Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

Singular majorants and minorants: enhanced design conditioning

D. R. Jensen & D. E. Ramirez

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
VOL. 87, NO. 9, 1827–1841
http://dx.doi.org/10.1080/00949655.2017.1289210

Taylor & Francis
Taylor & Francis Group

Check for updates

Singular majorants and minorants: enhanced design conditioning

D. R. Jensen[a] and D. E. Ramirez [b]

[a]Department of Statistics, Virginia Tech, Blacksburg, VA, USA; [b]Department of Mathematics, University of Virginia, Charlottesville, VA, USA

ABSTRACT
Lower and upper *spectral* bounds are known for matrices $\boldsymbol{X}'\boldsymbol{X}(k \times k)$ under Loewner [Uber monotone matrixfunktionen. Math Z. 1934;38:177–216] order, as are corresponding bounds for the factor $\boldsymbol{X}(n \times k)$ under an induced order. Least upper bounds for the latter give designs with dominating Fisher Information, with consequent gains in linear inference; see Jensen DR, Ramirez DE [Enhanced design efficiency through least upper bounds. J Stat Comput Simul. 2016;86:1798–1817]. The present study examines properties on ordering the *singular values* of a design matrix using *majorization* as in Marshall and Olkin [Inequalities: theory of majorization and its applications. New York: Academic Press; 1979]. The principal focus includes *conditioning* through condition numbers, variance inflation factors, and lengths and efficiencies of *OLS* solutions. Functions monotone under the induced order are identified; equivalence classes of designs are displayed preserving a dispersion matrix or its eigenvalues; a minimal element $\boldsymbol{X}_m(n \times k)$ is characterized; as are equivalence classes of (A, D, E)-optimal designs showing the latter not to be unique. Algorithms to achieve enhanced designs are given on modifying a single design, or on amalgamating two designs, with essential consequences in linear inference. A collateral procedure, based on mixtures of Fisher information matrices, serves effectively to ameliorate the ill effects of near collinearity. Case studies illustrate gains to be made in practice, to include a substantial improvement in the analysis of classically ill-conditioned data from the literature.

ARTICLE HISTORY
Received 12 July 2016
Accepted 27 January 2017

KEYWORDS
Conditioning; ordering by majorization; monotone functions; efficiency indices; design alterations; non-uniquely optimal designs

AMS SUBJECT CLASSIFICATION
06A06; 15A45; 62J05

1. Introduction

Extremal problems pervade much of applied probability and statistics, to include maximal, minimal, and optimal solutions as the common currency. Such solutions often shed new light on the structure of the system at hand. Specifically, given positive-definite $(k \times k)$ matrices $\boldsymbol{X}'\boldsymbol{X}$ under Loewner [1] order, lower and upper *spectral* bounds are known, as are lower and upper *singular* bounds for the $(n \times k)$ factor matrix $\boldsymbol{X}$ under an induced order. Details are found in [2,3], and are combined in [4] to the following effects. Given first-order designs $(\boldsymbol{X}, \boldsymbol{Z})$, their upper singular bound $\boldsymbol{X}_M$ serves to enhance both $\boldsymbol{X}$ and $\boldsymbol{Z}$, its Fisher information matrix dominating both, thus ordering essentials in Gauss–Markov inference. Such gains proceed on isolating elements from $\boldsymbol{Z}$ complementary to those of $\boldsymbol{X}$, and combining these into $\boldsymbol{X}_M$.

The present study induces yet another invariant order on the space $\mathbb{F}_{n\times k}$ of design matrices, but instead ordering their *singular values* through *majorization* as in [5]. This ordering serves in turn to gauge the degree of regularity or *smoothness* of a model. Consequences are drawn in regard to the conditioning of linear systems, as well as effects on essentials of Gauss–Markov inferences. Majorization has been invoked in earlier studies seeking designs optimal under various criteria or having other

CONTACT D. E. Ramirez der@virginia.edu University of Virginia, P.O. Box 400137, Charlottesville, VA22904, USA

stated characteristics. Selected examples include works of Chan and Li [6], Bhaumik [7], Zhang et al. [8], and Pericleous and Kounias [9], and numerous references cited in those studies. These typically entail majorization of eigenvalues of the Fisher information matrix or the dispersion matrix of *OLS* solutions. An exception is Zhang et al. [8] in using the majorization of pairwise coincidence vectors. The present study appears to be the first to undertake the majorization of the *singular values* of a design matrix. In consequence, instead of a *single* design optimal under a given criterion, as often sought in the literature, our methods enable the identification of equivalence classes of designs, each optimal under (A, D, E)-criteria. An outline follows as organized into three essential parts.

Part I seeks an order on the space $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$ on ordering singular values by majorization, together with informative equivalence classes. These are of independent interest as contributions to the structures of linear spaces. Part II draws on Part I as it applies to experimental design, to Gauss–Markov inferences, and to the conditioning of linear regression systems. Procedures for amalgamating designs $\boldsymbol{X}$ and $\boldsymbol{Z}$ give their *singular minorant* as the greatest lower bound, namely $\boldsymbol{X}_m = \boldsymbol{X} \wedge \boldsymbol{Z}$, or a variant as a mixture of their Fisher information matrices, both giving enhanced designs. Part III illustrates the findings through selected case studies, to include reducing by orders of magnitude the conditioning diagnostics in notoriously ill-conditioned data from the literature. Supporting topics are deferred to Appendices in order to expedite the presentation.

2. Preliminaries

Conventions for notation follow. Denote by $\mathbb{R}^k$ the Euclidean k-space ; by $\mathbb{R}^k_+$ its positive orthant; by $\mathbb{F}_{n\times k}$ the real $(n \times k)$ matrices of rank $k \le n$; by $\mathbb{S}_k$ the real symmetric $(k \times k)$ matrices, with $\mathbb{S}^0_k, \mathbb{S}^+_k$, and $\mathbb{D}_k$ as their positive semidefinite, positive definite, and diagonal varieties. The transpose, trace, and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\mathrm{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$ where defined; and special arrays include the unit vector $\boldsymbol{1}_k = [1, \dots, 1]' \in \mathbb{R}^k$, the unit matrix $\boldsymbol{I}_k$, and a typical diagonal matrix $\boldsymbol{D}_\alpha = \mathrm{Diag}(\alpha_1, \dots, \alpha_k) \in \mathbb{D}_k$. Transformation groups acting on $\mathbb{R}^k$ include the general linear group $\mathcal{G}_k$ and the real orthogonal group $\mathcal{O}_k$; and elements $\boldsymbol{H}$ of $\mathbb{H}_{n\times k}$ are semi-orthogonal in $\mathbb{F}_{n\times k}$ such that $\boldsymbol{H}\boldsymbol{H}'$ is idempotent of rank k and $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$. The *spectral decomposition* $\boldsymbol{A} = \sum_{i=1}^{k} \alpha_i \boldsymbol{q}_i\boldsymbol{q}_i' \in \mathbb{S}^+_k$ yields its *symmetric root* $\boldsymbol{A}^{1/2} = \sum_{i=1}^{k} \alpha_i^{1/2} \boldsymbol{q}_i\boldsymbol{q}_i'$. The *singular value decomposition* (*SVD*) of $\boldsymbol{X} \in \mathbb{F}_{n\times k}$ is $\boldsymbol{X} = \sum_{i=1}^{k} \kappa_i \boldsymbol{p}_i\boldsymbol{q}_i' = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ in which $\boldsymbol{P} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_k] \in \mathbb{H}_{n\times k}$ contains the *left singular vectors*, $\boldsymbol{Q} = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_k] \in \mathcal{O}_k$ contains the *right singular vectors*, and elements of $\boldsymbol{D}_{\boldsymbol{\kappa}} = \mathrm{Diag}(\kappa_1, \dots, \kappa_k)$ comprise the ordered *singular values* $\{\kappa_1 \ge \cdots \ge \kappa_k > 0\}$ of $\boldsymbol{X}$. Denote by $\mathrm{tr}^{\dagger}(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{D}_{\boldsymbol{\kappa}}) = \sum_{i=1}^{k} \kappa_i$. Moreover, $\sigma(\boldsymbol{X}) = \boldsymbol{\kappa}$ takes $\boldsymbol{X}$ into its ordered singular values; and $\mathbb{F}^{\tau}_{n\times k} = \{\boldsymbol{X} \in \mathbb{F}_{n\times k} : \mathrm{tr}^{\dagger}(\boldsymbol{X}) = \tau\}$ serves as reference in much that follows.

Standard usage refers to independent, identically distributed (iid) variates, their cumulative distribution function (cdf) and $\mathcal{L}(\boldsymbol{Y})$ as the distribution of $\boldsymbol{Y}$, with $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the Gaussian law on $\mathbb{R}^k$ having the mean $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\mu}$ and dispersion matrix $\mathrm{V}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$.

3. Part I. Fundamentals

This section establishes essentials of invariant orderings for matrix arrays through majorizing their singular values. It expands substantially on the material in Section 3.3 of Jensen [2]. Developments in this section are of interest independently of applications to be offered subsequently.

3.1. Ordered spaces

Invoke the order axioms: (i) complete, (ii) antisymmetric, (iii) reflexive, and (iv) transitive; then the object $\mathcal{A}$ in $(\mathcal{A}, \succeq_0)$ is said to be *linearly ordered* if the binary relation $\succeq_0$ satisfies (i)–(iv); is *partially ordered* if (ii)–(iv); and is *preordered* if (iii)–(iv). A partially ordered set is a *lower semi-lattice* if for elements (x, y) in $\mathcal{A}$, there is a *greatest lower bound* ($glb = x \wedge y$) in $\mathcal{A}$; an *upper semi-lattice* if there is a *least upper bound* (lub $= x \vee y$) in $\mathcal{A}$; and a *lattice* if both a lower and upper semi-lattice. Such spaces are central to this study.

(i) *Majorization.* In particular, take the simplex $\mathbb{C}_k = \{\boldsymbol{x} \in \mathbb{R}^k_+ \mid x_1 \geq \cdots \geq x_k\}$ and, for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{C}_k$, suppose that

$$\{x_1 + x_2 + \cdots + x_t \geq y_1 + y_2 + \cdots + y_t; 1 \leq t \leq k-1\}$$
$$\{x_1 + x_2 + \cdots + x_k = y_1 + y_2 + \cdots + y_k\}.$$

Then $\boldsymbol{x}$ is said to *majorize* $\boldsymbol{y}$, to be denoted as $\boldsymbol{x} \succeq \boldsymbol{y}$, and related as $\boldsymbol{x}\mathsf{P} = \boldsymbol{y}$ through a doubly stochastic matrix P. Alternatively, $\boldsymbol{y}$ may be recovered from $\boldsymbol{x}$ through a finite number of T-transforms; see [5]. We have the following.

Definition 3.1: (i) Let $\mathbb{C}^\tau_k = \{\boldsymbol{x} \in \mathbb{C}_k \mid \sum_{i=1}^k x_i = \tau\}$ together with the ordering $(\mathbb{C}^\tau_k, \succeq)$. The functions monotone increasing under $\succeq$ are called *Schur convex (S-convex)*, or *S-concave* if decreasing.

(ii) That $(\mathbb{C}^\tau_k, \succeq)$ is a lattice is shown by construction in [3], in Equation (2.1) for $\boldsymbol{\kappa} \wedge \boldsymbol{\omega}$ and Equation (2.2) for $\boldsymbol{\kappa} \vee \boldsymbol{\omega}$ in $(\mathbb{C}^\tau_k, \succeq)$.

(ii) *Invariant orderings* . The basic tenets follow. Consider a space $\mathcal{X}$ together with a group G of one-to-one transformations acting on $\mathcal{X}$. To induce an invariant ordering on $\mathcal{X}$, such ordering necessarily holds for a maximal invariant $\mathtt{M}(\mathbf{X})$ under G, taking $\mathbf{X} \in \mathcal{X} \to \mathtt{M}(\mathbf{X}) \in \mathfrak{M}$. Accordingly, if the latter is partially ordered as $(\mathfrak{M}, \succeq_M)$, the ordering $(\mathcal{X}, \succeq_\chi)$ to be induced on $\mathcal{X}$ is defined together with essentials as follows.

Proposition 3.1: *Given $\mathbf{X} \in \mathcal{X}$; a group G acting one-to-one on $\mathcal{X}$; and a maximal invariant $\mathtt{M}(\mathbf{X})$ taking $\mathbf{X}$ into the ordered set $(\mathfrak{M}, \succeq_M)$. Then an ordering $(\mathcal{X}, \succeq_\chi)$ is induced by $(\mathfrak{M}, \succeq_M)$ on stipulating that $\mathbf{X}_1 \succeq_\chi \mathbf{X}_2$ in $(\mathcal{X}, \succeq_\chi)$ if and only if $\mathtt{M}(\mathbf{X}_1) \succeq_M \mathtt{M}(\mathbf{X}_2)$ in $(\mathfrak{M}, \succeq_M)$.*

(i) *If $(\mathfrak{M}, \succeq_M)$ is partially ordered, then $(\mathcal{X}, \succeq_\chi)$ is preordered and is antisymmetric up to equivalence under G.*
(ii) *A real function f is monotone on $(\mathcal{X}, \succeq_\chi)$ if and only if f is a composition of the type $f(\mathbf{X}) = \psi(\mathtt{M}(\mathbf{X})) = [\psi \circ \mathtt{M}](\mathbf{X})$. with ψ a function in the class Ψ of all functions monotone on $(\mathfrak{M}, \succeq_M)$.*

Proof: See [2]. ■

(iii) *The ordered set* $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$. In particular, an ordering on $(\mathbb{F}^\tau_{n\times k}, \cdot)$ is to be invariant under left and right unitary operators $\{\boldsymbol{X} \to \boldsymbol{UXV}'\}$. From the singular decomposition $\boldsymbol{X} = \boldsymbol{PD_\kappa Q}'$, the singular values $\boldsymbol{\kappa} = [\kappa_1, \ldots, \kappa_k]$ comprise a maximal invariant as shown in [10]. Take $\sigma(\boldsymbol{X}) = [\kappa_1, \ldots, \kappa_k]$ and $\sigma(\boldsymbol{Z}) = [\omega_1, \ldots, \omega_k]$ to be their singular-value mappings. Accordingly, $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ is ordered by majorization of singular values (hence $\succeq_S$) as in the following, where the unitarily invariant condition numbers c_ϕ are identified in Appendix A.2.

Theorem 3.1: *Take $\boldsymbol{X} \succeq_S \boldsymbol{Z}$ in $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ to hold if and only if $\sigma(\boldsymbol{X}) \succeq \sigma(\boldsymbol{Z})$ in $(\mathbb{C}^\tau_k, \succeq)$.*

(i) *Since $(\mathbb{C}^\tau_k, \succeq)$ is partially ordered, then $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ is preordered and is antisymmetric up to equivalence under $\{\boldsymbol{X} \to \boldsymbol{UXV}'\}$.*
(ii) *The functions monotone on $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ correspond one-to-one with the Schur functions monotone on $(\mathbb{C}^\tau_k, \succeq)$, on viewing $f(\cdot)$ on $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ as the composition $f(\boldsymbol{X}) = \psi(\sigma(\boldsymbol{X})) = [\psi \circ \sigma](\boldsymbol{X})$ with $\sigma(\boldsymbol{X}) \in (\mathbb{C}^\tau_k, \succeq)$ and for some S-monotone function ψ on $(\mathbb{C}^\tau_k, \succeq)$.*
(iii) *The condition numbers c_ϕ on $(\mathbb{F}^\tau_{n\times k}, \succeq_S)$ of Appendix A.2 are Schur convex, i.e. if $\boldsymbol{X} \succeq_S \boldsymbol{Z}$, then $c_\phi(\boldsymbol{X}) \geq c_\phi(\boldsymbol{Z})$, so that $\boldsymbol{Z}$ is better conditioned than $\boldsymbol{X}$ as gauged by every unitarily invariant condition number $\{c_\phi(\cdot); \phi \in \Phi\}$.*

Proof: Conclusions (i) and (ii) follow directly on specializing Proposition 3.1, as sketched in Section 3.3 of Jensen [2]. From Appendix A.2, we have for $\{c_\phi(\cdot); \phi \in \Phi\}$ that

$$c_\phi(\boldsymbol{X}) = \|\boldsymbol{X}\|_\phi \|\boldsymbol{X}^\dagger\|_\phi = \phi(\kappa_1, \ldots, \kappa_k)\phi(\kappa_1^{-1}, \kappa_2^{-1}, \ldots, \kappa_k^{-1}),$$

where $\boldsymbol{X}^\dagger$ is the Moore–Penrose inverse and $\{\phi \in \Phi\}$ comprise the symmetric gauge functions on $\mathbb{R}^k$. The Schur convexity of $c_\phi(\boldsymbol{X})$ follows on applying Lemma 3.3 of Marshall and Olkin [11], to the effect that if $\boldsymbol{a} \succeq \boldsymbol{b}$ on $(\mathbb{C}_k^\tau, \succeq)$, then (a) $\phi(b_1, \ldots, b_k) \le \phi(a_1, \ldots, a_k)$ and (b) $\phi(b_1^{-1}, b_2^{-1}, \ldots, b_k^{-1}) \le \phi(a_1^{-1}, a_2^{-1}, \ldots, a_k^{-1})$, to give conclusion (iii). ■

(iv) *Ordered and extremal elements.* We seek *maximal* and *minimal* elements in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$. To these ends take $\bar{\kappa} = \tau/k$, $\bar{\boldsymbol{\kappa}} = [\bar{\kappa}, \ldots, \bar{\kappa}] \in (\mathbb{C}_k^\tau, \succeq)$, and $\boldsymbol{D}_{\bar{\kappa}} = \mathrm{Diag}(\bar{\kappa}, \ldots, \bar{\kappa})$; and let $\pi\boldsymbol{\kappa}$ be a permutation of the elements of $\boldsymbol{\kappa}$. Recall for $\boldsymbol{H} \in \mathbb{H}_{n\times k}$ that $\boldsymbol{H}\boldsymbol{H}'$ is idempotent of rank k and $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$, and $\mathcal{O}_k$ is the real orthogonal group. Next consider the following.

Definition 3.2: Ensembles of matrix expansions in $\mathbb{F}_{n\times k}$ of note are

$$SD_{\mathbb{H}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}' \mid \boldsymbol{L} \in \mathbb{H}_{n\times k}\}; \quad SD_{\mathbb{H}}^{\mathcal{O}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{R}' \mid (\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times k}, \mathcal{O}_k)\};$$
$$SD_{\pi\boldsymbol{\kappa}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_{\pi\boldsymbol{\kappa}}\boldsymbol{R}' \mid (\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times k}, \mathcal{O}_k)\}; \quad SD_{\bar{\boldsymbol{\kappa}}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{R}'; (\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times k}, \mathcal{O}_k)\}.$$

Remark 3.1: From Appendix Lemma A.1, these matrices all are in the form of singular decompositions.

In regard to the notions of *minimal*, and of lower and upper elements in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$, a first look applies Theorem 3.1 to the following effect. Begin with $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ and $\boldsymbol{Z} = \boldsymbol{P}_1\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Q}_1'$ such that neither $\boldsymbol{\kappa} \succeq \boldsymbol{\omega}$ nor $\boldsymbol{\omega} \succeq \boldsymbol{\kappa}$. That $(\mathbb{C}_k^\tau, \succeq)$ is a lattice is shown by construction in [3], in Equation (2.1) for $\boldsymbol{\kappa} \wedge \boldsymbol{\omega}$ and Equation (2.2) for and $\boldsymbol{\kappa} \vee \boldsymbol{\omega}$ in $(\mathbb{C}_k^\tau, \succeq)$, so that $\{\boldsymbol{\kappa} \wedge \boldsymbol{\omega} \preceq (\boldsymbol{\kappa}, \boldsymbol{\omega}) \preceq \boldsymbol{\kappa} \vee \boldsymbol{\omega}\}$. Define

$$\Theta = \{\boldsymbol{\theta} : \boldsymbol{\kappa} \wedge \boldsymbol{\omega} \preceq \boldsymbol{\theta} \preceq \boldsymbol{\kappa} \vee \boldsymbol{\omega}\} \subset \mathbb{C}_k^\tau. \tag{1}$$

Then we have the following, where $\boldsymbol{D}_\theta = \mathrm{Diag}(\theta_1, \ldots, \theta_k)$, $\mathbf{D}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}} = \mathrm{Diag}(\boldsymbol{\kappa} \wedge \boldsymbol{\omega})$, and $\mathbf{D}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}} = \mathrm{Diag}(\boldsymbol{\kappa} \vee \boldsymbol{\omega})$.

Theorem 3.2: *Take $\mathbf{X} = \mathbf{P}\mathbf{D}\mathbf{Q}'$ as the SVD for $\mathbf{X} \in \mathbb{F}_{n\times k}$; allow $\mathbf{D}$ to vary; identify $\mathbf{X}_\phi = \mathbf{P}\mathbf{D}_\phi\mathbf{Q}'$ for $\boldsymbol{\phi} \in (\mathbb{C}_k^\tau, \succeq)$; and let $\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{X}$. For $(\boldsymbol{\kappa}, \boldsymbol{\omega}) \in (\mathbb{C}_k^\tau, \succeq)$, let $\mathbf{D}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}} = \mathrm{Diag}(\boldsymbol{\kappa} \wedge \boldsymbol{\omega})$ and $\mathbf{D}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}} = \mathrm{Diag}(\boldsymbol{\kappa} \vee \boldsymbol{\omega})$, and $\mathbf{X}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}} = \mathbf{P}\mathbf{D}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}}\mathbf{Q}'$ and $\mathbf{X}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}} = \mathbf{P}\mathbf{D}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}}\mathbf{Q}'$. Then*

(i) *Bounds on $\boldsymbol{X_\theta} = \boldsymbol{P}\boldsymbol{D}_\theta\boldsymbol{Q}'$ are*

$$\boldsymbol{X}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}} \preceq_S \boldsymbol{X}_{\boldsymbol{\theta}} \preceq_S \boldsymbol{X}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}}, \quad \{\boldsymbol{\theta} \in \mathbb{C}_k^\tau : \boldsymbol{\kappa} \wedge \boldsymbol{\omega} \preceq \boldsymbol{\theta} \preceq \boldsymbol{\kappa} \vee \boldsymbol{\omega}\}, \tag{2}$$

where $\boldsymbol{X}_{\boldsymbol{\kappa}\wedge\boldsymbol{\omega}}$ is the singular minorant, and $\boldsymbol{X}_{\boldsymbol{\kappa}\vee\boldsymbol{\omega}}$ the singular majorant, of $\{\boldsymbol{X_\theta}; \theta \in \Theta\}$.

(ii) *The vector $\bar{\boldsymbol{\kappa}} = [\bar{\kappa}, \ldots, \bar{\kappa}]$ is uniquely minimal in $(\mathbb{C}_k^\tau, \succeq)$, in that $\bar{\boldsymbol{\kappa}}$ is majorized by every element $\boldsymbol{\omega} \in (\mathbb{C}_k^\tau, \succeq)$.*

(iii) *The ordering $\boldsymbol{X_\omega} \succeq_S \boldsymbol{X}_{\bar{\kappa}}$ in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$ holds for every $\boldsymbol{\omega} \in (\mathbb{C}_k^\tau, \succeq)$, so that $\boldsymbol{X}_{\bar{\kappa}}$ itself is minimal in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$.*

(iv) *$\boldsymbol{X}_{\bar{\kappa}}$ is not uniquely minimal in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$; instead, $SD_{\bar{\boldsymbol{\kappa}}}$ is an equivalence class of minima.*

(v) *$SD_{\mathbb{H}}$ is an equivalence class preserving $\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{X}$ such that, for $\boldsymbol{Z} \in SD_{\mathbb{H}}$, it follows that $\boldsymbol{Z}'\boldsymbol{Z} \equiv \boldsymbol{F}$.*

Proof: (i) Clearly $\{\boldsymbol{\kappa} \wedge \boldsymbol{\omega} \preceq \boldsymbol{\theta} \preceq \boldsymbol{\kappa} \vee \boldsymbol{\omega}\}$ by construction. Conclusion (i) follows from the equivalence of the orderings $(\mathbb{C}_k^\tau, \succeq)$ and $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$, together with the definitions of glb and lub in $(\mathbb{C}_k^\tau, \succeq)$.

(ii) Suppose instead that $\bar{\boldsymbol{\kappa}} \succeq \boldsymbol{\omega}$ for some $\boldsymbol{\omega} \in (\mathbb{C}_k^\tau, \succeq)$. Then $\bar{\kappa} \geq \omega_1$, and the successive inequalities $\{t\bar{\kappa} \geq \omega_1 + \cdots + \omega_t\}$ hold for $\{1 \leq t \leq k-1\}$. On the other hand, $\bar{\kappa} \leq \omega_i$ for some i in order that $\sum_{i=1}^k \omega_i = \tau$. But such $\boldsymbol{\omega}$ cannot belong to the ordered simplex $\mathbb{C}_k^\tau$, giving conclusion (ii) by contradiction.

(iii) That $\boldsymbol{X}_{\bar{\kappa}}$ is minimal in $(\mathbb{F}_{n\times k}^\tau, \succeq_S)$ follows directly from Theorem 3.1.

(iv) Take $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{R}' \in SD_{\bar{\boldsymbol{\kappa}}}$ for some $(\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times k}, \mathcal{O}_k)$, or equivalently, $\boldsymbol{Z} = \boldsymbol{Z}_{\bar{\kappa}}$. Apply conclusion (iii) again for $\boldsymbol{Z}_{\bar{\kappa}}$, so that $SD_{\bar{\boldsymbol{\kappa}}}$ contains equivalent minima independently of $\boldsymbol{L}$ and $\boldsymbol{R}$.

(v) Take $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ as reference, and $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}' \in SD_{\mathbb{H}}$. Then $\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\kappa}}^2\boldsymbol{Q}' = \boldsymbol{Z}'\boldsymbol{Z}$ holds independently of the choice for $\boldsymbol{L}$. ■

Remark 3.2: In correspondence with conclusion (ii), observe that $\boldsymbol{\kappa}_M = [\tau, 0, \ldots, 0]$ is maximal in $(\mathbb{C}_k^\tau, \succeq)$. However, this at best is of marginal interest here in that $\boldsymbol{X}_M = \boldsymbol{P}\mathrm{Diag}(\boldsymbol{\kappa}_M)\boldsymbol{Q}'$ has unit rank.

4. Part II. Efficiency and conditioning in linear inference

This portion of our study is designed to draw from the foundations of Part I with special reference to linear statistical inference, to include further derivations as needed. Details follow.

4.1. The models

Models $\{\boldsymbol{Y} = \beta_0\boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ and the model matrix $\boldsymbol{X}_0 = [\boldsymbol{1}_n, \boldsymbol{X}]$ with intercept are considered, taking the columns of $\boldsymbol{X}$ to be centred about their means. The Fisher information matrix is $\boldsymbol{F}_0 = \boldsymbol{X}_0'\boldsymbol{X}_0 = \mathrm{Diag}(n, \boldsymbol{X}'\boldsymbol{X})$; and the *OLS* solutions are $[\widehat{\beta_0}, \hat{\boldsymbol{\beta}}']$ with $\widehat{\beta_0} = \bar{Y}$ and $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Clearly, $\boldsymbol{X}$ serves as a design matrix and, from the special structure here, it suffices to focus on $\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{X}$ as the Fisher information matrix for $\boldsymbol{\beta}$, and its inverse as the dispersion matrix $\mathrm{V}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, exclusive of β_0. These are subject to the following.

Assumptions.

A1: $\mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\mathrm{V}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_n$;
A2: $\mathcal{L}(\boldsymbol{\epsilon}) = N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

Remark 4.1: In much that follows it suffices to take $\sigma^2 = 1$; otherwise, $\sigma^2 \neq 1$ may be reinstated as needed.

4.2. Efficiency indices

It is instructive to survey design efficiency criteria C_r as in Table 1, often employed in design evaluation including C_r-optimal designs. In wide usage are $\{A, D, E\}$ as the trace, determinant, and largest eigenvalue of $\boldsymbol{\Sigma}$, as in [13]. Determinants of historical note include the correlation determinant $|\boldsymbol{C}|$ from $\boldsymbol{\Sigma}$ as the *scatter coefficient* of Frisch [14], and the *generalized variance* $|\boldsymbol{\Sigma}|$ of Wilks [15]. To study local power in the analysis of variance, Wald [16] first considered E-efficiencies, but opted instead for D-efficiencies. Standard references are Federov [17], Silvey [18], and Pukelsheim [19], as well as *universal optimality* as surveyed recently in [20].

That *vector efficiencies* are multidimensional in concept is noted in [21]. Accordingly, Kiefer [22] advocated that designs be screened through multiple criteria. Scalar criteria as summarized in Table 1 are sorted into Group I depending on the eigenvalues of $\boldsymbol{\Sigma}$, and Group II depending otherwise on $\boldsymbol{\Sigma}$ itself. Corresponding efficiency indices are available for subsets of parameters. See [23], for example. Regarding D-efficiency, recall $\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq c_\alpha\}$ as confidence ellipsoids for $\boldsymbol{\beta}$ having volumes proportional to $D^{1/2}$. Special properties of determinants support a connection between the D-efficiency for $\boldsymbol{\beta}$, and the subset efficiencies D_1 for $\boldsymbol{\beta}_1$ and D_2 for $\boldsymbol{\beta}_2$ in the partitioned

Table 1. Efficiency criteria C_r for design $\boldsymbol{X}$ having the Fisher information matrix $\boldsymbol{X}'\boldsymbol{X}$ and $\mathrm{V}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, with eigenvalues $\lambda(\boldsymbol{\Sigma}) = [\lambda_1 \geq \cdots \geq \lambda_k]$, and with *minimization* as the operation yielding C_r-optimal designs.

Criterion	Group	Description	Comments
A	I	$\mathrm{tr}(\boldsymbol{\Sigma})$	A : Sum of $\mathrm{Var}(\hat{\beta}_i)$ for elements of $\hat{\boldsymbol{\beta}}$
D	I	$\lvert\boldsymbol{\Sigma}\rvert$	$D^{1/2}$:$\propto$ $\mathrm{Vol}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq c_\alpha\}$
E	I	$\lambda_1(\boldsymbol{\Sigma})$	E : Maximal variance of $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$, $\lvert\boldsymbol{c}\rvert = 1$
MV	II	$\max\{\mathrm{Var}(\hat{\beta}_i); 1 \leq i \leq k\}$	MV : Maximal variance of elements of $\hat{\boldsymbol{\beta}}$
T^{-1}	II	$1/\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})$	T : tr(Fisher information matrix)
$\boldsymbol{c}$	II	$\mathrm{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$	$\boldsymbol{c}'\boldsymbol{\beta}$: A distinguished linear function
$W(\boldsymbol{X})$	I	$\dfrac{k^k \prod_{i=1}^{k}(1/\kappa_i^2)}{[\sum_{i=1}^{k}(1/\kappa_i^2)]^k}$	W: Mauchly's [12] sphericity criterion

form $\boldsymbol{\beta}' = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2']$. Specifically, $D = D_1 D_2 \prod_{i=1}^{r}(1 - \rho_i^2)$, where r is the minimum subset size and $\prod_{i=1}^{r}(1 - \rho_i^2)$ is the *alienation coefficient* of Hotelling [24]. See [25]. Here, Mauchly's [12] criterion $W(\boldsymbol{X})$ serves to gauge the non-sphericity of contours of the Gaussian density of $\hat{\boldsymbol{\beta}}$, taking the value $W = 1.0$ when spherical and $W < 1.0$ otherwise.

4.3. Design efficiencies

We draw next on the foundations of Part I as they bear on topics in linear inference. Recall Definition 3.2 and the ensembles $SD_{\mathbb{H}}$, $SD_{\mathbb{H}}^{\mathcal{O}}$, and $SD_{\pi\boldsymbol{\kappa}}$.

Theorem 4.1: *Consider $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ as reference in $(\mathbb{F}_{n\times k}^{\tau}, \succeq_S)$, together with $\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{X}$ as the Fisher information matrix, the OLS solutions $\hat{\boldsymbol{\beta}}$, their dispersion matrix $V(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} = \boldsymbol{F}^{-1}$, and its eigenvalues $\lambda(\boldsymbol{\Sigma}) = [\lambda_1 \geq \cdots \geq \lambda_k]$.*

(i) *$SD_{\mathbb{H}}$ is an equivalence class having $\boldsymbol{F}$ and $\boldsymbol{\Sigma}$ matrices identical to those of $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$.*
(ii) *All matrices in $SD_{\mathbb{H}}^{\mathcal{O}}$ and $SD_{\pi\boldsymbol{\kappa}}$ have eigenvalues $\lambda(\boldsymbol{\Sigma})$ identical to those of $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$.*
(iii) *The Group I criteria of Table 1 are indistinguishable for all designs in $SD_{\mathbb{H}}^{\mathcal{O}}$ and $SD_{\pi\boldsymbol{\kappa}}$.*
(iv) *The Group II criteria of Table 1 are indistinguishable for all designs in $SD_{\mathbb{H}}$.*

Proof: Conclusion (i) follows since $\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{D}_{\kappa}^{2}\boldsymbol{Q}'$ and $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{D}_{\kappa}^{-2}\boldsymbol{Q}'$ are both independent of the choice for $\boldsymbol{L}$. For $\boldsymbol{X} \in SD_{\mathbb{H}}^{\mathcal{O}}$ we have $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{D}_{\kappa}^{-2}\boldsymbol{R}'$ independently of $\boldsymbol{L}$, with reciprocal eigenvalues $[1/\kappa_k^2 \geq \ldots \geq 1/\kappa_1^2]$ since $\boldsymbol{R}$ is orthogonal. Similarly, for $\boldsymbol{X} \in SD_{\pi\boldsymbol{\kappa}}$ and $\boldsymbol{F} = \boldsymbol{R}\boldsymbol{D}_{\pi\kappa}^{2}\boldsymbol{R}'$, its eigenvalues are squares of a permutation $\pi\boldsymbol{\kappa}$ of the singular values of $\boldsymbol{X}$, and $\lambda(\boldsymbol{\Sigma})$ are reciprocals of these as asserted in conclusion (ii). Conclusions (iii) and (iv) follow from (i) and (ii) and the definitions of the Group I and Group II criteria. ∎

It remains to examine the ordering $(\mathbb{F}_{n\times k}^{\tau}, \succeq_S)$ as it bears on the conditioning of *OLS* systems, and on properties of ensuing Gauss–Markov procedures. We have the following.

Theorem 4.2: *Consider matrices $(\boldsymbol{X}, \boldsymbol{Z}) \in (\mathbb{F}_{n\times k}^{\tau}, \succeq_S)$, together with invariant functions taking $\mathbb{F}_{n\times k}^{\tau} \to \mathbb{R}^1$; in particular, the condition numbers $c_\phi(\boldsymbol{X})$, the inverse operators $\Delta(\boldsymbol{X}) = |\boldsymbol{\Sigma}(\boldsymbol{X})|$, $\Gamma(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{X}))$, and the eigenvalues $\lambda(\boldsymbol{\Sigma}(\boldsymbol{X})) = [\lambda_1 \geq \ldots \geq \lambda_k]$. Then*

(i) *The inverse determinant operator $\boldsymbol{X} \to \Delta(\boldsymbol{X})$ on $(\mathbb{F}_{n\times k}^{\tau}, \succeq_S)$ is Schur convex, i.e. if $\boldsymbol{X} \succeq_S \boldsymbol{Z}$, then $\Delta(\boldsymbol{X}) \geq \Delta(\boldsymbol{Z})$.*
(ii) *The inverse trace operator $\boldsymbol{X} \to \Gamma(\boldsymbol{X})$ is Schur convex on $(\mathbb{F}_{n\times k}^{\tau}, \succeq_S)$, i.e. if $\boldsymbol{X} \succeq_S \boldsymbol{Z}$, then $\Gamma(\boldsymbol{X}) \geq \Gamma(\boldsymbol{Z})$.*

(iii) *The inverse extremal eigenvalues* $\boldsymbol{X} \to \lambda_1(\boldsymbol{\Sigma}(\boldsymbol{X}))$ *and* $\boldsymbol{X} \to \lambda_k(\boldsymbol{\Sigma}(\boldsymbol{X}))$ *are, respectively, Schur convex and Schur concave on* $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$, *i.e. if* $\boldsymbol{X} \succeq_S \boldsymbol{Z}$, *then* $\lambda_1(\boldsymbol{\Sigma}(\boldsymbol{X})) \geq \lambda_1(\boldsymbol{\Sigma}(\boldsymbol{Z}))$ *and* $\lambda_k(\boldsymbol{\Sigma}(\boldsymbol{X})) \leq \lambda_k(\boldsymbol{\Sigma}(\boldsymbol{Z}))$.

(iv) *Considered as functions on* $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$, *the* (A, D, E)*-efficiency indices for* $\boldsymbol{\Sigma}(\boldsymbol{X})$ *are minimized at* $\boldsymbol{X}_{\bar{\kappa}} = \boldsymbol{P}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{Q}'$. *Accordingly, the ensemble* $SD_{\bar{\boldsymbol{\kappa}}}$ *of Definition 3.2 comprises an equivalence class of* (A, D, E)*-optimal designs.*

(v) *For* $k = 2$ *Mauchly's criterion* $W(\boldsymbol{X})$ *is Schur concave on* $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$, *i.e. if* $\boldsymbol{X} \succeq_S \boldsymbol{Z}$, *then* $W(\boldsymbol{X}) \leq W(\boldsymbol{Z})$ *and the Gaussian contours of* $\hat{\boldsymbol{\beta}}(\boldsymbol{X})$ *are less spherical than* $\hat{\boldsymbol{\beta}}(\boldsymbol{Z})$.

Proof: For reference let $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ and $\boldsymbol{Z} = \boldsymbol{P}_1\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Q}_1'$. Objects to be ordered in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$ are invariant under left and right unitary operators $\boldsymbol{X} \to \boldsymbol{U}\boldsymbol{X}\boldsymbol{V}'$, thus depending on a maximal invariant. To continue, denote a generic function on $\mathbb{R}^k_+$ as $g(\boldsymbol{u}) = g(u_1, \ldots, u_k)$, and its partial derivatives as $\{g_{(i)}(\boldsymbol{u}) = \partial g(\boldsymbol{u})/\partial u_i; 1 \leq i \leq k\}$. On examining $(u_i - u_j)[g_{(i)}(\boldsymbol{u}) - g_{(j)}(\boldsymbol{u})]$ in reference to conclusion (i), we have $g(\boldsymbol{u}) = \prod_{i=1}^{k}(1/u_i^2)$. Since $g(\boldsymbol{u})$ is symmetric, it suffices to consider the case $(i, j) = (1, 2)$:

$$(u_1 - u_2)[g_{(1)}(\boldsymbol{u}) - g_{(2)}(\boldsymbol{u})] = 2(u_1 - u_2)^2 V/u_1^3u_2^3 > 0, \tag{3}$$

with $V = \prod_{i=3}^{k}(1/u_i^2)$. The Schur convexity of $g(\boldsymbol{u}) = \prod_{i=1}^{k}(1/u_i^2)$, and thus conclusion (i), now follow from Theorem A.4 of Marshall and Olkin [5, p.57]. To continue, the Schur convexity of $\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{X})) = \sum_{i=1}^{k}(1/\kappa_i^2)$ follows from Proposition C.1 of Marshall and Olkin [5, p.64], since $\kappa_i \to (1/\kappa_i^2)$ is convex, to give conclusion (ii). For conclusion (iii), $\boldsymbol{\Sigma}(\boldsymbol{X}) = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\kappa}}^{-2}\boldsymbol{Q}'$ with eigenvalues $[1/\kappa_k^2 \geq \cdots \geq 1/\kappa_1^2]$, and $\boldsymbol{\Sigma}(\boldsymbol{Z}) = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\omega}}^{-2}\boldsymbol{Q}'$ with eigenvalues $[1/\omega_k^2 \geq \cdots \geq 1/\omega_1^2]$. Conclusion (iii) follows as claimed since $\boldsymbol{\kappa} \succeq \boldsymbol{\omega}$ implies $\kappa_1 \geq \omega_1$ and $\kappa_k \leq \omega_k$. Conclusion (iv) follows from Conclusions (i)–(iii) together with Theorem 3.2(iii). Conclusion (v) follows as in (i) on evaluating Schur's condition

$$(u_1 - u_2)[W_{(1)}(\boldsymbol{u}) - W_{(2)}(\boldsymbol{u})] = -\frac{8u_1u_2(u_1 - u_2)^2(u_2 + u_1)^2}{(u_2^2 + u_1^2)^3} < 0 \tag{4}$$

to complete the proof for (v). A counter example to (v) for $k = 3$ is given in Appendix 3. ■

Illustration 4.1: Suppose that $k = 3$, $\boldsymbol{D}_{\boldsymbol{\kappa}} = \mathrm{Diag}(3, 2, 1)$; next modify $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}' \to \boldsymbol{Z} = \boldsymbol{P}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{Q}'$ with $\boldsymbol{D}_{\bar{\kappa}} = \mathrm{Diag}(2, 2, 2)$. Then $\boldsymbol{Z}$ has enhanced conditioning by Theorem 4.2. Specifically, $\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{X})) = [1 + \frac{1}{4} + \frac{1}{9}] = 1.36111$ and $\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{Z})) = 3 \times (\frac{1}{4}) = 0.7500$; $|\boldsymbol{\Sigma}(\boldsymbol{X})| = 1 \times \frac{1}{4} \times \frac{1}{9} = 0.029778$ and $|\boldsymbol{\Sigma}(\boldsymbol{Z})| = (\frac{1}{4})^3 = 0.015625$; and similarly $\lambda_1(\boldsymbol{\Sigma}(\boldsymbol{X})) = 1.00$ and $\lambda_1(\boldsymbol{\Sigma}(\boldsymbol{Z})) = 0.25$. Moreover, $\boldsymbol{P}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{Q}'$ is (A, D, E)-optimal by Theorem 4.2(iv). These conclusions hold for all $\boldsymbol{X} \in F^6_{n\times 3}$ having any n, $k = 3$, and $\tau = 6$, independently of the choice for $(\boldsymbol{P}, \boldsymbol{Q})$.

4.4. Information mixtures

A further approach to ill-conditioning in $\{\boldsymbol{Y} = \beta_0\boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$, with $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$, is to visualize a Fisher information matrix as a weighted average. To these ends, we again focus on $\boldsymbol{X}\boldsymbol{\beta}$, and we adopt the following convention for sums of diagonal matrices.

Definition 4.1: Let $\boldsymbol{D}_{\boldsymbol{a}} = \mathrm{Diag}(a_1, \ldots, a_k)$ and $\boldsymbol{D}_b = \mathrm{Diag}(b_1, \ldots, b_k)$. By $(\boldsymbol{D}_{\boldsymbol{a}} + \boldsymbol{D}_b)^{1/2}$ is meant the matrix $\mathrm{Diag}((a_1 + b_1)^{1/2}, \ldots, (a_k + b_k)^{1/2})$.

To continue, extend the notion of Fisher information $\boldsymbol{X} \rightarrow \boldsymbol{F}_I(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{X}$ to include

$$\begin{aligned} \boldsymbol{F}_I(\boldsymbol{X}_t) &= \boldsymbol{Q}[(1-t)\boldsymbol{D}_\kappa^2 + t\boldsymbol{D}_{\bar{\kappa}}^2]\boldsymbol{Q}'; \quad t \in [0,1]; \\ \boldsymbol{X}_t &= \boldsymbol{P}[(1-t)\boldsymbol{D}_\kappa^2 + t\boldsymbol{D}_{\bar{\kappa}}^2]^{1/2}\boldsymbol{Q}'; \quad t \in [0,1]; \\ \boldsymbol{X}_t &= \boldsymbol{P}\mathrm{Diag}([(1-t)\kappa_i^2 + t\bar{\kappa}^2]^{1/2}; 1 \leq i \leq k)\boldsymbol{Q}' \end{aligned} \tag{5}$$

offering the continuum $\{\boldsymbol{X}_t;\ t \in [0,1]\}$ as prospects for improved conditioning. Recall at $t=1$ that $\boldsymbol{X}_1 = \boldsymbol{P}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{Q}' = \boldsymbol{X}_{\bar{\kappa}}$ is minimal, so that $\boldsymbol{X}$ is drawn towards the minimal $\boldsymbol{X}_{\bar{\kappa}}$ through $\boldsymbol{X}_t$ as t increases in $[0,1]$.

Remark 4.2: This is an innovation on surrogate regression (SR), where $\mathrm{Diag}((\kappa_i^2 + c)^{1/2})$ in SR to be replaced by $\mathrm{Diag}((1-t)\kappa_i^2 + t\bar{\kappa}^2)^{1/2})$ to give $\boldsymbol{X}_t$.

The referenced SR has been studied in detail in [26], in comparison with ridge regression (RR) and *OLS*. Contrary to published claims that versions of RR 'act more like an orthogonal system', it was shown that RR typically exhibits erratic divergence from orthogonality as the ridge scalar evolves, often reverting back to *OLS* in the limit. In contrast, SR solutions converge monotonically to those from orthogonal systems.

Accordingly, we examine properties of $\{\boldsymbol{X}_t;\ t \in [0,1]\}$ corresponding to those for $\boldsymbol{X}_k$ in [26], in comparison with *OLS*. Let $\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}(\boldsymbol{X}_t) = (\boldsymbol{X}_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{Y}$; observe that $\hat{\boldsymbol{\beta}}_t$ is biased for $t > 0$; note that $\mathrm{V}(\hat{\boldsymbol{\beta}}_t) = \sigma^2\boldsymbol{\Sigma}_t$ with $\boldsymbol{\Sigma}_t = (\boldsymbol{X}_t\boldsymbol{X}_t)^{-1}$, and recall that biased estimators typically are assessed by their mean square error (MSE) at $\mathrm{E}(\hat{\boldsymbol{\beta}}_t) = \boldsymbol{\beta}_0$, namely,

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_t) = \mathrm{tr}(\mathrm{V}(\hat{\boldsymbol{\beta}}_t)) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

Essentials are

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\beta}}_t) - \boldsymbol{\beta} &= [(\boldsymbol{X}_t\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{X} - \boldsymbol{I}_k]\boldsymbol{\beta}, \\ \mathrm{V}(\hat{\boldsymbol{\beta}}_t) &= \sigma^2\boldsymbol{\Sigma}_t = \sigma^2\boldsymbol{Q}[(1-t)\boldsymbol{D}_\kappa^2 + t\boldsymbol{D}_{\bar{\kappa}}^2]^{-1}\boldsymbol{Q}', \\ \| \mathrm{E}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta} \|^2 &= \boldsymbol{\beta}'\boldsymbol{Q}[(1-t)\boldsymbol{D}_\kappa^2 + t\boldsymbol{D}_{\bar{\kappa}}^2]^{-1/2}(\boldsymbol{D}_\kappa^2)^{1/2} - \boldsymbol{I}_k]^2\boldsymbol{Q}'\boldsymbol{\beta}, \\ \mathrm{MSE}(\hat{\boldsymbol{\beta}}_t) &= \sum_{i=1}^k \frac{\sigma^2}{(1-t)\kappa_i^2 + t\bar{\kappa}^2} + \sum_{i=1}^k \theta_i^2 \left[\frac{\kappa_i}{\sqrt{(1-t)\kappa_i^2 + t\bar{\kappa}^2}} - 1\right]^2, \end{aligned} \tag{6}$$

where the squared bias in the second term on the right of Equation (6) is expressed in terms of the canonical parameters $\boldsymbol{\theta} = \boldsymbol{Q}'\boldsymbol{\beta}$.

To continue, in order to track properties of $\{\boldsymbol{X}_t;\ t \in [0,1]$, it is critical to examine the manner in which $\mathrm{MSE}(\hat{\boldsymbol{\beta}}_t)$ evolves with t, recalling that $\mathrm{MSE}(0) = \sigma^2\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1}$, the dispersion matrix of the *OLS* solution $\hat{\boldsymbol{\beta}}_t$ at $t=0$. In Section 5.3, we modify $\boldsymbol{X}_t$ to assure that $\mathrm{MSE}(\hat{\beta}_t)$ is decreasing at $t = 0$.

5. Part III. Case studies: enhanced designs

We seek to illustrate earlier developments, specifically, to modify a given design, or to construct new designs, so as to achieve enhanced design capabilities.

5.1. Case 1: equivalent designs

Regarding $\{\boldsymbol{Y} = \beta_0\mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$, take $\boldsymbol{X}_0 = [\mathbf{1}_n, \boldsymbol{X}]$ with $\boldsymbol{X}'$ as the first three rows of Table A1, Appendix 2, to include also its singular vectors $(\boldsymbol{P}', \boldsymbol{Q})$. To continue, construct $\boldsymbol{X}_1 = \boldsymbol{P}\,\mathrm{Diag}(\kappa_3, \kappa_2, \kappa_1)$

Table 2. Variances of *OLS* solutions; eigenvalues of the dispersion matrix $\boldsymbol{\Gamma} \in \{\boldsymbol{\Sigma}_0, \boldsymbol{\Xi}_0, \boldsymbol{\Omega}_0\}$ for the designs $\{\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_{\bar{\kappa}}\}$; and *A*, *D*, and *E* efficiencies for these designs.

	Design characteristics					
Design	$\boldsymbol{X}$	$\boldsymbol{X}_1$	$\boldsymbol{X}_{\bar{\kappa}}$	$\boldsymbol{X}$	$\boldsymbol{X}_1$	$\boldsymbol{X}_{\bar{\kappa}}$
Estimates	Variances			Eigenvalues of $\boldsymbol{\Gamma}$		
$\hat{\beta}_0$	0.12500	0.12500	0.12500	3.53636	3.53636	0.12500
$\hat{\beta}_1$	1.38170	1.83546	0.21629	0.22694	0.22694	0.21629
$\hat{\beta}_2$	0.69002	0.26120	0.21629	0.12500	0.12500	0.21629
$\hat{\beta}_3$	1.76012	1.73518	0.21629	0.06854	0.06854	0.21629

Diagnostic	A	D	E
$\boldsymbol{\Sigma}_0$	3.95684	0.00688	3.53636
$\boldsymbol{\Xi}_0$	3.95684	0.00688	3.53636
$\boldsymbol{\Omega}_0$	0.77387	0.00126	0.21629

$\boldsymbol{Q}' \in SD_{\pi\boldsymbol{\kappa}}$ in the terminology of Definition 3.2, together with $\boldsymbol{X}_{\bar{\kappa}} = \boldsymbol{P}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{Q}' \in SD_{\bar{\boldsymbol{\kappa}}}$ as in Definition 3.2, with $\boldsymbol{D}_{\bar{\kappa}} = \text{Diag}(\bar{\kappa}, \bar{\kappa}, \bar{\kappa})$ and $\bar{\kappa} = 2.1503$, giving both $\boldsymbol{X}_1$ and $\boldsymbol{X}_{\bar{\kappa}}$ as reported in Table A1.

Essentials of the output are reported in Table 2, where $(\boldsymbol{\Sigma}_0, \boldsymbol{\Xi}_0, \boldsymbol{\Omega}_0)$ refer to the full Fisher information matrix for each of the models $[\boldsymbol{1}_n, \boldsymbol{X}]$, $[\boldsymbol{1}_n, \boldsymbol{X}_1]$, and $[\boldsymbol{1}_n, \boldsymbol{X}_{\bar{\kappa}}]$, respectively, i.e. $\boldsymbol{\Sigma}_0 = \text{Diag}(1/n, (\boldsymbol{X}'\boldsymbol{X})^{-1})$. Clearly, $(\boldsymbol{\Sigma}_0, \boldsymbol{\Xi}_0, \boldsymbol{\Omega}_0)$ differ, their diagonals reported as variances. Nonetheless, as $\boldsymbol{X}$ and $\boldsymbol{X}_1$ belong to $SD_{\pi\boldsymbol{\kappa}}$, they have common eigenvalues by Theorem 4.1(ii), and thus identical (A, D, E)-efficiencies as asserted in Theorem 4.1(iii) and as seen in Table 2.

Noting that $\{\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_{\bar{\kappa}}\}$ all belong to $(F^{\tau}_{n\times 3}, \succeq_S)$, with $\tau = \kappa_1 + \kappa_2 + \kappa_3 = 6.4508$ and $n = 8$, it follows that $\boldsymbol{X}_{\bar{\kappa}}$ is (A, D, E)-optimal for all matrices $\{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{R}'; (\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times 3}, \mathcal{O}_3)\}$ whose singular values sum to 6.4508 by Theorem 4.2(iv), and for any n.

Remark 5.1: Despite their genesis, the designs $(\boldsymbol{X}, \boldsymbol{X}_1)$ are clearly disparate. Nonetheless, neither can be distinguished using the Group I criteria of Table 1. If information regarding β_2 is deemed critical, a choice between $\boldsymbol{X}$ and $\boldsymbol{X}_1$ would opt for $\boldsymbol{X}_1$ in view of the variances $\text{Var}(\hat{\beta}_2(\boldsymbol{X}_1)) = 0.26120\sigma^2$ and $\text{Var}(\hat{\beta}_2(\boldsymbol{X})) = 0.69002\sigma^2$.

5.2. Case 2: Nonunique optimality

We revisit claims for 'variance optimality' in Example 7.2 of Myers and Montgomery [27, p.285] in $n = 8$ experimental runs and $k = 3$ regressors. This represents a $\frac{1}{2}$ fraction of a 2^3 factorial design, replicated. The centred $\boldsymbol{X}$ in $[\boldsymbol{1}_n, \boldsymbol{X}]$ is transposed into the first three rows of Table 3. As before take $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$. The authors assert that

'This design is "variance optimal" for the model $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. That is, all coefficients in the above model have minimum variance over all designs with sample size $n = 8$'.

To proceed, construct $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ as in Table 3, taking $\boldsymbol{L}$ as listed. Direct computations show that $\text{V}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})) = \text{V}(\hat{\boldsymbol{\beta}}(\boldsymbol{Z})) = (\sigma^2/8)\boldsymbol{I}_3$, so that both distributions are isotropic, with Mauchly's [12] value $W = 1$. We infer that $\boldsymbol{X} = \boldsymbol{X}_{\bar{\kappa}}$ with $\bar{\kappa} = \sqrt{8}$ and $\boldsymbol{\Sigma} = (\sigma^2/8)\boldsymbol{I}_3$. Theorem 2(iii) gives $\boldsymbol{X}_{\bar{\kappa}}$ as minimal in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$, but not uniquely so as in Theorem 3.2(iv). In fact, returning to Definition 3.2, we have shown that $SD_{\bar{\kappa}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_{\bar{\kappa}}\boldsymbol{R}'; (\boldsymbol{L}, \boldsymbol{R}) \in (\mathbb{H}_{n\times k}, \mathcal{O}_k)\}$ is an equivalence class comprising all minima in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$. Indeed, that $\boldsymbol{Z}$ in Table 3 is a member of this class is seen on taking $\boldsymbol{L} \in \mathbb{H}_{n\times k}$ as in Table 3, together with $\boldsymbol{R} = \boldsymbol{Q} \in \mathcal{O}_k$ from the singular decomposition of $\boldsymbol{X}$.

In short, there are uncountably many designs in $SD_{\bar{\boldsymbol{\kappa}}}$, found on varying $(\boldsymbol{L}, \boldsymbol{R})$, all having identical 'minimum variance' among designs 'with sample size $n = 8$'. Moreover, such designs are indistinguishable under the design criteria of Table 1. In consequence, the 'variance optimal' designs, having

Table 3. Design matrix $\boldsymbol{X}$ as Example 7.2 of Myers and Montgomery [27]; $\boldsymbol{L}$ as alternative to $\boldsymbol{P}$; and the derived matrix $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_\kappa \boldsymbol{Q}'$; all of orders (8×3).

Design $\boldsymbol{X}'$							
1.0000	1.0000	−1.0000	−1.0000	−1.0000	−1.0000	1.0000	1.0000
−1.0000	−1.0000	1.0000	1.0000	−1.0000	−1.0000	1.0000	1.0000
−1.0000	−1.0000	−1.0000	−1.0000	1.0000	1.0000	1.0000	1.0000
Design $\boldsymbol{Z}' = [\boldsymbol{L}\boldsymbol{D}_\kappa \boldsymbol{Q}']'$							
1.4142	−1.4142	−1.4142	1.4142	0.0000	0.0000	0.0000	0.0000
1.4142	−1.4142	1.4142	−1.4142	0.0000	0.0000	0.0000	0.0000
−1.4142	−1.4142	1.4142	1.4142	0.0000	0.0000	0.0000	0.0000
$\boldsymbol{L}'$ in lieu of $\boldsymbol{P}'$							
0.5000	0.5000	−0.5000	−0.5000	0.0000	0.0000	0.0000	0.0000
0.5000	−0.5000	0.5000	−0.5000	0.0000	0.0000	0.0000	0.0000
0.5000	−0.5000	−0.5000	0.5000	0.0000	0.0000	0.0000	0.0000

identical design characteristics and hallmarks of conditioning, comprise an equivalence class in $(\mathbb{F}_{n\times k}, \succeq_S)$.

Further comparisons between $\boldsymbol{X}$ and $\boldsymbol{Z}$ may be drawn. The use of center runs often is advocated, whereas $\boldsymbol{X}$ has none, there are four center runs in $\boldsymbol{Z}$, thus three degrees of freedom for 'pure error'. In essence, all empirical evidence, initially distributed among the eight rows of $\boldsymbol{X}$, has been redistributed into the four non-center runs of $\boldsymbol{Z}$.

It is clear that equi-optimal designs having further attractive features may be constructed, for example, having two center runs. In this case, essential information contained in $\boldsymbol{X}$ would be redistributed to six points in the newly constructed $\boldsymbol{Z}_1$, as were redistributed from $\boldsymbol{X}$ to the four non-center runs of $\boldsymbol{Z}$.

5.3. Case 3: Hospital manpower

We revisit the Hospital Manpower data of Myers [28, p.132], itself notoriously ill-conditioned, and often reexamined in the literature for this reason. Again take $\{\boldsymbol{Y} = \beta_0 \boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$. This study seeks to ameliorate the ills of ill-conditioning, to include variance inflation factors (VIFs), condition numbers, *OLS* solutions $\hat{\boldsymbol{\beta}}$ having excessive lengths and variances, and related matters. Here, the condition number $\{c_1(\boldsymbol{M}); \boldsymbol{M} \in \mathbb{S}_k^+\}$ is the ratio of its largest to smallest eigenvalue, and the VIFs are computed as the diagonals of the inverse of the correlation matrix of the scaled and centred regressors. We expressly apply the tools of Section 4.4, constructing $\{\boldsymbol{X}_t;\ t \in [0,1]\}$ as in Equation (5).

The Hospital Manpower data have $n = 17$ runs and $k = 5$ regressors, with condition number $c_1 = 3.904 \times 10^{10}$ as the ratio of the largest to smallest eigenvalues of the Fisher information matrix $\boldsymbol{X}'\boldsymbol{X}$. From Appendix A.2, these are squares of the condition numbers $c_\phi(\boldsymbol{X})$ of Definition A.3, on taking $\phi(\boldsymbol{x}) = \max |x_i|$ in Definition A.2. Near singularity holds, with $\rho = 0.9999$ as the correlation between the first and third columns of the data matrix. Taking this matrix into its centred form $\boldsymbol{X}(17 \times 5)$, as prescribed by Belsley [29] to improve conditioning, gives $c_1 = 8.517 \times 10^8$, together with VIFs for $[\widehat{\beta}_1, \ldots, \widehat{\beta}_5]$ as $[9597.57,\ 7.94,\ 8933.09,\ 23.29,\ 4.28]$ under *OLS*. Even these are grossly in excess: A widely adopted threshold value is $\{\mathrm{VIF}(i) \le 10.0\}$ for data with acceptable conditioning. Accordingly, we choose $\boldsymbol{X}_{t_0}$ with $t_0 = 0.000295$ in order that $\max\{\mathrm{VIF}(i); 1 \le i \le 5\} = 10$. This is a decrease, in orders of magnitude, from the starting maximal VIF of 9597.57. The value t_0 was found using the Maple software package.

If the singular values for $\boldsymbol{X}_t$ in Equation (5) have the term $\bar{\kappa}^2$ replaced by $\overline{\kappa^2}$ then we have an alternate design $\boldsymbol{X}_{t_0}$ with $t_0 = 0.0000702$. For both designs, the values in Tables 4 and 5 agree to a least three significant digits. For this alternative design, Appendix 4 shows that $\mathrm{MSE}(\hat{\beta}_t)$ is decreasing at $t = 0$ and so satisfies the Admissibility Condition. In a forthcoming paper, we will compare various mixture estimators.

Table 4. Modified parameter estimates $\hat{\boldsymbol{\beta}}(t)$ with $t_0 = 0.0000702$.

t	$\hat{\beta}_1(t)$	$\hat{\beta}_2(t)$	$\hat{\beta}_3(t)$	$\hat{\beta}_4(t)$	$\hat{\beta}_5(t)$
0.00	−15.8517	0.0559	1.5896	−4.2187	−394.3141
$0.2t_0$	−3.7550	0.0731	0.9414	0.6549	−8.9341
$0.4t_0$	−2.6476	0.0735	0.9053	0.5329	−6.3195
$0.6t_0$	−2.1567	0.0738	0.8895	0.4614	−5.1605
$0.8t_0$	−1.8640	0.0739	0.8801	0.4130	−4.4693
t_0	−1.6643	0.0741	0.8737	0.3775	−3.9976

Table 5. Design criteria for $\boldsymbol{X}_t$ with $t_0 = 0.0000702$, to include the squared lengths $\|\hat{\boldsymbol{\beta}}(t)\|^2$, the condition numbers $c_1(\boldsymbol{M}_t)$ with $\boldsymbol{M}_t = \boldsymbol{X}_t'\boldsymbol{X}_t$, the maximal VIFs, the correlations $\rho(\boldsymbol{Y}, \hat{\boldsymbol{Y}}_t)$, and MSE$(\hat{\boldsymbol{\beta}}_t)$.

t	$\|\hat{\boldsymbol{\beta}}(t)\|^2$	$c_1(\boldsymbol{M}_t)$	max{VIF(i)}	$\rho(\boldsymbol{Y}, \hat{\boldsymbol{Y}}_t)$	MSE$(\hat{\boldsymbol{\beta}}_t)$
0.00	155,755.2257	8.517×10^8	9597.5708	0.9954	52,845.14
$0.2t_0$	95.2388	3.528×10^5	26.2453	0.9935	49.13
$0.4t_0$	48.0545	1.764×10^5	16.2575	0.9934	26.63
$0.6t_0$	32.2911	1.176×10^5	12.8172	0.9934	18.57
$0.8t_0$	24.4003	8.823×10^4	11.0646	0.9934	14.38
t_0	19.6625	7.058×10^4	10.0000	0.9934	11.80

It remains to examine further consequences of taking $\boldsymbol{X}_{t_0}$ *in lieu of* $\boldsymbol{X}$. Our revised estimators use the modified moment matrix in solving $\{\boldsymbol{X}_t'\boldsymbol{X}_t\hat{\boldsymbol{\beta}}(t) = \boldsymbol{X}_t'\boldsymbol{Y}\}$ as in SR. Values for $\hat{\boldsymbol{\beta}}(t)$ are given in Table 4 for selected values of t. The negative sign for the *OLS* $\hat{\beta}_4(0)$ has been observed to be 'curious', Myers [28, p.131], as sign reversal itself is often a disqualifying hallmark of ill-conditioning. In contrast, for $u \in \{0.2t_0,\ 0.4t_0,\ 0.6t_0,\ 0.8t_0,\ t_0\}$, the modified $\hat{\beta}_4(u) > 0$, reflecting a more substantive empirical model even for exceedingly small perturbations t. In addition to VIFs and signs of estimators, other concerns with ill-conditioning include excessive lengths of the *OLS* solutions, and the conditioning of the moment matrix $\boldsymbol{M}_t = \boldsymbol{X}_t'\boldsymbol{X}_t$. Values for $\|\hat{\boldsymbol{\beta}}(t)\|^2$ and $c_1(\boldsymbol{M}_t)$ are reported in Table 5 for selected values of t. To demonstrate the predictive utility of $\boldsymbol{X}_t$, the correlations $\rho(\boldsymbol{Y}, \hat{\boldsymbol{Y}}_t)$ between the observed and predicted responses are also reported. Moreover, as a measure of displacement of $\boldsymbol{X}_{t_0}$ from $\boldsymbol{X}$, the relative mean absolute deviation between these is found to be less than 1%. Computations were carried out using the Maple software package.

To demonstrate Theorem 4.3, we report in column 6 of Table 5 the values for MSE$(\hat{\boldsymbol{\beta}}_t)$ for values of t in column 1. Equation (5) for MSE$(\hat{\boldsymbol{\beta}}_t)$ requires knowledge of the unknown parameters σ^2 and $\boldsymbol{\theta} = \boldsymbol{Q}'\boldsymbol{\beta}$. For this computation, we have used $\boldsymbol{\beta}' = [-1.600, 0.075, 0.870, 0.380, -4.000]$ and $\sigma^2 = (637.75)^2$, the corresponding estimated error variance. As seen for $\|\hat{\boldsymbol{\beta}}(t)\|^2$ and max{VIF(i)}, the values for MSE$(\hat{\boldsymbol{\beta}}_t)$ decrease precipitously from the starting *OLS* value at $t = 0$. We note that while MSE$(\hat{\boldsymbol{\beta}}_t)$ decreases monotonically over $t \in [0, t_0]$, it is not monotone over the domain $[0, 1]$; instead, further computations show that it turns upward at $t = 0.756$.

In short, we find it to be remarkable in the extreme, that a centred matrix $\boldsymbol{X}_{t_0}$, so near to the highly ill-conditioned $\boldsymbol{X} \in \mathbb{F}_{n\times k}$, should exhibit such thoroughly enhanced conditioning characteristics.

6. Conclusions

We have studied the ordering $\boldsymbol{X} \succeq_S \boldsymbol{Z}$ for two design matrices in the space $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$, with ordering induced from the majorization ordering of their singular values in $(\mathbb{C}^{\tau}_k, \succeq)$. This continues the work of Jensen [3] who introduced the spectral glb and spectral lub for a pair of matrices in $(\mathbb{S}^+_k, \succeq_L)$. The spectral lub for two Fisher information matrices can yield a design with enhanced *OLS* efficiencies dominating both, as shown in [4]. Our study here has focused on two types of principal findings.

The first establishes equivalence classes of designs, these being indistinguishable under the standard design criteria of Table 1. Our equivalent designs typically are variations on the singular decomposition $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\kappa}\boldsymbol{Q}'$, for example on permuting the singular values, as shown in Theorem 3.2. Further examples follow on replacing $\boldsymbol{P}$ by some $\boldsymbol{L} \in \mathbb{H}_{n\times k}$ since the $\boldsymbol{F}_I$ matrix $\boldsymbol{X}'\boldsymbol{X}$ holds independently of $\boldsymbol{P}$.

The second principal finding seeks extreme designs in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$. Theorem 3.2 establishes extremal elements for $\boldsymbol{X} \in (\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$ using extremal elements for $\boldsymbol{\kappa} = [\kappa_1, \ldots, \kappa_k]$ in $(\mathbb{C}^{\tau}_k, \succeq)$. In particular, a design having equal singular values in $(\mathbb{C}^{\tau}_k, \succeq)$ is minimal in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$. Suppose that neither $\boldsymbol{X} \succeq_S \boldsymbol{Z}$ nor $\boldsymbol{Z} \succeq_S \boldsymbol{X}$. Then, lattice properties of $(\mathbb{C}^{\tau}_k, \succeq)$ enable the construction of the *singular minorant* $\boldsymbol{X}_m$, and the *singular majorant* $\boldsymbol{X}_M$, bounding an ensemble of designs including $\boldsymbol{X}$ and $\boldsymbol{Z}$. This provides a venue for amalgamating two designs into a single design superior to both in the ordering $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$.

Related studies show the sense in which these orderings will matter in practice. Theorem 4.2 shows that condition numbers, as well as the (A, D, E) diagnostics for $\mathrm{V}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})) = \boldsymbol{\Sigma}(\boldsymbol{X})$, are Schur convex on $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$. In consequence, *smoother* designs, in the sense of majorized singular values, have superior (A, D, E) efficiencies. Their equivalent minimal matrices comprise equivalence classes of (A, D, E)-optimal designs, showing the latter not to be unique. In the literature various algorithms serve to return a *single* design as optimal under a given criterion; it is reasonable ask how such algorithms manage to single out but one of the often uncountably many equivalently optimal candidates.

Theorem 4.2(v) regarding Mauchly's [12] sphericity criterion applies for $k = 2$; for $k > 2$ a counter example is given in Appendix 3. To the contrary, regardless of k, the distribution of $\hat{\boldsymbol{\beta}}(\boldsymbol{X})$ is less isotropic than $\hat{\boldsymbol{\beta}}(\boldsymbol{Z})$ by Mauchly's [12] $W(\cdot)$ criterion, under the alternative Loewner [1] ordering $\boldsymbol{X}'\boldsymbol{X} \succeq_L \boldsymbol{Z}'\boldsymbol{Z}$, as in Theorem 1 of Jensen [25].

A variation on surrogate and ridge regression, motivated by earlier developments in this study, replaces $\boldsymbol{X}$ by $\{\boldsymbol{X}_t = \boldsymbol{P}\mathrm{Diag}([(1-t)\kappa_i^2 + t\bar{\kappa}^2]^{1/2}; 1 \leq i \leq k)\boldsymbol{Q}'\}$ to give a continuum $\{\boldsymbol{X}_t;\ t \in [0, 1]\}$ of prospective modified designs. At $t = 0$ we have $\boldsymbol{X}_t = \boldsymbol{X}$, and at $t = 1.0$ the design $\boldsymbol{X}_t = \boldsymbol{X}_{\bar{\kappa}}$ is minimal in $(\mathbb{F}^{\tau}_{n\times k}, \succeq_S)$. Otherwise, $\{t \uparrow\}$ drives $\boldsymbol{X}$ towards this minimal design. The Hospital Manpower data of Myers [28, p.132], notoriously ill-conditioned, is amenable to this approach. Case Study 3 reports remarkable improvement in conditioning on choosing $t_0 = 0.0000702$ so as to achieve the maximal VIF at 10.0, a widely used threshold value beyond which a design is declared to be ill-conditioned. In addition, the relative mean absolute deviation between $\boldsymbol{X}$ and $\boldsymbol{X}_{t_0}$ is less than 1%. In short, we find estimators modified in this manner to be attractive alternatives to ridge and surrogate solutions in the analysis of ill-conditioned data.

Acknowledgments

Following the presentation of Jensen [3] to the *Joint AMS-IMS-SIAM Summer Research Conference on Stochastic Inequalities*, Seattle, Washington, July 1991, the late Professor Samuel Kotz remarked as most important the approximately three pages supporting, in retrospect, the present manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

D.E. Ramirez http://orcid.org/0000-0002-2419-4538

References

[1] Loewner C. Uber monotone matrixfunktionen. Math Z. 1934;38:177–216.
[2] Jensen DR. Invariant ordering and order preservation. In: Tong YL, editor. Inequalities in statistics and probability. IMS Lecture Notes/Monograph Series, Vol. 5. Hayward (CA): Institute of Mathematical Statistics; 1984.
[3] Jensen DR. Matrix extremes and related stochastic bounds. Stoc Inequal. 1993;22:133–144. (IMS Lecture Notes/Monograph Series).

[4] Jensen DR, Ramirez DE. Enhanced design efficiency through least upper bounds. J Stat Comput Simul. 2016;86:1798–1817.
[5] Marshall AW, Olkin I. Inequalities: theory of majorization and its applications. New York: Academic Press; 1979.
[6] Chan NN, Li K-H. Majorization for A–optimal designs. J Math Anal Appl. 1989;142:101–107.
[7] Bhaumik DK. Majorization and D-optimality for complete block designs under correlations. J Royal Statist Soc Ser B. 1995;57:139–143.
[8] Zhang A, Fang K-T, Li R, et al. Majorization framework for balanced lattice designs. Ann Statist. 2005;33:2837–2853.
[9] Pericleous K, Kounias S. The concept of majorization in experimental design. Commun Statist Theory Methods. 2012;41:2836–2850.
[10] Vinograde B. Canonical positive definite matrices under internal linear transformations. Proc Amer Math Soc. 1950;1:159–161.
[11] Marshall AW, Olkin I. Norms and inequalities for condition numbers. Pacific J Math. 1965;15:241–247.
[12] Mauchly JW. Significance test for sphericity of a normal n-variate distribution. Ann Math Statist. 1940;11:204–209.
[13] Kiefer J. Optimum experimental designs. J Roy Statist Soc Ser B. 1959;21:272–304.
[14] Frisch R. Correlation and scatter in statistical variables. Nordic Statist J. 1929;1:36–102.
[15] Wilks SS. Certain generalizations in the analysis of variance. Biometrika. 1932;24:471–494.
[16] Wald A. On the efficient design of statistical investigations. Ann Math Statist. 1943;14:134–140.
[17] Fedorov VV. Studden WJ and Klimko EM, translators, Theory of optimal experiments. New York: Academic Press; 1972.
[18] Silvey SD. Optimal design. London: Chapman and Hall; 1980.
[19] Pukelsheim F. Optimal design of experiments. New York: Wiley; 1993.
[20] Druilhet P. Conditions for optimality in experimental designs. Linear Algebra Appl. 2004;388:147–157.
[21] Jensen DR. Vector efficiency in multiparameter estimation. Linear Algebra Appl. 1991;151:143–155.
[22] Kiefer J. Optimum design: variation in structure and performance under change of criterion. Biometrika. 1975;62:277–288.
[23] Sibson R. D_A-optimality and duality. In: Gani J, Sarkadi K, Vincze I, editors. Progress in statistics, Proc. 9th European meeting of statisticians, Budapest, 1972, Vol. 2. Amsterdam: North-Holland; 1974.
[24] Hotelling H. Relations between two sets of variates. Biometrika. 1936;28:321–377.
[25] Jensen DR. Condition numbers and D-efficiency. Statist Probab Lett. 2004;66:267–274.
[26] Jensen DR, Ramirez DE. Surrogate models in ill-conditioned systems. J Statist Plann Infer. 2010;140:2069–2077.
[27] Myers RH, Montgomery DC. Response surface methodology: process and product optimization using designed experiments. New York: Wiley; 1995.
[28] Myers RH. Classical and modern regression with applications. 2nd ed. Boston (MA): PWS-KENT; 1990.
[29] Belsley DA. Centering, the constant, first-differencing, and assessing conditioning. In: Belsley DA, Kuh E, editors. Model reliability. Cambridge (MA): MIT Press; 1986. p. 117–153.

Appendix 1. Properties of $\mathbb{F}_{n\times k}$

A.1 The singular decomposition

For $\boldsymbol{X} \in \mathbb{F}_{n\times k}$ with $k < n$, equivalent versions of its SVD follow, with $\boldsymbol{D}_{\boldsymbol{\kappa}} = \mathrm{Diag}(\kappa_1, \dots, \kappa_k)$.

Definition A.1:

(i) $\boldsymbol{X} = \boldsymbol{P}_0\mathbf{D}\boldsymbol{Q}'$ with $\boldsymbol{P}_0 \in \mathcal{O}(n)$, $\boldsymbol{Q} \in \mathcal{O}_k$, and $\mathbf{D} = [\boldsymbol{D}_{\boldsymbol{\kappa}}, \mathbf{0}]'$.
(ii) $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\kappa}}\boldsymbol{Q}'$ on partitioning $\boldsymbol{P}_0 = [\boldsymbol{P}, \boldsymbol{P}_1]$ with $\boldsymbol{P} \in \mathbb{H}_{n\times k}$ as defined in Section 2.
(iii) Columns of $\boldsymbol{P} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_k]$ comprise the *left singular* vectors, and of $\boldsymbol{Q} = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_k]$ the *right singular* vectors.
(iv) Its Moore–Penrose pseudo-inverse is $\boldsymbol{X}^{\dagger} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\kappa}}^{-1}\boldsymbol{P}'$.

The following refers to the structure, synthesis, and identity of matrix decompositions in $\mathbb{F}_{n\times k}$.

Lemma A.1: *Let $\boldsymbol{P} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_k] \in \mathbb{H}_{n\times k}$ and $\boldsymbol{Q} \in \mathcal{O}_k$ such that $\{\boldsymbol{p}_i\boldsymbol{q}_i'; 1 \le i \le k\}$ comprise a collection of frames in $\mathbb{F}_{n\times k}$.*

(i) *Then the SVD $\boldsymbol{X} = \sum_{i=1}^{k} \kappa_i\boldsymbol{p}_i\boldsymbol{q}_i'$ lifts each frame $\{\boldsymbol{p}_i\boldsymbol{q}_i' \in \mathbb{F}_{n\times k}\}$ by κ_i.*
(ii) *Conversely, given $\boldsymbol{L} \in \mathbb{H}_{n\times k}$, $\boldsymbol{D}_{\boldsymbol{\omega}} = Diag(\omega_1, \dots, \omega_k)$, and $\boldsymbol{R} \in \mathcal{O}_k$, then the assemblage $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{R}'$ is the singular decomposition of $\boldsymbol{Z}$.*

Table A1. Design $\boldsymbol{X}$, derived matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_{\tilde{\kappa}}$; the left ($\boldsymbol{P}'$) and right ($\boldsymbol{Q}$) singular vectors; and the singular values $\boldsymbol{\kappa}$ = [3.8198, 2.0992, 0.5318] for Case Study 1.

	−1.0000	1.0000	−1.4142	1.4142	−1.0000	1.0000	0.0000	0.0000
$\boldsymbol{X}'$	0.0000	0.0000	1.0000	1.0000	0.0000	0.0000	−1.0000	−1.0000
	−0.8000	0.6000	−0.8000	2.0000	−0.8000	0.6000	−0.8000	0.0000
	0.4279	−1.0391	0.2809	0.7126	0.4279	−1.0391	−1.1080	1.3370
$\boldsymbol{X}_1'$	−0.0193	−0.3197	−1.3211	−0.3556	−0.0193	−0.3197	0.4993	1.8556
	−0.1811	0.8999	0.2484	−1.1704	−0.1811	0.8999	1.1798	−1.6954
	−0.3045	−0.1498	−0.5095	1.5719	−0.3045	−0.1498	−0.9855	0.8316
$\boldsymbol{X}_{\tilde{\kappa}}'$	0.0162	0.1501	1.1888	0.6961	0.0162	0.1501	−0.7762	−1.4414
	−0.7081	1.0106	−0.8536	0.3366	−0.7081	1.0106	0.5610	−0.6490
	−0.3308	0.2946	−0.3736	0.6594	−0.3308	0.2946	−0.1792	−0.0342
$\boldsymbol{P}'$	0.0960	−0.1138	0.6024	0.3785	0.0960	−0.1138	−0.5082	−0.4372
	0.0996	−0.3618	−0.1302	0.2927	0.0996	−0.3618	−0.3435	0.7055
			−0.7099	−0.1307	−0.6921			
$\boldsymbol{Q}$			0.3508	−0.9177	−0.1865			
			0.6108	0.3752	−0.6973			

Proof: Conclusion (i) is the conventional *SV D* expressed in terms of *frames*. To examine $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{R}'$, the expansion $\boldsymbol{Z}\boldsymbol{Z}' = \boldsymbol{L}\boldsymbol{D}^2_{\omega}\boldsymbol{L}' = \sum_{i=1}^{k} \omega_i^2 \boldsymbol{l}_i\boldsymbol{l}_i'$, and that $\boldsymbol{L}$ is semi-orthogonal, gives the spectral equations $\{\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{l}_j = \omega_j^2\boldsymbol{l}_j; 1 \leq j \leq k\}$, verifying $\boldsymbol{L}$ as the left singular vectors and $\boldsymbol{D}_{\boldsymbol{\omega}} = \mathrm{Diag}(\omega_1, \ldots, \omega_k)$ as the singular values. That $\boldsymbol{R}$ contains the right singular vectors follows similarly on expanding $\boldsymbol{Z}'\boldsymbol{Z}$, to give conclusion (ii). ■

A.2 Norms and condition numbers

The unitarily invariant norms on $\mathbb{F}_{n\times k}$ are invariant under left and right unitary operators $\boldsymbol{X} \to \boldsymbol{U}\boldsymbol{X}\boldsymbol{V}'$. The *symmetric gauge functions* $\phi : \mathbb{R}^k \to \mathbb{R}^1$ in the class Φ are defined in [5, p.96].

Definition A.2: (i) The class of *unitarily invariant norms* on $\mathbb{F}_{n\times k}$ is given by composition as

$$\{\|\boldsymbol{X}\|_{\phi} = \phi(\sigma(\boldsymbol{X})) = [\phi \circ \sigma]\boldsymbol{X}; \phi \in \Phi\},$$

with $\sigma(\boldsymbol{X})$ as its singular-value mapping.

(ii) Prominent examples are $\phi(\boldsymbol{x}) = \max |x_i|$ and $\phi(\boldsymbol{x}) = (\sum |x_i|^r)^{1/r}$, $r \geq 1$.

With $\boldsymbol{A} \in \mathbb{F}_{n\times k}$ the *conditioning* of the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has to do with the propagation of disturbances within the system into its solutions. For $\boldsymbol{A} \in \mathbb{F}_{n\times n}$ of full rank, the condition number $c_{\phi}(\boldsymbol{A})$ typically is defined as $c_{\phi}(\boldsymbol{A}) = \phi(\boldsymbol{A})\phi(\boldsymbol{A}^{-1})$, where ϕ ordinarily is a norm. See [5, p.270 ff]. If instead $\boldsymbol{A} \in \mathbb{F}_{n\times k}$ with *SVD* $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{a}}\boldsymbol{Q}'$, a solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{b}$, that is

$$\boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{a}}^{-1}\boldsymbol{P}'\boldsymbol{P}\boldsymbol{D}_{\boldsymbol{a}}\boldsymbol{Q}'\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{a}}^{-1}\boldsymbol{P}'\boldsymbol{b},$$

so that $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{a}}^{-1}\boldsymbol{P}'\boldsymbol{b}$ with $\boldsymbol{A}^{\dagger}$ as the Moore–Penrose inverse. Accordingly, we have the following.

Definition A.3: Elements of $\{c_{\phi}(\boldsymbol{X}) = \|\boldsymbol{X}\|_{\phi}\|\boldsymbol{X}^{\dagger}\|_{\phi}; \phi \in \Phi\}$ comprise the *unitarily invariant condition numbers* on $\mathbb{F}_{n\times k}$.

Appendix 2. Tables: Case study 1

Details are supplied here for completeness in regard to Case Study 1, beginning with $\boldsymbol{X}$ as listed in Table A1. The principals of its singular decomposition are given as $(\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{\kappa})$. Designs $\boldsymbol{X}_1$ and $\boldsymbol{X}_{\tilde{\kappa}}$, as listed in Table A1, are constructed from the singular decomposition of $\boldsymbol{X}$.

Appendix 3. A counter example

Disclaimer: Mauchly's [12] criterion is not Schur concave for $k > 2$.

Denote the criterion as $k^{-k}W(\boldsymbol{\Sigma}(\boldsymbol{X})) = |\boldsymbol{\Sigma}(\boldsymbol{X})|/[\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{X})]^k$, where $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\kappa}}^{-2}\boldsymbol{Q}'$, so that $k^{-k}W(\boldsymbol{\Sigma}(\boldsymbol{X})) = k^{-k}W(\boldsymbol{\kappa}) = \prod_{i=1}^{k}(1/\kappa_i^2)/[\sum_{i=1}^{k}(1/\kappa_i^2)]^k$. For $k = 3$ let $\boldsymbol{a}_1 = [11.1,\ 1.0,\ 1.0]$ and $\boldsymbol{a}_2 = [11.0,\ 1.1,\ 1.0]$ be surrogates for $\boldsymbol{\kappa}$. Clearly, $\boldsymbol{a}_1 \succeq \boldsymbol{a}_2$ in $(\mathbb{C}_k^{\tau}, \succeq)$. Similarly, let $\boldsymbol{b}_1 = [6.1,\ 4.9,\ 1.0]$ and $\boldsymbol{b}_2 = [5.9,\ 5.1,\ 1.0]$ be further surrogates for $\boldsymbol{\kappa}$, with $\boldsymbol{b}_1 \succeq \boldsymbol{b}_2$ in $(\mathbb{C}_k^{\tau}, \succeq)$. Direct computations for $\boldsymbol{a}_1 \succeq \boldsymbol{a}_2$ give $3^{-3}W(\boldsymbol{a}_1) = 0.00100$ and $3^{-3}W(\boldsymbol{a}_2) = 0.00111$. However, the ordering reverses at $\boldsymbol{b}_1 \succeq \boldsymbol{b}_2$, namely, $3^{-3}W(\boldsymbol{b}_1) = 0.000917$ and $3^{-3}W(\boldsymbol{b}_2) = 0.000909$.

Appendix 4. Admissibility Condition

Take the derivative $(\mathrm{d}/\mathrm{d}t)\mathrm{MSE}(\hat{\boldsymbol{\beta}}_t)$ at Equation (5) and evaluate it at $t=0$ to get (using Maple)

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{MSE}(0) = -\sigma^2 \sum_{i=1}^{k} \frac{(-\kappa_i^2 + \overline{\kappa^2})}{\kappa_i^4} = \sigma^2 \left[\sum_{i=1}^{k} \frac{1}{\kappa_i^2} - \overline{\kappa^2} \sum_{i=1}^{k} \frac{1}{\kappa_i^4} \right] \leq 0.$$

Inequality at the last step is a consequence of Chebychev's inequality: With $\{a_1 \geq \cdots \geq a_k > 0\}$ and $\{0 < b_1 \leq \ldots \leq b_k\}$, Chebyshev's inequality gives

$$k(a_1 b_1 + \cdots + a_k b_k) \leq (a_1 + \cdots + a_k)(b_1 + \cdots + b_k)$$

$$k\left(\kappa_1^2 \frac{1}{\kappa_1^4} + \cdots + \kappa_k^2 \frac{1}{\kappa_k^4}\right) \leq (\kappa_1^2 + \cdots + \kappa_k^2)\left(\frac{1}{\kappa_1^4} + \cdots + \frac{1}{\kappa_k^4}\right)$$

so that $\sum_{i=1}^{k}(1/\kappa_i^2) \leq \bar{\kappa}^2 \sum_{i=1}^{k}(1/\kappa_i^4)$.