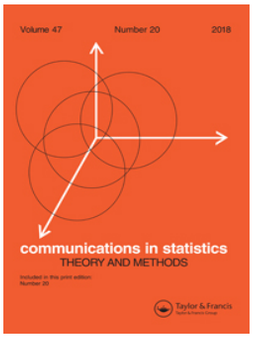

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

Designs enhancing Fisher information

D. R. Jensen & D. E. Ramirez

To cite this article: D. R. Jensen & D. E. Ramirez (2018) Designs enhancing Fisher information, Communications in Statistics - Theory and Methods, 47:20, 4895-4904, DOI: 10.1080/03610926.2017.1380831

To link to this article: https://doi.org/10.1080/03610926.2017.1380831

Accepted author version posted online: 19 Sep 2017.
Published online: 08 Nov 2017.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2018, VOL. 47, NO. 20, 4895–4904
https://doi.org/10.1080/03610926.2017.1380831

Designs enhancing Fisher information

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
Given a model $\{\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}\}$ with Fisher information matrix $\boldsymbol{\Xi} = \boldsymbol{X}'\boldsymbol{X}$, a principal objective is to find information enhancing transformations $\boldsymbol{T}$ for which $\boldsymbol{T\Xi T}' \succeq_L \boldsymbol{\Xi}$ under the positive definite ordering, so as to improve essentials in linear inference. This is achieved through properties of congruences together with basic orderings of linear spaces. These foundations in turn support a new class of geometric mixture models on "mixing" the original design with another to assume the role of "target," to the following effects. Ridge, surrogate, and other solutions are often used to mitigate the effects of ill-conditioned models. Instead, in this study an ill-conditioned design matrix $\boldsymbol{X}$ is mixed with a well-conditioned design as target, leveraging the former toward the latter as the mixing parameter evolves, thus offering an alternative approach to ill-conditioning. The methodology is demonstrated with case studies from the literature, where the geometric mixtures are compared with the ridge, surrogate, and recently found arithmetic mixture models.

ARTICLE HISTORY
Received 13 April 2017
Accepted 12 September 2017

KEYWORDS
Arithmetic mixture designs; conditioning; efficiency indices; geometric mixture designs; ridge and surrogate designs.

MATHEMATICS SUBJECT CLASSIFICATION
62J05; 62J20

1. Introduction

Consider $\{\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}\}$ with $\boldsymbol{X}$ in $\mathbb{F}_{n\times p}$ as a design matrix. The Fisher Information Matrix ($\mathfrak{F}_{\mathrm{M}}$), namely $\mathfrak{I}(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{X}$, assumes a central role in the analysis and interpretation of linear models. As a first objective we seek information-enhancing transformations $T(\boldsymbol{X})$ whose $\mathfrak{F}_{\mathrm{M}}$ matrices $\mathfrak{I}(T(\boldsymbol{X}))$ dominate $\mathfrak{I}(\boldsymbol{X})$ under the positive definite ordering $(\mathbb{S}_p^+, \succeq_L)$ of Loewner (1934), as this in turn will ensure the ordering of essential design diagnostics. A first intuitive view suggests that this might be achieved on rescaling the columns of $\boldsymbol{X} \to \boldsymbol{XD}_\delta$ with $\boldsymbol{D}_\delta$ as the diagonal matrix $\{\boldsymbol{D}_\delta = \mathrm{Diag}(\delta_1, \ldots, \delta_p); \delta_i \geq 1\}$. However, this fails, as the following counter example demonstrates.

Counter Example 1. Take $p = 2$; $\boldsymbol{D}_\delta = \mathrm{Diag}(10,\ 1.1)$, and $\boldsymbol{X}'\boldsymbol{X} = \left[\begin{smallmatrix} 1.0 & -1.0 \\ -1.0 & 2.0 \end{smallmatrix}\right]$. Then the difference is $\boldsymbol{D}_\delta\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_\delta - \boldsymbol{X}'\boldsymbol{X} = \left[\begin{smallmatrix} 99.00 & -10.00 \\ -10.00 & 0.42 \end{smallmatrix}\right]$ with determinant -58.42, so that $\boldsymbol{D}_\delta\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_\delta \not\succeq_L \boldsymbol{X}'\boldsymbol{X}$. This despite having scaled the first column of $\boldsymbol{X}$ by the substantial factor of 10.

The matter clearly is more delicate, which we nonetheless achieve through congruent transformations and basic ordering properties of linear spaces. In particular, designs having the enhanced $\mathfrak{F}_{\mathrm{M}}$ matrices are exhibited explicitly through their singular value representations. This achievement leads to our second objective, namely, to construct geometric mixtures (GM) of $\boldsymbol{X}$ and another design to be designated as a "target". For the case of ill-conditioned systems,

CONTACT D. E. Ramirez der@virginia.edu Department of Mathematics, University of Virginia, Charlottesville, VA, USA.

the target may be chosen to be well conditioned, providing yet another venue for mitigating the ill effects of ill-conditioned models through biased estimation.

Uses for the discovery of $\mathfrak{F}_{\mathrm{M}}$-enhanced designs include the following. In the planning stage a prospective design may be reassessed as to whether a nearby design may be found to be more informative, but still satisfying design constraints. Additionally, such design points may be reserved for use in subsequent experiments. Otherwise the user may elect to undertake a non standard analysis of the original design, featuring smaller variances but biased estimators, taking the mean square error (MSE) to quantify the Bias–Variance tradeoff. This paper is organized as follows. Conventions for the study are established in Section 2. The principal findings are given in Section 3 to include foundations. The case studies that illustrate the concepts are given in Section 4. Section 5 presents the discussion and summary of the paper. Supporting materials are deferred in the Appendix.

2. Preliminaries

2.1. Notation

Identify $\mathbb{R}^p$ as Euclidean p-space; $\mathbb{F}_{n\times p}$ as the real $(n \times p)$ matrices of rank $p < n$; and $\mathbb{S}_p$ as the real symmetric $(p \times p)$ matrices, with $\mathbb{S}_p^0$, and $\mathbb{S}_p^+$ as their positive semidefinite and positive definite varieties. The transpose, trace, and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\mathrm{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$; and special arrays are the unit vector $\mathbf{1}_p = [1, \ldots, 1]' \in \mathbb{R}^p$, the unit matrix $\boldsymbol{I}_p$, and a typical diagonal matrix $\boldsymbol{D}_\alpha = \mathrm{Diag}(\alpha_1, \ldots, \alpha_p)$. Groups acting on $\mathbb{R}^p$ include the real orthogonal group $\mathcal{O}_p$. The *spectral decomposition* of $\boldsymbol{A}$ is $\boldsymbol{A} = \sum_{i=1}^p \alpha_i \boldsymbol{q}_i \boldsymbol{q}_i' \in \mathbb{S}_p^+$ with $\lambda(\boldsymbol{A}) = \{\alpha_1 \geq \cdots \geq \alpha_p > 0\}$ as its eigenvalues. The *singular decomposition* of $\boldsymbol{X} \in \mathbb{F}_{n\times p}$ is $\boldsymbol{X} = \sum_{i=1}^p \xi_i \boldsymbol{p}_i \boldsymbol{q}_i' = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ in which $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_p]$ contains the *left singular vectors* as semiorthogonal in $\mathbb{H}_{n\times p}$, $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p] \in \mathcal{O}_p$ contains the *right singular vectors,* and elements of $\boldsymbol{D}_\xi = \mathrm{Diag}(\xi_1, \ldots, \xi_p)$ are its *singular values* given by $\sigma(\boldsymbol{X}) = \{\xi_1 \geq \cdots \geq \xi_p > 0\}$. Moreover, for a vector $\boldsymbol{a} = [\, a_1, \ldots, a_p \,]$, and for $\boldsymbol{b} = [\, b_1, \ldots, b_p \,]$ having non zero elements, by $\boldsymbol{a} \cdot \boldsymbol{b}$ is meant $[\, a_1 b_1, \ldots, a_p b_p \,]$ and by $\boldsymbol{a}/\boldsymbol{b} = [\, a_1/b_1, \ldots, a_p/b_p \,]$.

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 $\mathbb{R}^p$ having the mean $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\mu}$ and dispersion matrix $\mathrm{V}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$.

3. The principal findings

3.1. Overview

The models here are $\{\boldsymbol{Y}_0 = \beta_0 \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ with intercept, the columns of $\boldsymbol{X} \in \mathbb{F}_{n\times p}$ having been centered about their means such that $\mathbf{1}_n' \boldsymbol{X} = \mathbf{0}$. Throughout $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ is given in its singular value representation. In addition, elements of $\boldsymbol{Y}_0$ also are centered such that

$$\boldsymbol{Y}_0 - \overline{Y}\mathbf{1}_n = \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

The conventional assumptions A1. $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{V}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$; A2. $\mathcal{L}(\boldsymbol{\epsilon}) = N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ are set to apply, where σ^2 is taken to be unity unless stipulated otherwise. The OLS solutions are $\widehat{\boldsymbol{\beta}}_L = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.

We seek a design $\{\boldsymbol{X} \to \boldsymbol{X}_{\boldsymbol{\theta}}\}$ altered so as to enhance its $\mathfrak{F}_{\mathrm{M}}$ matrix $\mathfrak{I}(\boldsymbol{X}_{\boldsymbol{\theta}}) = \boldsymbol{X}_{\boldsymbol{\theta}}'\boldsymbol{X}_{\boldsymbol{\theta}}$ beyond that of $\mathfrak{I}(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{X}$ as intrinsic to the original model. The following is basic.

Definition 1. In reference to $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$, ensembles of matrix expansions in $\mathbb{F}_{n\times p}$ of note are

$$SD_{\mathbb{H}} = \{\boldsymbol{X} = \boldsymbol{L}\boldsymbol{D}_\xi \boldsymbol{Q}' \,|\, \boldsymbol{L} \in \mathbb{H}_{n\times p}\}; \quad SD_{\mathbb{H}}^{\boldsymbol{\xi}} = \{\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\theta}} \boldsymbol{Q}' \,|\, \boldsymbol{L} \in \mathbb{H}_{n\times p};\ \theta_i \geq \xi_i;\ 1 \leq i \leq p\}$$

3.2. Information-Enhanced designs

We next reconfigure $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ into designs $\boldsymbol{X}_{\boldsymbol{\theta}}$ having enhanced $\mathfrak{F}_{\mathtt{M}}$ matrices. These in turn are characterized within a large class, to include their explicit singular value expansions. All admit numerical evaluations through routinely available algorithms. To these ends take $\boldsymbol{\Xi}^{\frac{1}{2}} = \boldsymbol{Q}\boldsymbol{D}_\xi \boldsymbol{Q}'$ as the spectral square root of $\boldsymbol{\Xi} = \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{D}_\xi^2 \boldsymbol{Q}'$.

A fundamental result, and the key to our approach, rests on the existence of $\mathtt{W} \in \mathbb{F}_{k\times k}$ having singular values $\{\sigma(\mathtt{W}) = \boldsymbol{\omega} = (\omega_1, \ldots, \omega_p);\ \omega_i \geq 1\}$. Then $\boldsymbol{\Omega}$ represents a $\mathfrak{F}_{\mathtt{M}}$-enhancement of $\boldsymbol{\Xi}$ if and only if $\boldsymbol{\Omega} = \boldsymbol{\Xi}^{\frac{1}{2}}\mathtt{W}'\mathtt{W}\boldsymbol{\Xi}^{\frac{1}{2}}$ from Theorem A.1, Appendix, with $\mathtt{W}'\mathtt{W} = \boldsymbol{Q}_1\boldsymbol{D}_\omega^2\boldsymbol{Q}_1'$ as its spectral form, so that

$$\begin{aligned}\boldsymbol{\Omega} &= (\boldsymbol{Q}\boldsymbol{D}_\xi \boldsymbol{Q}')\left(\boldsymbol{Q}_1\mathbf{D}_\omega^2\boldsymbol{Q}_1'\right)(\boldsymbol{Q}\boldsymbol{D}_\xi \boldsymbol{Q}') \succeq_L \boldsymbol{\Xi} \\ &= \boldsymbol{Q}\boldsymbol{D}_\xi \boldsymbol{Q}'\boldsymbol{Q}\mathbf{D}_\omega^2\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{D}_\xi \boldsymbol{Q}' = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\xi}\cdot\boldsymbol{\omega}}^2\boldsymbol{Q}' \succeq_L \boldsymbol{\Xi}\end{aligned}$$

That $\boldsymbol{Q}$ may be substituted for $\boldsymbol{Q}_1$ follows since $\boldsymbol{Q}_1\boldsymbol{D}_\omega^2\boldsymbol{Q}_1' \succeq_L \boldsymbol{I}_k$ if and only if $\boldsymbol{Q}\boldsymbol{D}_\omega^2\boldsymbol{Q}' \succeq_L \boldsymbol{I}_k$. These steps in turn serve to establish conclusion (i) of the following pivotal result.

Theorem 1. *Consider the design $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ together with $\mathfrak{I}(\boldsymbol{X}) = \boldsymbol{\Xi}$, and let $\boldsymbol{D}_{\boldsymbol{\xi}\cdot\boldsymbol{\omega}}^2 = \boldsymbol{D}_{\boldsymbol{\theta}}^2$. Then*

(i) $\boldsymbol{\Omega}$ comprises a $\mathfrak{F}_{\mathtt{M}}$-enhancement of $\boldsymbol{\Xi}$ if and only if $\boldsymbol{\Omega} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\theta}}^2\boldsymbol{Q}'$ for $\{\theta_i^2 \geq \xi_i^2;\ 1 \leq i \leq p\}$.

(ii) The expansion $\boldsymbol{\Omega} = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\theta}}^2\boldsymbol{Q}'$ is the spectral decomposition for $\boldsymbol{\Omega}$, and $\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{Q}'$ the singular value expansion of $\boldsymbol{X}_{\boldsymbol{\theta}}$ having $\mathfrak{I}(\boldsymbol{X}_{\boldsymbol{\theta}}) = \boldsymbol{\Omega}$;

(iii) The class $SD_{\mathbb{H}}^{\boldsymbol{\xi}}$ of Definition 1 comprises an equivalence class of designs $\boldsymbol{X}_{\boldsymbol{\theta}} \in SD_{\mathbb{H}}^{\boldsymbol{\xi}}$ whose $\mathfrak{F}_{\mathtt{M}}$-matrices dominate that of $\boldsymbol{X}$ under the ordering $(\mathbb{S}_p^+, \succeq_L)$.

(iv) In particular, taking $\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{Q}'$ preserves the left- and right-singular vectors of $\boldsymbol{X}$.

Proof. Conclusion (i) is from $\boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{\xi}\cdot\boldsymbol{\omega}}^2\boldsymbol{Q}'$ since elements of $\boldsymbol{\omega}$ satisfy $\{\omega_i \geq 1;\ 1 \leq i \leq p\}$. With $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p]$, conclusion (ii) follows on verifying $\{(\boldsymbol{q}_i, \theta_i^2);\ 1 \leq i \leq p\}$ as the eigenvector–eigenvalue pairs for $\boldsymbol{\Omega}$. Next recall that the $\mathfrak{F}_{\mathtt{M}}$-matrix $\mathfrak{I}(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{X}$ does not depend on $\boldsymbol{P}$, and similarly $\mathfrak{I}(\boldsymbol{X}_{\boldsymbol{\theta}})$, so that $\mathfrak{I}(\boldsymbol{L}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{Q}') = \mathfrak{I}(\boldsymbol{P}\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{Q}')$ for any $\boldsymbol{L} \in \mathbb{H}_{n\times p}$, i.e., the same $\mathfrak{I}$ matrix holds for all. Accordingly, it suffices to choose any $\boldsymbol{L} \in \mathbb{H}_{n\times p}$, or equivalently any $\boldsymbol{X}_{\boldsymbol{\theta}} \in SD_{\mathbb{H}}^{\boldsymbol{\xi}}$ of Definition 1, to establish conclusion (iii) and then (iv) as a special case. □

3.3. Geometric mixtures

Arithmetic mixtures (AM) of $\mathfrak{F}_{\mathtt{M}}$ matrices, namely $\{(1-t)\boldsymbol{X}'\boldsymbol{X} + t\boldsymbol{Z}'\boldsymbol{Z};\ t \in [0, 1]\}$, were developed to advantage in Jensen and Ramirez (2017) for combating collinearity. An allied concept follows.

Definition 2. The collection of designs $\{\boldsymbol{X}_{\boldsymbol{\theta}}(t) = \boldsymbol{P}\mathrm{Diag}(\boldsymbol{\xi}^{1-t}\boldsymbol{\omega}^t)\boldsymbol{Q}';\ t \in [0, 1]\}$ is said to embody a *geometric mixture* (GM) of the vectors $(\boldsymbol{\xi}, \boldsymbol{\omega})$, with $\boldsymbol{P}\mathbf{D}_{\boldsymbol{\omega}}\boldsymbol{Q}'$ as the target design.

To illustrate, Table 1 refers to the ill-conditioned Body Fat Data from Neter et al. (1996) with $n = 20$, $p = 3$ regressors, and singular values $\sigma(\boldsymbol{X})$ as in the first row of Table 1. In an attempt

Table 1. Singular values $[\xi_1(t), \xi_2(t), \xi_3(t)]; t \in [0, 1]$ of the geometric mixture designs $\{\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{PD}(\xi_i^{(1-t)}\omega_i^t)\boldsymbol{Q}'\}$ as t varies for the Body Fat Data with target $\boldsymbol{\omega}' = [\xi_1, \xi_1, \xi_1]$, and condition number $\kappa(t) = \xi_1(t)/\xi_p(t)$.

t	$\xi_1(t)$	$\xi_2(t)$	$\xi_3(t)$	$\kappa(t)$
0.0	1.43752	0.96582	0.02696	53.32
0.2	1.43752	1.04578	0.05971	24.08
0.4	1.43752	1.13235	0.13226	10.87
0.6	1.43752	1.22610	0.29297	4.91
0.8	1.43752	1.32761	0.64897	2.22
1.0	1.43752	1.43752	1.43752	1.00

to improve conditioning, as well as its $\mathfrak{F}_{\mathrm{M}}$ matrix, take the target to be $\boldsymbol{X}_{\boldsymbol{\omega}}$ having $\sigma(\boldsymbol{X}_{\boldsymbol{\omega}}) = [\omega, \omega, \omega]$ with $\omega = \xi_1$ and condition number $\kappa = 1.0$ as listed in Table 1. In consequence, the GM model $\{\boldsymbol{X}_{\boldsymbol{\theta}}(t) = \boldsymbol{P}\mathrm{Diag}(\boldsymbol{\xi}^{1-t}\boldsymbol{\omega}^t)\boldsymbol{Q}';\ t \in [0, 1]\}$ has condition numbers $\kappa(t)$ decreasing monotonically with increasing t in $[0, 1]$ as in the final column of Table 1.

Remark 1. In practice $\boldsymbol{X}_{\boldsymbol{\omega}} = \boldsymbol{PD}_{\boldsymbol{\omega}}\boldsymbol{Q}'$ may be considered as a "target". Then the geometric mixture provides a continuum, leveraging the beginning design $\boldsymbol{X}_{\xi}$ more and more toward its target as $\{t \uparrow \in [0, 1]\}$. By construction, the altered design $\boldsymbol{X}_{\boldsymbol{\theta}}$ tends to the "target" $\boldsymbol{X}_{\boldsymbol{\omega}}$ as $t \uparrow 1$ when the latter is well-conditioned. This desirable property does not hold for ridge and surrogate designs where as $k \to \infty$ the perturbed design $\boldsymbol{X}_k$ is infeasible in the limit with entries tending to infinity.

3.4. Properties of the solutions

Linear estimators from the altered design $\boldsymbol{X}_{\boldsymbol{\theta}}$ are $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}} = (\boldsymbol{X}_{\boldsymbol{\theta}}'\boldsymbol{X}_{\boldsymbol{\theta}})^{-1}\boldsymbol{X}_{\boldsymbol{\theta}}'\boldsymbol{Y}$. Essential properties are listed in Table 2. These follow directly from first principles, albeit sometimes tedious determinations. Standard criteria for design evaluation are $[\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}]$ where for design $\boldsymbol{X}$, we have $\kappa(\boldsymbol{X}) = \xi_1/\xi_p$ as the ratio of its extreme singular values; if $\mathrm{V}(\widehat{\boldsymbol{\beta}}_L) = \boldsymbol{\Sigma} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$ under OLS, then $\mathtt{A} = \mathrm{tr}(\boldsymbol{\Sigma})$, $\mathtt{D} = |\boldsymbol{\Sigma}|$, and $\mathtt{E} = \lambda_1(\boldsymbol{\Sigma})$, its largest eigenvalue. These values are identified in Table 2 as they apply for $\boldsymbol{X}_{\boldsymbol{\theta}}$.

3.5. MSE Considerations

In estimating $\boldsymbol{\beta}$ using $\widetilde{\boldsymbol{\beta}}$ with bias $B(\widetilde{\boldsymbol{\beta}}) = \mathrm{E}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})$, its mean square error is $\mathtt{MSE}(\widetilde{\boldsymbol{\beta}}) = \mathrm{tr}\,(\mathrm{V}(\widetilde{\boldsymbol{\beta}})) + (\boldsymbol{\beta}_0 - \boldsymbol{\beta})'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$, effecting the variance-bias trade-off under

Table 2. Conversion of $\boldsymbol{X} = \boldsymbol{PD}_{\xi}\boldsymbol{Q}'$ to the $\mathfrak{F}_{\mathrm{M}}$ enhanced $\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{PD}_{\boldsymbol{\theta}}\boldsymbol{Q}'$, together with the resulting linear estimators and their essential properties.

$\mathfrak{F}_{\mathrm{M}}$-Enhanced regression	Properties
$\boldsymbol{X} \to \boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{PD}_{\boldsymbol{\theta}}\boldsymbol{Q}'$	$\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) = \boldsymbol{QD}_{\xi}^{t}\mathbf{D}_{\boldsymbol{\omega}}^{-t}\boldsymbol{Q}'\boldsymbol{\beta}$
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}} = (\boldsymbol{X}_{\boldsymbol{\theta}}'\boldsymbol{X}_{\boldsymbol{\theta}})^{-1}\boldsymbol{X}_{\boldsymbol{\theta}}'\boldsymbol{Y}.$	$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{QD}_{\xi}^{-2(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-2t}\boldsymbol{Q}'$
$(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}} - \widehat{\boldsymbol{\beta}}_L) = \boldsymbol{Q}[\boldsymbol{D}_{\xi}^{-(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-t} - \boldsymbol{D}_{\xi}^{-1}]\boldsymbol{P}'\boldsymbol{Y}$	$\lambda(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})^{\sim} = \{\xi_i^{-2(1-t)}\omega_i^{-2t}; 1 \le i \le p\}$
Efficiency indices $[\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}]$	
$\kappa(\boldsymbol{X}_{\boldsymbol{\theta}}) = \frac{\max\{\xi_i^{1-t}\omega_i^t\}}{\min\{\xi_i^{1-t}\omega_i^t\}}$	$\mathtt{A} = \mathrm{tr}(\boldsymbol{D}_{\xi}^{-2(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-2t})$
$\mathtt{D} = \lvert(\boldsymbol{D}_{\xi}^{-2(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-2t})\rvert$	$\mathtt{E} = \max\{\xi_i^{-2(1-t)}\omega_i^{-2t};\ 1 \le i \le p\}$

squared error loss. Equivalently, taking $\boldsymbol{\psi} = \boldsymbol{Q}'\boldsymbol{\beta}$ in canonical form with $\boldsymbol{Q}$ orthogonal, and $\widetilde{\boldsymbol{\psi}} = \boldsymbol{Q}'\widetilde{\boldsymbol{\beta}}$, it suffices that $\mathtt{MSE}(\widetilde{\boldsymbol{\beta}}) = \mathtt{MSE}(\widetilde{\boldsymbol{\psi}})$. To these ends consider the canonical form in which $\boldsymbol{\psi} = \boldsymbol{Q}'\boldsymbol{\beta}, \boldsymbol{\psi}_{\boldsymbol{\theta}} = \boldsymbol{Q}'\boldsymbol{\beta}_{\boldsymbol{\theta}}$, together with $\widehat{\boldsymbol{\psi}}$ and $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}$. Accordingly, on returning to Table 2 we have the following pivotal result.

Theorem 2. *Consider the* MSE *for* $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}$ *as in Table 2. Then*

(i) $\mathtt{MSE}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) = \mathtt{MSE}(\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}) = \sigma^2 tr\,(\boldsymbol{D}_{\xi}^{-2}\boldsymbol{D}_{\xi}^{2t}\mathbf{D}_{\boldsymbol{\omega}}^{-2t}) + (\boldsymbol{\psi}_{\boldsymbol{\theta}} - \boldsymbol{\psi})'\boldsymbol{D}_{\xi}^{2t}\mathbf{D}_{\boldsymbol{\omega}}^{-2t}(\boldsymbol{\psi}_{\boldsymbol{\theta}} - \boldsymbol{\psi})$.

(ii) For $\boldsymbol{\omega} = [\xi_1, \ldots, \xi_1]$, *the elements of* $(\boldsymbol{D}_{\xi}^{2t}\mathbf{D}_{\boldsymbol{\omega}}^{-2t})$ *are all less than or equal to unity for* $t \in (0,\ 1)$.

Proof. On reinstating the scalar σ^2 in Table 2, we have the bias $\boldsymbol{B}$ and dispersion matrices in equivalent forms as

$$\boldsymbol{B} = [\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) - \boldsymbol{\beta}] = \boldsymbol{Q}\left(\boldsymbol{D}_{\xi}^{t}\mathbf{D}_{\boldsymbol{\omega}}^{-t} - \boldsymbol{I}_p\right)\boldsymbol{Q}'\boldsymbol{\beta} \to \left(\boldsymbol{D}_{\xi}^{t}\mathbf{D}_{\boldsymbol{\omega}}^{-t} - \boldsymbol{I}_p\right)\boldsymbol{\psi}$$
$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) = \sigma^2\boldsymbol{Q}\boldsymbol{D}_{\xi}^{-2(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-2t}\boldsymbol{Q}' \to \mathrm{V}(\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}) = \sigma^2\boldsymbol{D}_{\xi}^{-2(1-t)}\mathbf{D}_{\boldsymbol{\omega}}^{-2t}$$

as in conclusion (i), giving conclusion (ii) directly. □

Remark 2. Thus with $\boldsymbol{\omega} = [\,\xi_1, \ldots, \xi_1]$ the variance term decreases from $\sigma^2\sum_{i=1}^{p}\frac{1}{\xi_i^2}$ at $t = 0$ to $\sigma^2\frac{p}{\xi_1^2}$ at $t = 1$, while the bias term increases from 0 to ∞. As MSE is a function of the unknown parameters σ^2 and $\boldsymbol{\beta}$, finding the minimal MSE currently remains insolvable. Nonetheless, knowing the *structure* of MSE as in Theorem 2 may prove helpful to prospective users.

4. Case studies

4.1. Basics

We illustrate with near collinear data, the columns of $\boldsymbol{X}$ having been both centered and scaled. Hallmarks of ill-conditioning include the Variance Inflation Factors (VIFs) as ratios of actual to "ideal" variances of estimators had the columns of $\boldsymbol{X}$ been linearly independent; and noted by Marquardt and Snee (1975) as "the best single measure of the conditioning of the data." Rules-of-thumb that VIFs be accorded of consequence are those exceeding 10 or even 4; see, for example, Myers (1990), O'Brien (2007) and Sengupta and Bhimasankaram (1997).

To continue, take $\mathtt{V}_{\mathtt{M}} = \max\{\mathtt{VIF};\ 1 \le i \le p\}$ to gage the overall ill-conditioning of the system. Several design features to be reported for each modified $\boldsymbol{X}_{\boldsymbol{\theta}}$ with dispersion matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are

$$[\,\mathtt{V}_{\mathtt{M}},\ \kappa,\ \mathtt{A},\ \mathtt{D},\ \mathtt{E},\ \mathtt{M},\ \mathtt{MAD},\ \Delta] \tag{2}$$

where $[\kappa,\ \mathtt{A},\ \mathtt{D},\ \mathtt{E}]$ are as defined but now $\mathtt{D} = \log|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|$; M is Mauchly's (1940) criterion for the sphericity of the Gaussian contours of $\mathcal{L}(\widehat{\boldsymbol{\beta}})$. In addition, $\mathtt{MAD}(\boldsymbol{X}_{\boldsymbol{\theta}}) = \frac{1}{np}\sum|x_{ij}(\theta) - x_{ij}|$ gages the discrepancy between $\boldsymbol{X} = [x_{ij}]$ and its modified $\boldsymbol{X}_{\boldsymbol{\theta}} = [\,x_{ij}(\theta)]$, and $\Delta(\boldsymbol{X}_{\boldsymbol{\omega}}) = \sum|\xi_j(\theta) - \xi_j|$ for singular values. Clearly $[\kappa,\ \mathtt{A},\ \mathtt{D},\ \mathtt{E}]$ ideally would be small, whereas Mauchly's $\mathtt{M}(\cdot)$ would increase toward unity for t increasing in $[\,0, 1\,]$, and thus toward more nearly spherical contours, in increasingly well conditioned data.

Of the designs to be examined. Hadi (2011) identified the Surrogate procedures of Jensen and Ramirez (2008) as among the principal techniques for mitigating collinearity. The Surrogate $\boldsymbol{X}_k$ is listed in Table 3, as are $\boldsymbol{Z}_t$ and $\boldsymbol{G}_t$ as the arithmetic (AM) and geometric (GM) mixtures of their respective $\mathfrak{F}_{\mathtt{M}}$ matrices. Choices for $k \in [\,0, \infty)$ in $\boldsymbol{X}_k$, and for the mixing

Table 3. Altered designs studied numerically, including the Surrogate Design $\boldsymbol{X}_k$; the AM mixture $\boldsymbol{Z}_t$ with $\overline{\xi^2} = (\sum_{i=1}^{p} \xi_i^2)/p$; and the GM mixture $\boldsymbol{G}_t$, together with their $\mathfrak{F}_{\mathrm{M}}$ matrices.

Item	Design	Fisher Information Matrix
$\boldsymbol{X}_k$	$\boldsymbol{P}\,\mathrm{Diag}((\xi_i^2+k)^{\frac{1}{2}})\boldsymbol{Q}'$	$\boldsymbol{Q}(\boldsymbol{D}_\xi^2+k\boldsymbol{I}_p)\boldsymbol{Q}'$
$\boldsymbol{Z}_t$	$\boldsymbol{P}\,\mathrm{Diag}([(1-t)\xi_i^2+t\overline{\xi^2}]^{\frac{1}{2}}\ \boldsymbol{Q}'$	$\boldsymbol{Q}\,\mathrm{Diag}([(1-t)\xi_i^2+t\overline{\xi^2}])\,\boldsymbol{Q}'$
$\boldsymbol{G}_t$	$\boldsymbol{P}\mathrm{Diag}(\xi_i^{1-t}\omega_i^t)\boldsymbol{Q}'$	$\boldsymbol{Q}\,\mathrm{Diag}(\xi_i^{2(1-t)}\omega_i^{2t})\,\boldsymbol{Q}'$

Table 4. A small design $\boldsymbol{X}' = \boldsymbol{G}_0'$; and its GM mixture $\boldsymbol{G}_t'$ at $t = 0.0666$.

	−0.3536	0.3536	−0.5000	0.5000	−0.3536	0.3536	0.0000	0.0000
	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000	− 0.5000	− 0.5000
$\boldsymbol{G}_0'$	− 0.2965	0.2224	− 0.2965	0.7412	− 0.2965	0.2224	− 0.2965	0.0000
	− 0.3566	0.3614	− 0.5039	0.4921	− 0.3566	0.3614	0.0107	− 0.0083
$\boldsymbol{G}_t'$	0.0021	− 0.0001	0.5144	0.5033	0.0021	− 0.0001	− 0.5069	− 0.5148
	− 0.2944	0.2152	− 0.2985	0.7473	− 0.2944	0.2152	− 0.3037	0.0134

parameters $t \in [0, 1]$, are derived numerically so as to achieve maximal $\mathrm{V_M} = 10$ consistently across case studies, to place them on common ground. All computations are supported by the Maple software package.

Remark 3. (i) Ridge solutions are not included, subsumed instead by Surrogate solutions, as reported here, having uniformly smaller residual sums of squares point-wise for each $k > 0$; see Jensen and Ramirez (2010a). (ii) In addition, although intended to ameliorate ill-conditioning over a wide range of k, Ridge often must be abandoned in favor of OLS for values of k exceeding a threshhold k_0 as stipulated in Jensen and Ramirez (2010b). This in turn would further complicate computations for Ridge solutions. (iii) Finally, the original definition of VIF as given in Marquardt (1970) is in dispute, allowing values less than unity and thus "inconsistent with the theoretical definition of VIF." See García et al. (2015a, 2015b) and Salmerón et al. (2016).

Case studies follow.

(1) **A Small Design.** The data having $n = 8$ and $p = 3$ are from Jensen and Ramirez (2017) as in Table 4. The purpose here is to exhibit the design matrix $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}'$ as $\boldsymbol{G}_0$, its GM mixture with $\boldsymbol{X}_\omega = \boldsymbol{P}\mathrm{Diag}(\xi_1, \xi_1, \xi_1)\boldsymbol{Q}'$ as $\boldsymbol{G}_t$, so that the reader may gage the proximity of $\boldsymbol{G}_0$ to $\boldsymbol{G}_t$ visually, in addition to the diagnostics $\mathrm{MAD}(\boldsymbol{G}_t) = 0.0060$ and $\Delta(\boldsymbol{G}_t) = 0.0504$ as in Table 5. Observe that $t = 0.0666$ is determined to achieve $\mathrm{V_M} = 10$. This in turn reflects a negligible displacement of $\boldsymbol{G}_t$ from $\boldsymbol{G}_0$, as may be seen on comparing their respective elements in Table 4. Enhancement of $\mathfrak{I}(\boldsymbol{X}) = \boldsymbol{\Xi}$ is tantamount to diminishing the corresponding dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Xi}^{-1}$, a compelling reason to enhance $\mathfrak{I}(\boldsymbol{X})$. Nonetheless, despite the negligible shift $t = 0.0666$, reductions in the diagnostics $[\kappa, \mathrm{A}, \mathrm{D}, \mathrm{E}]$ amount to $[12.25\%,\ 21.74\%,\ 11.76\%,\ 22.89\%]$, respectively. In addition to decreases in the diagnostics $[\kappa, \mathrm{A}, \mathrm{D}, \mathrm{E}]$, conditioning of the dispersion matrix has reduced from 49.28 for OLS under $\boldsymbol{G}_0$, to 37.95 for $\boldsymbol{G}_t$.

Table 5. A small design $\boldsymbol{X}' = \boldsymbol{G}_0'$; its GM mixture $\boldsymbol{G}_t'$ at $t = 0.0666$; and diagnostics for each.

Diagnostics	$\mathrm{V_M}$	κ	A	D	E	M	MAD	Δ
$\boldsymbol{G}_0$	12.81	7.02	26.63	2.55	25.12	0.0183	0	0
$\boldsymbol{G}_t$	10.00	6.16	20.84	2.25	19.37	0.0282	0.0060	0.0504

To illustrate that $SD_{\mathbb{H}}$ from Definition 1 is a class of designs all with a common moment matrix, consider the design $\boldsymbol{G}_0 = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ from Table 4. For $\boldsymbol{L} \in \mathbb{H}_{n\times p}$ we set $\boldsymbol{L} = \boldsymbol{P}\boldsymbol{R}$ with $\boldsymbol{R} \in \mathcal{O}_p$ as the $(p \times p)$ varimax rotation matrix from the Minitab Factor Analysis procedure applied to $\boldsymbol{P}$. The matrix $\boldsymbol{W} = \boldsymbol{L}\boldsymbol{D}_\xi \boldsymbol{Q}' \in SD_{\mathbb{H}}$ satisfies $\boldsymbol{W}'\boldsymbol{W} = \boldsymbol{X}'\boldsymbol{X}$ and is given by

$$\boldsymbol{W}' = \begin{bmatrix} -0.3315 & 0.3721 & -0.6251 & 0.2033 & -0.3315 & 0.3721 & 0.2516 & 0.0890 \\ -0.1159 & 0.2693 & 0.3569 & 0.2724 & -0.1159 & 0.2693 & -0.1612 & -0.7749 \\ -0.2304 & 0.2690 & -0.6760 & -0.1278 & -0.2304 & 0.2690 & 0.4404 & 0.2863 \end{bmatrix}$$

to be compared with $\boldsymbol{G}_0$ of Table 4.

Surveys of case studies from the literature are listed next. Comparative summaries from their numerical analyses are given subsequently.

(2) **Acetylene Data.** Taken from Marquardt and Snee (1975), the Reduced Quadratic Model has $n = 16$ and $p = 5$ with explanatory variables x_1 : reactor temperature; x_2 : ratio of H_2 to n-heptone; x_3 : contact time; x_1x_2 : interaction; x_1^2 : squared temperature; and y : the conversion percentage of n-heptone to acetylene. The conditioning improves from $\kappa(\boldsymbol{X}) = 218.33 \to \{\kappa(\boldsymbol{X}_k) = 7.22, \kappa(\boldsymbol{Z}_t) = 7.22, \kappa(\boldsymbol{G}_t) = 7.08\}$.

(3) **Body Fat Data.** The data are given in Neter et al. (1996) with $n = 20$ and $p = 3$ explanatory variables, namely x_1 : tricep skinfold thickness; x_2 : thigh circumference; x_3 : midarm circumference; and y : the amount of body fat. The conditioning improves from $\kappa(\boldsymbol{X}) = 53.33 \to \{\kappa(\boldsymbol{X}_k) = 6.17, \kappa(\boldsymbol{Z}_t) = 6.17, \kappa(\boldsymbol{G}_t) = 6.17\}$.

(4) **French Economy Data.** Analysis is given in Chatterjee and Hadi (2006) to model the French Economy for years 1949 to 1959 with $n = 11$ and $p = 3$ explanatory variables, namely x_1 : domestic production; x_2 : stock formation; x_3 : domestic consumption; and y : imports. The conditioning improves from $\kappa(\boldsymbol{X}) = 27.24 \to \{\kappa(\boldsymbol{X}_k) = 6, 16, \kappa(\boldsymbol{Z}_t) = 6.16, \kappa(\boldsymbol{G}_t) = 6.16\}$.

(5) **Hospital Manpower Data.** As reported in Myers (1990), the Hospital Manpower Data comprise records at $n = 17$ U. S. Naval Hospitals with $p = 5$ regressors, specifically x_1 : average daily patient load; x_2 : monthly X-ray exposures; x_3 : monthly occupied bed days; x_4 : eligible population in the area divided by 1000; x_5 : average length of patients' stay in days; and y : monthly man-hours. The conditioning improves from $\kappa(\boldsymbol{X}) = 278.87 \to \{\kappa(\boldsymbol{X}_k) = 7.68, \kappa(\boldsymbol{Z}_t) = 7.68, \kappa(\boldsymbol{G}_t) = 7.97\}$.

(6) **Number of Active Metropolitan Physicians.** The Standard Metropolitan Statistical Area (SMSA) data have $n = 141$ and $p = 3$ from the website.[1] The variables are x_1 : total population (in thousands); x_2 : land area (in square miles); x_3 : total personal income (in millions of dollars); and y : number of active physicians. The conditioning improves from $\kappa(\boldsymbol{X}) = 29.22 \to \{\kappa(\boldsymbol{X}_k) = 6.20, \kappa(\boldsymbol{Z}_t) = 6.20, \kappa(\boldsymbol{G}_t) = 6.19\}$.

(7) **Summary.** Essentials from these studies are summarized in Table 6 with reference to their $[\,\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}\,]$ efficiency indices. The well-conditioned $\boldsymbol{P}\mathrm{Diag}(\xi_1, \ldots, \xi_1)\boldsymbol{Q}'$ is the target design for the GM mixture $\boldsymbol{G}_t$ where values for ξ_1 are listed in Table 6 for each case study. For the AM mixture, the target design for $\boldsymbol{Z}_t$ is the well conditioned $\boldsymbol{P}\mathrm{Diag}([\overline{\xi^2}]^{\frac{1}{2}}, \ldots, [\overline{\xi^2}]^{\frac{1}{2}})\boldsymbol{Q}'$.

The parameter $k \in [\,0, \infty\,)$ for $\boldsymbol{X}_k$, and the mixing parameters $t \in [\,0, 1\,]$ for $\boldsymbol{Z}_t$ and $\boldsymbol{G}_t$, were chosen to give $\mathtt{V}_{\mathtt{M}} = 10$ in order to place the altered designs on comparable scales.

In regard to the $[\,\mathtt{A}, \mathtt{D}, \mathtt{E}\,]$ efficiencies, summary comparisons among $\{\boldsymbol{X}_k, \boldsymbol{Z}_t, \boldsymbol{G}_t\}$ are given in Table 7 where $\boldsymbol{X} \approx \boldsymbol{Z}$ identifies $(\boldsymbol{X}, \boldsymbol{Z})$ on balance to be comparable. Recalling that smaller

[1] [https://onlinecourses. science.psu.edu/stat857/sites/onlinecourses.science.psu. edu.stat857/files/smsa.data]

Table 6. Comparison of [κ, A, D, E] efficiency indices across choices among $(\boldsymbol{X}_k, \boldsymbol{Z}_t, \boldsymbol{G}_t)$ together with ξ_1 for the five case studies.

$\boldsymbol{X}_k$				$\boldsymbol{Z}_t$				$\boldsymbol{G}_t$			
κ^2	A	D	E	κ^2	A	D	E	κ^2	A	D	E
Acetylene ($\xi_1 = 1.8222$)											
52.07	39.30	5.78	15.38	52.07	41.86	6.10	16.38	50.15	22.05	2.43	15.10
Body Fat ($\xi_1 = 1.4375$)											
38.11	19.45	2.15	17.96	38.11	20.52	2.31	18.95	38.07	19.60	1.83	18.42
French Economy ($\xi_1 = 1.4139$)											
38.00	19.97	2.15	18.53	38.00	21.00	2.30	19.48	37.99	20.24	1.94	19.00
Hospital Manpower ($\xi_1 = 2.0487$)											
59.09	30.28	5.45	13.84	59.09	32.46	5.80	14.83	63.57	18.13	0.77	15.15
SMSA ($\xi_1 = 1.4255$)											
38.48	19.93	2.16	18.47	38.48	20.96	2.32	19.42	38.30	20.08	1.92	18.85

Table 7. Summary comparisons of $\{\boldsymbol{X}_k, \boldsymbol{Z}_t, \boldsymbol{G}_t\}$ with regard to the [A, D, E] efficiency criteria.

Case Study	$\kappa(\boldsymbol{X})$	Preference Order
Acetylene Data	218.33	$\boldsymbol{G}_t \prec \boldsymbol{X}_k \prec \boldsymbol{Z}_t$
Body Fat Data	53.33	$(\boldsymbol{X}_k \approx \boldsymbol{G}_t) \prec \boldsymbol{Z}_t$
French Economy Data	27.24	$\boldsymbol{X}_k \approx \boldsymbol{Z}_t \approx \boldsymbol{G}_t$
Hospital Manpower Data	278.87	$\boldsymbol{G}_t \prec \boldsymbol{X}_k \prec \boldsymbol{Z}_t$
SMSA Data	29.22	$\boldsymbol{X}_k \approx \boldsymbol{Z}_t \approx \boldsymbol{G}_t$

[A, D, E] values reflect greater efficiencies, we use the designation $\boldsymbol{X} \prec \boldsymbol{Z}$ to reflect smaller thus more efficient [A, D, E] values for $\boldsymbol{X}$. Also listed are values $\kappa(\boldsymbol{X})$ for conditioning of the original data. For the Acetylene and Hospital Manpower Data, more highly ill-conditioned at the outset, the GM mixture $\boldsymbol{G}_t$ is preferred in efficiency to the Surrogate $\boldsymbol{X}_k$, in turn preferred to the AM mixture $\boldsymbol{Z}_t$. For the intermediate $\kappa(\boldsymbol{X}) = 53.33$ in the Body Fat Data, $\boldsymbol{X}_k$ and $\boldsymbol{G}_t$ on balance are comparable but preferred to $\boldsymbol{Z}_t$. On the other hand, $(\boldsymbol{X}_k, \boldsymbol{Z}_t, \boldsymbol{G}_t)$ are comparable under the smaller initial conditioning of the French Economy and SMSA Data.

4.2. Choices for VIF

In the foregoing studies, as noted, the perturbation parameters $\{k,\ t\}$ were chosen numerically so as to standardize $\mathrm{V_M} = \max\{\mathrm{VIF};\ 1 \le i \le p\}$ to the common value $\mathrm{V_M} = 10$ as suggested in O'Brien (2007), for example. We acknowledge a Reviewer's suggestion to consider this further. Accordingly, Table 8 shows for the Acetylene Data the corresponding values for $\{\kappa^2, \mathrm{A}, \mathrm{D}, \mathrm{E}\}$ with the perturbation parameter, either k or t, chosen so as to achieve the reduction of $\mathrm{V_M}$

Table 8. Comparisons of $\{\boldsymbol{X}_k, \boldsymbol{Z}_t, \boldsymbol{G}_t\}$ with regard to the indices $\{\kappa^2, \mathrm{A}, \mathrm{D}, \mathrm{E}\}$ with constraints $\mathrm{V_M} = \{5,\ 20.\}$

$\mathrm{V_M}$	Index	$\mathbf{X}_k$	$\mathbf{Z}_t$	$\mathbf{G}_t$
5	κ^2	23.47	23.47	22.75
20	κ^2	111.57	111.57	106.39
5	A	19.12	21.96	11.30
20	A	77.08	79.39	42.95
5	D	3.57	4.26	0.73
20	D	7.66	7.81	4.05
5	E	6.77	7.76	6.85
20	E	33.30	34.30	32.04

to $\{5, 20\}$ in addition to $\mathrm{V_M} = 10$ as in Table 6. For nearly all of the row comparisons, the ordering $\boldsymbol{G}_t \prec \boldsymbol{X}_k \prec \boldsymbol{Z}_t$ holds as in Table 7 for the case $\mathrm{V_M} = 10$.

5. Conclusions

Beginning with the Fisher Information Matrix $(\mathfrak{F}_{\mathrm{M}})\,\boldsymbol{\Xi} = \boldsymbol{Q}\boldsymbol{D}_{\xi}^2\boldsymbol{Q}'$ from the design $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_{\xi}\boldsymbol{Q}'$, foundations for this study rest on characterizing the congruences $\boldsymbol{T}$ for which $\boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}' \succeq_L \boldsymbol{\Xi}$ as ordered in Loewner (1934). In this case, $\boldsymbol{T}$ can be decomposed as $\boldsymbol{T} = \boldsymbol{\Xi}^{\frac{1}{2}}\mathbf{W}\boldsymbol{\Xi}^{-\frac{1}{2}}$ in which the singular values of $\mathbf{W}$ satisfy $\{\sigma_i(\mathbf{W}) \geq 1;\ 1 \leq i \leq p\}$ from Theorem 2. Theorem 1 shows that $\boldsymbol{\Omega} = \boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}' \succeq_L \boldsymbol{\Xi}$ for $\boldsymbol{\Omega}$ having the spectral decomposition $\boldsymbol{\Omega} = \boldsymbol{Q}\boldsymbol{D}_{\xi\cdot\omega}^2\boldsymbol{Q}'$. Thus enhancing the second moment matrix is tantamount to increasing the singular values $\{\xi_i \to \xi_i\omega_i;\ 1 \leq i \leq p\}$. The geometric mixture (GM) design has the decomposition $\boldsymbol{G}_t = \boldsymbol{P}\mathrm{Diag}(\xi_i^{1-t}\omega_i^t)\boldsymbol{Q}'$ together with its $\mathfrak{F}_{\mathrm{M}}$ matrix $\boldsymbol{\Xi}_t = \boldsymbol{G}_t'\boldsymbol{G}_t$.

We have chosen the target design $\boldsymbol{G}_1 = \boldsymbol{P}\mathrm{Diag}(\omega_1, \ldots, \omega_p)\boldsymbol{Q}'$ with $\{\omega_i = \xi_1;\ 1 \leq i \leq p\}$ to satisfy $\{\xi_i^{1-t}\omega_i^t = \xi_i(\frac{\xi_1}{\xi_i})^t \geq \xi_i;\ 1 \leq i \leq p\}$, to assure that $\boldsymbol{\Xi}_t \succeq_L \boldsymbol{\Xi}$. Additionally, with this target design, if $\{0 \leq t_1 \leq t_2 \leq 1\}$, then $\{\boldsymbol{\Xi}_1 \succeq_L \boldsymbol{\Xi}_{t_2} \succeq_L \boldsymbol{\Xi}_{t_1} \succeq_L \boldsymbol{\Xi}\}$, so that $\{\boldsymbol{\Xi}_t;\ t \in [0, 1]\}$ is monotone increasing under the matrix ordering $\prec$ as $t \uparrow \in [0, 1]$. In consequence their inverses, as dispersion matrices for $\widehat{\boldsymbol{\beta}}(t)$, are reverse ordered so that the [A, D, E] criteria all decrease monotonically with increasing t. Using five data sets from the literature, we have demonstrated the manner in which the GM designs serve to mitigate ill-conditioning as well as to improve the efficiency indices [A, D, E]. We have compared the GM designs to the Surrogate designs $\boldsymbol{X}_k$ and to the corresponding Arithmetic Mixture (AM) designs $\boldsymbol{Z}_t$. The perturbation parameters (k, t) were chosen to achieve the maximal variance inflation factor $\mathrm{V_M} = 10$ to facilitate comparisons among the designs. In all our case studies, the GM procedure produced either a superior design with smaller efficiency indices, or a design nearly equivalent in its efficiencies.

ORCID

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

References

Chatterjee, S., and A. Hadi. 2006. *Regression analysis by example*. 4th ed. Hoboken, NJ: John Willey & Sons.

García, C., J. García, M. D. M. López, and R. Salmerón. 2015a. Collinearity: Revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics* 42:648–61.

García, J., R. Salmerón, C. García, and M. D. M. López. 2015b. Standardization of variables and collinearity diagnostic in ridge regression. *International Statistical Review* 84:245–66.

Hadi, A. S. 2011. Ridge and surrogate ridge regressions. In *International encyclopedia of statistical science*, ed. M. Lovric, Berlin: Springer.

Jensen, D. R., and D. E. Ramirez. 1990. Dispersion-diminishing transformations. *Communications in Statistics–Theory and Methods* 19:3259–66.

Jensen, D. R., and D. E. Ramirez. 2008. Anomalies in the foundations of ridge regression. *International Statistical Review* 76:89–105.

Jensen, D. R., and D. E. Ramirez. 2010a. Surrogate models in ill-conditioned systems. *Journal of Statistical Planning and Inference* 140:2069–77.

Jensen, D. R., and D. E. Ramirez. 2010b. Tracking MSE efficiencies in ridge regression. *Advances and Applications in Statistics Science* 1:381–98.

Jensen, D. R., and D. E. Ramirez. 2017. Singular majorants and minorants: Enhanced design conditioning. *Journal of Statistical Computation and Simulation* 87:1–15.
Loewner, C. 1934. Über monotone Matrixfunktonen. *Mathematische Zeitschrift* 38:177–216.
Marquardt, D. W. 1970. Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics* 12:591–612.
Marquardt, D. W., and R. D. Snee. 1975. Ridge regression in practice. *Journal of the American Statistical Association* 29:3–20.
Mauchly, J. W. 1940. Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics* 11:204–9.
Myers, R. H. 1990. *Classical and modern regression with applications*. 2nd ed. Boston, MA: PWS-Kent Publishing.
Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. *Applied linear statistical models*. 4th ed. Chicago, IL: Irwin.
O'Brien, R. M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 41:673–90.
Salmerón, R., J. García, M. D. M. López, and C. García. 2016. Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *Journal of Applied Statistics* 43:1821–49.
Sengupta, D., and P. Bhimasankaram. 1997. On the roles of observations in collinearity in the linear model. *Journal of the American Statistical Association* 92:1024–32.

Appendix

Considered here are similarity transformations $T(\boldsymbol{\Xi}) = \boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}'$ on $(\mathbb{S}_p^+, \succeq_L)$ ordered as in Loewner (1934). We have the following.

Definition A.1.

(i) Call $T(\boldsymbol{\Xi}) = \boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}'$ as *order-increasing* if $\boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}' \succeq_L \boldsymbol{\Xi}$, and as *order-decreasing* if $\boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}' \prec \boldsymbol{\Xi}$ under $\succeq_L$.

(ii) Specifically,

$$\mathrm{D}^{\uparrow}(\boldsymbol{\Xi}) = \{\boldsymbol{G} \in \mathbb{S}_k^+ : \boldsymbol{G} \succeq_L \boldsymbol{\Xi}\} \text{ are matrices dominating } \boldsymbol{\Xi};$$
$$\mathrm{D}^{\downarrow}(\boldsymbol{\Xi}) = \left\{\boldsymbol{L} \in \mathbb{S}_k^+ : \text{Ł} \prec \boldsymbol{\Xi}\right\} \text{ are matrices dominated by } \boldsymbol{\Xi};$$

(iii) Let $\tau(\boldsymbol{\Xi}) = \{\boldsymbol{T} : \boldsymbol{T} = \boldsymbol{\Xi}^{\frac{1}{2}}\mathtt{W}'\boldsymbol{\Xi}^{-\frac{1}{2}}\}$ with $\mathtt{W} \in \mathbb{F}_{k\times k}$; so that $\boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}' = \boldsymbol{\Xi}^{\frac{1}{2}}\mathtt{W}'\mathtt{W}\boldsymbol{\Xi}^{\frac{1}{2}}$.

Against this background the following determination is essential.

Theorem A.1. *For $\boldsymbol{\Xi} \in \mathbb{S}_k^+$ and $\boldsymbol{T} \in \mathbb{F}_{k\times k}$ consider the similarity transformations $T(\boldsymbol{\Xi}) = \boldsymbol{T}\boldsymbol{\Xi}\boldsymbol{T}'$ together with $\mathtt{W} \in \mathbb{F}_{k\times k}$ having singular values $\sigma(\mathtt{W}) = (\sigma_1, \ldots, \sigma_p)$.*

(i) $T(\boldsymbol{\Xi})$ is order-increasing if and only if $\{\sigma_i(\mathtt{W}) \geq 1;\ 1 \leq i \leq p\}$;
(ii) $T(\boldsymbol{\Xi})$ is order-decreasing if and only if $\{\sigma_i(\mathtt{W}) \leq 1;\ 1 \leq i \leq p\}$;
(iii) $\boldsymbol{\Omega} \in D^{\uparrow}(\boldsymbol{\Xi})$ if and only if $\{\sigma_i(\mathtt{W}) \geq 1;\ 1 \leq i \leq p\}$;
(iv) $\boldsymbol{\Omega} \in D^{\downarrow}(\boldsymbol{\Xi})$ if and only if $\{\sigma_i(\mathtt{W}) \leq 1;\ 1 \leq i \leq p\}$.

Proof. Proofs for (ii) and (iv) are given verbatim for Theorems 1 and 2 of Jensen and Ramirez (1990) as dispersion-diminishing operations with $\boldsymbol{\Sigma}$ as $\boldsymbol{\Xi}$. The dual conclusions (i) and (iii) follow on taking inverses step-by-step in the proofs for Theorems 1 and 2 of (1990). □