Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
EMS2015-2118

Revisiting Some Design Criteria

Diarmuid O'Driscoll
Head, Department of Mathematics and Computer Studies
Mary Immaculate College
Ireland

Donald E. Ramirez
Department of Mathematics
University of Virginia
USA

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

O'Driscoll, D. and Ramirez, D.E. (2015). **"Revisiting Some Design Criteria",** Athens: ATINER'S Conference Paper Series, No: **EMS2015-2118.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

ISSN: 2241-2891
03/02/2017

Revisiting Some Design Criteria

Diarmuid O'Driscoll
Head, Department of Mathematics and Computer Studies
Mary Immaculate College
Ireland

Donald E. Ramirez
Department of Mathematics
University of Virginia
USA

Abstract

We address the problem that the *A* (trace) design criterion is not scale invariant and often is in disagreement with the *D* (determinant) design criterion. We consider the canonical moment matrix *CM* and use the trace of its inverse as the canonical trace *CA* design criterion and use the determinant of its inverse as the canonical determinant *CD* design criterion. For designs which contain higher order terms, we note that the determinant of the canonical moment matrix gives a measure of the collinearity between the lower order terms and the higher order terms.

Introduction

We consider a linear regression $\mathbf{Y} = \mathbf{X}\beta+\varepsilon$ with $\mathbf{X}$ a full rank $n \times p$ matrix and $L(\varepsilon) = N(\mathbf{0}, \sigma^2 I_n)$. The Least Squares Estimator is $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{Y}$ with variance-covariance matrix $Cov(\widehat{\beta}) = \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$, where $Var(\epsilon_i) = \sigma^2$. The diagonal terms of the matrix $Cov(\widehat{\beta})$ are the variances of the Least Squares estimators $\widehat{\beta}_i$, $0 \leq i \leq p-1$ and it is desirable to keep these variances as small as possible and to have the off diagonal terms (covariance between the estimators) as close to zero as possible. If the experimenter knows, before the experiment takes place, where he wishes to make predictions $\widehat{\mathbf{Y}}(\mathbf{x}_0) = \mathbf{x}_0'\widehat{\beta}$, then the scaled prediction variance $SPV = nVar(\widehat{\mathbf{Y}}(\mathbf{x}_0))/\sigma^2 = n\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$ is an excellent measure of the design efficiency. Here $Var(\widehat{\mathbf{Y}}(\mathbf{x}_0))$ is the variance of the estimated response with n to allow for comparisons of designs with varying sample sizes. Unfortunately, it is often the case that the experimenter does not know where in the design space he will need to predict. Thus an overall design criterion is required.

Goos and Leemans (2004) state that "Not only courses, but also textbooks on experimental design, (for example, Kuehl 2000; Montgomery 2000; Neter, Kutner, Nachtsheim, and Wasserman 1996; Oehlert 2000; Weber and Skillings 1999) pay little attention to the design of experiments involving quantitative variables. Typically, at most one chapter or section is spent on this kind of experiment, which is often referred to as a response surface experiment. The optimal design of experiments receives even less attention."

The standard optimality criteria (A, D, and E) are useful measures for comparing competing designs. The criteria are all based on the eigenvalues $\{\lambda_i > 0 : 1 \leq i \leq p\}$ of $(\mathbf{X}'\mathbf{X})^{-1}$ with $A(\mathbf{X}) = tr((\mathbf{X}'\mathbf{X})^{-1}) = \sum_{i=1}^{n} \lambda_i, D(\mathbf{X}) = \det((\mathbf{X}'\mathbf{X})^{-1}) = \prod_{i=1}^{n} \lambda_i$ and $E(\mathbf{X}) = \max\{\lambda_i : 1 \leq i \leq n\}$. Each determines a "size" of $(\mathbf{X}'\mathbf{X})^{-1}$. Authors have proposed other techniques to complement these existing criteria, such as the fraction of design space technique (*FDS*) of Zahran *et al.* (2003).

To illustrate the optimal design using determinants (D-optimal), Goos and Leemans (2004) provide the following example for a simple linear regression model.

"Physicians often use the so-called diving reflex to reduce abnormally rapid heartbeats in humans by submerging the patients' faces in cold water. Suppose that a research physician would like to conduct an experiment in order to investigate the effects of the water temperature on the pulse rates of six small children. One intuitive way to approach the problem is to select six temperatures and to assign each of the children in a random fashion to these temperatures. A reasonable set of temperatures, measured in Farenheit, for this problem might be 45, 50, 55, 60, 65 and 70. The reduction in pulse can be measured for each child (e.g., in beats/minute) and a regression line can be fitted to the data." In this case the design matrix

$$\mathbf{X_6} = \begin{bmatrix} 1 & 45 \\ 1 & 50 \\ 1 & 55 \\ 1 & 60 \\ 1 & 65 \\ 1 & 70 \end{bmatrix} \tag{1}$$

gives

$$(\mathbf{X_6}'\mathbf{X_6})^{-1} = \begin{bmatrix} 7.724 & -.131 \\ -.131 & .002 \end{bmatrix} \text{ and } \det(\mathbf{X_6}'\mathbf{X_6})^{-1} = .000381 \tag{2}$$

In an effort to improve the design, we consider

$$\mathbf{X_3} = \begin{bmatrix} 1 & 45 \\ 1 & 45 \\ 1 & 57.5 \\ 1 & 57.5 \\ 1 & 70 \\ 1 & 70 \end{bmatrix} \tag{3}$$

which gives

$$(\mathbf{X_3}'\mathbf{X_3})^{-1} = \begin{bmatrix} 5.457 & -0.092 \\ -0.092 & .002 \end{bmatrix} \text{ and } \det(\mathbf{X_3}'\mathbf{X_3})^{-1} = .000267 \tag{4}$$

Goos and Leemans (2004) use “Solver” in Microsoft Excel to show that the D-optimal design is found by changing the temperatures to 45,45,45,70,70 and 70. In this case the D-optimal design matrix is

$$\mathbf{X_2} = \begin{bmatrix} 1 & 45 \\ 1 & 45 \\ 1 & 45 \\ 1 & 70 \\ 1 & 70 \\ 1 & 70 \end{bmatrix} \tag{5}$$

which gives

$$(\mathbf{X_2}'\mathbf{X_2})^{-1} = \begin{bmatrix} 3.693 & -0.061 \\ -0.61 & .001 \end{bmatrix} \text{ and } \det(\mathbf{X_2}'\mathbf{X_2})^{-1} = .000178 \tag{6}$$

The relative D-efficiency of two designs is defined as the ratio of the two determinants raised to the power of $1/p$ where p is the number of unknown model parameters. In this example the relative D-efficiency between designs (1) and (5) is 0.6835.

Consider designs which contain only the constant and main effects. We say that the design criterion C is scale invariant when for any two designs $\mathbf{X}$ and $\mathbf{Z}$, with the same column rank p and $\mathbf{D}$ a $p\times p$ diagonalmatrix, that if $C(\mathbf{X}) \leq C(\mathbf{Z})$ then $C(\mathbf{XD}) \leq C(\mathbf{ZD})$. A criticism of the A and E design criteria is that these techniques are not scale invariant and thus investigators may differ on the choice of a design based on the units they will be using; for example, design $\mathbf{X}$ may be considered a better design than $\mathbf{Z}$ using English units, but the reverse using the metric system. While the D design is scale invariant, it favors ill-conditioned designs with very oblique moment matrices, as in Jensen (2004), and often the optimal D design is infeasible. To avoid the scale invariance issue, $\mathbf{X}$ can be assumed to have been standardized with the moment matrix $\mathbf{X'X}$ having unity on the diagonal. We do not suggest this standardization for response surface designs as this standardization destroys the nature of the quadratic terms and does not guarantee agreement between the trace and determinant criteria. We offer an alternative design criterion for response surface designs. In this paper, we consider the canonical moment matrix **CM** and its associated trace CA and determinant CD criteria. For designs which contain higher order terms, we note that the determinant of the canonical moment matrix gives a measure of the collinearity between the lower order terms and the higher order terms.

The variance inflation factor VIF measures the penalty for adding one non-orthogonal additional variable to a linear regression model and it can be computed as a ratio of determinants. The extension of VIF to a measure of the penalty for adding a subset of variables to a model is the generalized variance inflation factor $GVIF$ of Fox and Monette (1992). We give the relationship between $GVIF$ and CD to study response surface designs; in particular, as the penalty for adding the quadratic terms to the main effects.

Simple Linear Regression and the A Criterion

With a simple linear regression the model is Y $\mathbf{Y} = \mathbf{X}\beta+\varepsilon$ with $\mathbf{Y}$ the $(n \times 1)$ vector of responses, $\mathbf{X}$ the $(n \times p)$ experimental full rank design matrix, the $(p\times1)$ vector of linear parameters and ε the $(n\times1)$ vector of errors. For $p = 2$, denote

$$\mathbf{X}(\alpha) = \begin{bmatrix} 1 & \alpha x_1 \\ 1 & \alpha x_2 \\ . & . \\ 1 & \alpha x_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ 1 & x_n \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} = \mathbf{XD}$$

where the moment matrix of this design is $\mathbf{M}(\mathbf{X}(\alpha)) = \mathbf{X}'(\alpha)\mathbf{X}(\alpha)$. The A criterion measures the "goodness" of the design by the trace of the inverse of the moment matrix as

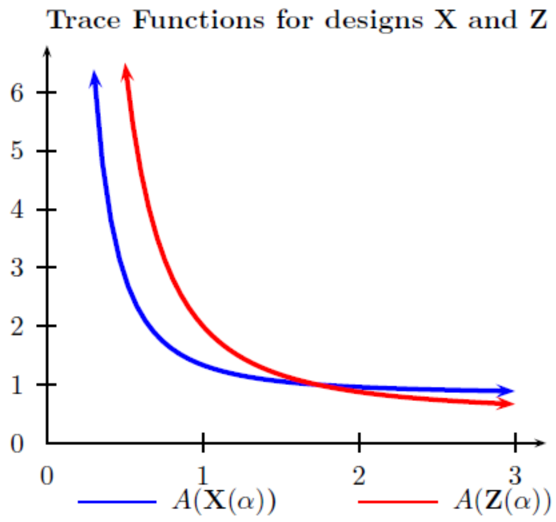
$$A(\mathbf{X}(\alpha)) = tr\ (\mathbf{M}(\mathbf{X}(\alpha))^{-1}) \tag{7}$$

but is not scale invariant. For example, consider the models having three designs points with

$$\mathbf{X}(\alpha) = \begin{bmatrix} 1 & 0 \\ 1 & \alpha \\ 1 & 2\alpha \end{bmatrix} \text{ and } \mathbf{Z}(\alpha) = \begin{bmatrix} 1 & 0 \\ 1 & \alpha \\ 1 & 0 \end{bmatrix} \tag{8}$$

The trace functions for the designs $\mathbf{X}(\alpha)$ and $\mathbf{Z}(\alpha)$ in (8) are shown in Figure 1.

Figure 1. $A(\mathbf{X}(\alpha))$, $A(\mathbf{Z}(\alpha))$



With α = 1, we choose the design **X** over the design **Z**; however, with α = 2, we would choose the design **Z** over the design **X** as

$$\begin{array}{rr} A(\mathbf{X}(1)) = 1.333 & A(\mathbf{Z}(1)) = 2.000 \\ A(\mathbf{X}(\sqrt{3})) = 1.000 & A(\mathbf{Z}(\sqrt{3})) = 1.000 \\ A(\mathbf{X}(2)) = 0.9583 & A(\mathbf{Z}(2)) = 0.8750 \end{array}$$

Simple Linear Regression with the E Criterion

Another popular design criterion is based on the largest eigenvalue of the inverse of the moment matrix as

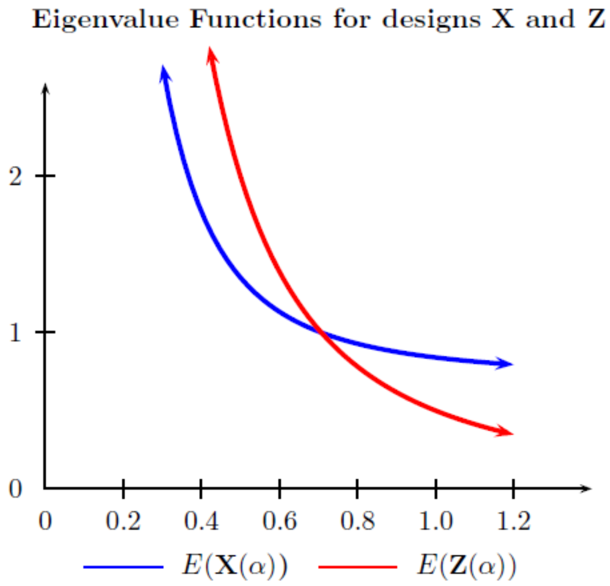
$$E(\mathbf{X}) = \max\{\text{eigenvalues of } \mathbf{M}(\mathbf{X}(\alpha))^{-1}\}. \tag{9}$$

This design criterion also is not scale invariant. For example, consider the models having four design points with

$$\mathbf{X}(\alpha) = \begin{bmatrix} 1 & 0 \\ 1 & \alpha \\ 1 & 2\alpha \\ 1 & 3\alpha \end{bmatrix} \quad \mathbf{Z}(\alpha) = \begin{bmatrix} 1 & 0 \\ 1 & \alpha \\ 1 & 0 \\ 1 & -\alpha \end{bmatrix}. \tag{10}$$

The eigenvalue functions for the designs $\mathbf{X}(\alpha)$ and $\mathbf{Z}(\alpha)$ in (10) are shown in Figure 2.

Figure 2. $E(\mathbf{X}(\alpha))$, $E(\mathbf{Z}(\alpha))$



With α = 1/2, we choose the design **X** over the design **Z**; however, with α = 1, we would choose the design **Z** over the design **X** as

$$\begin{array}{rr} E(\mathbf{X}(\frac{1}{2})) = 1.352 & E(\mathbf{Z}(\frac{1}{2})) = 2.000 \\ E(\mathbf{X}(\frac{\sqrt{2}}{2})) = 1.000 & E(\mathbf{Z}(\frac{\sqrt{2}}{2})) = 1.000 \\ E(\mathbf{X}(1)) = 0.8405 & E(\mathbf{Z}(1)) = 0.5000 \end{array}$$

Hyperellipticity Index

For a positive-definite matrix **A** with eigenvalues $\{\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p > 0\}$, the measure of sphericity is the hyperellipticity index which is the ratio of the arithmetic mean of the eigenvalues to the geometric mean as

$$ellip(\mathbf{A}) = \frac{\overline{\lambda}}{(\lambda_1 \cdots \lambda_p)^{1/p}} = \frac{tr(\mathbf{A})/p}{(\det(\mathbf{A}))^{1/p}} \geq 1.$$

Orthogonal designs with **X'X** a scalar matrix have *ellip*(**X'X**) = 1. If two designs have the same *A* value, then the optimal *D* design, by necessity, will have a larger hyperellipticity index and thus it will be more ill-conditioned. Conversely, if two designs have the same *D* value, then the optimal *A* design is more spherical.

The Standardized Moment Matrix (p = 2)

For designs which contain only the constant and main effects, we can avoid the scale invariance issue by requiring the design matrices to have all columns of unit length. Thus in the case of simple linear regression with design matrix **X**, the standardized moment matrix has the form of the correlation matrix

$$\mathbf{SM}(\mathbf{X}) = \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}$$

Following Hotelling (1936), we call $\gamma_{\mathbf{X}}^2 = \gamma^2$ the *canonical index* for the design $(0 \leq \gamma_{\mathbf{X}}^2 \leq 1)$. The eigenvalues of $\mathbf{SM}(\mathbf{X})^{-1}$ are given by $\frac{1}{1-\gamma}$ and $\frac{1}{1+\gamma}$. We define the scale free design crite
ria using the trace and the determinant of $\mathbf{SM}(\mathbf{X})^{-1}$ and for p = 2

$$\begin{aligned} SA(\mathbf{X}) &= tr(\mathbf{SM}(\mathbf{X})^{-1}) = \frac{2}{1-\gamma_{\mathbf{X}}^2} \\ SD(\mathbf{X}) &= \det(\mathbf{SM}(\mathbf{X})^{-1}) = \frac{1}{1-\gamma_{\mathbf{X}}^2}, \end{aligned} \tag{11}$$

with $SA(\mathbf{X})/SD(\mathbf{X}) = 2$. Thus the minimum possible value for $n\times 2$ design has $SA = 2$ and $SD = 1$.

For the three point designs in Eq. (8),

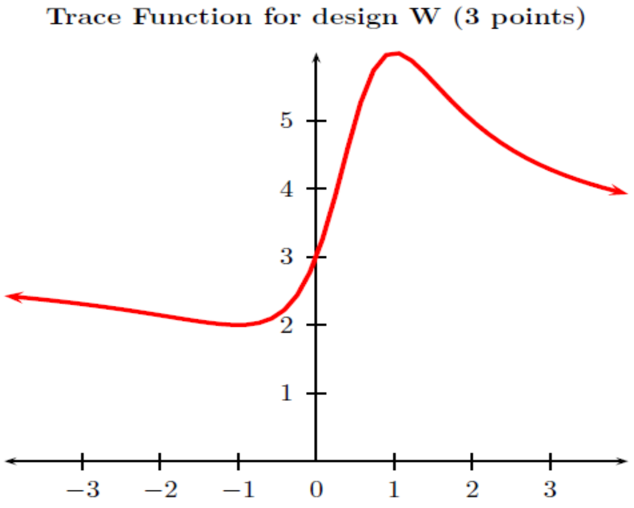
$$SA(\mathbf{X}) = 5 \qquad SA(\mathbf{Z}) = 3$$
$$SD(\mathbf{X}) = 2.5 \qquad SD(\mathbf{Z}) = 1.5$$

Both **X** and **Z** contain the design points 0 and 1. Using *SA* (equivalently *SD*), the optimal value for a third point for the design

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & t \end{bmatrix}$$

has $t = -1$, as can be seen from the graph of $SA(\mathbf{W}(\mathbf{t}))$ in **Figure 3**, with $SA(\mathbf{W}) = 2$ and $SD(\mathbf{W}) = 1$.

Figure 3. *SA*(**W**(**t**)) *for the Standardized Moment Matrix*



Similarly, for the four point designs in Eq. (10),

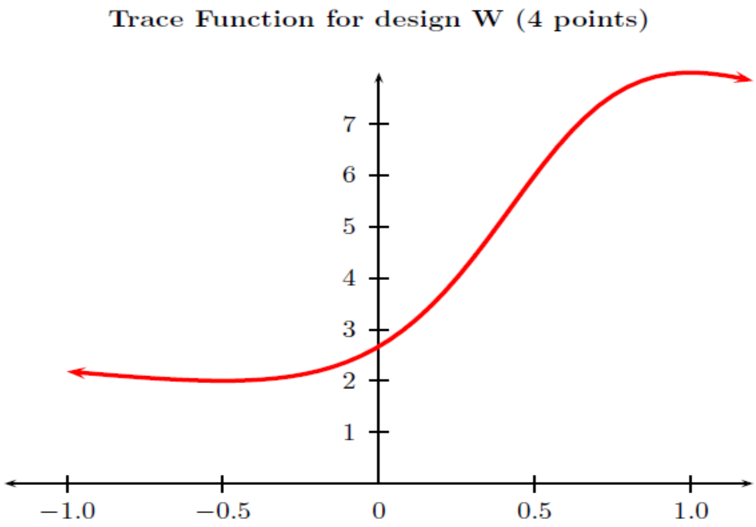
$$SA(\mathbf{X}) = 5.60 \qquad SA(\mathbf{Z}) = 2.00\ .$$

Both **X** and **Z** contain the design points 0 and 1. Using *SA* (equivalently *SD*), the optimal value for the design **W** with a repeated value *t*

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & t \\ 1 & t \end{bmatrix}$$

has t = −1/2 (**Figure 4**) with *SA*(**W**) = 2 and *SD*(**W**) = 1.

Figure 4. *SA*(**W**(**t**)) *for the Standardized Moment Matrix*



The Standardized Moment Matrix (p = 3)

We increase the complexity of the models under review by considering the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ with $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2]$. For each column of **X**, let $(\mathbf{X}'\mathbf{X})_{i,i}^{1/2}$ be the Euclidean length of the column and set $\mathbf{D}_\mathbf{X}$ to be the diagonal matrix with entries on the $\mathbf{D}_\mathbf{X}[i,i] = (\mathbf{X}'\mathbf{X})_{i,i}^{-1/2}$. *The standardized moment matrix* is given by

$$\mathbf{SM}(\mathbf{X}) = \mathbf{D}_\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{D}_\mathbf{X}$$

with $SA(\mathbf{X}) = tr(\mathbf{SM}(\mathbf{X})^{-1})$ and $SD(\mathbf{X}) = \det(\mathbf{SM}(\mathbf{X})^{-1})$.

The criterion $SA(\mathbf{X})$ has been constructed to be scale invariant for first order designs. If **S** is a diagonal matrix which will change the scaling of **X** to **XS**, then the standardized moment matrix for XS is given by $\mathbf{SM}(\mathbf{XS}) = \mathbf{D}_{\mathbf{XS}}(\mathbf{SX'XS})\mathbf{D}_{\mathbf{XS}} = (\mathbf{D}_\mathbf{X}\mathbf{S}^{-1})\mathbf{SX'XS}(\mathbf{S}^{-1}\mathbf{D}_\mathbf{X}) = \mathbf{D}_\mathbf{X}(\mathbf{X'X})\mathbf{D}_\mathbf{X} = \mathbf{SM}(\mathbf{X})$. For the linear model $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, this change of scale of **X** is a change of scale in to $\mathbf{Y}$ $(\mathbf{XS})(\mathbf{S}^{-1}\beta) + \varepsilon$.

Unfortunately, $SA(\mathbf{X})$ lacks the desirable property of being in agreement with SD(X). For example, with

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & -1.05 & 1.1025 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \tag{12}$$

the design criteria are

$$\begin{array}{cc} A(\mathbf{X}) = 1.19 & A(\mathbf{Z}) = 2.75 \\ D(\mathbf{X}) = 0.0113 & D(\mathbf{Z}) = 0.125 \\ SA(\mathbf{X}) = 8.92 & SA(\mathbf{Z}) = 9.25 \\ SD(\mathbf{X}) = 5.01 & SD(\mathbf{Z}) = 4.50 \end{array}$$

so we would choose **X** using the trace criteria A, the determinant criteria D or the scale free trace SA, but we would choose **Z** using the scale free determinant SD.

Optimal D Designs and Oblique Designs

The D criterion is given by

$$D(\mathbf{X}) = \det((\mathbf{X'X})^{-1}).$$

It has the desirable property of being scale invariant. One criticism of $D(\mathbf{X})$ is that it favors oblique moment matrices (see Jensen (2004)).

The 9 × 3 central composite design for the quadratic model in Table A1 of Appendix A has a free parameter *a* which is classically set equal to $\sqrt{2}.$. The optimal *D* design has *a* = 1 with $D(\mathbf{X}) = 0$, but an infeasible solution.

Variance Inflation Factors and the Metric Number

Considering designs which contain only the constant and main effects, set $\mathbf{D}_X$ to be the diagonal matrix with entries on the diagonal $\mathbf{D}_X[i,i] = (\mathbf{X}'\mathbf{X})^{-1/2}_{i,i}$.

When the design has been standardized $\mathbf{X} \to \mathbf{X}\mathbf{D}_X$, the *A* criterion is the sum of the variance inflation factors VIF_i as the VIF_s are the diagonal entries of the inverse of the standardized moment matrix $\mathbf{D}_X^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}_X^{-1}$. Note that we follow Steward (1987) and do not necessarily center the explanatory variables.

For our linear model $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ view $\mathbf{X} = [\mathbf{X}_{[p]}, \mathbf{x}_p]$ with $\mathbf{x}_p$ the p^{th} column of $\mathbf{X}$ and $\mathbf{X}_{[p]}$ the matrix formed by the remaining columns. The variance inflation factors measure the effect of adding column $\mathbf{x}_p$ to $\mathbf{X}_{[p]}$. For notational convenience, we demonstrate with the last column p. An ideal column would be orthogonal to the previous columns with the entries in the off diagonal elements of the p^{th} row and p^{th} column of $\mathbf{X'X}$ all zeros. Denote by $\mathbf{M}_p$ the idealized moment matrix

$$\mathbf{M}_p = \begin{bmatrix} \mathbf{X}'_{[p]}\mathbf{X}_{[p]} & \mathbf{0}_{p-1} \\ \mathbf{0}'_{p-1} & \mathbf{x}'_p\mathbf{x}_p \end{bmatrix}$$

The *metric number* associated to $\mathbf{x}_p$ is defined by

$$MN(\mathbf{x}_p) = \sqrt{\frac{\det(\mathbf{X}'\mathbf{X})}{\det(\mathbf{M}_p)}}$$

The metric number has been used in Garcia *et al.* (2011) as a measure of collinearity. A similar measure of collinearity is mentioned in Footnote 2 in Wichers (1975) and Theorem 1 of Berk (1977). The geometry for the metric number has been shown in Garcia *et al.* (1999). The case study in Garcia et al. (2011) suggests the functional relationship between $MN(\mathbf{x}_p)$ and the variance inflation factor for $\widehat{\beta}_p$ as

$$VIF(\widehat{\beta}_p) = \frac{1}{MN(\mathbf{x}_p)^2}$$

We show that this relationship holds and so the metric number $MN(\mathbf{x}_p)$is also functionally equivalent to the variance inflation factors VIF_p, and equivalently, to the collinearity indices κ_p of Stewart (1987).

To evaluate $VIF(\widehat{\beta}_p)$, transform **X'X** into standardized form **R**. With **R** = $\mathbf{D}_X(\mathbf{X'X})\mathrm{D}_X$, the $V\,IF_s$ are the diagonal entries of $\mathbf{R}^{-1}$ = D $\mathbf{D}_X^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}_X^{-1}$. It remains to note that the inverse $\mathbf{R}^{-1}$ can be computed using cofactors $\mathbf{C}_{i,j}$; and, in particular,

$$\begin{aligned} VIF(\widehat{\beta}_p) &= \mathbf{R}^{-1}_{p,p} = [\mathbf{D}_X^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}_X^{-1}]_{p,p} && (13)\\ &= (\mathbf{x}_p'\mathbf{x}_p)^{1/2}\frac{\det(\mathbf{C}_{p,p})}{\det(\mathbf{X}'\mathbf{X})}(\mathbf{x}_p'\mathbf{x}_p)^{1/2} = \frac{\mathbf{x}_p'\mathbf{x}_p\det(\mathbf{X}_{[p]}'\mathbf{X}_{[p]})}{\det(\mathbf{X}'\mathbf{X})} && (14)\\ &= \frac{\det(\mathbf{M}_p)}{\det(\mathbf{X}'\mathbf{X})} = \frac{1}{MN(\mathbf{x}_p)^2}. && (15) \end{aligned}$$

Generalized Variance Inflation Factors *GV IF*

The generalized variance inflation factors are an extension of the classical variance inflation factors *V IF* from Eq. (15). For a linear model $\mathbf{Y} = \mathbf{X}\beta+\varepsilon$, view $\mathbf{X} = [\mathbf{X}_1,\mathbf{X}_2]$ partitioned with $\mathbf{X}_1$ of dimension $n \times r$ usually consisting of the lower order terms and $\mathbf{X}_2$ of dimension $n \times s$ usually consisting of the higher order terms. The idealized moment matrix for the (r, s) partitioning of **X** is

$$\mathbf{M}_{(r,s)} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0}_{r\times s} \\ \mathbf{0}_{s\times r} & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}$$

Following Eq. (15), the *generalized variance inflation factor* is a measure of the effect of adding $\mathbf{X}_2$ to the design $\mathbf{X}_1$. That is for $\mathbf{X}_2|\mathbf{X}_1$

$$GVIF(\mathbf{X}_2|\mathbf{X}_1) = \frac{\det(\mathbf{M}_{(r,s)})}{\det(\mathbf{X}'\mathbf{X})} = \frac{\det(\mathbf{X}_1'\mathbf{X}_1)\det(\mathbf{X}_2'\mathbf{X}_2)}{\det(\mathbf{X}'\mathbf{X})} \qquad (16)$$

as in Eq. (10) of Fox and Monette (1992), who compared the sizes of the joint confidence regions for β for partitioned designs. Note that when $\mathbf{X} = [\mathbf{X}_{[p]}, \mathbf{x}_p]$, $GV\,IF[\mathbf{X}_{[p]}, \mathbf{x}_p] = V\,IF_p$ in Eq. (15). Eq. (16) is in spirit of the efficiency comparisons in linear inferences introduced in Theorems 4 and 5 of Jensen and Ramirez (1993). For the simple linear regression model with $p = 2$, Eq. (16) gives $VIF = \frac{1}{1-\gamma^2}$ as required. Fox and Monette suggested that $\mathbf{X}_1$ contain the variables which are of "simultaneous interest" while $\mathbf{X}_2$ contain additional variables selected by the investigator. We set X1 for the constant and main effects and set $\mathbf{X}_2$ the (optional) quadratic terms with values from $\mathbf{X}_1$.

The *CA* and *CD* Criteria

We partition the design $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2]$ with $\mathbf{X}_1$ consisting of the constant and the main effects and X2 consisting of the quadratic terms. In general, $\mathbf{X}_1$ will be of dimension $n \times r$ and will consist of the lower order terms and $\mathbf{X}_2$ will be of dimension $n \times s$ and consist of the higher order terms. Let

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{22} \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1/2} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2'\mathbf{X}_2)^{-1/2} \end{bmatrix}$$

Following the structure of the standardized moment matrix in the case of simple linear regression, we define the canonical moment matrix as

$$CM([\mathbf{X}_1|\mathbf{X}_2]) = \mathbf{L}[\mathbf{X}_1|\mathbf{X}_2]'[\mathbf{X}_1|\mathbf{X}_2]\mathbf{L} \tag{17}$$

$$= \begin{bmatrix} \mathbf{I}_{r\times r} & (\mathbf{X}_1'\mathbf{X}_1)^{-1/2}(\mathbf{X}_1'\mathbf{X}_2)(\mathbf{X}_2'\mathbf{X}_2)^{-1/2} \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1/2}(\mathbf{X}_2'\mathbf{X}_1)(\mathbf{X}_1'\mathbf{X}_1)^{-1/2} & \mathbf{I}_{s\times s} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I}_{r\times r} & CM(\mathbf{X})_{1,2} \\ CM(\mathbf{X})_{2,1} & \mathbf{I}_{s\times s} \end{bmatrix}.$$

We define the *canonical A trace* and *canonical D determinant* design criteria as the trace and determinant of the inverse of the canonical moment matrix with

$$CA([\mathbf{X}_1|\mathbf{X}_2]) = tr(CM([\mathbf{X}_1|\mathbf{X}_2])^{-1}) \tag{18}$$
$$CD([\mathbf{X}_1|\mathbf{X}_2]) = \det(CM([\mathbf{X}_1|\mathbf{X}_2])^{-1}).$$

For the linear model $\mathbf{Y} = [\mathbf{X}_1|\mathbf{X}_2][\beta'_1|\beta'_2]' + \varepsilon$, the transformation $\mathbf{X} \to \mathbf{XL}$ is transformation $\beta \to \mathbf{L}^{-1}\beta$ with $\mathbf{Y} = (\mathbf{XL})(\mathbf{L}^{-1}\beta) + \varepsilon = (\mathbf{X}_1\mathbf{L}_{11})(\mathbf{L}_{11}^{-1}\beta_1) + (\mathbf{X}_2\mathbf{L}_{22})(\mathbf{L}_{22}^{-1}\beta_2) + \varepsilon$. The canonical $CD([\mathbf{X}_1|\mathbf{X}_2])$ has the structure

$$CD([\mathbf{X}_1|\mathbf{X}_2]) = \frac{\det((\mathbf{X}'\mathbf{X})^{-1})}{\det((\mathbf{X}_1'\mathbf{X}_1)^{-1})\det((\mathbf{X}_2'\mathbf{X}_2)^{-1})} = GVIF(\mathbf{X}_2|\mathbf{X}_1) \tag{19}$$

and thus is a measure of collinearity between the covariance matrices for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for the lower order and higher order terms and $\widehat{\beta}$ for the full model.

For the four point designs given in Eq.(12) viewed $y = \beta_0+\beta_1x_1+ \beta_2x^2$ with $\mathbf{X} = [\mathbf{1}, \mathbf{x}|\mathbf{x}^2]$,

$$\mathbf{X_1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & -1.055 \end{bmatrix} \quad \mathbf{X_2} = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 1.1025 \end{bmatrix}$$

$$\mathbf{L_{11}} = \begin{bmatrix} 0.535 & -0.981 \\ -0.981 & 0.430 \end{bmatrix} \quad \mathbf{L_{22}} = \begin{bmatrix} 0.234 \end{bmatrix} \quad \mathbf{L_{11}}(\mathbf{X}_1'\mathbf{X}_2)\mathbf{L_{22}} = \begin{bmatrix} 0.585 \\ 0.649 \end{bmatrix}$$

$$CM([\mathbf{X}_1|\mathbf{X}_2]) = \begin{bmatrix} 1 & 0 & 0.585 \\ 0 & 1 & 0.649 \\ 0.585 & 0.649 & 1 \end{bmatrix}.$$

Similarly

$$CM([\mathbf{Z}_1|\mathbf{Z}_2]) = \begin{bmatrix} 1 & 0 & 0.846 \\ 0 & 1 & -0.204 \\ 0.846 & -0.204 & 1 \end{bmatrix}$$

Note that these matrices have the form

$$CM([\mathbf{1}, \mathrm{x}|\mathrm{x}^2]) = \begin{bmatrix} 1 & 0 & \beta_1 \\ 0 & 1 & \beta_2 \\ \beta_1 & \beta_2 & 1 \end{bmatrix}$$

with eigenvalues $\{1, 1 \pm \sqrt{\beta_1^2 + \beta_2^2}\}$. The eigenvalues of $CM([\mathbf{1}, \mathbf{x}|\mathbf{x}^2])^{-1}$ are the reciprocals of these values. Set

$$\gamma_{\mathrm{X}}^2 = \beta_1^2 + \beta_2^2$$

the *canonical index*. The trace and determinant of $CM([\mathbf{1}, \mathbf{x}|\mathbf{x}^2])^{-1}$ are functions of γ_{X}^2 and with $p = 3$,

$$\begin{aligned} CA([\mathbf{1}, \mathrm{x}|\mathrm{x}^2]) &= 1 + \frac{2}{1 - \gamma_{\mathrm{X}}^2} \geq p \\ CD([\mathbf{1}, \mathrm{x}|\mathrm{x}^2]) &= \frac{1}{1 - \gamma_{\mathrm{X}}^2} \geq 1 \\ \frac{CA([\mathbf{1}, \mathrm{x}|\mathrm{x}^2])}{CD([\mathbf{1}, \mathrm{x}|\mathrm{x}^2])} &= 3 - \gamma_{\mathrm{X}}^2. \end{aligned}$$

Thus the canonical *CA* criterion is in agreement with the canonical *CD* criterion, both agreeing that there is a little less collinearity between the lower and higher order terms with the **Z** design with

$$\begin{array}{cc} \gamma_{\mathrm{X}}^2 = 0.7636 & \gamma_{\mathrm{Z}}^2 = 0.7576 \\ CA([\mathbf{1}, \mathrm{x}|\mathrm{x}^2]) = 9.460 & CA([\mathbf{1}, \mathrm{z}|\mathrm{z}^2]) = 9.250 \\ CD([\mathbf{1}, \mathrm{x}|\mathrm{x}^2]) = 4.230 & CD([\mathbf{1}, \mathrm{z}|\mathrm{z}^2]) = 4.125 \end{array}$$

Central Composite and Factorial Designs

In this section we will compare the Central Composite Design **X** and the Factorial Design **Z**. The design points are shown in Table A1 of Appendix A. Both designs are 9 × 6 and are used with the response model

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

We partition the design matrix $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2]$ with $\mathbf{X}_1$ the lower order terms consisting of the constant and linear columns and $\mathbf{X}_2$ the higher order terms consisting of the quadratic columns. The design Z is similarly partitioned. Both of these designs have an unique non-zero singular value γ for the off-diagonal array in the canonical matrix with $CD = \frac{1}{1-\gamma^2}$ and $CA = 4 + \frac{2}{1-\gamma^2}$. For these designs, with $a = \sqrt{2}$, **Z** has less collinearity between the lower and higher order terms than **X**:

$$\begin{array}{cc} \gamma_{\mathbf{X}}^2 = 0.88889 & \gamma_{\mathbf{Z}}^2 = 0.80000 \\ CA([\mathbf{X}_1|\mathbf{X}_2]) = 22 & CA([\mathbf{Z}_1|\mathbf{Z}_2]) = 14 \\ CD([\mathbf{X}_1|\mathbf{X}_2]) = 9 & CD([\mathbf{Z}_1|\mathbf{Z}_2]) = 5 \end{array}$$

A surprising result is that the classical choice of $a = \sqrt{2}$has the most collinearity between the lower and higher order terms as measured by

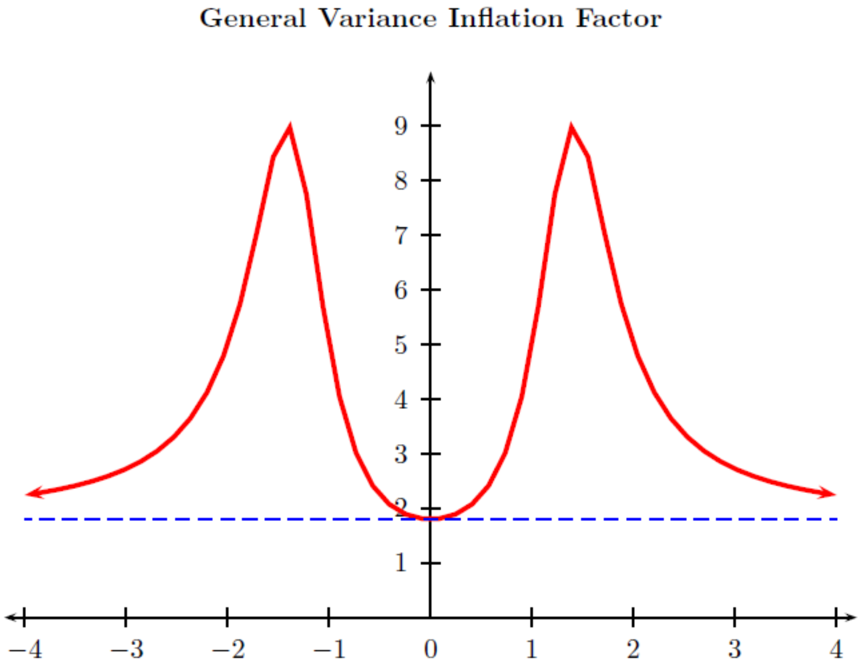
$$GVIF(a) = CD(a) = \frac{9(4 + a^4)}{5a^4 - 16a^2 + 20}$$

with

a	0	1	$\sqrt{2}$	1.5	1.75	∞
$\gamma_{\mathbf{X}}^2$	0.444	0.800	0.889	0.886	0.851	0.444
$GVIF$	1.80	5.00	9.00	8.76	6.72	1.80

The plot of the generalized variance inflation factors, $GVIF(a)$, for the Central Composite Design is shown in **Figure 5**.

Figure 5. *GV IF(a) for the Central Composite Design*



In O'Driscoll and Ramirez (2015), we have continued using GVIF as a design criterion and have studied the hybrid designs (H310 and H311B) of Roquemore (1976), the design from Box and Behnken (1960), the minimal design of Box and Draper (1970) and the small composite design of Hartley (1955).

Integrated Variance *IV* Optimality for the Quadratic Response Model

For computing optimal designs for quadratic models, Jones and Goos (2012) assert that "it makes sense to focus attention on the predictive capability of the designs" and thus they advocated the use of the *IV* -optimality criterion also called *V* -optimality, *I*-optimality or *Q*-optimality as noted in Myers and Montgomery (2002).

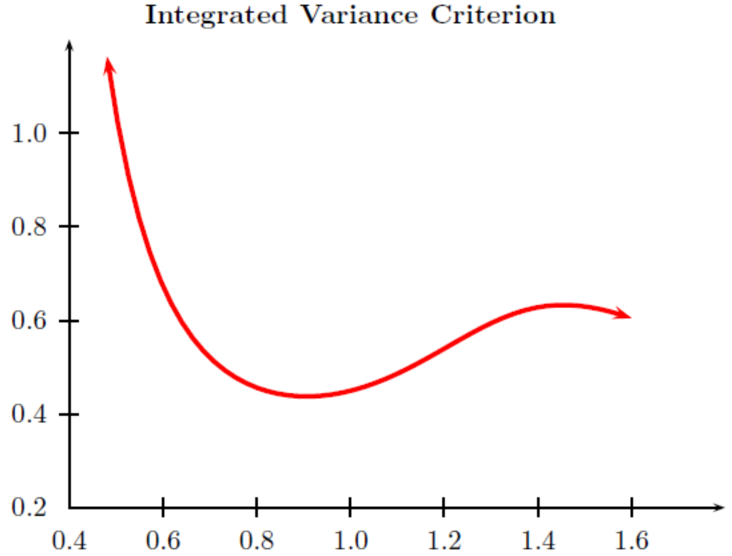
Denote by $\mathbf{X}_1$ the 9 × 3 lower order terms for the Central Composite Design from Table A1 with $a \in [\frac{1}{2}, \sqrt{2}]$ to be chosen later; and denote by **X** the 9 × 6 full quadratic response model. The 6 × 6 variance-covariance matrix for the quadratic response model is denoted by $\mathbf{\Sigma}_F = \sigma^2(\mathbf{X'X})^{-1}$. For a given $\mathbf{t} = (x_1, x_2)$ in the design space, the variance of the predicted value $\widehat{y}(\mathbf{t})$ is $Var(\widehat{y}(\mathbf{t})) = \sigma^2\mathbf{x}'\mathbf{\Sigma}_F\mathbf{x}$ where $\mathbf{x} = [1, x_1, x_2, x_1^2, x_2^2, x_1x_2]'$ denotes the augmentation of the design point t for use in the quadratic response model. The Integrated Variance *IV* criterion is a measure of the prediction performance over a region *R* of interest and is given by

$$IV(\mathbf{\Sigma}_F) = \frac{1}{(2b)^2}\int_{-b}^{b}\int_{-b}^{b}[\mathbf{x}'\mathbf{\Sigma}_F\mathbf{x}]dx_1dx_2$$

for $R = [-b, b] \times [-b, b]$. We note that in the extreme case with b $\rightarrow 0$, *IV* $\rightarrow Var(\widehat{\beta}_0)$ the (1, 1) entry of $\mathbf{\Sigma}_F = \sigma^2(\mathbf{X'X})^{-1}$. We will set $R = [-1, 1] \times [-1, 1]$ as our region of interest and seek the *IV* -optimal design $\mathbf{X}_1$ with $a \in [\frac{1}{2}, \sqrt{2}]$.

We follow Borkowski (2003, p. 75) who noted, with symbolic software such as Maple, that the *IV* criterion can be numerically evaluated. We find the optimal value $a = 0.90630$ with the graph of *IV* in **Figure 6**, showing that the Factorial Design with $a = 1$ and $IV = 0.450$ is favored over the popular Central Composite Design with $a = \sqrt{2}$ with $IV = 0.631$.

Figure 6. *IV* ($\mathbf{\Sigma}_F$) *for the Quadratic Response Model*



Summary

We have discussed some standard optimal design criteria for first order models and have noted that, except for *D*-optimality, they are not necessarily scale in-variant. For quadratic response models we have shown that the metric number (equivalently, the generalized variance inflation factor) is a measure of collinearity for subsets of variables and that it is an extension of the Variance Inflation Factors for single variables. These optimal design criteria are important considerations for the researcher in planning the design of the experiment. Using the Goos and Leemans (2004) experiment discussed in Section 1, we considered three separate designs: $\mathbf{X}_6$ with six unique values from Eq. (1), $\mathbf{X}_3$ with three unique values from Eq. (3), and $\mathbf{X}_2$ with 2 unique values from Eq. (5). The design $\mathbf{X}_2$ is the preferred design using the *D*, *A*, and *E* criteria, and also with the integrated variance IV_1 criterion for the simple regression model $y = \beta_0 + \beta_1 t$ over the Region of Interest R = [45, 70].

However, if the researcher believes the responses follow a quadratic model $y = \beta_0 + \beta_1 t + \beta_2 t^2$ then, using the quadratic version IV_2 of the integrated variance criterion, X_2 is infeasible and $\mathbf{X}_6$ would be preferred.

	D	A	E	IV_1	IV_2
$\mathbf{X}_6$	$3.810 10^{-4}$	7.726	7.726	0.2857	0.3973
$\mathbf{X}_3$	$2.616 10^{-4}$	5.541	5.541	0.2500	0.4259
$\mathbf{X}_2$	$1.778 10^{-4}$	3.694	3.694	0.2222	NA

Appendix A

Table A1.

The Lower Order Matrix for the Central Composite Design
with Center Run with $a = \sqrt{2}$, n = 9 and
The Lower Order Matrix for the Factorial Design with Center Run, n = 9.

Central Composite Design

$\mathbf{1}$	$\mathbf{x}_1$	$\mathbf{x}_2$
1	1	1
1	1	−1
1	−1	1
1	−1	−1
1	a	0
1	$-a$	0
1	0	a
1	0	$-a$
1	0	0

Factorial Design

$\mathbf{1}$	$\mathbf{x}_1$	$\mathbf{x}_2$
1	1	1
1	1	−1
1	−1	1
1	−1	−1
1	1	0
1	−1	0
1	0	1
1	0	−1
1	0	0

References

Berk, K. (1977). Tolerance and condition in regression computations, J.*Amer. Stat. Assoc.*, 72, 863-866.

Borkowski, J. (2003). A comparison of a prediction variance criteria for response surface designs, J. *Quality Tech*., 35, 70-77.

Box, G.E.P., Behnken, D.W. (1960). Some new three-level designs for the study of quantitative variables, *Technometrics*, 2, 455-475.

Box, M.J., Draper, N.R. (1974). On minimum-point second order design, *Technometrics*, 16, 613-616.

Fox, J., Monette, G. (1992). Generalized collinearity diagnostics, *J. Amer. Stat. Assoc.*, 87, 178-183.

Garcia, J., And´ujar, A., Soto, J. (1999). Acerca de la detecci´on de la colinealidad en los modelos de regresion lineal. *XIII Reunion Asepelt-Espana*. Universidad de Burgos.

Garcia, C.B., Garcia, J. and Soto, J. (2011). The raise method: An alternative procedure to estimate the parameters in presence of collinearity, *Quality and Quantity*, 45, 403-423.

Goos, P., Leemans, H. (2004). Teaching optimal design of experiments using a spreadsheet, *Journal of Statistics Education*, 12(3).

Hartley, H.O. (1959). Smallest composite design for quadratic response surfaces, *Biometrics*, 15, 611-624.

Hotelling, H. (1936). Relations between two sets of variates, *Biometrika*, 28, 321-377.

Jensen, D.R. (2004). Condition numbers and D-efficiency. *Stat. and Prob. Letters*, 66, 267-274.

Jensen, D.R., Ramirez, D.E. (1993). Efficiency comparisons in linear inference, *J. Stat. Plan. Infer.*, 37, 51-68.

Jones, B., Goos, P. (2012). I-optimal versus D-optimal split-plot response surface designs, *Journal of Quality Technology*, April, 85-101.

Myers, R. H, Montgomery, D. C. (2002). *Response Surface Methodology*, 2nd. ed., John Wiley & Sons, Inc., New York.

O'Driscoll, D., Ramirez, D.E. (2015). Response surface designs using the generalized variance inflation factors, *Cogent Mathematics*, 2, 1-11.

Roquemore, K.G. (1976). Hybrid designs for quadratic response surfaces, *Technometrics*, 18, 419-423.

Stewart, G.W. (1987). Collinearity and least squares regression, *Stat. Sci.*,2, 68-84.

Wichers, R. (1975). The detection of multicollinearity: A comment, *Review Econ. Stat*., 57, 366-368.

Zahran, A., Anderson-Cook, C.M., Myers, R.H. (2003). Fraction of design space to assess prediction capability of response surface designs, *J. Qual. Tech*., 35, 377-386.