Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 723–736
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

New criteria for robust integer-valued designs in linear models

Adeniyi J. Adewale*, Douglas P. Wiens

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alta., Canada T6G 2G1

Received 17 October 2005; received in revised form 27 March 2006; accepted 27 March 2006
Available online 24 April 2006

Abstract

We investigate the problem of designing for linear regression models, when the assumed model form is only an approximation to an unknown true model, using two novel approaches. The first is based on a notion of averaging of the mean-squared error of predictions over a neighbourhood of contaminating functions. The other is based on the usual D-optimal criterion but subject to bias-related constraints in order to ensure robustness to model misspecification. Both approaches are integer-valued constructions in the spirit of Fang and Wiens [2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. J. Amer. Statist. Assoc. 95(451), 807–818]. Our results are similar to those that have been reported using a minimax approach even though the rationale for the designs presented here are based on the notion of averaging, rather than maximizing, the loss over the contamination space. We also demonstrate the superiority of an integer-valued construction over the continuous designs using specific examples. The designs which protect against model misspecification are clusters of observations about the points that would have been the design points for classical variance-minimizing designs.

MSC: Primary 62K05; 62F35; secondary 62J05

Keywords: Bias-constrained; Contamination; D-optimal design; Finite design space; Minave design; Misspecification; Polynomial regression; Simulated annealing

1. Statistical model

We consider the general regression model with additive errors: $Y = E(Y|\mathbf{x}) + \varepsilon$. Suppose an experimenter is faced with a set $\mathscr{S} = \{\mathbf{x}_i\}_{i=1}^N$ of possible design points from which he is interested in choosing n, not necessarily distinct, points at which to observe Y. The experimenter makes $n_i \geqslant 0$ observations at $\mathbf{x}_i$ such that $\sum_{i=1}^N n_i = n$. The design problem is to choose $n_1, \ldots, n_N$ in an optimal manner. Equivalently, the objective is to find an optimal probability distribution $\{p_i\}_{i=1}^N$, with $p_i = n_i/n$, on the design space $\mathscr{S} = \{\mathbf{x}_i\}_{i=1}^N$. The resulting design is said to be *integer valued*.

The experimenter believes that the mean response $E(Y|\mathbf{x})$ may be approximated, but not perfectly described, by a linear combination $\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$ of p regressors $(z_1(\mathbf{x}), \ldots, z_p(\mathbf{x}))^{\mathrm{T}} = \mathbf{z}(\mathbf{x})$. Since $E(Y|\mathbf{x}) = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$ is just an approximation

* Corresponding author. Tel.: +1 780 4924406; fax: +1 7804926826.
E-mail addresses: aadewale@ualberta.ca (A.J. Adewale), doug.wiens@ualberta.ca (D.P. Wiens).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.03.010

to the true model, the "best" $\boldsymbol{\beta}_0$ for predicting the mean response is defined to be the minimizer of the average-squared error of the approximation:

$$\boldsymbol{\beta}_0 = \arg\min_{\mathbf{t}} \frac{1}{N} \sum_{i=1}^{N} \left(E\left[Y|\mathbf{x}_i\right] - \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\mathbf{t} \right)^2. \tag{1}$$

Define $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}_0$, so that the model becomes

$$Y_{ij} = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}_0 + f\left(\mathbf{x}_i\right) + \varepsilon_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i, \tag{2}$$

where ε_{ij} is the random error associated with the jth observation chosen at the ith design point and $var\left(\varepsilon_{ij}\right) = \sigma^2$.

From (1), we have

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}\left(\mathbf{x}_i\right) f\left(\mathbf{x}_i\right) = \mathbf{0}, \tag{3}$$

and in order to ensure that the bias in the least-squares estimate $\hat{\boldsymbol{\beta}}$ remains within bounds, we place a bound on the misspecification, that is

$$\frac{1}{N} \sum_{i=1}^{N} f^2\left(\mathbf{x}_i\right) \leqslant \tau^2. \tag{4}$$

Let the class of contamination functions $f(\mathbf{x})$ satisfying (3) and (4) be denoted $\mathscr{F}$. We define the loss I as the average mean-squared error (AMSE) of $\hat{Y}(\mathbf{x}) = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}$ as an estimate of $E[Y|\mathbf{x}]$:

$$\begin{aligned} I &= \frac{1}{N} \sum_{i=1}^{N} E\left\{ \hat{Y}\left(\mathbf{x}_i\right) - E\left[Y|\mathbf{x}_i\right] \right\}^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(E\left[\hat{Y}\left(\mathbf{x}_i\right)\right] - \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} \right)^2 + \frac{1}{N} \sum_{i=1}^{N} var\left[\hat{Y}\left(\mathbf{x}_i\right)\right] + \frac{1}{N} \sum_{i=1}^{N} f^2\left(\mathbf{x}_i\right). \end{aligned} \tag{5}$$

In Section 2 of this article we develop our design criteria. An algorithm to obtain designs minimizing the corresponding loss functions is described in Section 3, with examples and applications in Sections 4 and 5.

2. Loss functions

Fang and Wiens (2000) used a minimax approach to construct integer-valued designs. The optimal design in the minimax sense is the design that minimizes the maximum, over the misspecification neighbourhood $\mathscr{F}$, value of the loss. The minimax approach aims to obtain the best design for the worst possible case of model misspecification. Here, we introduce new criteria for robust designs which may have more intuitive appeal to practitioners. Rather than minimizing the maximum loss we instead choose the design which minimizes the *average* value of the loss over the misspecification neighbourhood. The averaging requires a parameterization of $\mathscr{F}$ which is carried out in Section 2.1 below. This approach can be seen as a generalization of Läuter's (1974, 1976) approach. Läuter accommodated model uncertainty in the choice of design by averaging design criterion functions over a finite set of plausible models. Here, we have an infinite set of plausible models defined by (2)–(4). While Läuter's criterion is based on variance only, in the spirit of Box and Draper (1959) we base our design criteria on possible bias engendered by the model misspecification as well as on variance.

2.1. "Minave" mean-squared error model-robust design criteria

Given the misspecification neighbourhood, we seek integer-valued designs that minimize the average (over $\mathscr{F}$) value of the loss. Let $\{p_i = n_i/n\}_{i=1}^{N}$ be an integer-valued design on $\mathscr{S}$, $\mathbf{P}$ the $N \times N$ diagonal matrix with diagonal elements

$\{p_i\}$, $\mathbf{Z}$ the $N \times p$ matrix, assumed to be of full rank, with rows $\mathbf{z}^{\mathrm{T}}(\mathbf{x}_1), \ldots, \mathbf{z}^{\mathrm{T}}(\mathbf{x}_N)$. Define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))^{\mathrm{T}}$. In this notation, the AMSE defined in (5) can be written as

$$I = \frac{1}{N}\left\{\frac{\sigma^2}{n}\operatorname{tr}\left[\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right] + \mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \mathbf{f}^{\mathrm{T}}\mathbf{f}\right\}.$$

Note that we are assuming that the design is feasible for the full parameter vector $\boldsymbol{\beta}$, or equivalently that it has a minimum of p distinct support points $\mathbf{x}_i$ in $\mathscr{S}$ such that the vectors $\mathbf{z}(\mathbf{x}_i)$ are linearly independent. This implies the nonsingularity of $\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}$.

We obtain the singular value decomposition of $\mathbf{Z}$, $\mathbf{Z} = \mathbf{U}_{N\times p}\boldsymbol{\Lambda}_{p\times p}\mathbf{V}^{\mathrm{T}}_{p\times p}$ say, with $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_p$ and $\boldsymbol{\Lambda}$ diagonal and invertible. The $N \times p$ matrix $\mathbf{U}$ is augmented by $\tilde{\mathbf{U}}_{N\times(N-p)}$ such that $\left[\mathbf{U}\vdots\tilde{\mathbf{U}}\right]_{N\times N}$ is orthogonal. Then by (3) and (4), we have that there is an $N - p \times 1$ vector $\mathbf{c}$, with $\|\mathbf{c}\| \leqslant 1$, satisfying $\mathbf{f} = \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{c}$, and then

$$I = \frac{1}{N}\left\{\frac{\sigma^2}{n}\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2 N\operatorname{tr}\left[\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{P}\mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{U}}\mathbf{c}\mathbf{c}^{\mathrm{T}}\right] + \tau^2 N\mathbf{c}^{\mathrm{T}}\mathbf{c}\right\}.$$

For details of this development see Fang and Wiens (2000). We define our design criterion as I, with $\mathbf{f}$ integrated over $\mathbf{c}$:

$$I_{\mathrm{ave}} = \frac{\sigma^2}{nN}\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2\int_{\|\mathbf{c}\|\leqslant 1}\left(\operatorname{tr}\left[\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{P}\mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{U}}\mathbf{c}\mathbf{c}^{\mathrm{T}}\right] + \mathbf{c}^{\mathrm{T}}\mathbf{c}\right)\mathrm{d}\mathbf{c}.$$

Theorem 1. *Define*

$$\kappa_{N,p} = \frac{\pi^{(N-p)/2}}{((N-p)/2+1)\,\Gamma((N-p)/2)} = \int_{\|\mathbf{c}\|\leqslant 1}\mathbf{c}^{\mathrm{T}}\mathbf{c}\,\mathrm{d}\mathbf{c}.$$

The average of I, the AMSE over the misspecification neighbourhood $\mathscr{F}$, is given by $I_{\mathrm{ave}} = \left(\sigma^2/n + \tau^2\kappa_{N,p}\right)\mathscr{L}_{\mathrm{ave}}$, where

$$\mathscr{L}_{\mathrm{ave}} = \rho\,\frac{\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right]}{N} + (1-\rho)\left(1 + \frac{\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\right] - p}{N-p}\right) \tag{6}$$

for $\rho = \sigma^2/n/\left(\sigma^2/n + \tau^2\kappa_{N,p}\right)$.

Proof. Note that $\int_{\|\mathbf{c}\|\leqslant 1}\mathbf{c}\mathbf{c}^{\mathrm{T}}\,\mathrm{d}\mathbf{c} = (\kappa_{N,p}/(N-p))\mathbf{I}_{N-p}$. We then calculate that

$$\begin{aligned} I_{\mathrm{ave}} &= \frac{\sigma^2}{nN}\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2\operatorname{tr}\left[\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{P}\mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{U}}\int_{\|\mathbf{c}\|\leqslant 1}\mathbf{c}\mathbf{c}^{\mathrm{T}}\,\mathrm{d}\mathbf{c}\right] + \tau^2\kappa_{N,p} \\ &= \frac{\sigma^2}{nN}\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right] + \tau^2\,\frac{\kappa_{N,p}}{N-p}\operatorname{tr}\left[\mathbf{P}\mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{P}\left(\mathbf{I}_N - \mathbf{U}\mathbf{U}^{\mathrm{T}}\right)\right] + \tau^2\kappa_{N,p}, \end{aligned}$$

which reduces to (6).

Since $\left(\sigma^2/n + \tau^2\kappa_{N,p}\right)$ does not depend on the design, we use the loss $\mathscr{L}_{\mathrm{ave}}$ for the design construction. We call $(1/N)\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\right]$ and $1 + \left(\operatorname{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\right] - p\right)/(N-p)$ the variance and bias discrepancies, respectively; $\mathscr{L}_{\mathrm{ave}}$ is a weighted average of them. The term ρ may be interpreted as the relative contribution of variance to mean-squared error. Values of ρ near 1 mean that variance is considered much more important than bias in the mind of the experimenter, while values of $1 - \rho$ near 1 indicate a correspondingly large concern with bias. Thus, ρ can be understood as the prior belief of the experimenter as to the nature of the true response function. Our "Minave" design is the design that minimizes $\mathscr{L}_{\mathrm{ave}}$ for a given value of ρ.

2.2. Bias-constrained D-optimal model-robust design criteria

The design criteria we consider here are motivated by a problem we encountered in the context of robust designs for generalized linear models. In the presence of model misspecification, the asymptotic covariance matrix of estimated model parameters in generalized linear models does not in general simply equal the inverse of the information matrix. The asymptotic covariance matrix is rather a function of the information matrix and the variance of the response Y under the true model. Thus, robust designs for generalized linear models would seem to require not only knowledge of the model parameters as in classical designs for generalized linear models but also knowledge of the variance of the response under the true but unknown model. The approach reported here is motivated by a requirement to generate robust designs without assuming this latter knowledge.

Here, we develop and apply this approach in the context of linear models; considerations for generalized linear models are reported in Adewale and Wiens (2005). We explore designs based on constrained maximization of the determinant of the information matrix of the fitted model parameters. Robustness is achieved via the constraint which represents some bound on a measure of the bias engendered by the model misspecification. Montepiedra and Fedorov (1997) proposed similar criteria using convex design theory when the true model is known specifically but the fitted model is, for the sake of parsimony, a simplified version of the true model. Their consideration is in the same setting as that of Box and Draper (1959) but instead of using a function of the mean-squared error as loss function they worked with constrained variance or bias, in order to keep the mathematics within the context of convex theory where the criteria would be amenable to Kiefer's equivalence theorem. Our construction differs in three senses: the true model is assumed unknown, we focus on finite design space and our designs are integer-valued. We investigate two kinds of constraints:

1. bounding the norm of the bias in the estimated regression parameter vector; and
2. bounding the average (over $\mathscr{F}$) of the average norm (over $\mathscr{S}$) of the bias in the predicted response.

This first constraint might be suitable when the focus of the design is on parameter estimation while the second would be recommended when the interest is in prediction.

The bias in $\hat{\boldsymbol{\beta}}$ is calculated as

$$bias\left(\hat{\boldsymbol{\beta}}\right) = E\left[\hat{\boldsymbol{\beta}}\right] - \boldsymbol{\beta}_0 = \left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{f}.$$

A development similar to that preceding Theorem 1 yields that the average-squared norm of the bias in $\hat{\boldsymbol{\beta}}$ is

$$aver_{\mathscr{F}}\left(\left\|bias\left(\hat{\boldsymbol{\beta}}\right)\right\|^2\right) = aver_{\mathscr{F}}\left(\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-2}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{f}\right)$$

and the average-squared norm of the bias in the predicted response is

$$aver_{\mathscr{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|bias\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\hat{\boldsymbol{\beta}}\right)\right\|^2\right) = aver_{\mathscr{F}}\left(\frac{1}{N}\left\{\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \mathbf{f}^{\mathrm{T}}\mathbf{f}\right\}\right).$$

In each case we average over $\mathscr{F}$ through the vector $\mathbf{c}$ and obtain the following results.

Theorem 2. *The average-squared norm of the bias in $\hat{\boldsymbol{\beta}}$ over the misspecification neighbourhood $\mathscr{F}$ is given by*

$$aver_{\mathscr{F}}\left(\left\|bias\left(\hat{\boldsymbol{\beta}}\right)\right\|^2\right) = \frac{\kappa_{N,p}\tau^2 N}{N-p}\,\mathrm{tr}\left\{\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1} - \mathbf{I}\right]\boldsymbol{\Lambda}^{-2}\right\}.$$

Theorem 3. *The average-squared norm of bias in predicted response over $\mathscr{F}$ and $\mathscr{S}$ is given by*

$$aver_{\mathscr{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|bias\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\hat{\boldsymbol{\beta}}\right)\right\|^2\right) = \kappa_{N,p}\tau^2\left(1 + \frac{\mathrm{tr}\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\right] - p}{N-p}\right).$$

The classical D-optimal design is the design that maximizes the determinant of the information matrix of the model parameters or, equivalently, the design that minimizes the determinant of the covariance matrix of the model parameters. Here, our robust design is the design that maximizes the determinant of the information matrix (divided by $\det(\mathbf{\Lambda}^2)$, which does not depend on the design) subject to either of the two bias-related constraints. That is:

Criterion 1 (*Robust estimation*). Choose $\mathbf{P}$ so as to maximize $\det\{\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\}$ subject to

$$\operatorname{tr}\left\{\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}-\mathbf{I}\right]\mathbf{\Lambda}^{-2}\right\}\leqslant\alpha. \tag{7}$$

Criterion 2 (*Robust prediction*). Choose $\mathbf{P}$ so as to maximize $\det\{\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\}$ subject to

$$\operatorname{tr}\left\{\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^2\mathbf{U}\right)\right\}\leqslant\beta. \tag{8}$$

The resulting design from the estimation criterion depends on α, while that of the prediction criterion depends on β. For values of α and β exceeding those attained by the classical D-optimal design, the "robust" and classical designs coincide.

3. Numerical algorithms

3.1. Simulated annealing algorithm for Minave designs

The optimization problem for Minave designs is an unconstrained nonlinear integer optimization problem. We consider models with p regressors $(z_1(x),\ldots,z_p(x))^{\mathrm{T}}$ where $x\in[-1,1]$. Given the desired number of observations (n) to be taken and the number of points in the design space (N), we seek designs that minimize the Minave loss function (6). Our algorithm accommodates all (n, N) combinations. If the problem of interest demands a design which is symmetric about the origin, (n, N) must be chosen such that symmetry is possible. For instance, symmetric designs would not be obtainable when N is even and n is odd. We take $\mathscr{S}$ to be the set $\{x_i=-1+2(i-1)/(N-1)\}_{i=1}^{N}$ of equally spaced points in $[-1, 1]$.

Simulated annealing is employed to search for optimal designs. The simulated annealing algorithm seeks to assign integers $n_i\geqslant 0$ to each of the design points x_i in such a way that $\mathscr{L}_{\mathrm{ave}}$ is a minimum. Simulated annealing is a direct search optimization algorithm which has been quite successful at finding the global extremum of a function, possibly nonsmooth, that has many local extrema. The algorithm is a biased random walk consisting of three steps. The first step is a specification of the initial state, $\mathbf{n}_0$, of the process based on which the corresponding initial value of the objective function $\mathscr{L}_{\mathrm{ave}}(\mathbf{n}_0)$ is calculated. The second step is the random choice of the next state of the process from the optimization space. The last step is a prescription of the basis of acceptance or rejection of the new state.

For the algorithm to be successful, each step has to be empirically tailored to the context of the problem at hand. We randomly choose p points in $\mathscr{S}$ and the n observations are randomly allocated to these p locations. This constitutes the initial state of the simulated annealing process. When the interest is in symmetric designs we randomly pick $\lceil p/2\rceil$ points of the points of $\mathscr{S}$ in $[-1, 0)$ and randomly allocate $\lceil n/2\rceil$ observations to these points. If n is odd then N has to be odd for symmetry, in this case replace $\lceil n/2\rceil$ by $\lfloor n/2\rfloor$ and assign one observation to the point $\{0\}$. A symmetric initial design is obtained by assigning the number of observations for locations in $[-1, 0)$ to their corresponding mirror image about the origin. Fang and Wiens (2000) assumed that one of (n, N) is a multiple of the other and then chose the initial design to be as uniform as possible. In the examples presented, we considered this approach to choosing the initial design as well. With either choice of initial design the algorithm converges to the same design.

To generate a new design, we perturb the current state as described in Fang and Wiens (2000). Define $\mathbf{v}$ to be the $N\times 1$ current allocation vector. For symmetric designs redefine $\mathbf{v}$ to be the $[N/2]\times 1$ vector consisting of the initial segment $(n_1,\ldots,n_{[N/2]})$ of the current allocation vector. Let $J_+=\{i\,|\,v_i>0\}$, $J_0=\{i\,|\,v_i=0\}$ with cardinalities $j_+\geqslant 1$ and j_0. If $j_+\geqslant 2$, generate a Bernoulli random variable

$$B=\begin{cases}1 & \text{with probability } j_0/(j_0+j_+),\\ 0 & \text{with probability } j_+/(j_0+j_+),\end{cases}$$

choose two indices (t_1, t_2) from J_+, at random without replacement, choose an index t_0 from J_0, at random and modify the selected components of $\mathbf{v}$ as follows:

$$v_{t_0} = v_{t_0} + B, \quad v_{t_1} = v_{t_1} - 1 \quad \text{and} \quad v_{t_2} = v_{t_2} + 1 - B. \tag{9}$$

If $j_+ = 1$, choose t_0 from J_0 at random, let t_1 be the index in the singleton set J_+, and then replace (9) by

$$v_{t_0} = v_{t_0} + 1, \quad v_{t_1} = v_{t_1} - 1.$$

This completes the perturbation scheme for general designs. For symmetric designs, we complete the scheme as follows. If N is even, then let $\mathbf{n} = (n_1, \ldots, n_N) = (v_1, \ldots, v_{N/2}, v_{N/2}, \ldots, v_1)$. If N is odd, then generate a uniform random variable u. If $u < 1/N$, with probability $\frac{1}{2}$ increase $n_{[N/2]+1}$ by 2 then randomly and symmetrically reduce the remaining n_i by 2; with probability $1/2$ reduce $n_{[N/2]+1}$ by 2 then randomly and symmetrically increase the remaining n_i. This step is omitted if $n_{[N/2]+1} < 2$. We then construct $\mathbf{n}$ as described above, with the inclusion of the new frequency $n_{[N/2]+1}$.

The value of the loss $\mathscr{L}_{\text{ave}} = \mathscr{L}_{\text{ave}}(\tilde{\mathbf{n}})$ is evaluated at the new state $\tilde{\mathbf{n}}$ and the state is accepted with probability π, defined as

$$\pi = \begin{cases} 1 & \text{if } \Delta\mathscr{L}_{\text{ave}} \leqslant 0, \\ \exp(-\Delta\mathscr{L}_{\text{ave}}/T) & \text{if } \Delta\mathscr{L}_{\text{ave}} \geqslant 0, \end{cases}$$

where $\Delta\mathscr{L}_{\text{ave}} = \mathscr{L}_{\text{ave}}(\tilde{\mathbf{n}}) - \mathscr{L}_{\text{ave}}(\mathbf{n})$. Thus, a favourable state ($\Delta\mathscr{L}_{\text{ave}} \leqslant 0$) is accepted with certainty and an unfavourable state is accepted according to a separate Bernoulli experiment with success probability $\exp(-\Delta\mathscr{L}_{\text{ave}}/T)$. We choose T such that initially the inequality $0.5 < \exp(-\Delta\mathscr{L}_{\text{ave}}/T) < 0.9$ is satisfied; this follows a suggestion from Bohachevsky et al. (1986). As long as $\exp(-\Delta\mathscr{L}_{\text{ave}}/T) > 0$ an unfavourable state could be accepted, thus providing the possibility of the path leading out of local minima. To ensure that the process settles at a global minimum we progressively decrease T. Fang and Wiens (2000) decrease T by a factor of 0.9 after each 100 iterations. We found that it is better to adapt the change of T to the specific problem of interest. In some cases we were able to decrease T by a factor of 0.95 after every 20th iteration.

In Fig. 1 we present the simulated annealing trajectory, for one of the examples of the next section, to illustrate our algorithm. As seen in this plot the simulated annealing process walks from one design to the next irrespective of whether the step taken is progressive (when it moves to a design with a lesser loss) or detrimental (when it moves to a design

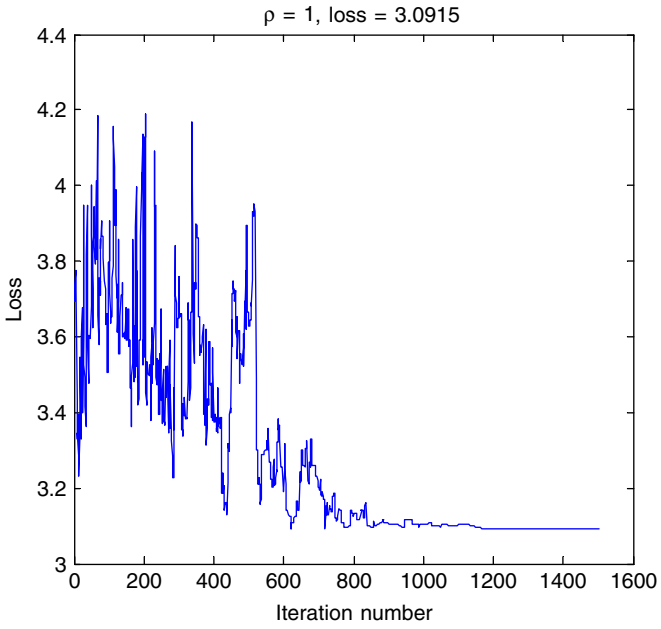

Fig. 1. Simulated annealing trajectory for cubic regression design with $N = 40$, $n = 20$ and $\rho = 1$.

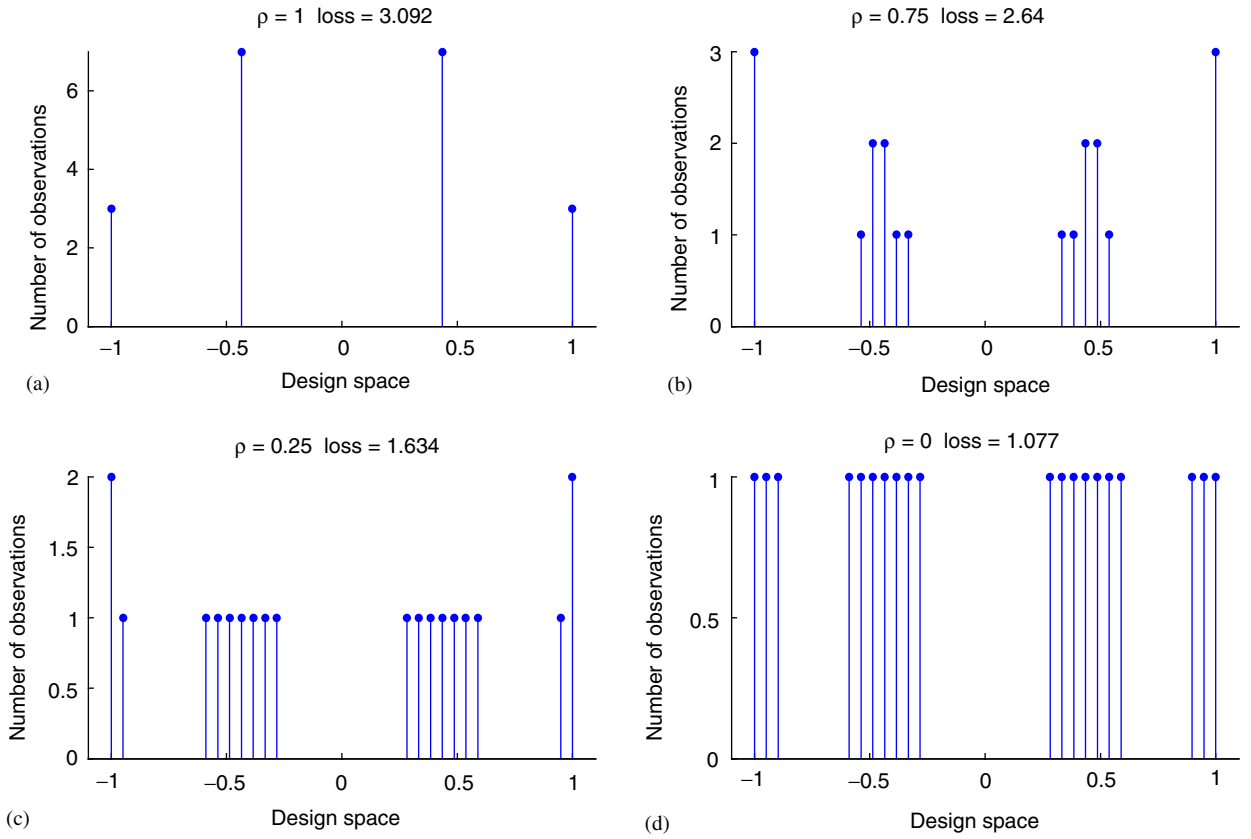

Fig. 2. Integer-valued designs for cubic regression with $N = 40$ and $n = 20$ for (a) $\rho = 1$, (b) $\rho = 0.75$, (c) $\rho = 0.25$ and (d) $\rho = 0$.

with a higher loss). The ability to accept detrimental steps prevents the process from being trapped in local extrema. After a large number of iterations, in this case 1500, it is expected that the algorithm has settled at a design with (near) minimum loss. In the examples below we satisfied ourselves of the convergence to a minimum by varying the number of iterations, the factor by which the temperature parameter is decreased and by running the algorithm many times.

3.2. Simulated annealing algorithm for bias-constrained D-optimal designs

Designing using our bias-constrained criteria is an integer optimization problem with a nonlinear constraint. The approach taken is to convert the problem into an unconstrained integer optimization problem such that the preceding algorithm for unconstrained designs becomes applicable. We achieved this by adding a penalty function to the objective function. Since the optimal design of interest is that which maximizes the objective function, we choose a negative-valued penalty function. Thus, the constrained optimization problems (7) and (8) are converted to

$$\text{maximize } \det\left\{\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right\} + g(x)I_{[x>\alpha]},$$

where

$$x = \operatorname{tr}\left\{\left[\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{2}\mathbf{U}\right)\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-1} - \mathbf{I}\right]\boldsymbol{\Lambda}^{-2}\right\} \tag{10}$$

and

$$\text{maximize } \det\left\{\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right\} + g(x)I_{[x>\beta]},$$

where

$$x = \mathrm{tr}\left\{\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right)^{-2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{2}\mathbf{U}\right)\right\}, \tag{11}$$

respectively, where $g(\cdot)$ is a negative-valued penalty function and $I_{[\cdot]}$ is an indicator function. Given $g(\cdot)$, the optimization problems become amenable to the algorithm we used in the preceding section for unconstrained integer optimization. We found that the algorithm was generally unsuccessful when the penalty function is chosen to be a constant, that is when $g(x) = -c$, where c is a positive real number. The choice of $g(\cdot)$ that we have found useful for solving (10) and (11) is instead $g(x) = -c \cdot x$, $c > 0$. In the examples we reported we chose $c = 1$; other values of c worked equally well.

4. Examples: polynomial regression

Example 1. For the purpose of comparison with the minimax designs of Fang and Wiens (2000), we consider approximate cubic regression. For $\rho = 1$ (so that $\mathscr{L}_{\mathrm{ave}}$ coincides with the average variance of the predicted values), $n = 20$ and $N = 40$ we obtain the design placing 3 of the 20 observations at each of ± 1 and 7 at each of ± 0.436. This agrees with the result of Fang and Wiens (2000) when the value of their design parameter υ, measuring the importance of variance relative to bias, is taken to be very large. See Fig. 2 for this and other obtained designs, all for $N = 40$. In contrast, when $\rho = 0$ the designs are as uniform as is allowed by the (n, N) combination.

In Figs. 3 and 4, we present designs for $n = 40$ and 60, respectively, for values of ρ ranging from 1 (all-variance design) to 0 (all-bias design). A comparison of the results for cases when $n < N$, $n = N$ and $n > N$ underscores the advantage of an exact integer-valued design over an approximate continuous design. Having obtained the design: $\pm 1(3)$,

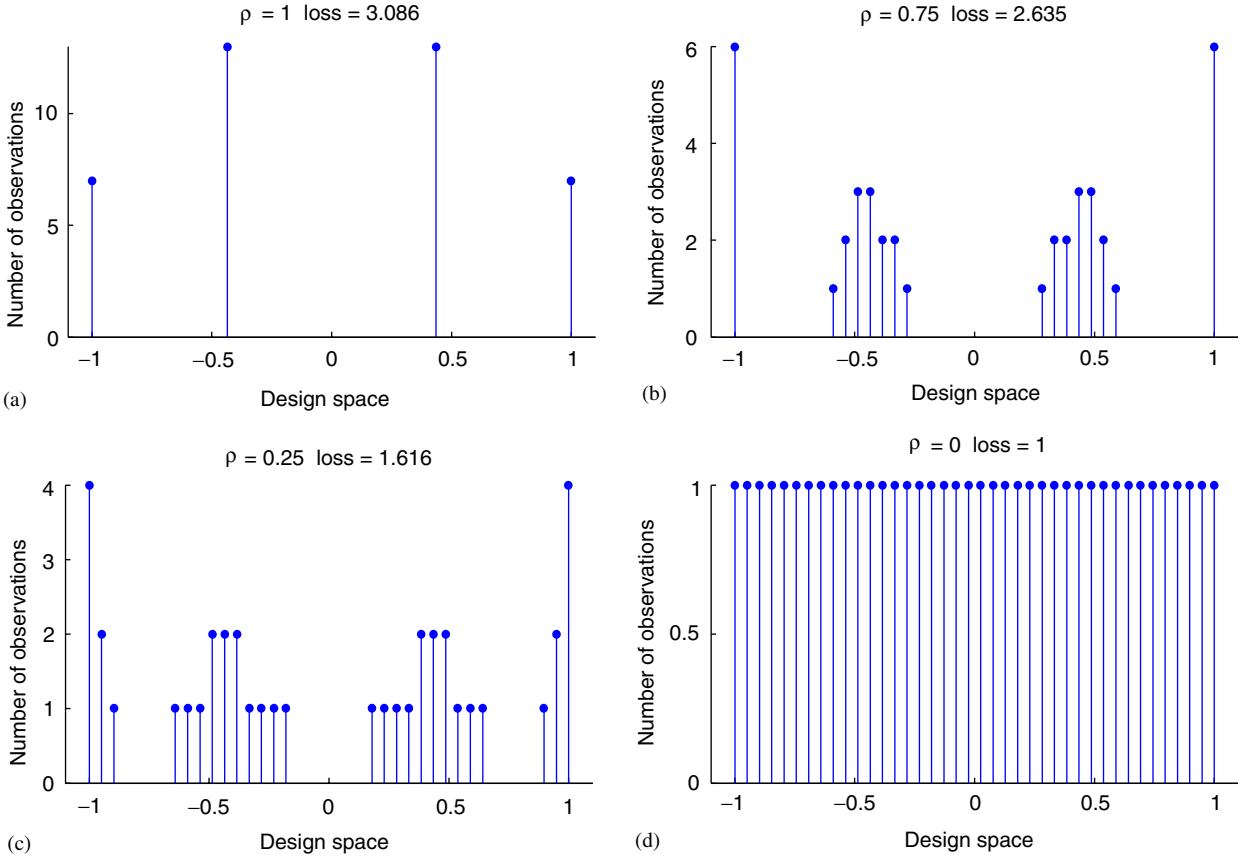

Fig. 3. Integer-valued designs for cubic regression with $N = 40$ and $n = 40$ for (a) $\rho = 1$, (b) $\rho = 0.75$, (c) $\rho = 0.25$ and (d) $\rho = 0$.

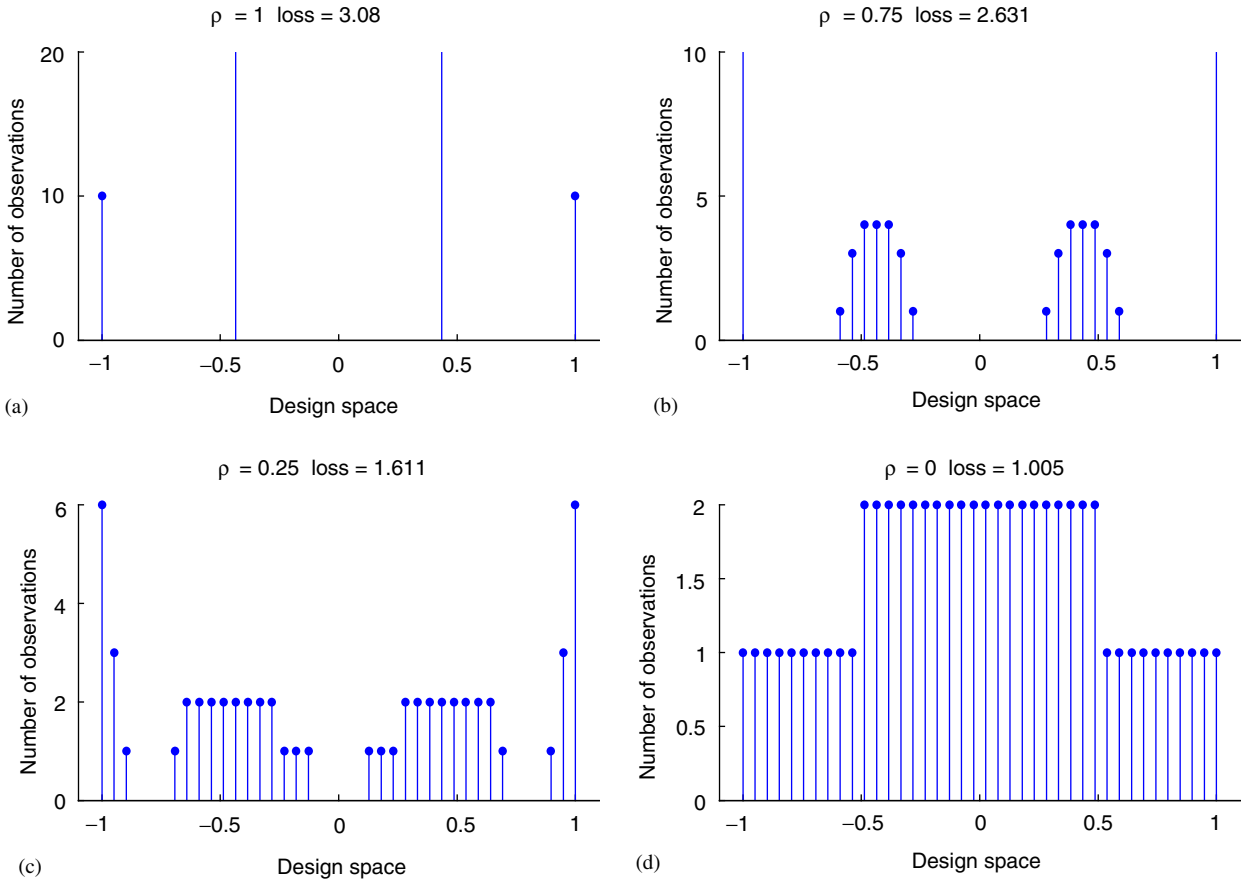

Fig. 4. Integer-valued designs for cubic regression with $N = 40$ and $n = 60$ for (a) $\rho = 1$, (b) $\rho = 0.75$, (c) $\rho = 0.25$ and (d) $\rho = 0$.

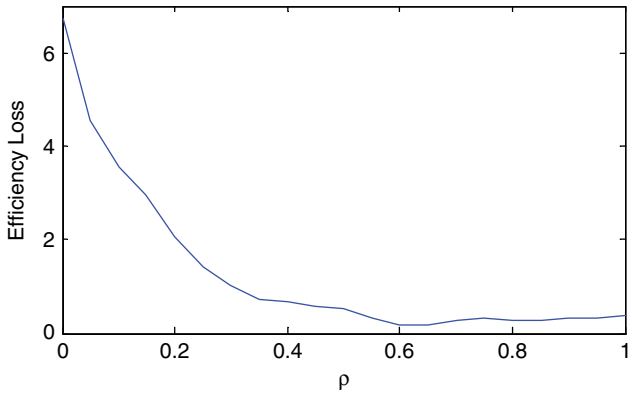

Fig. 5. Plot of efficiency loss (versus ρ) due to mere replications of the design points for $n = 20$ as opposed to an explicit construction of the designs for $n = 60$.

$\pm 0.436(7)$ for $n = 20$ and $\rho = 1$, without an explicit investigation of the case when $n = 40$ and $\rho = 1$, one would have been tempted to suggest the design: $\pm 1(6)$, $\pm 0.436(14)$ which would have been 99.8% efficient still because the design points are the same in this case. In the same vein, having obtained the design for $n = 20$ and $\rho = 0.2$ (Fig. 3(b)) to obtain the design for $n = 60$ and $\rho = 0.2$ one would probably multiply the number of observations taken at each of the design points obtained for $n = 20$ by 3. However, an explicit construction of the design for $n = 60$ and $\rho = 0.2$

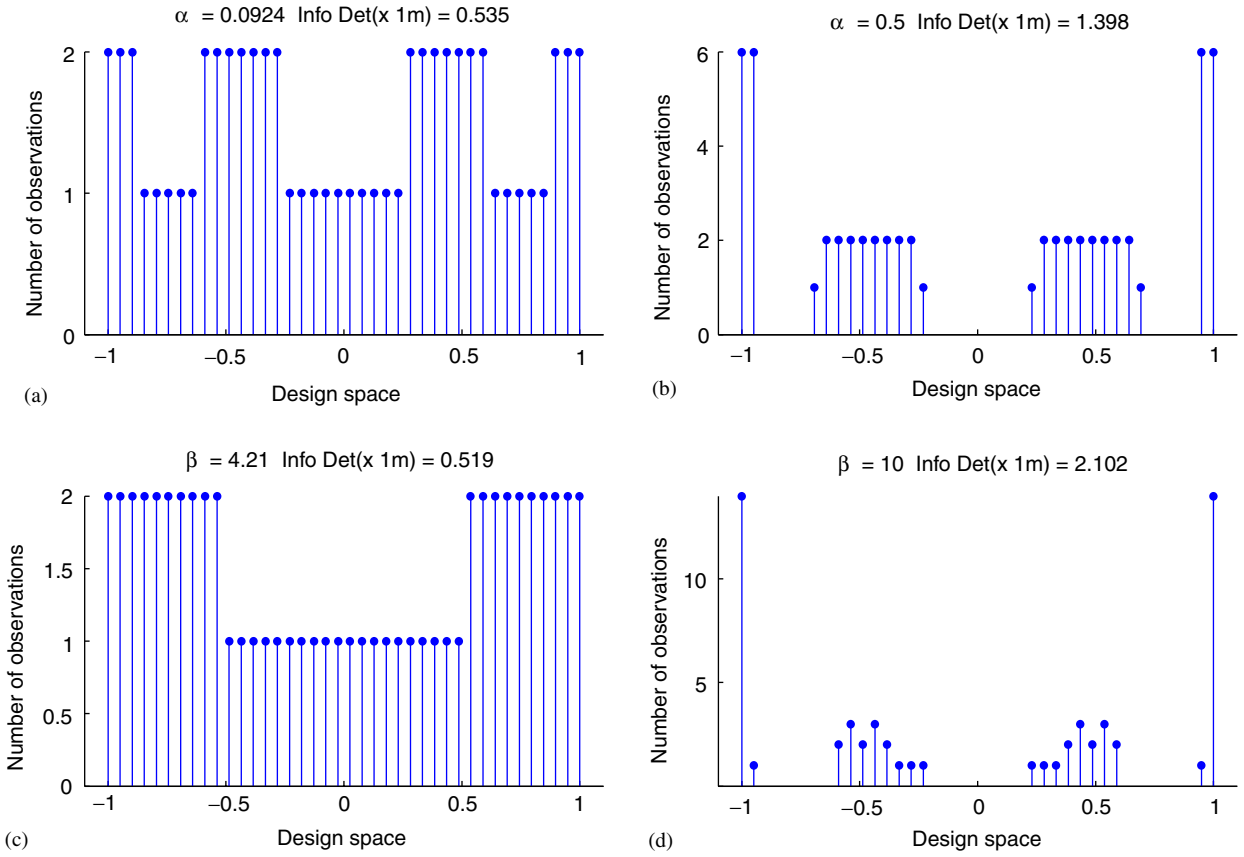

Fig. 6. Robust designs for parameter estimation with (a) $\alpha = 0.0924$, (b) $\alpha = 0.5$ and prediction with (c) $\beta = 4.2068$ and (d) $\beta = 10$.

resulted in a different design. We found that a naive replication would have resulted in about 2% loss in efficiency. The efficiency loss is much more here because the design for $n = 60$ and $\rho = 0.2$ includes other points that are not included in the design for $n = 20$ and $\rho = 0.2$. Fig. 5 is a plot of efficiency loss due to a naive replication of the design points for $n = 20$ as opposed to an explicit construction of design for $n = 60$ for varying values of our design parameter, ρ. The highest loss of efficiency occurs when $\rho = 0$ (so that $\mathscr{L}_{\text{ave}}$ coincides with the average bias of the predicted values). This comparison emphasizes the need to be cautious in implementing approximate continuous designs and the superiority of an integer-valued (exact) construction for specific applications.

The resulting designs from the two bias-constrained criteria coincide with the classical D-optimal design, that is, the design that maximizes the determinant of the information matrix, for sufficiently large values of α or β. The classical D-optimal design is the design that takes the assume model to be exact. Gaffke (1987), Huang (1987), Haines (1987) and Chen and Huang (2000) have constructed exact D-optimal designs, with "optimality" holding when the fitted model is exactly correct. In contrast, the constraints imposed here protects against misspecifications in the assumed model within a finite design space framework. The least attainable α (resp., β), say $\alpha_0\left(\beta_0\right)$, corresponds to the value of the constraint for an all-bias design. When n is a multiple of N, $\alpha_0\left(\beta_0\right)$ corresponds to the bias of the, minimum bias, uniform design. For the estimation criterion it follows from (3) that $\alpha_0 = 0$ and for the prediction criterion $\beta_0 = p$, the number of parameters in the model. When n is not a multiple of N, the design with minimum bias has $\alpha_0 > 0$ for the estimation criterion and $\beta_0 > p$ for the prediction criterion.

Example 2 (*Bias-constrained D-optimal designs*). We again use the cubic regression model to illustrate these criteria. Guest (1958) and Hoel (1958) independently reported an approach based on zeroes of derivatives of Legendre polynomials for obtaining the D-optimal support points for polynomials regression. Using this approach, Pukelsheim (1993,

p. 218) presented the D-optimal design over the interval $[-1, 1]$ for cubic regression as the design placing equal mass at each of $\pm 0.447, \pm 1$. For large α or β, whichever is appropriate, our simulated annealing algorithm results in a design taking equal number of observations at each of $\pm 0.436, \pm 1$. Our algorithm attains the closest approximation to that given above since the points ± 0.436 are the nearest, in our design space, to ± 0.447. In panel (a) of Fig. 6 we present robust parameter estimation design when $N = 40$, $n = 60$ with α set to a value ($\alpha = 0.0924$) just slightly greater than the least admissible α ($\alpha_0 = 0.0923$), for this setting. A corresponding design for the robust prediction design criterion is presented in Fig. 6(c), for this $\beta_0 = 4.2067$ and $\beta = 4.2068$. Design for another value of α (parameter estimation design) and β (prediction design) are presented in panels (b) and (d), respectively. Notable is the fact that each constraint shapes the designs differently.

5. Designs for multiple regression

The algebra of Sections 1 and 2 applies to general regression models. However, the simulated annealing algorithm described in Section 3 may not be readily useful in situations where there are two or more independent variables. This is because the design space is then multidimensional, raising the usual problems related to the curse of dimensionality. Here, we consider first and second order multiple regression models. Suppose $\mathbf{x} = (x_1, \ldots, x_q)^{\mathrm{T}}$ and $\mathbf{z}(\mathbf{x})$ has elements $1, x_1, \ldots, x_q$ and possibly second-order terms $x_i x_j$ $(1 \leqslant i \leqslant j \leqslant q)$. We define the design space as the q-fold Cartesian product $\mathscr{S} = \mathscr{S}_1 \cdots \mathscr{S}_q$, where each $\mathscr{S}_j$ has N_0 equally spaced points in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Thus, the design space consists of $N = N_0^q$ points. If, for instance, $N_0 = 40$ as in the examples for polynomial regression, then for $q = 2$, the design space consists of $40^2 = 1600$ points and for $q = 3$, the number of points increases to $40^3 = 64,000$. The computational complexity would thus increase with the dimensionality and thus the algorithm of Section 3 may not produce the optimal design in any reasonable amount of time. However, since $\mathscr{S}$ is symmetric about $\mathbf{0}$ and invariant under permutations of the coordinate axes, Fang and Wiens (2000) proposed that the designs in $\mathscr{S}$ can be generated by symmetrically choosing n_0 points in the x_1-axis and then forming the q-fold Cartesian product of these points with themselves, whence $n = n_0^q$. The algorithm of Section 3 then becomes applicable in choosing the n_0 points. Here, we propose to adopt an alternative approach, previously employed by Heo et al. (2001). In this approach we continue to assume symmetry of designs about $\mathbf{0}$ and exchangeability of the coordinate axes. (We can often arrange symmetry through an affine transformation of the independent variables, in which case we do not lose generality.) Invariance under permutations of the axes would be reasonable when there is no a priori reason to prefer one coordinate over another. To describe the approach, suppose the problem of interest requires n design points. We choose $n_0 = n/8$ points (x_1, x_2) in $\left\{0 \leqslant x_2 \leqslant x_1 \leqslant \frac{1}{2}\right\}$ and then

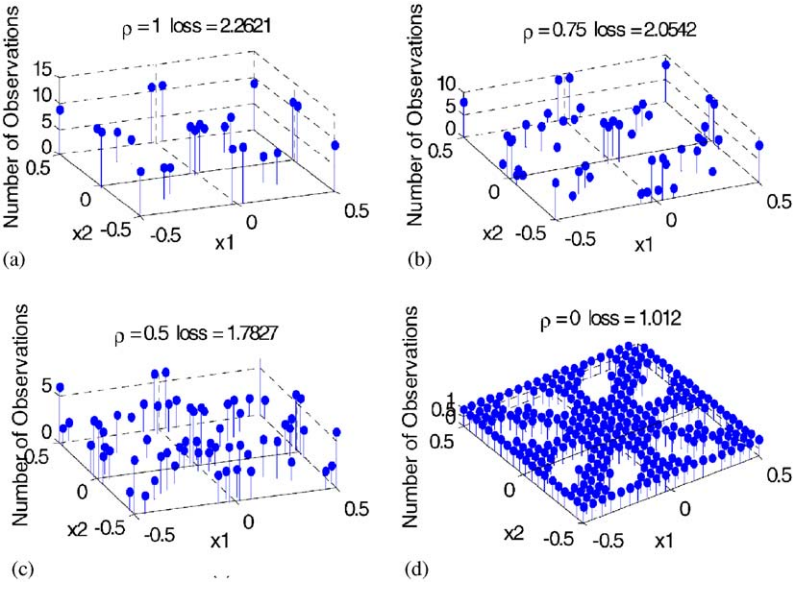

Fig. 7. Minave designs for partial second-order model (a) $\rho = 1$, (b) $\rho = 0.75$, (c) $\rho = 0.5$, (d) $\rho = 0$.

Table 1
Minave designs for partial second-order model

ρ	(Design points; frequencies) in half-quadrant $0 \leqslant x_1 \leqslant x_2 \leqslant 1/2$	Loss
1	(0.0263, 0.0263; 9), (0.2368, 0.2368; 5), (0.2895, 0.2895; 6), (0.5, 0.0263; 11), (0.5, 0.5; 9)	2.2621
0.75	(0.0263, 0.0263; 8), (0.1842, 0.1842; 1), (0.2368, 0.2368; 5), (0.2895, 0.2895; 5), (0.3421, 0.3421; 1), (0.4474, 0.0263; 1), (0.5, 0.0263; 9), (0.5, 0.0789; 2), (0.5, 0.5; 8)	2.0542
0.5	(0.0263, 0.0263; 6), (0.0789, 0.0263; 2), (0.1842, 0.1842; 2), (0.2368, 0.2368; 4), (0.2895, 0.2895; 4), (0.3421, 0.3421; 2), (0.4474, 0.0263; 3), (0.5, 0.0263; 6), (0.5, 0.0789; 3), (0.5, 0.4474; 2), (0.5, 0.5; 6)	1.7827
0	(0.0263, 0.0263; 1), (0.0789, 0.0263; 1), (0.0789, 0.0789; 1), (0.1316, 0.0263; 1), (0.1316, 0.0789; 1), (0.1316, 0.1316; 1), (0.1842, 0.0263; 1), (0.1842, 0.0789; 1), (0.1842, 0.1316; 1), (0.1842, 0.1842; 1), (0.2368, 0.0263; 1), (0.2368, 0.2368; 1), (0.2895, 0.0263; 1), (0.2895, 0.1842; 1), (0.2895, 0.2368; 1), (0.2895, 0.2895; 1), (0.3421, 0.2895; 1), (0.3421, 0.3421; 1), (0.3947, 0.0263; 1), (0.3947, 0.0789; 1) (0.3947, 0.3421; 1), (0.3947, 0.3947; 1), (0.4474, 0.0263; 1), (0.4474, 0.0789; 1) (0.4474, 0.1316; 1), (0.4474, 0.3421; 1), (0.4474, 0.4474; 1), (0.5, 0.0263; 1), (0.5, 0.0789; 1), (0.5, 0.1316; 1), (0.5, 0.1842; 1), (0.5, 0.5; 1)	1.0120

Table 2
Minave designs for full second-order model

ρ	(Design points; frequencies) in half-quadrant $0 \leqslant x_1 \leqslant x_2 \leqslant 1/2$	Loss
1	(0.0263, 0.0263; 6), (0.2368, 0.0263; 6), (0.2368, 0.2368; 6), (0.2895, 0.0263; 6), (0.2895, 0.2895; 6), (0.3421, 0.1842; 1), (0.5, 0.0263; 6), (0.5, 0.2368; 6), (0.5, 0.2895; 6), (0.5, 0.5; 6)	3.9320
0.75	(0.0263, 0.0263; 7), (0.2368, 0.0263; 5), (0.2368, 0.2368; 5), (0.2895, 0.0263; 6), (0.2895, 0.2895; 6), (0.3421, 0.1842; 3), (0.5, 0.0263; 6), (0.5, 0.2368; 6), (0.5, 0.2895; 5), (0.5, 0.5; 6)	3.2863
0.5	(0.0263, 0.0263; 6), (0.1842, 0.0263; 1), (0.1842, 0.1842; 1), (0.2368, 0.0263; 5), (0.2368, 0.2368; 5), (0.2895, 0.0263; 4), (0.2895, 0.0789; 1), (0.2895, 0.2368; 1), (0.2895, 0.2895; 4), (0.3421, 0.0263; 1), (0.3421, 0.1842; 1), (0.3421, 3421; 1), (0.4474, 0.2368; 2), (0.5, 0.0263; 6), (0.5, 0.1842; 1), (0.5, 0.2368; 4), (0.5, 0.2895; 4), (0.5, 0.3421; 1), (0.5, 0.5; 6)	2.6173
0	{*all points on the design space with one observation taken a point*	1.0000

obtained the remaining $7n_0$ points by symmetry and exchangeability. Given the finite design space with $N = N_0^2$ points generated by the two-fold Cartesian product of equally spaced N_0 points in $\left[-\frac{1}{2}, \frac{1}{2}\right]$, this requires a search over the $N_0\,(N_0+1)\,/8$ points satisfying $\left\{0 \leqslant x_2 \leqslant x_1 \leqslant \frac{1}{2}\right\}$. We again employ simulated annealing.

Example 3 (*Partial second-order model with two interacting regressors*). We consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + f(\mathbf{x}) + \varepsilon$$

on $\mathscr{S} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$ with two interacting regressors, that is, $z(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)^{\mathrm{T}}$. This model is exchangeable in x_1 and x_2. For $N_0 = 20$ and $n_0 = 40$, we obtain designs corresponding to $\rho = 1$ (minimizing variance only), 0.75, 0.5 and 0 (minimizing bias only). The variance-minimizing design has five support points. The number of support points increases with increasing bias in the model. In general, the pattern of results agree with previous work in the literature. The designs are clusters of observations about the points that would have been the design points for the variance-minimizing design and the spread of the clusters increases with increasing bias (see Fig. 7 and Tables 1 and 2).

Example 4 (*Full second-order model with* $q = 2$). Here, we consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + f(\mathbf{x}) + \varepsilon$$

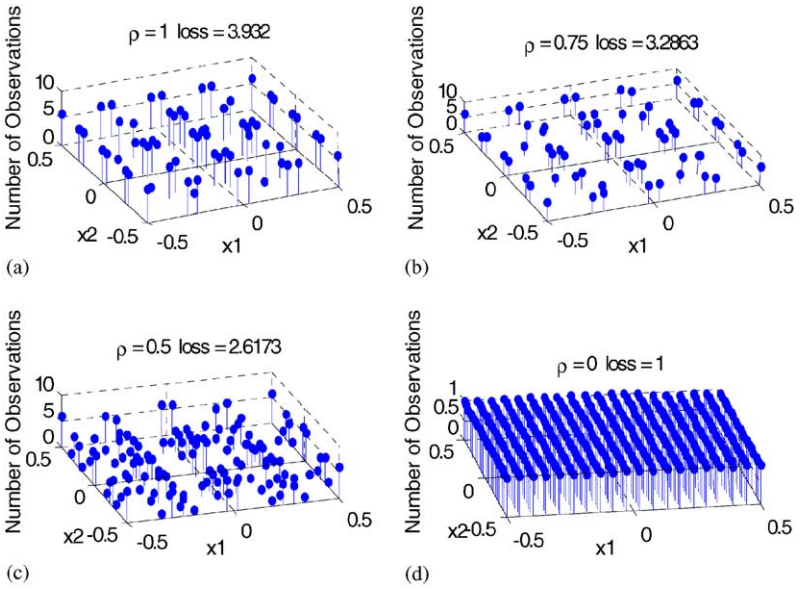

Fig. 8. Minave designs for full second-order model (a) $\rho = 1$, (b) $\rho = 0.75$, (c) $\rho = 0.5$, (d) $\rho = 0$.

on $\mathscr{S} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$, $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$ and $z(\mathbf{x}) = \left(1, x_1, x_2, x_1x_2, x_1^2, x_2^2\right)^{\mathrm{T}}$. This model is an extension of that considered in the last example. It is also exchangeable in x_1 and x_2. Given $N_0 = 20$ and $n_0 = 55$, we obtained designs corresponding to $\rho = 1$ (minimizing variance only), 0.75, 0.5 and 0 (minimizing bias only). The pattern of results is again similar to those in the previous example. The number of distinct design points for each value of ρ is greater than that for the corresponding design for partial second-order model since we have more parameters here. See design plots in Fig. 8.

6. Concluding remarks

We have investigated new criteria for the construction of robust integer-valued regression designs. These designs are robust against misspecifications in the assumed model form. We recommend the use of the Minave and robust prediction constrained criteria when the focus of the design is prediction while the robust estimation criterion is recommended when the design interest is estimation. Simulated annealing has been used successfully to seek integer designs both for unconstrained and constrained optimizations. The results obtained from the use of the three criteria follow a consistent pattern.

The patterns of the results are still similar to those obtained using a minimax approach even though the rationale for the designs presented here are based on the notion of averaging, rather than maximizing, the loss over the contamination space. The key message remains that the designs that protect against the general forms of model misspecification may be approximated by taking clusters of observations about those points that would have otherwise serve as the design points for variance-minimizing designs. We have also demonstrated the superiority of an integer-valued construction over the approximate continuous method in that as the desired number of observations increase our design does not merely produce replicates at the points that have been included in the design when fewer observations were taken; it instead incorporates additional design points.

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We appreciate the suggestions of an anonymous referee.

References

Adewale, A.J., Wiens, D.P., 2005. Robust designs for misspecified logistic models. University of Alberta Statistics Centre Technical Report 05-08.
Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. Technometrics 28, 209–217.
Box, G.E.P., Draper, N.R., 1959. A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622–654.
Chen, R.-B., Huang, M.-N.L., 2000. Exact D-optimal designs for weighted polynomial regression model. Comput. Statist. Data Anal. 33, 137–149.
Fang, Z., Wiens, D.P., 2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. J. Amer. Statist. Assoc. 95 (451), 807–818.
Gaffke, N., 1987. On D-optimality of exact linear regression designs with minimum support. J. Statist. Planning Infer. 15, 189–204.
Guest, P.G., 1958. The spacing of observations in polynomial regression. Ann. Math. Statist. 29 (1), 294–299.
Haines, L.M., 1987. The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. Technometrics 29, 439–447.
Heo, G., Schmuland, B., Wiens, D.P., 2001. Restricted minimax robust designs for misspecified regression models. Canad. J. Statist. 29, 117–128.
Hoel, P.G., 1958. Efficiency problems in polynomial estimation. Ann. Math. Statist. 29 (4), 1134–1145.
Huang, M.-N.L., 1987. Exact D-optimal design for polynomial regression. Bull. Inst. Math. Acad. Sinica 15, 59–71.
Läuter, E., 1974. Experimental design in a class of models. Math. Oper. Statistik 5, 379–396.
Läuter, E., 1976. Optimal multipurpose designs for regression models. Math. Oper. Statistik 7, 51–68.
Montepiedra, G., Fedorov, V.V., 1997. Minimum bias designs with constraints. J. Statist. Planning Infer. 63, 97–111.
Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, New York.