of A indexed by $\mathcal{J}$ and $A_{\mathcal{I}\mathcal{J}}$ denotes the submatrix of A comprising the rows and columns of A indexed by $\mathcal{I}$ and $\mathcal{J}$, respectively. The space of $n \times n$ symmetric matrices will be denoted by $\mathcal{S}^n$. If $A \in \mathcal{S}^n$ is positive semidefinite (resp., definite), we write $A \succeq 0$ (resp., $A \succ 0$). The cone of positive semidefinite (resp., definite) matrices is denoted by $\mathcal{S}^n_+$ (resp., $\mathcal{S}^n_{++}$). For $A, B \in \mathcal{S}^n$, $A \succeq B$ (resp., $A \succ B$) means $A - B \succeq 0$ (resp., $A - B \succ 0$). The trace of a real square matrix A is denoted by $\mathrm{tr}(A)$. We denote by I the identity matrix, whose dimension should be clear from the context.

A function $f : \mathcal{S}^n \to \Re$ is said to be increasing (resp., decreasing) if for any $A \succeq B$, it holds that

$$f(A) \geq f(B) \quad (\text{resp., } f(A) \leq f(B)).$$

2 The multiplicative algorithm

In this section we review the multiplicative algorithm introduced in [29] for solving problem (1) and discuss its convergence. In particular, we first describe the multiplicative algorithm as follows, which is specified through a power parameter $\lambda \in (0,1]$.

Multiplicative Algorithm:

1. **Start:** Let a positive $w^0 \in \Omega$ and $\lambda \in (0,1]$ be given.
2. **For** $k = 0, 1, \dots$

$$w_i^{k+1} = w_i^k \frac{(d_i(w^k))^\lambda}{\sum_{j=1}^n w_j^k (d_j(w^k))^\lambda}, \quad i = 1,\dots,n, \tag{2}$$

where $d_i(w) = -\mathrm{tr}(\nabla\Phi(\mathcal{M}(w))A_i)$ and $\nabla\Phi(\mathcal{M}(w))$ is the gradient of Φ at $\mathcal{M}(w)$.
End (for)

Remark 2.1. *The above algorithm is the same as the one described in [41], in the sense that both algorithms generate exactly the same sequence $\{w^k\}$ provided the initial points w^0 are identical.* ∎

We now state a global convergence result recently established by Yu [41, Theorem 2] for the multiplicative algorithm when applied to solve the following problem, which is closely related to (1):

$$\begin{array}{rl} \mathrm{val} := \sup\limits_{w} & -\Phi(\mathcal{M}(w)) \\ \text{s.t.} & w \in \Omega,\ \mathcal{M}(w) \succ 0. \end{array} \tag{3}$$

Observe that (1) and (3) are equivalent (i.e., the optimal value being negative of each other) if there exists an optimal solution w^* of (1) with $\mathcal{M}(w^*) \succ 0$, or if Φ is convex in

$$\mathcal{S}^m_+(K) := \{X \in \mathcal{S}^m_+ :\ \mathrm{Range}(K) \subseteq \mathrm{Range}(X)\}.$$

and (3) has an optimal solution.

Proposition 2.1. *Let $\{w^k\}$ be the sequence generated from the above multiplicative algorithm. Suppose the following assumptions hold:*

(a) *for any feasible point w of* (3), $\nabla\Phi(\mathcal{M}(w)) \preceq 0$ *and* $\nabla\Phi(\mathcal{M}(w))A_i \neq 0$ *for* $i = 1, \ldots, n$;

(b) *for any feasible point w of* (3), *if* $T(w) \neq w$, *then* $\Phi(\mathcal{M}(T(w))) < \Phi(\mathcal{M}(w))$, *where*

$$[T(w)]_i \;:=\; w_i \frac{(d_i(w))^\lambda}{\sum_{j=1}^n w_j (d_j(w))^\lambda}, \quad i = 1, \ldots, n;$$

(c) Φ *is strictly convex and* $\nabla\Phi$ *is continuous in* $\mathcal{S}^m_{++}$;

(d) *for any* $\{X^k\} \subset \mathcal{S}^m_{++}$, *if* $X^k \to X^*$ *and* $\{\Phi(X^k)\}$ *is decreasing, then* $X^* \succ 0$.

Then $\Phi(\mathcal{M}(w^k)) \to -\mathrm{val}$ *monotonically, and moreover, any accumulation point of* $\{w^k\}$ *is an optimal solution of* (3).

Remark 2.2. *Notice that the assumptions in the above proposition imply that any accumulation point w^* of $\{w^k\}$ satisfies $\mathcal{M}(w^*) \succ 0$. Hence, if the assumptions in Proposition 2.1 hold and Φ is convex in $\mathcal{S}^m_+(K)$, then* (1) *is equivalent to* (3) *and any accumulation point of the sequence $\{w^k\}$ generated from the above multiplicative algorithm solves* (1). ■

Using Proposition 2.1 and some technical results developed in [41], one can establish the convergence of the above multiplicative algorithm when applied to problem (1) with A-, D- and pth mean criterion for $p \in (-1, 0)$ and $K = I$, which is summarized as follows.

Corollary 2.1. *Assume that $K = I$ and $A_i \neq 0$ for $i = 1, \ldots, n$. Then the multiplicative algorithm converges for any $\lambda \in (0, 1]$ when applied to problem* (1) *with D- and pth mean criterion for $p \in (-1, 0)$. Also, it converges for A-criterion when $\lambda \in (0, 1)$.*

As seen from Proposition 2.1 and Corollary 2.1, the multiplicative algorithm converges for a large class of optimality criteria Φ. Nevertheless, for some important convex optimality criteria, the assumptions stated in Proposition 2.1 may not hold and hence there is no theoretical guarantee for its convergence. Indeed, as observed in [41, Section 5], the assumption (b) with $\lambda = 1$ does not hold for pth mean criterion with $p = -2$. Moreover, for such a criterion, our numerical experiments in Section 5 demonstrate that the multiplicative algorithm appears not to converge when $p < -1$.

Due to the aforementioned potential drawbacks of the multiplicative algorithm, we will propose an IP method for solving problem (1) with a broad class of optimality criteria Φ including A-, D- and pth mean criterion in subsequent sections.

3 IP method for a class of convex optimality criteria

In this section, we propose an IP method for solving (1) with a class of convex optimality criteria $\Phi = \Psi \circ \mathcal{C}_K$. We make the following assumption on Ψ throughout this section.

Assumption 3.1. *The function Ψ is convex, decreasing, twice continuously differentiable and bounded below in $\mathcal{S}^m_{++}$. Moreover, for any bounded sequences $\{X^k\} \subseteq \mathcal{S}^m_{++}$ with $\lambda_{\min}(X^k) \to 0$, one has $\Psi(X^k) \to \infty$.*

Remark 3.1. *We now make some brief comments on the above assumptions.*

(a) *Assumption 3.1 is fairly reasonable. Indeed, all optimality criteria described in Section 1 satisfy this assumption.*

(b) *Since the feasible set is not necessarily closed, problem* (1) *with a general convex optimality criterion may not have an optimal solution. However, when the optimality criterion satisfies Assumption 3.1, it must have an optimal solution as shown in Theorem 3.1(a). We refer the readers to [25, Chapter 5] for more discussion on conditions guaranteeing existence of solutions for problem* (1).

(c) *In contrast to Proposition 2.1, we do not require the existence of a positive definite optimal moment matrix $\mathcal{M}(w^*)$. Indeed, Assumption 3.1 may hold even when problem* (1) *does not have a positive definite optimal moment matrix. For instance, the design problem*

$$\begin{array}{ll} \min\limits_{w,X} & \begin{pmatrix}1\\0\end{pmatrix}^T X^\dagger \begin{pmatrix}1\\0\end{pmatrix} \\ \text{s.t.} & X = \begin{pmatrix}w_1 & 0\\ 0 & w_2\end{pmatrix}, w_1+w_2=1, w_1,w_2 \geq 0, \\ & \begin{pmatrix}1\\0\end{pmatrix} \in \mathrm{Range}(X), \end{array}$$

has a unique optimal solution at $(w_1, w_2) = (1, 0)$. The corresponding optimal moment matrix is not positive definite; thus, the assumption (d) of Proposition 2.1 does not hold. However, it is easy to check that Assumption 3.1 is satisfied for this design problem (with $\Psi(t) = 1/t$). In general, the assumption (d) of Proposition 2.1 is likely not satisfied when K is not invertible, while our Assumption 3.1 is independent of K.

■

Under Assumption 3.1, it is not hard to show that the function $\Phi(\mathcal{M}(\cdot))$ is bounded below on the feasible set of (1). Also, it is routine to show that the function Φ is twice continuously differentiable in $\mathcal{S}^m_{++}$. Furthermore, it can be shown that Φ is convex in $\mathcal{S}^m_+(K)$ by considering suitable Schur complements (see, for example, [23, Section 6]). We include a short proof below for the convenience of readers. Before proceeding, we state the following well-known fact, which concerns the Schur complement of a positive semidefinite submatrix (see, for example, [25, Lemma 3.12]).

Lemma 3.1. *Let $A \in \mathcal{S}^k$, $B \in \Re^{m\times k}$ and $C \in \mathcal{S}^m$. Then the matrix $\begin{pmatrix}A & B^T\\ B & C\end{pmatrix}$ is positive semidefinite if and only if $A \succeq B^T C^\dagger B$, $C \succeq 0$ and* $\mathrm{Range}(B) \subseteq \mathrm{Range}(C)$.

Proposition 3.1. *The optimality criterion Φ is convex in $\mathcal{S}^m_+(K)$.*

Proof. First of all, it can be shown that the set $\mathcal{S}^m_+(K)$ is convex (see, for example, [25, Section 3.3]). In addition, notice that for any $X \in \mathcal{S}^m_+(K)$, we have

$$\begin{aligned} \Phi(X) &= \Psi((K^T X^\dagger K)^{-1}) \;=\; \inf_U \left\{\Psi(U):\ (K^T X^\dagger K)^{-1} \succeq U \succ 0\right\} \\ &= \inf_U \left\{\Psi(U):\ U^{-1} \succeq K^T X^\dagger K, U \succ 0\right\} \\ &= \inf_U \left\{\Psi(U):\ \begin{pmatrix}U^{-1} & K^T\\ K & X\end{pmatrix} \succeq 0, U \succ 0\right\} \\ &= \inf_U \left\{\Psi(U):\ X \succeq KUK^T, U \succ 0\right\}, \end{aligned} \tag{4}$$

where the second equality follows from the fact that Ψ is decreasing, the fourth and last equalities follow from Lemma 3.1, while the third equality holds because $K^T X^\dagger K$ is invertible for $X \in \mathcal{S}^m_+(K)$ when K has full column rank. Convexity of Φ in $\mathcal{S}^m_+(K)$ now follows from [27, Theorem 5.7]. ■

Observe that $\mathcal{M}(w) \succ 0$ whenever $w > 0$. Thus, under Assumption 3.1, the function Φ is twice continuously differentiable for any positive $w \in \Omega$. It is hence natural to develop an IP method to solve (1) since such a method keeps all iterates in the relative interior of Ω until convergence. To proceed, we first reformulate the problem by eliminating the equality constraint. The resulting equivalent problem is given by

$$\begin{array}{rl} f^* = \inf\limits_{\tilde{w}} & f(\tilde{w}) := \Phi(\mathcal{M}(P\tilde{w}+q)) \\ \text{s.t.} & e^T\tilde{w} \leq 1,\ \tilde{w} \geq 0, \\ & \mathrm{Range}(K) \subseteq \mathrm{Range}(\mathcal{M}(P\tilde{w}+q)), \end{array} \tag{5}$$

where $P \in \Re^{n\times(n-1)}$ and $q \in \Re^n$ are such that

$$P\tilde{w} + q = \begin{pmatrix} \tilde{w} \\ 1 - e^T\tilde{w} \end{pmatrix} \qquad \forall \tilde{w} \in \Re^{n-1}. \tag{6}$$

We next develop an IP method for solving problem (5) instead. First, we need to build a suitable barrier function. Given any $\tilde{w} > 0$ satisfying $e^T\tilde{w} < 1$, one can observe that $P\tilde{w} + q > 0$ and hence $\mathcal{M}(P\tilde{w} + q) \succ 0$, which leads to $\mathrm{Range}(K) \subseteq \mathrm{Range}(\mathcal{M}(P\tilde{w} + q))$. This implies that any barrier function that takes into account the first two inequality constraints of (5) is sufficient for the development of IP method. Here we naturally choose the logarithmic barrier function and then solve the barrier subproblem in the form of

$$\min_{\tilde{w}} f_\mu(\tilde{w}) := f(\tilde{w}) - \mu \sum_{i=1}^{n-1} \log(\tilde{w}_i) - \mu \log\left(1 - e^T\tilde{w}\right) \tag{7}$$

for a sequence of parameters $\mu \downarrow 0$. In view of Assumption 3.1, we see that any level set of f_μ is compact. Moreover, f_μ is strictly convex. Thus, there exists a unique minimizer to (7) for any $\mu > 0$. Furthermore, it follows from Assumption 3.1 that f_μ is twice continuously differentiable and its Hessian is positive definite in its domain. Therefore, problem (7) can be suitably solved by the Newton's method with a line search whose stepsize is chosen by Armijo rule.

We are now ready to present our IP method for solving problem (5).

IP Method:

1. **Start:** Let a strictly feasible $\tilde{w}^0$, $0 < \beta, \gamma, \eta, \sigma < 1$ and $\mu_1 > 0$ be given. Let $\epsilon(\mu)$ be an increasing function of μ so that $\lim_{\mu\downarrow 0} \epsilon(\mu) = 0$. Set $\tilde{w} = \tilde{w}^0$ and $k = 1$.

2. **While** $\|\nabla f_{\mu_k}(\tilde{w})\| > \epsilon(\mu_k)$ **do**

   (a) Compute the Newton direction

   $$d := -(\nabla^2 f_{\mu_k}(\tilde{w}))^{-1} \nabla f_{\mu_k}(\tilde{w}). \tag{8}$$

   (b) Let $\alpha_{\max}(\tilde{w}) := \max\{\alpha : \ \tilde{w}[\alpha] \geq 0, \ e^T\tilde{w}[\alpha] \leq 1\}$, where $\tilde{w}[\alpha] := \tilde{w} + \alpha d$.

   (c) Let α be the largest element of $\{\bar{\alpha}(\tilde{w}), \beta\bar{\alpha}(\tilde{w}), \beta^2\bar{\alpha}(\tilde{w}), \cdots\}$ satisfying

   $$f_{\mu_k}(\tilde{w}[\alpha]) \leq f_{\mu_k}(\tilde{w}) + \sigma\alpha(\nabla f_{\mu_k}(\tilde{w}))^T d,$$

   where $\bar{\alpha}(\tilde{w}) := \min\{1, \eta\alpha_{\max}(\tilde{w})\}$.

   (d) Set $\tilde{w} \leftarrow \tilde{w}[\alpha]$.

   End (while)

3. Set $\tilde{w}^k \leftarrow \tilde{w}$, $\mu_{k+1} \leftarrow \gamma\mu_k$, $k \leftarrow k + 1$, and go to step 2.

In standard convergence analysis of IP methods, the feasible sets are usually assumed to be closed and the objective functions are twice continuously differentiable in a neighborhood of the feasible sets (see, for example, [17]). Nevertheless, these two conditions do not necessarily hold for our problem (5). In particular, the objective function is not necessarily continuous up to the boundary of the feasible region [25, Section 3.16]. Hence, it is not immediately clear the sequence generated by our method will accumulate at a global minimizer of (5). Thus, we discuss convergence of our IP method below. We first present convergence results regarding the outer iterations of our IP method and then discuss the convergence of its inner iterations.

For notational convenience, in the remainder of this section, we associate with each $\tilde{w} \in \Re^{n-1}$ a unique $w \in \Re^n$ by letting $w := P\tilde{w} + q$. Analogously, we associate with each $w \in \Re^n$ a unique $\tilde{w} \in \Re^{n-1}$ by letting $\tilde{w}_i = w_i$ for $i = 1, \ldots, n-1$. Also, we let $\Phi_{\mathcal{M}}(w) := \Phi(\mathcal{M}(w))$.

We first observe that if problem (1) has an optimal solution w^* with $\mathcal{M}(w^*) \succ 0$, then there exists a Lagrange multiplier $u^* \geq 0$ such that (w^*, u^*) satisfies the following KKT system:

$$\begin{array}{rcl} P^T(\nabla\Phi_{\mathcal{M}}(w) - u) & = & 0, \\ e^T w & = & 1, \\ u \circ w & = & 0, \\ (w, u) & \geq & 0. \end{array} \tag{9}$$

Given a strictly feasible point $\tilde{w} \in \Re^{n-1}$ of problem (7), we notice that

$$\nabla f_\mu(\tilde{w}) = P^T(\nabla\Phi_{\mathcal{M}}(w) - \mu w^{-1}). \tag{10}$$

Then it is not hard to observe that for each $\mu > 0$, the w associated with the approximate solution $\tilde{w}$ of (7) obtained by the Newton's method detailed in step 2 above together with $u := \mu w^{-1}$ satisfies the following perturbed KKT system:

$$\begin{array}{rcl} P^T(\nabla\Phi_{\mathcal{M}}(w) - u) & = & v, \\ e^T w & = & 1, \\ u \circ w & = & \mu e, \\ (w, u) & > & 0 \end{array} \tag{11}$$

for some $v \in \Re^{n-1}$. The convergence regarding the outer iterations of our IP method is related to the limiting behavior of the solutions of system (11) as $(\mu, v) \to (0_+, 0)$, that is, $(\mu, v) \to (0, 0)$ with $\mu > 0$.

We first claim that system (11) has a unique solution for any $(\mu, v) \in \Re_{++} \times \Re^{n-1}$. Indeed, it is easy to observe that (w, u) is a solution of (11) if and only if $\tilde{w} \in \Re^{n-1}$ is an optimal solution of

$$\min_{\tilde{w}} f_\mu(\tilde{w}) - v^T\tilde{w}. \tag{12}$$

Since the objective function of (12) is strictly convex and it has compact level sets, problem (12) has a unique optimal solution, which immediately implies that system (11) has a unique solution. From now on, we denote by $(w(\mu, v), u(\mu, v))$ the unique solution of (11). Our main theorem below discusses the limiting behavior of $(w(\mu, v), u(\mu, v))$ as $(\mu, v) \to (0_+, 0)$. The proof of this theorem can be found in the appendix.

Theorem 3.1. *Let $(w(\mu, v), u(\mu, v))$ be defined above for $(\mu, v) \in \Re_{++} \times \Re^{n-1}$. Then the following statements hold:*

(a) $\lim\limits_{(\mu,v)\to(0_+,0)} \Phi(\mathcal{M}(w(\mu, v))) = f^*$ *and any accumulation point of $w(\mu, v)$ as $(\mu, v) \to (0_+, 0)$ is an optimal solution of* (1).

(b) *Suppose in addition that problem* (1) *has an optimal solution w^* with $\mathcal{M}(w^*) \succ 0$. Then any accumulation point of $w(\mu, v)$ as $(\mu, v) \xrightarrow[\Xi_C]{} (0, 0)$, i.e., $(\mu, v) \to (0, 0)$ with $(\mu, v) \in \Xi_C := \{(\mu, v) : \ \|v\|_\infty < C\mu\}$ for some given $C > 0$, is an optimal solution of* (1) *with maximum cardinality.*

As an immediate consequence of Theorem 3.1, we have the following global convergence result regarding the outer iterations of our IP method, whose simple proof is omitted.

Corollary 3.1. *Let $\{\mu_k\}$ and $\{\tilde{w}^k\}$ be the sequences generated in the IP method. Let $w^k = P\tilde{w}^k + q$ for all k. Then the following statements hold:*

(a) $\lim\limits_{k\to\infty} \Phi(\mathcal{M}(w^k)) = f^*$ *and any accumulation point of $\{w^k\}$ is an optimal solution of* (1).

(b) *Suppose in addition that problem* (1) *has an optimal solution w^* with $\mathcal{M}(w^*) \succ 0$ and $\epsilon(\mu_k) = O(\mu_k)$. Then any accumulation point of $\{w^k\}$ is an optimal solution of* (1) *with maximum cardinality.*

We emphasize that in Corollary 3.1 (a), we do *not* require existence of an optimal solution w^* with $\mathcal{M}(w^*) \succ 0$. On the other hand, if such an optimal solution does exist, for example, when $K = I$, then Corollary 3.1 (b) states that the accumulation point (with $\epsilon(\mu_k) = O(\mu_k)$) must be an optimal solution of (1) that has the largest number of non-zero entries among all the optimal solutions of (1).

Before ending this section, we establish a convergence result regarding the inner iterations of our IP method.

Proposition 3.2. *Let $\mu_k > 0$ and $\epsilon(\mu_k) > 0$ be given. Then the Newton's method detailed in step 2 of the IP method starting from any strictly feasible point $\tilde{w}^{\text{init}}$ of (5) generates a point $\tilde{w}^k$ satisfying $\|\nabla f_{\mu_k}(\tilde{w}^k)\| \leq \epsilon(\mu_k)$ within a finite number of iterations.*

Proof. First, observe that all iterates generated by the Newton's method lie in the compact level set $\Upsilon := \{\tilde{w} : \ f_{\mu_k}(\tilde{w}) \leq f_{\mu_k}(\tilde{w}^{\text{init}})\}$. Furthermore, it holds that $\tilde{w} > 0$ and $1 - e^T\tilde{w} > 0$ for all $\tilde{w} \in \Upsilon$. This together with the assumption that $\mathcal{M}(\Omega) \cap \mathcal{S}^m_{++} \neq \emptyset$ implies that $\mathcal{M}(\Upsilon) \subset \mathcal{S}^m_{++}$. Thus ∇f_{μ_k} and $\nabla^2 f_{\mu_k}$ are continuous in Υ. Using this observation and the strong convexity of f_{μ_k} in Υ, there exist $\underline{\lambda}$, $\overline{\lambda} > 0$ such that $\underline{\lambda} I \preceq \nabla^2 f_{\mu_k}(\tilde{w}) \preceq \overline{\lambda} I$ for all $\tilde{w} \in \Upsilon$. This relation along with the continuity of ∇f_{μ_k} and $\nabla^2 f_{\mu_k}$ implies that $d = -(\nabla^2 f_{\mu_k}(\tilde{w}))^{-1}\nabla f_{\mu_k}(\tilde{w})$ is continuous in Υ. In view of this result and the definition of $\bar{\alpha}(\tilde{w})$, it is not hard to show that $\bar{\alpha}(\tilde{w})$ is positive and continuous in Υ. This fact together with the compactness of Υ yields $\underline{\alpha} := \inf\{\bar{\alpha}(\tilde{w}) : \ \tilde{w} \in \Upsilon\} > 0$. Thus, all iterates $\tilde{w}$ generated by the Newton's method satisfy $\underline{\lambda} I \preceq \nabla^2 f_{\mu_k}(\tilde{w}) \preceq \overline{\lambda} I$ and $\bar{\alpha}(\tilde{w}) \in [\underline{\alpha}, 1]$. The remaining proof follows the same arguments as in the proof of [22, Theorem 3.13]. ∎

4 IP method for classical optimality criteria

In this section, we discuss how to apply our IP method to solve problem (1) with A-, D- and pth mean criterion. In particular, we will demonstrate how the Newton direction (8) can be efficiently computed for each criterion.

Before proceeding, we introduce some notations that will be used in this section (see, for example, [31] for more details). Given matrices A and B in $\Re^{m\times n}$, $A \otimes B$ denotes the Kronecker product of A and B, while $A \circ B$ denotes the Hadamard (entry-wise) product of A and B. In addition, $\mathbf{vec}(A)$ denotes the column vector formed by stacking columns of A one by one. For any $m\times m$ symmetric matrix U, we define the vectors $\mathbf{svec}(U) \in \Re^{m(m+1)/2}$ and $\mathbf{svec}_0(U) \in \Re^{m(m+1)/2}$ as

$$\begin{aligned}
\mathbf{svec}(U) &= (u_{11}, \sqrt{2}u_{21}, \ldots, \sqrt{2}u_{m1}, u_{22}, \sqrt{2}u_{32}, \ldots, \sqrt{2}u_{m2}, \ldots, u_{mm})^T. \\
\mathbf{svec}_0(U) &= (u_{11}, u_{21}, \ldots, u_{m1}, u_{22}, u_{32}, \ldots, u_{m2}, \ldots, u_{mm})^T.
\end{aligned}$$

It is not hard to observe that $\mathbf{svec}$ is an isometry between $\mathcal{S}^m$ and $\Re^{m(m+1)/2}$ and moreover,

$$\operatorname{tr}(UV) = \mathbf{svec}(U)^T\mathbf{svec}(V) \quad \forall U, V \in \mathcal{S}^m. \tag{13}$$

We denote the inverse map of $\mathbf{svec}$ by $\mathbf{smat}$. Clearly, they are adjoint of each other, namely,

$$u^T\mathbf{svec}(V) = \operatorname{tr}(\mathbf{smat}(u)V) \quad \forall u \in \Re^{m(m+1)/2}, V \in \mathcal{S}^m.$$

The symmetric Kronecker product of any two (not necessarily symmetric) matrices $G, H \in \Re^{m\times m}$ is a square matrix of order $m(m+1)/2$ such that

$$(G \otimes_s H)\mathbf{svec}(U) = \frac{1}{2}\mathbf{svec}(GUH^T + HUG^T) \quad \forall U \in \mathcal{S}^m. \tag{14}$$

As mentioned in [31], $G \otimes_s H$ can be expressed in terms of the standard Kronecker product of G and H as follows:

$$G \otimes_s H = \frac{1}{2}Q(G \otimes H + H \otimes G)Q^T,$$

where $Q \in \Re^{m(m+1)/2 \times m^2}$ is such that

$$Q\,\mathbf{vec}(U) = \mathbf{svec}(U), \quad Q^T\mathbf{svec}(U) = \mathbf{vec}(U) \quad \forall U \in \mathcal{S}^m. \tag{15}$$

It is easy to observe that the above Q exists and is unique. Moreover, $QQ^T = I$.

Throughout this section, for each optimality criterion Φ, we define the associated function ϕ as follows:

$$\phi(x) = \Phi(\mathbf{smat}(x)) \tag{16}$$

for any $x \in \Re^{m(m+1)/2}$, provided that $\Phi(\mathbf{smat}(x))$ is well-defined. It is clear to observe that ϕ is convex due to the convexity of Φ. Define

$$M := [\mathbf{svec}(A_1)\ldots\mathbf{svec}(A_n)].$$

Clearly, $M \in \Re^{m(m+1)/2 \times n}$.

With the notations above, the function f_μ defined in (7) can be rewritten as

$$f_\mu(\tilde{w}) = \phi(M(P\tilde{w}+q)) - \mu\sum_{i=1}^{n-1}\log(\tilde{w}_i) - \mu\log\left(1 - e^T\tilde{w}\right).$$

By the chain rule, the gradient and Hessian of f_μ are given by

$$\begin{aligned}
\nabla f_\mu(\tilde{w}) &= P^TM^T\nabla\phi(Mw) - \mu P^Tw^{-1}, \\
\nabla^2 f_\mu(\tilde{w}) &= P^TM^T\nabla^2\phi(Mw)MP + \frac{\mu}{(1-e^T\tilde{w})^2}ee^T + \mu\mathscr{D}(\tilde{w}^{-2}),
\end{aligned} \tag{17}$$

where $w = P\tilde{w} + q$.

The main computational effort of our IP method lies in computing the Newton direction d by solving the system $\nabla^2 f_\mu(\tilde{w})d = -\nabla f_\mu(\tilde{w})$ (see (8)). In applications, n can be significantly larger than m^2. Since the rank of $\nabla^2\phi(Mw)$ is at most $m(m+1)/2$, the first matrix in (17) has "low" rank compared to $\nabla^2 f_\mu(\tilde{w})$. It is generally more efficient to compute the Newton direction via the Sherman-Morrison-Woodbury formula, *without* explicitly forming the Hessian matrix. To this end, suppose that $\nabla^2\phi(Mw)$ has rank r. Let VDV^T be the *partial* eigenvalue decomposition of $\nabla^2\phi(Mw)$, where D is the $r\times r$ diagonal matrix whose diagonal consists of r largest eigenvalues of $\nabla^2\phi(Mw)$, and the columns of V are the corresponding eigenvectors.[1] Due to the convexity of ϕ, one can observe that $\nabla^2\phi(Mw) = VDV^T$. It then follows from (17) that

$$\begin{aligned}
\nabla^2 f_\mu(\tilde{w}) &= (P^TM^TV)D(V^TMP) + \frac{\mu}{(1-e^T\tilde{w})^2}ee^T + \mu\mathscr{D}(\tilde{w}^{-2}) \\
&= \begin{pmatrix} P^TM^TV & e \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & \frac{\mu}{(1-e^T\tilde{w})^2} \end{pmatrix}\begin{pmatrix} V^TMP \\ e^T \end{pmatrix} + \mu\mathscr{D}(\tilde{w}^{-2}),
\end{aligned}$$

which together with the Sherman-Morrison-Woodbury formula yields the Newton direction

$$d = -\left(\nabla^2 f_\mu(\tilde{w})\right)^{-1}\nabla f_\mu(\tilde{w}) = -\left[\frac{1}{\mu}\mathscr{D}(\tilde{w}^2) - \frac{1}{\mu^2}\mathscr{D}(\tilde{w}^2)\begin{pmatrix} P^TM^TV & e \end{pmatrix}W\begin{pmatrix} V^TMP \\ e^T \end{pmatrix}\mathscr{D}(\tilde{w}^2)\right]\nabla f_\mu(\tilde{w}),$$

where

$$W = \left(\begin{pmatrix} D^{-1} & 0 \\ 0 & \frac{(1-e^T\tilde{w})^2}{\mu} \end{pmatrix} + \frac{1}{\mu}\begin{pmatrix} V^TMP \\ e^T \end{pmatrix}\mathscr{D}(\tilde{w}^2)\begin{pmatrix} P^TM^TV & e \end{pmatrix}\right)^{-1}.$$

When $n \gg m^2$, the above approach is much more efficient than solving the Newton system directly by performing Cholesky factorization of $\nabla^2 f_\mu(\tilde{w})$. We remark that the ideas of using Sherman-Morrison-Woodbury formula to solve specially structured Newton systems have been explored in literature (see, for example, [2, 16]).

[1] The partial eigenvalue decomposition can be efficiently computed by the package PROPACK [21].

As seen from above, $\nabla\phi(Mw)$ and $\nabla^2\phi(Mw)$ are needed to compute Newton direction. Furthermore, since the Hessian tends to become more ill-conditioned as $\mu \to 0$, it is more desirable to explicitly determine the rank r of $\nabla^2\phi(Mw)$ a priori than to use the numerical rank obtained from the Matlab built-in function in each iteration. For the rest of this section, we will discuss how to evaluate $\nabla\phi(Mw)$ and $\nabla^2\phi(Mw)$ for A-, D- and pth mean criterion, and determine the rank r of $\nabla^2\phi(Mw)$ used in the aforementioned partial eigenvalue decomposition of $\nabla^2\phi(Mw)$. The latter quantity turns out to be independent of $w > 0$.

4.1 IP method for pth mean criterion

Recall from Section 1 that in $\mathcal{S}^m_{++}$, the pth mean criterion Φ becomes

$$\Phi(X) = \mathrm{tr}((K^T X^{-1} K)^{-p}) \tag{18}$$

for some $p < 0$ and $K \in \Re^{m\times k}$ with full column rank. It is easy to check that Assumption 3.1 holds for Φ. Hence, problem (1) with this criterion can be suitably solved by our IP method proposed in Section 3.

Based on the above discussion, we know that our IP method needs the gradient and Hessian of the associated function ϕ for computing Newton direction, where ϕ is defined by (16). We next discuss how to compute them. Before proceeding, we state the following classical result (see, for example, [11, Proposition 4.3]) that will be used subsequently.

Lemma 4.1. *Let $g : \Re \to \Re$ be a differentiable function and let $g^{\square} : \mathcal{S}^m \to \mathcal{S}^m$ be defined by*

$$g^{\square}(Y) := V \begin{pmatrix} g(d_1) & & & \\ & g(d_2) & & \\ & & \ddots & \\ & & & g(d_m) \end{pmatrix} V^T,$$

where $V\mathscr{D}(d)V^T$ is an eigenvalue decomposition of Y for some $d \in \Re^m$. Then the function $g^{\square}$ is well-defined, i.e., it is independent of the choice of V and d, and is also differentiable. Moreover, let $S^{g,d} \in \mathcal{S}^m$ be a symmetric matrix whose (i,j)th entry is given by

$$s^{g,d}_{ij} := \begin{cases} \dfrac{g(d_i) - g(d_j)}{d_i - d_j} & \text{if } d_i \neq d_j, \\ g'(d_i) & \text{otherwise.} \end{cases}$$

Then the directional derivative of $g^{\square}$ at Y along the direction $H \in \mathcal{S}^m$ is given by

$$V(S^{g,d} \circ (V^T H V))V^T.$$

Proposition 4.1. *Let Φ be defined in (18) and the associated ϕ be defined in (16). Let $Q \in \Re^{m(m+1)/2\times m^2}$ be defined in (15). Then the gradient and Hessian of ϕ at any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$ are given by*

$$\nabla\phi(x) = p\mathbf{svec}(X^{-1}K(K^T X^{-1} K)^{-p-1} K^T X^{-1}), \tag{19}$$

$$\begin{aligned}\nabla^2\phi(x) = Q(&-p[(X^{-1}KV) \otimes (X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV) \otimes (X^{-1}KV)]^T \\ &- p\, X^{-1} \otimes G - p\, G \otimes X^{-1})Q^T,\end{aligned} \tag{20}$$

respectively, where $X = \mathbf{smat}(x)$, $V\mathscr{D}(d)V^T$ is an eigenvalue decomposition of $K^T X^{-1} K$ for some $d \in \Re^m$, $g(t) = t^{-p-1}$, and $G = X^{-1}K[K^T X^{-1} K]^{-p-1} K^T X^{-1}$. In particular, when $K = I$, the above gradient and Hessian reduce to

$$\nabla\phi(x) = p\mathbf{svec}(X^{p-1}), \tag{21}$$

$$\nabla^2\phi(x) = (V \otimes_s V)\mathscr{D}(\mathbf{svec}_0(S^{g,d}))(V \otimes_s V)^T, \tag{22}$$

where $g(t) = pt^{p-1}$ and $V\mathscr{D}(d)V^T$ is an eigenvalue decomposition of X for some $d \in \Re^m$.

Proof. To derive the gradient of ϕ, we fix an arbitrary $x \in \mathbf{svec}(\mathcal{S}^m_{++})$. Let $X = \mathbf{smat}(x)$. For all sufficiently small $h \in \Re^{m(m+1)/2}$, we have $X + H \succ 0$, where $H = \mathbf{smat}(h)$, and moreover,

$$(X+H)^{-1} = X^{-1} - X^{-1}HX^{-1} + o(H). \tag{23}$$

Using (23) and Lemma 4.1 with $g(t) = t^{-p}$ and $Y = K^T X^{-1} K$, we obtain that

$$\begin{aligned}\Phi(X+H) &= \mathrm{tr}((K^T[X+H]^{-1}K)^{-p}) = \mathrm{tr}((K^TX^{-1}K - K^TX^{-1}HX^{-1}K + o(H))^{-p}) \\ &= \Phi(X) - \mathrm{tr}(V(S^{g,d} \circ (V^TK^TX^{-1}HX^{-1}KV))V^T) + o(H), \end{aligned} \tag{24}$$

where $V\mathscr{D}(d)V^T$ is an eigenvalue decomposition of Y. Letting $R := -K^TX^{-1}HX^{-1}K$ and using the fact that $V^TV = I$ and $s^{g,d}_{ii} = -pd_i^{-p-1}$ for all i, we further have

$$\begin{aligned}\mathrm{tr}(V(S^{g,d} \circ (V^TRV))V^T) &= \mathrm{tr}(S^{g,d} \circ (V^TRV)) = \sum_{i=1}^m s^{g,d}_{ii} \sum_{j,k} v_{ji} r_{jk} v_{ki} \\ &= -p\sum_{j,k}\left(\sum_{i=1}^m v_{ji} d_i^{-p-1} v_{ki}\right) r_{jk} = -\mathrm{tr}(p(K^TX^{-1}K)^{-p-1}R) \\ &= \mathrm{tr}(pX^{-1}K(K^TX^{-1}K)^{-p-1}K^TX^{-1}H). \end{aligned} \tag{25}$$

In view of the definitions of ϕ, Φ, X and H, it follows from (24), (25) and (13) that

$$\phi(x+h) - \phi(x) = \Phi(X+H) - \Phi(X) = h^T\left(p\mathbf{svec}(X^{-1}K(K^TX^{-1}K)^{-p-1}K^TX^{-1})\right) + o(h),$$

which yields (19). And (21) immediately follows from (19) by letting $K = I$.

We next derive the Hessian of ϕ at any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$. To proceed, we first recall the following well-known results (see, for example, page 243 and Lemma 4.3.1 of [20]):

$$\mathbf{vec}(ABC) = (C^T \otimes A)\,\mathbf{vec}(B), \qquad (A \otimes B)^T = A^T \otimes B^T. \tag{26}$$

Let X, h and H be defined as above. Using (23) and Lemma 4.1 with $g(t) = t^{-p-1}$ and $Y = K^TX^{-1}K$, we have

$$\begin{aligned}\nabla\Phi(X+H) &= p(X+H)^{-1}K[K^T(X+H)^{-1}K]^{-p-1}K^T(X+H)^{-1} \\ &= p(X^{-1} - X^{-1}HX^{-1})K[K^T(X^{-1} - X^{-1}HX^{-1})K]^{-p-1}K^T(X^{-1} - X^{-1}HX^{-1}) + o(H) \\ &= \nabla\Phi(X) - p(X^{-1}K)V(S^{g,d} \circ (V^TK^TX^{-1}HX^{-1}KV))V^T(K^TX^{-1}) \\ &\quad - pGHX^{-1} - pX^{-1}HG + o(H), \end{aligned} \tag{27}$$

where G is defined as above. Since X is symmetric, it follows from (26) that

$$\begin{aligned}&\mathbf{vec}((X^{-1}KV)(S^{g,d} \circ (V^TK^TX^{-1}HX^{-1}KV))(V^TK^TX^{-1})) \\ &= [(X^{-1}KV) \otimes (X^{-1}KV)]\,\mathbf{vec}(S^{g,d} \circ (V^TK^TX^{-1}HX^{-1}KV)) \\ &= [(X^{-1}KV) \otimes (X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))\,\mathbf{vec}(V^TK^TX^{-1}HX^{-1}KV)) \\ &= [(X^{-1}KV) \otimes (X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV) \otimes (X^{-1}KV)]^T\,\mathbf{vec}(H). \end{aligned} \tag{28}$$

In addition, since G is symmetric, we further have that

$$\mathbf{vec}(GHX^{-1} + X^{-1}HG) = \left[X^{-1} \otimes G + G \otimes X^{-1}\right]\mathbf{vec}(H). \tag{29}$$

In addition, by virtue of (15), (16), the definition of X and H, and the fact that **svec** is the adjoint operator of **smat**, one can have

$$\nabla\phi(x+h) - \nabla\phi(x) = \mathbf{svec}(\nabla\Phi(X+H) - \nabla\Phi(X)) = Q\,\mathbf{vec}(\nabla\Phi(X+H) - \nabla\Phi(X)).$$

This relation together with (15), (27)–(29) and the definition of H yields

$$\begin{aligned}
\nabla\phi(x+h)-\nabla\phi(x) &= Q(-p[(X^{-1}KV)\otimes(X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV)\otimes(X^{-1}KV)]^T \\
&\quad -p\,X^{-1}\otimes G-p\,G\otimes X^{-1})\,\mathbf{vec}(H)+o(Q\,\mathbf{vec}(H)) \\
&= Q(-p[(X^{-1}KV)\otimes(X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV)\otimes(X^{-1}KV)]^T \\
&\quad -p\,X^{-1}\otimes G-p\,G\otimes X^{-1})Q^T\mathbf{svec}(H)+o(\mathbf{svec}(H)) \\
&= Q(-p[(X^{-1}KV)\otimes(X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV)\otimes(X^{-1}KV)]^T \\
&\quad -p\,X^{-1}\otimes G-p\,G\otimes X^{-1})Q^Th+o(h),
\end{aligned}$$

and hence (20) holds.

For the case when $K=I$, $\nabla^2\phi$ can be directly derived as follows. We know from (21) that $\nabla\Phi(X)=pX^{p-1}$. Letting $g(t)=p\,t^{p-1}$ and $V\mathscr{D}(d)V^T$ be an eigenvalue decomposition of X, it follows from Lemma 4.1 that

$$\nabla\Phi(X+H)=\nabla\Phi(X)+V(S^{g,d}\circ(V^THV))V^T+o(H).$$

In view of (14), one can see that

$$\begin{aligned}
\mathbf{svec}(V(S^{g,d}\circ(V^THV))V^T) &= (V\otimes_s V)\mathbf{svec}(S^{g,d}\circ(V^THV)) \\
&= (V\otimes_s V)[\mathbf{svec}_0(S^{g,d})\circ\mathbf{svec}(V^THV)] \\
&= (V\otimes_s V)(\mathbf{svec}_0(S^{g,d})\circ[(V\otimes_s V)^T\mathbf{svec}(H)]) \\
&= (V\otimes_s V)\mathscr{D}(\mathbf{svec}_0(S^{g,d}))(V\otimes_s V)^T\mathbf{svec}(H).
\end{aligned}$$

Using these relations and a similar proof as above, we can see that (22) holds. ■

As mentioned earlier, we need to know the rank of $\nabla^2\phi(x)$ for performing the partial eigenvalue decomposition of $\nabla^2\phi(x)$ which is used to compute Newton direction. In the next proposition, we determine the rank of $\nabla^2\phi(x)$ at any $x\in\mathbf{svec}(\mathcal{S}^m_{++})$.

Proposition 4.2. *Let Φ be defined in* (18) *and the associated ϕ be defined in* (16)*. Then the rank of $\nabla^2\phi(x)$ is $m(m+1)/2-(m-k)(m-k+1)/2$ for any $x\in\mathbf{svec}(\mathcal{S}^m_{++})$.*

Proof. Let $x\in\mathbf{svec}(\mathcal{S}^m_{++})$ be arbitrarily chosen. Define $X=\mathbf{smat}(x)$. Let G, V, d and $S^{g,d}$ be defined in Proposition 4.1 with $g(t)=t^{-p-1}$. For convenience, we define

$$\begin{aligned}
M_1 &= [(X^{-1}KV)\otimes(X^{-1}KV)]\mathscr{D}(\mathbf{vec}(S^{g,d}))[(X^{-1}KV)\otimes(X^{-1}KV)]^T, \\
M_2 &= X^{-1}\otimes G+G\otimes X^{-1}.
\end{aligned}$$

To determine the rank of $\nabla^2\phi(x)$, it suffices to know the dimension of the null space of $\nabla^2\phi(x)$, denoted by $\mathrm{Null}(\nabla^2\phi(x))$. Notice that ϕ is a twice differentiable convex function in $\mathbf{svec}(\mathcal{S}^m_{++})$. Thus, $\nabla^2\phi(x)\succeq 0$. It implies that $h\in\mathrm{Null}(\nabla^2\phi(x))$ if and only if $h^T\nabla^2\phi(x)h=0$. We will subsequently show that

$$h^T\nabla^2\phi(x)h=0 \;\Leftrightarrow\; K^TX^{-1}H=0, \tag{30}$$

where $H=\mathbf{smat}(h)$. It then follows that

$$h\in\mathrm{Null}(\nabla^2\phi(x))\Leftrightarrow\; K^TX^{-1}H=0.$$

Notice that K^TX^{-1} has full row rank. Thus, there exist nonsingular matrices E_1 and E_2 such that $K^TX^{-1}=E_1\begin{pmatrix}I & 0\end{pmatrix}E_2$, where I is the identity matrix of order k. It then follows that

$$K^TX^{-1}H=0 \;\Leftrightarrow\; \begin{pmatrix}I & 0\end{pmatrix}U=0,$$

where $U=E_2HE_2^T\in\mathcal{S}^m$. It is easy to see that the dimension of $\{U\in\mathcal{S}^m:\; \begin{pmatrix}I & 0\end{pmatrix}U=0\}$ is $(m-k)(m-k+1)/2$. Since E_2 is invertible, we conclude that the dimension of $\{H\in\mathcal{S}^m:$

$K^TX^{-1}H = 0\}$ is also $(m-k)(m-k+1)/2$. Since **smat** is a one-to-one map between $\Re^{m(m+1)/2}$ and $\mathcal{S}^m$, the dimension of $\mathrm{Null}(\nabla^2\phi(x))$ is $(m-k)(m-k+1)/2$, and hence the rank of $\nabla^2\phi(x)$ is $m(m+1)/2-(m-k)(m-k+1)/2$. To complete the proof, we next show that (30) holds by considering two cases $p \le -1$ or $-1 < p < 0$.

We start with the first case $p \le -1$. Notice that all entries of $S^{g,d}$ are nonnegative and thus $M_1 \succeq 0$. Also, $M_2 \succeq 0$. It then follows from (20) and (15) that $h^T\nabla^2\phi(x)h = 0$ if and only if

$$\mathbf{vec}(H)^T M_1\, \mathbf{vec}(H) = 0, \quad \mathbf{vec}(H)^T M_2\, \mathbf{vec}(H) = 0. \tag{31}$$

By (29), the second equality of (31) becomes

$$\mathrm{tr}(HX^{-1}K(K^TX^{-1}K)^{-p-1}K^TX^{-1}HX^{-1}) = 0,$$

which is equivalent to

$$\begin{aligned} &\mathrm{tr}(X^{-\frac{1}{2}}HX^{-1}K(K^TX^{-1}K)^{-\frac{p+1}{2}}(K^TX^{-1}K)^{-\frac{p+1}{2}}K^TX^{-1}HX^{-\frac{1}{2}}) = 0 \\ &\Leftrightarrow\ (K^TX^{-1}K)^{-\frac{p+1}{2}}K^TX^{-1}HX^{-\frac{1}{2}} = 0 \ \Leftrightarrow\ K^TX^{-1}H = 0. \end{aligned} \tag{32}$$

Moreover, $K^TX^{-1}H = 0$ implies that the first equality of (31) holds. Therefore, (31) holds if and only if $K^TX^{-1}H = 0$. It follows that (30) holds for $p \le -1$.

We next show that (30) also holds for $-1 < p < 0$. Indeed, for such p, all entries of $S^{g,d}$ are negative and hence $-M_1 \succeq 0$. Using Proposition 4.1, we see that $h^T\nabla^2\phi(x)h = 0$ if and only if

$$\mathbf{vec}(H)^T(M_1 + M_2)\, \mathbf{vec}(H) = 0. \tag{33}$$

We claim that

$$\frac{1}{2}\,\mathbf{vec}(H)^T M_2\, \mathbf{vec}(H) \ge -\,\mathbf{vec}(H)^T M_1\, \mathbf{vec}(H). \tag{34}$$

Indeed, letting $W = (K^TX^{-1}K)^{-1}$ and using Lemma 3.1, we have

$$W^{-1} = K^TX^{-1}K \ \Rightarrow\ \begin{pmatrix} X & K \\ K^T & W^{-1} \end{pmatrix} \succeq 0 \ \Rightarrow\ X \succeq KWK^T.$$

The latter relation together with the definitions of M_2, G and (29) implies that

$$\begin{aligned} \tfrac{1}{2}\,\mathbf{vec}(H)^T M_2\, \mathbf{vec}(H) \ &= \ \mathrm{tr}(HX^{-1}HX^{-1}KW^{p+1}K^TX^{-1}) \\ &= \ \mathrm{tr}([X^{-1}H(X^{-1}KW^{p+1}K^TX^{-1})^{\frac{1}{2}}]^T X[X^{-1}H(X^{-1}KW^{p+1}K^TX^{-1})^{\frac{1}{2}}]) \\ &\ge \ \mathrm{tr}([X^{-1}H(X^{-1}KW^{p+1}K^TX^{-1})^{\frac{1}{2}}]^T KWK^T[X^{-1}H(X^{-1}KW^{p+1}K^TX^{-1})^{\frac{1}{2}}]) \\ &= \ \mathrm{tr}(HX^{-1}KWK^TX^{-1}HX^{-1}KW^{p+1}K^TX^{-1}) \end{aligned} \tag{35}$$

Let $Z = V^TK^TX^{-1}HX^{-1}KV$. Notice that $W = V\mathscr{D}(d^{-1})V^T$. Using this relation, the definition of Z and (26), we have

$$\begin{aligned} \mathrm{tr}(HX^{-1}KWK^TX^{-1}HX^{-1}KW^{p+1}K^TX^{-1}) \ &= \ \mathrm{tr}(HX^{-1}KV\mathscr{D}(d^{-1})Z\mathscr{D}(d^{-p-1})V^TK^TX^{-1}) \\ = \ \mathrm{tr}(\mathscr{D}(d^{-1})Z\mathscr{D}(d^{-p-1})Z) \ &= \ \mathbf{vec}(Z)^T[\mathscr{D}(d^{-1}) \otimes \mathscr{D}(d^{-p-1})]\,\mathbf{vec}(Z), \end{aligned}$$

which together with (35) yields

$$\frac{1}{2}\,\mathbf{vec}(H)^T M_2\, \mathbf{vec}(H) \ \ge \ \mathbf{vec}(Z)^T[\mathscr{D}(d^{-1}) \otimes \mathscr{D}(d^{-p-1})]\,\mathbf{vec}(Z). \tag{36}$$

Also, by the definitions of M_1, Z and (26), we obtain that

$$-\,\mathbf{vec}(H)^T M_1\, \mathbf{vec}(H) = \mathbf{vec}(Z)^T\mathscr{D}(\mathbf{vec}(-S^{g,d}))\,\mathbf{vec}(Z). \tag{37}$$

Since $1 > p+1 > 0$ and $d_i > 0$ for all i, it is not hard to show that

$$d_i^{-1}d_j^{-p-1} \ge -\frac{d_i^{-p-1}-d_j^{-p-1}}{d_i - d_j},$$

whenever $d_i \ne d_j$. Thus, $\mathscr{D}(d^{-1}) \otimes \mathscr{D}(d^{-p-1}) \succeq \mathscr{D}(\mathbf{vec}(-S^{g,d}))$, which together with (36) and (37) implies that (34) holds. It then follows from (34), (33) and the fact $M_2 \succeq 0$ that $\mathbf{vec}(H)^T M_2\, \mathbf{vec}(H) = 0$. The rest of proof is similar to the case $p \le -1$. ∎

4.2 IP method for A-criterion

Recall from Section 1 that in $\mathcal{S}^m_{++}$, the A-criterion Φ becomes

$$\Phi(X) = \mathrm{tr}(K^T X^{-1} K) \tag{38}$$

for some $K \in \Re^{m\times k}$ with full column rank. Since A-criterion is a special case of pth mean criterion, the IP method discussed in Sections 3 and 4.1 can be suitably applied to solve problem (1) with A-criterion. We next show that by exploiting the special structure, we can obtain a more compact representation of the associated Hessian matrix that is used to compute Newton direction for our IP method.

Proposition 4.3. *Let Φ be defined in* (38) *and the associated ϕ be defined in* (16)*. Then the gradient and Hessian of ϕ at any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$ are given by*

$$\nabla\phi(x) = -\mathbf{svec}(X^{-1}KK^TX^{-1}), \tag{39}$$

$$\nabla^2\phi(x) = 2X^{-1}\otimes_s(X^{-1}KK^TX^{-1}), \tag{40}$$

where $X = \mathbf{smat}(x)$.

Proof. (39) follows immediately from (19) with $p = -1$. We now prove (40). Let $x \in \mathbf{svec}(\mathcal{S}^m_{++})$ be arbitrarily chosen, and let $X = \mathbf{smat}(x)$. For all sufficiently small $h \in \Re^{m(m+1)/2}$, we observe $X + H \succ 0$, where $H = \mathbf{smat}(h)$. In view of the definitions of X and H, it then follows from (39), (23) and (14) that

$$\begin{aligned}
\nabla\phi(x+h) - \nabla\phi(x) &= -\mathbf{svec}\left((X+H)^{-1}KK^T(X+H)^{-1} - X^{-1}KK^TX^{-1}\right) \\
&= \mathbf{svec}(X^{-1}HX^{-1}KK^TX^{-1} + X^{-1}KK^TX^{-1}HX^{-1}) + o(\mathbf{svec}(H)) \\
&= 2X^{-1}\otimes_s(X^{-1}KK^TX^{-1})h + o(h),
\end{aligned}$$

which proves (40). ∎

Since the A-criterion is a special case of the pth mean criterion, it follows from Proposition 4.2 that the rank of $\nabla^2\phi(x)$ is also $m(m+1)/2 - (m-k)(m-k+1)/2$ for every $x \in \mathbf{svec}(\mathcal{S}^m_{++})$.

4.3 IP method for D-criterion

Recall from Section 1 that in $\mathcal{S}^m_{++}$, the D-criterion Φ becomes

$$\Phi(X) = \log\det(K^TX^{-1}K) \tag{41}$$

for some $K \in \Re^{m\times k}$ with full column rank. It is easy to verify that Assumption 3.1 is satisfied. Hence, problem (1) with this criterion can be suitably solved by the IP method studied in Section 3. In the next proposition, we provide formulas for computing gradient and Hessian of the associated function ϕ that are used in the IP method. The proof is similar to that of Proposition 4.3 and is thus omitted.

Proposition 4.4. *Let Φ be defined in* (41) *and the associated ϕ be defined in* (16)*. Then the gradient and Hessian of ϕ at any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$ are given by*

$$\begin{aligned}
\nabla\phi(x) &= -\mathbf{svec}(X^{-1}KWK^TX^{-1}), \\
\nabla^2\phi(x) &= 2X^{-1}\otimes_s(X^{-1}KWK^TX^{-1}) - (X^{-1}KWK^TX^{-1})\otimes_s(X^{-1}KWK^TX^{-1}),
\end{aligned} \tag{42}$$

where $X = \mathbf{smat}(x)$ and $W = (K^TX^{-1}K)^{-1}$.

We next determine the rank of $\nabla^2\phi(X)$ at any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$.

Proposition 4.5. *Let Φ be defined in* (41) *and the associated ϕ be defined in* (16)*. Then the rank of $\nabla^2\phi(x)$ is $m(m+1)/2-(m-k)(m-k+1)/2$ for any $x \in \mathbf{svec}(\mathcal{S}^m_{++})$.*

Proof. Let $x \in \mathbf{svec}(\mathcal{S}^m_{++})$ be arbitrarily chosen. Define $X = \mathbf{smat}(x)$. As in the proof of Proposition 4.2, to determine the rank of $\nabla^2\phi(x)$, it suffices to know the dimension of $\mathrm{Null}(\nabla^2\phi(x))$. Notice that ϕ is a twice differentiable convex function in $\mathbf{svec}(\mathcal{S}^m_{++})$. Thus, $\nabla^2\phi(x) \succeq 0$. It implies that $h \in \mathrm{Null}(\nabla^2\phi(x))$ if and only if $h^T\nabla^2\phi(x)h = 0$. In view of (14) and (42), it is not hard to verify that $h^T\nabla^2\phi(x)h = 0$ if and only if

$$2\,\mathrm{tr}(HX^{-1}HX^{-1}KWK^TX^{-1}) - \mathrm{tr}(HX^{-1}KWK^TX^{-1}HX^{-1}KWK^TX^{-1}) = 0,$$

where $H = \mathbf{smat}(h)$. In addition, we can observe that (35) also holds for $p = 0$, and hence

$$\mathrm{tr}(HX^{-1}HX^{-1}KWK^TX^{-1}) \;\geq\; \mathrm{tr}(HX^{-1}KWK^TX^{-1}HX^{-1}KWK^TX^{-1}).$$

Furthermore,

$$\begin{aligned} & \mathrm{tr}(HX^{-1}KWK^TX^{-1}HX^{-1}KWK^TX^{-1}) \\ = \;& \mathrm{tr}\left([(X^{-1}KWK^TX^{-1})^{\frac{1}{2}}H]X^{-1}KWK^TX^{-1}[H(X^{-1}KWK^TX^{-1})^{\frac{1}{2}}]\right) \;\geq\; 0. \end{aligned}$$

The above relations imply that $h^T\nabla^2\phi(x)h = 0$ if and only if

$$\mathrm{tr}(HX^{-1}HX^{-1}KWK^TX^{-1}) = 0,$$

which together with definition of W and the same arguments used in (32) implies that

$$h^T\nabla^2\phi(x)h = 0 \;\Leftrightarrow\; K^TX^{-1}H = 0.$$

The rest of the proof follows similarly as that of Proposition 4.2. ∎

5 Computational results

In this section, we conduct numerical experiments to test the performance of the IP method discussed in this paper for solving problem (1) with A-, D- and pth mean criterion and also compare its performance with the multiplicative algorithm.

We develop Matlab codes for our IP method to solve (1) with A-, D- and pth mean criterion. We also implement the multiplicative algorithm in Matlab for solving (1) with A-, D- and pth mean criterion. To benchmark the performance of our IP method, we also report the computational results using a general SDP solver, namely, SDPT3 [32, 36] (Version 4.0) on solving a linear SDP reformulation of (1) with A-criterion (see [14, Page 532]) and a log-determinant SDP reformulation of (1) with D-criterion (see [23, Equation (10)]). We shall mention that it is not clear whether problem (1) with pth mean criterion can be reformulated into a problem that can be efficiently solved by SDPT3. As SDPT3 implements an infeasible path-following algorithm, we project the approximate solution w found by SDPT3 onto the unit simplex to obtain an approximate optimal feasible solution for problem (1) and the final objective value reported in our tests is based on the latter solution. [2] All computations in this section are performed in Matlab 7.14.0 (2012a) on a workstation with an Intel Xeon E5410 CPU (2.33 GHz) and 8GB RAM running Red Hat Enterprise Linux (kernel 2.6.18).

For our IP method, we set $\tilde{w}^0 = \frac{1}{n}e \in \Re^{n-1}$, $\mu_1 = 10$, $\beta = \gamma = 0.5$, $\sigma = 0.1$ and $\eta = 0.95$. In addition, we set $\epsilon(\mu) = \max\{\mu, 1e-10\}$ and terminate the algorithm once $\mu_k \leq 1e-10$. On the

[2] Such projection makes a difference when SDPT3 terminates early at a solution that is highly infeasible, which could be a consequence of "near infeasibility" of the linear SDP reformulation; see the first three rows of Table 1.

other hand, for the multiplicative algorithm, similarly as in [41], we set $\lambda = 1$, $w^0 = \frac{1}{n}e \in \Re^n$, and terminate the algorithm when it reaches 10000 iterations or

$$\max_{1 \le i \le n} d_i(w^k) \le (1+\delta) \sum_{i=1}^{n} w_i^k d_i(w^k)$$

holds with $\delta = 2e-4$, where $d_i(w)$ is defined in (2) [3]. Furthermore, for SDPT3, we use the default tolerance. Finally, we use the mex files skron, smat and svec from the SDPT3 package for efficient operations on symmetric matrices in our implementation of the IP method and the multiplicative algorithm.

In our tests below, we consider the following four design spaces:

$$\begin{array}{ll} \chi_1(n) & = \{x_i = (e^{-s_i}, s_i e^{-s_i}, e^{-2s_i}, s_i e^{-2s_i})^T, 1 \le i \le n\}, \\ \chi_2(n) & = \{x_i = (1, s_i, s_i^2, s_i^3)^T, 1 \le i \le n\}, \\ \chi_3(n) & = \{x_{(i-1)\lceil\sqrt{n}\rceil + j} = (1, r_i, r_i^2, t_j, r_i t_j)^T, 1 \le i, j \le \lceil\sqrt{n}\rceil\}, \\ \chi_4(n) & = \{x_i = (t_i, t_i^2, \sin(2\pi t_i), \cos(2\pi t_i))^T, 1 \le i \le n\}, \end{array}$$

where $s_i = \frac{3i}{n}$, $r_i = \frac{2i}{n} - 1$ and $t_i = \frac{i}{n}$. The space $\chi_1(n)$ represents the linearization of a compartmental model [4]. The space $\chi_2(n)$ corresponds to polynomial regression. The third space, as described in [42], represents a response surface with a nonlinear effect and an interaction, while the fourth space is the quadratic/trigonometric example proposed in [40]. The test sets χ_1, χ_3 and a variant of the test set χ_2 are also used in [42].

In our first test, for each design space, we set $A_i = x_i x_i^T$ for $i = 1, \ldots, n$, with $n = 10000$, 50000, 100000 for χ_1, χ_2, χ_4, and $n = 10000$, 40000, 90000 for χ_3. For each n and each design space, we randomly generate 30 different matrices $K \in \Re^{m\times 3}$ (i.e., we set $k = 3$), each having i.i.d. Gaussian entries of mean 0 and variance 1. We then apply our IP method and the multiplicative algorithm to solve problem (1) with A-, D- and pth mean criterion on these instances and also apply SDPT3 to solve (1) with A- and D-criterion. The computational results averaged over the 30 instances are reported in Tables 1–4. In particular, the performance of our IP method, the multiplicative algorithm and SDPT3 are reported under the columns named "IP", "MUL" and "SDPT3", respectively. In addition, the CPU time abbreviated as "cpu" is in seconds and the objective value abbreviated as "obj" is rounded off to six significant digits. We see that our IP method significantly outperforms the multiplicative algorithm in terms of CPU time, and gives a smaller objective value in all instances. Moreover, our IP method also outperforms SDPT3 in CPU time and gives a smaller objective value in most instances. Furthermore, it is worth pointing out that SDPT3 reports infeasibility and hence *early terminates* when solving some instances for χ_1 with A-criterion, possibly due to bad scaling of $\mathcal{M}(w)$. This accounts for its significantly larger objective values in Table 1 corresponding to χ_1. Finally, for pth mean criterion with $p < -1$, our IP method achieves significantly better objective values than the multiplicative algorithm, where the objective value of the latter algorithm is chosen to be the minimum over all iterations (see Table 4). This phenomenon is actually not surprising since the multiplicative algorithm is only known to converge for $p \in (-1, 0)$, but it may not converge when $p < -1$.

In our second test, we consider the case when $K = I$. The instances used in this test are the same as those in the first test except $K = I$. We also apply our IP method and the multiplicative algorithm to solve problem (1) with A-, D- and pth mean criterion on these instances and apply SDPT3 to solve (1) with A- and D-criterion. The computational results are reported in Tables 5–8. We again observe that our IP method outperforms the multiplicative algorithm in terms of objective value in all instances, and is generally much faster on large instances. Furthermore, our IP method is usually faster than SDPT3 and produces comparable or smaller objective values.

[3] We also tried $\delta = 1e-4$, but the multiplicative algorithm tends to take a long time for relatively little improvement on some instances.

Table 1: Computational results for A-criterion with random K

		cpu			obj		
χ_i	n	MUL	IP	SDPT3	MUL	IP	SDPT3
1	10000	13.76	0.69	1.93	193041	191410	211735
1	50000	62.13	3.75	10.64	154584	153219	172514
1	100000	135.26	7.38	19.22	208599	206787	242633
2	10000	17.69	0.77	1.90	215.754	212.356	212.356
2	50000	89.89	4.22	10.94	188.509	185.55	185.551
2	100000	163.28	8.12	21.83	242.414	237.823	237.824
3	10000	33.74	1.02	2.36	54.8551	54.7332	54.7332
3	40000	140.85	5.24	14.17	49.0008	48.9784	48.9791
3	90000	322.52	10.75	35.50	50.8124	50.7906	50.7906
4	10000	13.42	0.90	1.90	572.779	558.088	558.088
4	50000	58.41	4.46	9.89	501.924	487.99	487.991
4	100000	139.62	9.03	20.36	343.827	337.003	337.023

Table 2: Computational results for D-criterion with random K

		cpu			obj		
χ_i	n	MUL	IP	SDPT3	MUL	IP	SDPT3
1	10000	1.47	0.95	1.47	19.7352	19.7347	19.7356
1	50000	5.40	4.73	6.03	19.9312	19.9307	19.933
1	100000	14.17	9.31	12.34	19.7973	19.7968	19.7987
2	10000	2.10	0.79	1.57	5.95269	5.95229	5.9523
2	50000	20.60	4.07	6.78	5.30436	5.3039	5.3039
2	100000	51.43	8.33	13.32	5.08652	5.08608	5.08609
3	10000	4.31	1.05	1.77	6.58713	6.58694	6.58694
3	40000	18.86	4.28	9.11	6.65124	6.65104	6.65103
3	90000	66.40	9.93	21.48	6.74346	6.74327	6.74382
4	10000	1.67	0.90	1.41	7.40587	7.40535	7.40535
4	50000	13.46	4.35	6.09	7.65401	7.6535	7.6535
4	100000	39.01	7.90	12.03	8.66619	8.66575	8.66574

Table 3: Computational results for pth mean criterion with random K for some $p \in (-1, 0)$

		$p = -0.25$				$p = -0.75$			
		cpu		obj		cpu		obj	
χ_i	n	MUL	IP	MUL	IP	MUL	IP	MUL	IP
1	10000	6.39	0.85	25.4567	25.4558	9.17	0.79	5187.73	5187.23
1	50000	36.84	4.18	25.1902	25.1894	45.72	4.09	7128.51	7126.32
1	100000	82.59	8.29	25.1312	25.1304	118.04	8.11	7207.59	7205.68
2	10000	6.19	0.76	5.68067	5.68046	13.67	0.79	46.4144	46.4008
2	50000	28.10	3.89	6.00911	6.00886	73.65	4.28	58.2028	58.1903
2	100000	71.37	8.00	6.12458	6.12434	152.84	8.30	60.9256	60.9108
3	10000	5.47	0.98	5.58387	5.58379	3.73	1.07	24.8691	24.868
3	40000	18.97	4.18	5.56727	5.56718	17.21	4.66	23.5463	23.5451
3	90000	58.35	9.91	5.45907	5.45899	56.66	10.98	24.7679	24.7664
4	10000	1.84	0.90	7.30484	7.30456	2.88	0.90	118.871	118.859
4	50000	8.84	4.63	7.27622	7.27589	10.98	4.97	108.079	108.066
4	100000	31.14	8.81	7.30129	7.30102	45.27	9.70	128.676	128.662

6 Concluding remarks

In this paper we propose an IP method for solving problem (1) with a broad class of convex optimality criteria and establish its global convergence. We demonstrate how the Newton direction

Table 4: Computational results for pth mean criterion with random K for some $p < -1$

		$p = -1.1$				$p = -1.2$			
		cpu		obj		cpu		obj	
χ_i	n	mul	IP	mul	IP	mul	IP	mul	IP
1	10000	5.07	0.70	611960	602294	4.66	0.68	1.46813e+06	1.43891e+06
1	50000	20.64	3.59	541355	532904	19.71	3.59	2.0649e+06	2.02777e+06
1	100000	48.85	7.47	371942	365201	51.33	7.35	1.79042e+06	1.75803e+06
2	10000	6.25	0.80	373.376	359.802	5.34	0.79	650.345	629.288
2	50000	21.38	4.12	492.463	476.123	21.15	4.13	667.047	645
2	100000	61.53	8.44	302.083	288.88	62.32	8.41	539.641	514.087
3	10000	19.16	1.11	74.7421	71.4397	20.40	1.18	95.224	88.142
3	40000	71.26	4.76	69.2354	65.8478	68.95	4.79	127.857	116.933
3	90000	204.48	11.78	69.2994	65.8143	160.07	11.85	109.287	100.475
4	10000	6.75	0.92	961.75	910.571	7.17	0.96	1640.19	1524.98
4	50000	27.46	4.91	903.773	846.954	36.12	5.02	1631.8	1520.71
4	100000	75.39	9.85	824.269	776.036	75.90	9.82	1710.28	1596.1

Table 5: Computational results for A-criterion with $K = I$

		cpu			obj		
χ_i	n	mul	IP	SDPT3	mul	IP	SDPT3
1	10000	13.69	0.74	2.29	54286.3	53848.3	53848.4
1	50000	62.08	4.17	12.23	54245.2	53807.3	54103.8
1	100000	133.65	7.37	27.46	54240.1	53802.1	54103.8
2	10000	16.53	0.81	1.82	73.4521	72.4443	72.4443
2	50000	75.99	4.26	11.03	73.391	72.385	72.3853
2	100000	164.18	8.60	20.23	73.3837	72.3778	72.3777
3	10000	1.58	0.93	2.13	21.6203	21.6191	21.6191
3	40000	12.81	4.38	11.38	21.2826	21.2812	21.2812
3	90000	36.66	9.14	30.21	21.1721	21.1706	21.1706
4	10000	12.84	0.96	1.58	174.279	170.775	170.775
4	50000	59.76	5.19	9.51	174.276	170.775	170.775
4	100000	128.73	9.93	17.13	174.277	170.775	170.776

Table 6: Computational results for D-criterion with $K = I$

		cpu			obj		
χ_i	n	mul	IP	SDPT3	mul	IP	SDPT3
1	10000	1.11	1.02	0.87	20.5125	20.5119	20.5125
1	50000	4.86	4.67	3.86	20.5098	20.5091	20.5091
1	100000	14.77	9.13	7.59	20.5094	20.5087	20.5088
2	10000	1.92	0.74	1.01	0.410745	0.410221	0.41022
2	50000	16.74	3.80	4.75	0.409964	0.409267	0.40926
2	100000	55.28	6.96	8.89	0.409795	0.409154	0.409145
3	10000	1.76	0.89	1.16	5.14292	5.14267	5.14267
3	40000	15.90	3.99	6.53	5.08236	5.08212	5.08211
3	90000	47.07	8.70	15.38	5.06226	5.06202	5.06201
4	10000	1.35	1.02	0.94	7.25257	7.25189	7.25189
4	50000	11.16	5.04	4.17	7.25253	7.2519	7.25189
4	100000	35.09	9.86	8.14	7.25246	7.2519	7.25189

can be efficiently computed when the method is applied to (1) with classical optimality criteria. Our computational results show that the IP method outperforms the widely used multiplicative

Table 7: Computational results for pth mean criterion with $K = I$ for some $p \in (-1, 0)$

		$p = -0.25$				$p = -0.75$			
		cpu		obj		cpu		obj	
χ_i	n	mul	IP	mul	IP	mul	IP	mul	IP
1	10000	7.48	0.91	23.3728	23.372	3.53	0.82	3635.71	3635.29
1	50000	42.29	4.21	23.3683	23.3675	24.12	4.37	3633.58	3633.2
1	100000	91.80	8.29	23.3677	23.367	57.01	8.86	3633.31	3632.94
2	10000	3.43	0.74	5.58855	5.58838	2.55	0.80	27.4836	27.4811
2	50000	20.56	4.00	5.58796	5.58771	11.69	4.14	27.4691	27.4653
2	100000	69.43	7.67	5.58785	5.58763	37.49	8.60	27.467	27.4634
3	10000	1.65	0.77	6.70457	6.70448	1.56	0.99	14.1435	14.1429
3	40000	14.46	4.24	6.68235	6.68225	13.22	3.82	13.9841	13.9834
3	90000	42.01	8.24	6.675	6.67491	37.85	9.25	13.9318	13.9311
4	10000	1.75	0.92	7.25984	7.25955	1.43	0.92	52.2922	52.286
4	50000	8.97	4.58	7.25988	7.25956	6.05	4.52	52.2937	52.286
4	100000	30.50	8.85	7.25983	7.25957	20.67	9.20	52.2927	52.2861

Table 8: Computational results for pth mean criterion with $K = I$ for some $p < -1$

		$p = -1.1$				$p = -1.2$			
		cpu		obj		cpu		obj	
χ_i	n	mul	IP	mul	IP	mul	IP	mul	IP
1	10000	4.62	0.72	162818	159210	4.64	0.67	485415	471459
1	50000	18.15	3.64	162740	159077	18.28	3.63	482380	471030
1	100000	49.28	7.47	162732	159060	47.57	7.56	485149	470975
2	10000	4.59	0.79	108.922	108.171	4.65	0.80	165.133	162.297
2	50000	18.44	4.19	109.588	108.072	19.23	4.27	164.314	162.134
2	100000	47.95	8.66	109.495	108.06	45.78	8.87	165.458	162.114
3	10000	36.72	1.02	25.9565	25.7793	36.40	1.04	31.8264	30.8276
3	40000	142.62	4.17	25.599	25.3307	139.85	4.50	31.5254	30.2362
3	90000	328.00	9.60	25.5115	25.1841	322.93	9.68	31.46	30.0431
4	10000	6.51	0.95	297.604	277.597	7.99	0.89	497.138	453
4	50000	27.17	4.83	297.686	277.597	33.88	4.89	497.287	453
4	100000	63.40	9.57	297.696	277.597	81.00	10.02	497.306	453

algorithm in both speed and solution quality. The codes for this paper, including our implementation of the multiplicative algorithm and our codes generating inputs for SDPT3, are available online at www.math.sfu.ca/∼zhaosong.

Finally, we would like to remark that the performance of our IP method depends on whether the Newton direction can be computed accurately and efficiently. In our implementation, we observe that for pth mean criterion with large $|p|$, as well as for the design space $\{x_i = (1, s_i, s_i^2, s_i^3, s_i^4)^T, 1 \leq i \leq n\}$ with $n \geq 50000$ and some random $K \in \Re^{m\times 3}$, the Newton direction cannot be computed accurately due to numerical errors and hence our IP method fails to terminate with a good approximate solution, compared with the multiplicative algorithm. Indeed, it is known [37, 38] that the performance of a barrier method deteriorates as $\mu \to 0$. It is conceivable that such issues would not arise if a primal-dual IP method was used instead. However, it is much more involved to develop a primal-dual IP method for solving (1): since the feasible set of (1) is not closed in general, one would have to develop a primal-dual IP method on an equivalent nonlinear semidefinite programming reformulation of (1). We leave this as a future research direction.

Appendix

We present the proof of Theorem 3.1 in this appendix.

Proof. In this proof, we denote by $\tilde{w}(\mu, v)$ the vector obtained from $w(\mu, v)$ by dropping the last entry for all $(\mu, v) \in \Re_{++} \times \Re^{n-1}$. Notice that $\tilde{w}(\mu, v)$ is the unique optimal solution of (12).

We now prove part (a). Let

$$\bar{f}^* := \inf_w \{\Phi_{\mathcal{M}}(w) : \ e^T w = 1, w > 0\}. \tag{43}$$

We first show that $\lim\limits_{(\mu,v)\to(0_+,0)} \Phi_{\mathcal{M}}(w(\mu, v)) = \bar{f}^*$.

Given an arbitrary $\epsilon > 0$, there exists a positive $\tilde{w}$ satisfying $e^T\tilde{w} < 1$ such that $f(\tilde{w}) < \bar{f}^* + \epsilon/2$. Then we have that for any $v \in \Re^{n-1}$,

$$f_\mu(\tilde{w}(\mu, v)) - v^T\tilde{w}(\mu, v) \le f_\mu(\tilde{w}) - v^T\tilde{w}. \tag{44}$$

On the other hand, note that $\tilde{w}(\mu, v) > 0$ and $e^T\tilde{w}(\mu, v) < 1$. Hence,

$$-\sum_{i=1}^{n-1} \log(\tilde{w}_i(\mu, v)) - \log\left(1 - e^T\tilde{w}(\mu, v)\right) > 0$$

and $f(\tilde{w}(\mu, v)) \ge \bar{f}^*$. In view of these inequalities, (44) and the fact that $\|\tilde{w}(\mu, v)\|_1 \le 1$ and $\|\tilde{w}\|_1 \le 1$, one can obtain that for any $(\mu, v) \in \Re_{++} \times \Re^{n-1}$,

$$\begin{aligned}
\bar{f}^* \le f(\tilde{w}(\mu, v)) &= f_\mu(\tilde{w}(\mu, v)) + \mu\sum_{i=1}^{n-1} \log(\tilde{w}_i(\mu, v)) + \mu\log\left(1 - e^T\tilde{w}(\mu, v)\right) \\
&\le f_\mu(\tilde{w}(\mu, v)) \ \le \ f_\mu(\tilde{w}) + v^T\tilde{w}(\mu, v) - v^T\tilde{w} \\
&\le f(\tilde{w}) - \mu\sum_{i=1}^{n-1} \log(\tilde{w}_i) - \mu\log\left(1 - e^T\tilde{w}\right) + 2\|v\|_\infty \\
&\le \bar{f}^* + \frac{\epsilon}{2} - \mu\sum_{i=1}^{n-1} \log(\tilde{w}_i) - \mu\log\left(1 - e^T\tilde{w}\right) + 2\|v\|_\infty.
\end{aligned}$$

Thus, there exists some $\delta > 0$ such that $\bar{f}^* \le f(\tilde{w}(\mu, v)) \le \bar{f}^* + \epsilon$ whenever $\|(\mu, v)\| < \delta$, $\mu > 0$. Hence, $\Phi_{\mathcal{M}}(w(\mu, v)) = f(\tilde{w}(\mu, v)) \to \bar{f}^*$ as $(\mu, v) \to (0_+, 0)$.

We next show that $f^* = \bar{f}^*$. Clearly, $f^* \le \bar{f}^*$. We now suppose for contradiction that $f^* < \bar{f}^*$. By the definitions of f^* and $\bar{f}^*$, there exist w^1 and w^2 which are feasible points of (1) and (43), respectively, so that $\Phi_{\mathcal{M}}(w^1) < (f^* + \bar{f}^*)/2$ and $\Phi_{\mathcal{M}}(w^2) < \bar{f}^* + (\bar{f}^* - f^*)/2$. Let $w = (w^1 + w^2)/2$. Clearly, $w > 0$, $e^T w = 1$ and $\mathrm{Range}(K) \subseteq \mathrm{Range}(\mathcal{M}(w))$ due to $\mathcal{M}(w) \succ 0$. By convexity of Φ in $\mathcal{S}^m_+(K)$, we obtain that $\Phi_{\mathcal{M}}(w) \le (\Phi_{\mathcal{M}}(w^1) + \Phi_{\mathcal{M}}(w^2))/2 < \bar{f}^*$, which is a contradiction to the definition of $\bar{f}^*$. Thus, $\lim\limits_{(\mu,v)\to(0_+,0)} \Phi_{\mathcal{M}}(w(\mu, v)) = \bar{f}^* = f^*$.

Now suppose that w^* is an accumulation point of $w(\mu, v)$ as $(\mu, v) \to (0_+, 0)$. We next show that w^* is an optimal solution of (1). Indeed, it follows from (4) that for any feasible point w of (1),

$$\Phi_{\mathcal{M}}(w) = \inf_U \left\{\Psi(U) : \ \mathcal{M}(w) \succeq KUK^T, U \succ 0\right\}. \tag{45}$$

In view of (45), for each $(\mu, v) \in \Re_{++} \times \Re^{n-1}$, there exists $U(\mu, v) \succ 0$ such that

$$\Phi_{\mathcal{M}}(w(\mu, v)) + \|(\mu, v)\| > \Psi(U(\mu, v)) \quad \text{and} \quad \mathcal{M}(w(\mu, v)) \succeq KU(\mu, v)K^T. \tag{46}$$

From the second relation in (46), we see that $\mathrm{tr}(\mathcal{M}(w(\mu, v))) \ge \lambda_{\min}(K^TK)\mathrm{tr}(U(\mu, v))$, from which it follows that $U(\mu, v)$ is bounded and thus it has an accumulation point as $(\mu, v) \to (0_+, 0)$. Let

U^* be such an accumulation point. In view of the first relation in (46) and the assumption on Ψ, we see that $U^* \succ 0$. Moreover, we obtain by taking limit in (46) upon $(\mu, v) \to (0_+, 0)$ that

$$\lim_{(\mu,v)\to(0_+,0)} \Phi_{\mathcal{M}}(w(\mu,v)) \geq \Psi(U^*), \quad \mathcal{M}(w^*) \succeq KU^*K^T. \tag{47}$$

The second relation in (47) together with Lemma 3.1 implies that

$$\mathcal{M}(w^*) \succeq KU^*K^T \;\Rightarrow\; \begin{pmatrix} U^{*-1} & K^T \\ K & \mathcal{M}(w^*) \end{pmatrix} \succeq 0 \;\Rightarrow\; \mathrm{Range}(K) \subseteq \mathrm{Range}(\mathcal{M}(w^*)).$$

Hence, w^* is a feasible point of (1). In view of (45), the first relation in (47) and the result $\lim\limits_{(\mu,v)\to(0_+,0)} \Phi_{\mathcal{M}}(w(\mu,v)) = f^*$, we have

$$\Phi_{\mathcal{M}}(w^*) \leq \Psi(U^*) \leq \lim_{(\mu,v)\to(0_+,0)} \Phi_{\mathcal{M}}(w(\mu,v)) = f^*.$$

Thus, w^* is an optimal solution of (1). This proves part (a).

We next show that part (b) holds. Let $w^\star$ be an optimal solution of (1) with maximum cardinality. Then it follows immediately from assumption that $\mathcal{M}(w^\star) \succ 0$. Thus, there exists a corresponding Lagrange multiplier $u^\star$ so that $(w^\star, u^\star)$ satisfies (9). Let $\tilde{w}^\star$ be the vector obtained from $w^\star$ by dropping the last entry. In view of (6) and the first equation of (9) and (11), we observe that for any $(\mu, v) \in \Xi_C$,

$$\begin{aligned}
&(w(\mu,v) - w^\star)^T (u(\mu,v) - u^\star) \\
=\;& (P\tilde{w}(\mu,v) - P\tilde{w}^\star)^T (u(\mu,v) - u^\star) \\
=\;& (\tilde{w}(\mu,v) - \tilde{w}^\star)^T P^T (\nabla\Phi_{\mathcal{M}}(w(\mu,v)) - \nabla\Phi_{\mathcal{M}}(w^\star)) - (\tilde{w}(\mu,v) - \tilde{w}^\star)^T v \\
=\;& (w(\mu,v) - w^\star)^T (\nabla\Phi_{\mathcal{M}}(w(\mu,v)) - \nabla\Phi_{\mathcal{M}}(w^\star)) - (\tilde{w}(\mu,v) - \tilde{w}^\star)^T v \\
\geq\;& -2C\mu,
\end{aligned}$$

where the last inequality holds since Φ is convex in $\mathcal{S}^m_{++}$, $w(\mu,v), w^\star \in \Omega$ and $\|v\|_\infty < C\mu$. Using this inequality and the third equation in (9) and (11), we see that

$$w^{\star T} u(\mu,v) + w(\mu,v)^T u^\star \;\leq\; w^{\star T} u^\star + w(\mu,v)^T u(\mu,v) + 2C\mu \;=\; (2C+n)\mu. \tag{48}$$

Dividing both sides of the above inequality by μ and using the third equation of (11), we obtain that

$$\sum_{i=1}^n \frac{w_i^\star}{w_i(\mu,v)} + \sum_{i=1}^n \frac{u_i^\star}{u_i(\mu,v)} \;\leq\; 2C+n. \tag{49}$$

Since $(w^\star, u^\star) \geq 0$ and $(w(\mu,v), u(\mu,v)) > 0$, it follows from (49) that for all i,

$$w_i(\mu,v) \;\geq\; \frac{w_i^\star}{2C+n}, \qquad u_i(\mu,v) \;\geq\; \frac{u_i^\star}{2C+n}. \tag{50}$$

It immediately implies that the ith entry of any accumulation point $w^\diamond$ of $w(\mu,v)$ as $(\mu,v) \xrightarrow[\Xi_C]{} (0,0)$ must be positive whenever $w_i^\star > 0$. Since $w^\diamond$ is an optimal solution of (1) by part (a), we conclude that part (b) holds.

■

References

[1] S. D. Ahipasaoglu, P. Sun and M. J. Todd. Linear convergence of a modified Frank-Wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimization Methods and Software* 23, pp. 5–19 (2008).

[2] E. D. Andersen, C. Roos, T. Terlaky, T. Trafalis and J. P. Warners. The use of low-rank updates in interior-point methods. AdvOl-Report No. 2000/9, February 2000, Hamilton, Ontario, Canada.

[3] A. Atkinson, A. Donev and R. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press (2007).

[4] A. C. Atkinson, K. Chaloner, A. M. Herzberg and J. Juritz. Optimum experimental designs for properties of a compartmental model. *Biometrics* 49, pp. 325–337 (1993).

[5] C. L. Atwood. Sequences converging to D-optimal designs of experiments. *Annals of Statistics* 1, pp. 342–352 (1973).

[6] C. L. Atwood. Convergent design sequences for sufficiently regular optimality criteria. *Annals of Statistics* 4, pp. 1124–1138 (1976).

[7] C. L. Atwood. Convergent design sequences for sufficiently regular optimality criteria II: singular case. *Annals of Statistics* 8, pp. 894–913 (1980).

[8] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM (2001).

[9] D. Böhning. A vertex-exchange-method in D-optimal design theory. *Metrika* 33, pp. 337–347 (1986).

[10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press (2004).

[11] X. Chen, H. Qi and P. Tseng. Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems. *SIAM Journal on Optimization* 13, pp. 960–985 (2002).

[12] H. Dette, A. Pepelyshev and A. Zhigljavsky. Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics and Data Analysis* 53, pp. 312–320 (2008).

[13] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York (1972).

[14] V. V. Fedorov and J. Lee. Design of experiments in statistics. In H. Wolkowicz, R. Saigal and L. Vandenberghe (Eds.), *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer Academic Publishers (2000).

[15] J. Fellman. On the allocation of linear observations (Thesis). *Commentationes Physico-Mathematicae* 44, pp. 27–78 (1974).

[16] M. C. Ferris and T. S. Munson. Interior-point methods for massive support vector machines. *SIAM Journal on Optimization* 13, pp. 783–804 (2003).

[17] A. Forsgren, P. E. Gill and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Review* 44, pp. 525–597 (2002).

[18] R. Harman and L. Pronzato. Improvements on removing nonoptimal support points in D-optimum design algorithms. *Statistics and Probability Letters* 77, pp. 90–94 (2007).

[19] R. Harman and M. Trnovská. Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices. *Mathematica Slovaca* 59, pp. 693–704 (2009).

[20] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press (2008).

[21] R. M. Larsen. PROPACK - Software for large and sparse SVD calculations. Available at http://sun.stanford.edu/∼rmunk/PROPACK/.

[22] Z. Lu and Y. Zhang. An augmented Lagrangian approach for sparse principal component analysis. *Mathematical Programming* 135, pp. 149–193 (2012).

[23] D. Papp. Optimal designs for rational function regression. *Journal of the American Statistical Association* 107, pp. 400–411 (2012).

[24] A. Pázman. *Foundations of Optimum Experimental Design.* Reidel, Dordrecht (1986).

[25] F. Pukelsheim. *Optimal Design of Experiments.* John Wiley and Sons Inc., New York (1993).

[26] P. Richtarik. Simultaneously solving seven optimization problems in relative scale. Technical Report (2009).

[27] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton (1970).

[28] G. Sagnol. Computing optimal designs of multiresponse experiments reduces to second-order cone programming. *Journal of Statistical Planning and Inference* 141, pp. 1684–1708 (2010).

[29] S. D. Silvey, D. M. Titterington and B. Torsney. An algorithm for optimal designs on a finite design space. *Communications in Statistics – Theory and Methods* 14, pp. 1379–1389 (1978).

[30] D. M. Titterington. Algorithms for computing D-optimal design on finite design spaces. In *Proceedings of the 1976 Conference on Information Science and Systems*, John Hopkins University, 3, pp. 213–216 (1976).

[31] M. J. Todd, K. C. Toh and R. H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization* 8, pp. 769–796 (1998).

[32] K. C. Toh, M. J. Todd and R. H. Tütüncü. SDPT3 — a Matlab software package for semidefinite programming. *Optimization Methods and Software* 11, pp. 545–581 (1999).

[33] B. Torsney. A moment inequality and monotonicity of an algorithm. In Kortanek, K.O. and Fiacco, A.V. (Eds.), *Proceedings of the International Symposium on Semi-Infinite Programming and Applications*, Lecture Notes in Economics and Mathematical Systems 215. University of Texas at Austin, pp. 249–260 (1983).

[34] B. Torsney. W-iterations and ripples therefrom. In Pronzato, L., Zhigljavsky, A. (Eds.), *Optimal Design and Related Areas in Optimization and Statistics.* Springer-Verlag, New York, pp. 1–12 (2007).

[35] B. Torsney and R. Martín-Martín. Multiplicative algorithms for computing optimum designs. *Journal of Statistical Planning and Inference* 139, pp. 3947–3961 (2009).

[36] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming* Series B 95, pp. 189–217 (2003).

[37] M. H. Wright. Why a pure primal Newton barrier step may be infeasible. *SIAM Journal on Optimization* 5, pp. 1–12 (1995).

[38] M. H. Wright. Ill-conditioning and computational error in interior methods for nonlinear programming. *SIAM Journal on Optimization* 9, pp. 84–111 (1999).

[39] C. F. Wu and H. P. Wynn. The convergence of general step-length algorithms for regular optimum design criteria. *Annals of Statistics* 6, pp. 1273–1285 (1978).

[40] H. P. Wynn. Results in the theory and construction of D-optimum experimental designs. *Journal of the Royal Statistical Society* Series B 34, pp. 133–147 (1972).

[41] Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs. *Annals of Statistics* 38, pp. 1593–1606 (2010).

[42] Y. Yu. D-optimal designs via a cocktail algorithm. *Statistics and Computing* 21, pp. 475–481 (2011).