Jittering and Clustering: Strategies for the Construction of Robust Designs

Douglas P. Wiens[a,*]

[a]*Mathematical & Statistical Sciences, University of Alberta, Edmonton, Canada, T6G 2G1*

May 10, 2024

Abstract

We discuss, and give examples of, methods for randomly implementing some minimax robust designs from the literature. These have the advantage, over their deterministic counterparts, of having bounded maximum loss in large and very rich neighbourhoods of the, almost certainly inexact, response model fitted by the experimenter. Their maximum loss rivals that of the theoretically best possible, but not implementable, minimax designs. The procedures are then extended to more general robust designs. For two-dimensional designs we sample from contractions of Voronoi tessellations, generated by selected basis points, which partition the design space. These ideas are then extended to k-dimensional designs for general k.

Keywords: central composite design, deterministic design, minimax, random design, robustness, tessellation, Voronoi.
2010 MSC: Primary 62F35, Secondary 62K05

1. Introduction and summary

In this article we investigate various methods of implementing experimental designs, robust against model inadequacies. We begin with a review of the 'minimax' theory of robustness of design, and of some minimax designs from the literature. For this we initially follow Atkinson (1996) and view a design as any probability measure on the design space. It will be seen that the designs which protect against a large class of alternative response models are necessarily absolutely continuous, and so lose their optimality when approximated by implementable, discrete (deterministic) designs. Two remedies for this and other issues are proposed, suggested by work of Waite and Woods (2022), who propose and study *random design* strategies.

The first remedy is a random design strategy termed *jittering*. The designs are obtained by uniform sampling from small neighbourhoods of an optimal set $t^* = \{t_i | i = 1, ..., n\}$ of points, chosen to represent the minimax design density. Both completely random and stratified random – i.e. random within each neighbourhood – are considered. We assess

*E-mail: doug.wiens@ualberta.ca. Supplementary material is at www.ualberta.ca/~dwiens/.

these designs by looking at the sample distributions of the mean squared prediction errors incurred; with respect to these measures both sampling strategies typically lead to designs very nearly optimal, with the stratification strategy clearly outperforming its completely random counterpart.

We then investigate a strategy leading to *cluster* designs, motivated by the observation that robust designs for a particular response model tend to place their mass near those points $t_i \in t^*$ at which classically optimal designs, focussed solely on variance minimization, are replicated – but with their support points spread out in clusters of nearby points, rather than being replicated. In clustering the idea is to sample from densities concentrated near the t_i. An advantage to this method over jittering is that there is no need for the minimax design to already have been derived.

Both these approaches parallel the 'random translation design strategy' of Waite and Woods (2022), who sample uniformly in small neighbourhoods of a chosen set of points, but with some significant differences. The choice of t^* in jittering allows for designs whose maximum expected loss rivals that of the minimax, absolutely continuous design. In clustering, both the support of the non-uniform densities from which we sample, and the extent of their concentration near the t_i, are governed by a user-chosen parameter ν, representing the bias/variance trade-off desired by a user seeking robustness against model misspecifications.

We start by applying these ideas in several one-dimensional cases for which minimax designs – in continuous or discrete design spaces – have previously been derived. The framework is that the experimenter will fit a polynomial response, and our random designs have points assigned at random but in a structured manner near the t^*. The densities from which we sample are chosen to capture the salient properties of the minimax designs (in jittering) or classically optimal but deterministic designs (in clustering). The structure we impose – especially that of stratification – is shown, through a number of examples, to lead to efficient designs approximating the variance minimizing properties of the deterministic designs concentrated on the t^*. But the randomness, leading to the clustering effect – this alone is known to increase robustness – ensures that the bias is bounded as well, even in continuous design spaces in which the bias of deterministic designs can be unbounded.

We then consider two-dimensional clustering applications in which intervals containing the t_i are replaced by less regular regions formed by shrinking Voronoi tessellations generated by t^*. We sample from spherical beta densities centred on the t_i, and suggest tuning constants which again result in both efficiency and robustness. We finish with recommendations for the construction of k-dimensional designs for $k \geq 3$.

The examples were prepared using MATLAB; the code is available on the author's website.

2. Minimax robustness of design

The theory of robustness of design was largely initiated by Box and Draper (1959), who investigated the robustness of some classical experimental designs in the presence of certain model inadequacies, e.g. designs optimal for a low order polynomial response

when the true response was a polynomial of higher order. Huber (1975) derived *minimax* designs for straight line regression; these minimize the maximum integrated mean squared error, with the maximum taken over a large class of alternative responses. Wiens (1990, 1992) extended these results to multiple regression responses and in a variety of other directions – see Wiens (2015) for a summary of these and other approaches to robustness of design. Specifically, the general problem is phrased in terms of an approximate regression response

$$E\left[Y\left(\boldsymbol{x}\right)\right] \approx \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}, \tag{1}$$

for p regressors $\boldsymbol{f}$, each functions of q independent variables $\boldsymbol{x}$, and a parameter $\boldsymbol{\theta}$. Since (1) is an approximation the interpretation of $\boldsymbol{\theta}$ is unclear; we *define* this target parameter by

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \int_{\mathcal{X}} \left(E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\eta}\right)^2 \mu\left(d\boldsymbol{x}\right), \tag{2}$$

where $\mu\left(d\boldsymbol{x}\right)$ represents either Lebesgue measure or counting measure, depending upon the nature of the *design space* $\mathcal{X}$. We then define $\psi\left(\boldsymbol{x}\right) = E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}$. This results in the class of responses $E\left[Y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta} + \psi\left(\boldsymbol{x}\right)$, with – by virtue of (2) – ψ satisfying the orthogonality requirement

$$\int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\psi\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) = \boldsymbol{0}. \tag{3}$$

Assuming that $\mathcal{X}$ is rich enough that the matrix $\boldsymbol{A} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right)$ is invertible, the parameter defined by (2) and (3) is unique.

We identify a design with its design measure – a probability measure $\xi\left(d\boldsymbol{x}\right)$ on $\mathcal{X}$. Define

$$\boldsymbol{M}_{\xi} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right), \quad \boldsymbol{b}_{\psi,\xi} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\psi\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right),$$

and assume ξ is such that $\boldsymbol{M}_{\xi}$ is invertible. The covariance matrix of the least squares estimator $\hat{\boldsymbol{\theta}}$, assuming homoscedastic errors with variance σ_{ε}^2, is $\left(\sigma_{\varepsilon}^2/n\right)\boldsymbol{M}_{\xi}^{-1}$, and the bias is $E\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \boldsymbol{M}_{\xi}^{-1}\boldsymbol{b}_{\psi,\xi}$; together these yield the mean squared error (*mse*) matrix

$$\text{MSE}\left[\hat{\boldsymbol{\theta}}\right] = \frac{\sigma_{\varepsilon}^2}{n}\boldsymbol{M}_{\xi}^{-1} + \boldsymbol{M}_{\xi}^{-1}\boldsymbol{b}_{\psi,\xi}\boldsymbol{b}_{\psi,\xi}'\boldsymbol{M}_{\xi}^{-1}$$

of the parameter estimates, whence the *mse* of the fitted values $\hat{Y}\left(\boldsymbol{x}\right) = \boldsymbol{f}'\left(\boldsymbol{x}\right)\hat{\boldsymbol{\theta}}$ is

$$\text{MSE}\left[\hat{Y}\left(\boldsymbol{x}\right)\right] = \frac{\sigma_{\varepsilon}^2}{n}\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{M}_{\xi}^{-1}\boldsymbol{f}\left(\boldsymbol{x}\right) + \left(\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{M}_{\xi}^{-1}\boldsymbol{b}_{\psi,\xi}\right)^2.$$

A loss function that is commonly employed is the *integrated mse* of the predictions:

$$\begin{aligned} \text{IMSE}\left(\xi|\psi\right) &= \int_{\mathcal{X}} \text{MSE}\left[\hat{Y}\left(\boldsymbol{x}\right)\right] d\boldsymbol{x} \\ &= \frac{\sigma_{\varepsilon}^2}{n} tr\left(\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\right) + \boldsymbol{b}_{\psi,\xi}'\boldsymbol{M}_{\xi}^{-1}\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\boldsymbol{b}_{\psi,\xi} + \int_{\mathcal{X}} \psi^2\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right). \end{aligned} \tag{4}$$

The dependence on ψ is eliminated by adopting a *minimax* approach, according to which one first maximizes (4) over a neighbourhood of the assumed response. This neighbourhood is constrained by (3) and by a bound $\int_{\mathcal{X}} \psi^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) \leq \tau^2/n$, required so that errors due to bias and to variation remain of the same order, asymptotically.

Huber (1975) took $\mathcal{X}$ to be an interval of the real line and assumed that the minimax design measure had a density $m(x)$; Wiens (1992) justified this assumption by proving that any design whose design space $\mathcal{X}$ has positive Lebesgue measure, and which places positive mass on a set of Lebesgue measure zero, necessarily has $\sup_{\psi} \text{IMSE}(\xi|\psi) = \infty$. Thus in order that a design on an interval, hypercube, etc. have finite maximum loss, it must be absolutely continuous. For such a design $\max_{\psi} \text{IMSE}(\xi|\psi)$ is $\left(\sigma_\varepsilon^2 + \tau^2\right)/n$ times

$$I_\nu(\xi) = (1-\nu)\, tr\boldsymbol{A}\boldsymbol{M}_\xi^{-1} + \nu ch_{\max}\boldsymbol{K}_\xi\boldsymbol{H}_\xi^{-1}, \tag{5}$$

where

$$\boldsymbol{H}_\xi = \boldsymbol{M}_\xi\boldsymbol{A}^{-1}\boldsymbol{M}_\xi,\ \boldsymbol{K}_\xi = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\,\boldsymbol{f}'(\boldsymbol{x})\,m^2(\boldsymbol{x})\,d\boldsymbol{x},$$

$ch_{\max}$ denotes the maximum eigenvalue and $\nu = \tau^2 \Big/ \left(\sigma_\varepsilon^2 + \tau^2\right) \in [0,1]$, representing the relative importance, to the experimenter, of errors due to bias rather than to variance. Our examples in this article use $\nu = .5$; other values tell much the same story.

With

$$\begin{aligned} \boldsymbol{G}_\xi &= \boldsymbol{K}_\xi - \boldsymbol{H}_\xi \\ &= \int_{\mathcal{X}} \left[\left(m(\boldsymbol{x})\,\boldsymbol{I}_p - \boldsymbol{M}_\xi\boldsymbol{A}^{-1}\right)\boldsymbol{f}(\boldsymbol{x})\right]\left[\left(m(\boldsymbol{x})\,\boldsymbol{I}_p - \boldsymbol{M}_\xi\boldsymbol{A}^{-1}\right)\boldsymbol{f}(\boldsymbol{x})\right]' d\boldsymbol{x}, \\ \mathbf{r}_\xi(\boldsymbol{x}) &= \frac{\tau}{\sqrt{n}}\boldsymbol{G}_\xi^{-1/2}\left(m(\boldsymbol{x})\,\boldsymbol{I}_p - \boldsymbol{M}_\xi\boldsymbol{A}^{-1}\right)\boldsymbol{f}(\boldsymbol{x}), \end{aligned} \tag{6}$$

the least favourable contaminant is

$$\psi_\xi(\boldsymbol{x}) = \mathbf{r}_\xi'(\boldsymbol{x})\beta_\xi, \tag{7}$$

where β_ξ is the unit eigenvector belonging to the maximum eigenvalue of $\boldsymbol{G}_\xi^{1/2}\boldsymbol{H}_\xi^{-1}\boldsymbol{G}_\xi^{1/2} + \boldsymbol{I}_p$. See Wiens (2015) for details and further references.

2.1. Random designs

In the following sections we construct distributions $\Phi(\boldsymbol{x})$, with densities $\phi(\boldsymbol{x})$, and propose randomly choosing design points from Φ. An n-point design $D = \{\boldsymbol{x}_i\}_{i=1}^n$ chosen in this way has design measure $\delta = n^{-1}\sum \delta_{\boldsymbol{x}_i}$, where $\delta_{\boldsymbol{x}_i}$ is point mass at $\boldsymbol{x}_i \sim \Phi$. By the preceding any such design has unbounded IMSE once it is chosen. Of interest however is the *expected* IMSE against a common alternative ψ; for this we take the least favourable contaminant ψ_Φ, given by (6) and (7) but with ξ replaced by Φ. In the Appendix we show that

$$E_\Phi\left[\text{IMSE}(\delta|\psi_\Phi)\right] = \left(\sigma_\varepsilon^2 + \tau^2\right)/n \times J_\nu(\Phi), \tag{8}$$

where

$$\begin{aligned} J_\nu(\Phi) &= E_\Phi\left[j_\nu(\delta)\right], \text{ for} \\ j_\nu(\delta) &= (1-\nu)\, tr\boldsymbol{A}\boldsymbol{M}_\delta^{-1} + \nu\gamma_\delta, \text{ and} \\ \gamma_\delta &= \beta_\Phi'\boldsymbol{G}_\Phi^{-1/2}\left(\boldsymbol{M}_\phi\boldsymbol{M}_\delta^{-1} - \boldsymbol{M}_\Phi\boldsymbol{A}^{-1}\right)\boldsymbol{A}\left(\boldsymbol{M}_\delta^{-1}\boldsymbol{M}_\phi - \boldsymbol{A}^{-1}\,\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi + 1. \end{aligned} \tag{9}$$

Here $\boldsymbol{M}_\delta \overset{def}{=} \frac{1}{n}\sum_{\boldsymbol{x}_i\in D}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$ and $\boldsymbol{M}_\phi \overset{def}{=} \frac{1}{n}\sum_{\boldsymbol{x}_i\in D}\boldsymbol{f}(\boldsymbol{x}_i)\phi(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$; β_Φ is the unit eigenvector belonging to the maximum eigenvalue of $\boldsymbol{G}_\Phi^{1/2}\boldsymbol{H}_\Phi^{-1}\boldsymbol{G}_\Phi^{1/2} + \boldsymbol{I}_p$.

Note that both $\boldsymbol{M}_\delta$ and $\boldsymbol{M}_\phi$ are random. The expectation in (9) can be estimated by averaging over a large number of realizations of δ – we do this in Sections 4 and 5. In the special case that $\phi(\boldsymbol{x})$ is constant on its support – as is the case in §3 – $\boldsymbol{M}_\delta^{-1}\boldsymbol{M}_\phi$ is a constant multiple of $\boldsymbol{I}_p$, γ_δ is non-random, and these formulas simplify considerably – see (12).

An efficient design strategy should result in $J_\nu(\Phi)$ being close to $I_\nu(\Phi)$, with the $j_\nu(\delta)$ being concentrated near their expectation.

A referee has pointed out that a more natural measure is perhaps the maximizer ψ_0 of $E_\delta\left[\text{IMSE}(\delta|\psi)\right]$; this turns out to be computationally infeasible in all but the simplest scenarios. And see §3.1, where we argue that a contaminant less favourable than ψ_Φ is difficult to imagine. See also Figure 4(d)-(f).

3. Jittering

There are obvious issues in implementing an absolutely continuous design measure within this framework, since any discrete approximation necessarily suffers from the drawback, as above, that the maximum loss is infinite. Noting that in this case the least favourable contaminating function ψ is largely concentrated on a set of measure zero – an unlikely eventuality against which to seek protection – Wiens (1992, p. 355) states that "Our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse." He places one observation at each of the quantiles

$$t_i = \xi^{-1}\left(\frac{i-1/2}{n}\right),\ i = 1, ..., n, \tag{10}$$

which is the n-point design closest to ξ in Kolmogorov distance (Fang and Wang 1994; see Xu and Yuen 2011 for other possibilities).

Despite the disclaimer above, such discrete implementations have become controversial; see in particular Bischoff (2010). In this article we investigate a resolution to these difficulties offered by Waite and Woods (2022), who propose randomly sampling the design points from uniform densities highly concentrated in small neighbourhoods of an optimally chosen set of deterministic points. In our case we propose random sampling from a piecewise uniform density

$$\phi_n(x; c) = \frac{1}{2c}\sum_{i=1}^{n} I\left[t_i - \frac{c}{n} \le x \le t_i + \frac{c}{n}\right], \tag{11}$$

for chosen $c \in (0, 1)$.

We illustrate the method in the context of straight line regression – $\mathcal{X} = [-1, 1]$ and $\boldsymbol{f}(x) = (1, x)'$ – for which Huber (1975) obtained the minimax density

$$m(x) = 3\left(x^2 - \alpha\right)^+ / d(\alpha),$$

with α chosen to minimize (5), which in terms of

$$\mu_2(\alpha) = \int_{-1}^{1} x^2 m(x)\, dx,\ \kappa_0(\alpha) = \int_{-1}^{1} m^2(x)\, dx,\ \kappa_2(\alpha) = \int_{-1}^{1} x^2 m^2(x)\, dx,$$

is

$$K_\nu(\alpha) = 2(1-\nu)\left(1 + \frac{1}{3\mu_2}\right) + 2\nu \max\left(\kappa_0, \frac{\kappa_2}{3\mu_2^2}\right).$$

Apart from minor modifications resulting from the change in the support to $[-1, 1]$ from $[-1/2, 1/2]$, the details of the construction of m are as in Huber (1975). We assume that $\max\left(\kappa_0, \kappa_2/3\mu_2^2\right) = \kappa_0$ and check this once m is obtained. We find

$$d(\alpha) = \begin{cases} 2(1-3\alpha), & \alpha \le 0, \\ 2\left(1-\sqrt{\alpha}\right)^2\left(1+2\sqrt{\alpha}\right), & \alpha \ge 0, \end{cases}$$

with α and ν related by

$$\nu^{-1} = \begin{cases} 1 + \frac{9(3-5\alpha)^2}{25(1-3\alpha)^3}, & \alpha \le 0, \\ 1 + \frac{9\left(3+6\sqrt{\alpha}+4\alpha+2\alpha^{3/2}\right)^2}{25\left(1-\sqrt{\alpha}\right)^2\left(1+2\sqrt{\alpha}\right)^3}, & \alpha \ge 0. \end{cases}$$

The limiting cases are (i) $\alpha \to -\infty$, $\nu \to 1$, $m(x) \to .5$ (the uniform density), (ii) $\alpha = 0$, $\nu = 25/106$, $m(x) = 3x^2/2$, and (iii) $\alpha \to \infty$, $\nu \to 0$, $m(x) \to$ point masses of $1/2$ at ± 1.

It is a fortuitous consequence of the choice of *imse* as loss that for all $\nu \in [0, 1]$, $\max\left(\kappa_0, \kappa_2/3\mu_2^2\right) = \kappa_0$, the choice used in the derivation of the minimizing density m. For other common choices – D-, A- and E-optimality for instance – the situation is far more complicated. See Daemi and Wiens (2013).

3.1. Jittered designs for SLR

In the construction of the sampling density (11) for this example we will take $\alpha \le 0$ – the case of most interest from a robustness standpoint – and then for m as above, the symmetrically placed points t_i are determined by

$$t_i^3 - 3\alpha t_i = (1-3\alpha)\left(\frac{2i-1-n}{n}\right),\ i = 1, ..., n.$$

This equation has an explicit solution furnished by Cardano's formula:

$$t_i = \left(-s/2 + \sqrt{\Delta}\right)^{1/3} + \left(-s/2 - \sqrt{\Delta}\right)^{1/3},$$

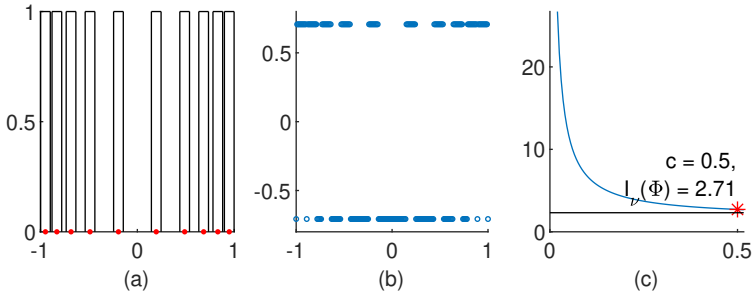

Figure 1: (a) Jittered design density $\phi_n(x;c)$ for an approximately linear univariate response using $\nu = .5$, $c = .5$, $n = 10$. (b) $\sqrt{n}\psi_\Phi(x;c)/\tau$. (c) $I_\nu(\Phi)$ vs. c; horizontal line at $I_\nu(\xi) = 2.31$.

for

$$s = -(1-3\alpha)\left(\frac{2i-1-n}{n}\right),\ \Delta = s^2/4 - \alpha^3 > 0.$$

From (10), and the bowl-shape of $m(x)$, one infers that the distances between adjacent t_i are smallest near ± 1, largest near 0. Thus the intervals of support of ϕ_n will be non-overlapping, and within $[-1,1]$, as long as $t_1 - c/n \geq -1$, i.e. $c \leq n(1+t_1)$. Note that the interpretation of c is that it is the proportion of the design space being randomly sampled.

In Figure 1(a) we plot $\phi_n(x;c)$, when placing equal weight on protection against bias versus variance ($\nu = .5$), 50% of the design space to be sampled from Φ ($c = .5$) and $n = 10$. The t_i are the quantiles arising from $m(x) \propto \left(x^2 + .325\right)$.

A comparison of the maximum loss (5) of ξ versus that of the design measure Φ corresponding to ϕ is obtained from

$$\begin{aligned} I_\nu(\xi) &= 2(1-\nu)\left(1 + \frac{1}{3\mu_2(\alpha)}\right) + \nu\left(1 + \frac{5}{4}(3\mu_2(\alpha)-1)^2\right), \\ I_\nu(\Phi) &= 2(1-\nu)\left(1 + \frac{1}{3\lambda_2(c)}\right) + \frac{\nu}{c}\max\left(1, \frac{1}{3\lambda_2(c)}\right), \end{aligned}$$

where

$$\mu_2(\alpha) = \frac{3-5\alpha}{5(1-3\alpha)},\ \text{and}\ \lambda_2(c) = \int_{-1}^{1} x^2 \phi_n(x;c)\,dx = \frac{1}{n}\sum_{i=1}^{n} t_i^2 + \frac{c^2}{3n^2}.$$

As noted in §2.1, $E_\Phi\left[\text{IMSE}(\delta|\psi_\Phi)\right]$ simplifies considerably for these jittered designs and then $J_\nu(\Phi)$ is very similar to $I_\nu(\Phi)$, plotted in Figure 1(c). We show in the Appendix that in this case (9) becomes

$$J_\nu(\Phi) = (1-\nu)E_\Phi\left[tr\boldsymbol{A}\boldsymbol{M}_\delta^{-1}\right] + \nu\gamma_0, \text{where } \gamma_0 = \frac{1}{c}\max\left(1, \frac{1}{3\lambda_2(c)}\right), \tag{12}$$

and that, with $I_S(x;c) = I(\phi_n(x;c) > 0)$, the least favourable contaminant for Φ is

$$\psi_\Phi(x;c) = \frac{\tau}{\sqrt{n}}\left(\frac{I_S(x;c) - \frac{1}{\gamma_0}}{\sqrt{2c\left(1-\frac{1}{\gamma_0}\right)}}\right) \cdot \left(\frac{x}{\sqrt{\lambda_2}}\right)^{I(\lambda_2(c)<1/3)}. \tag{13}$$

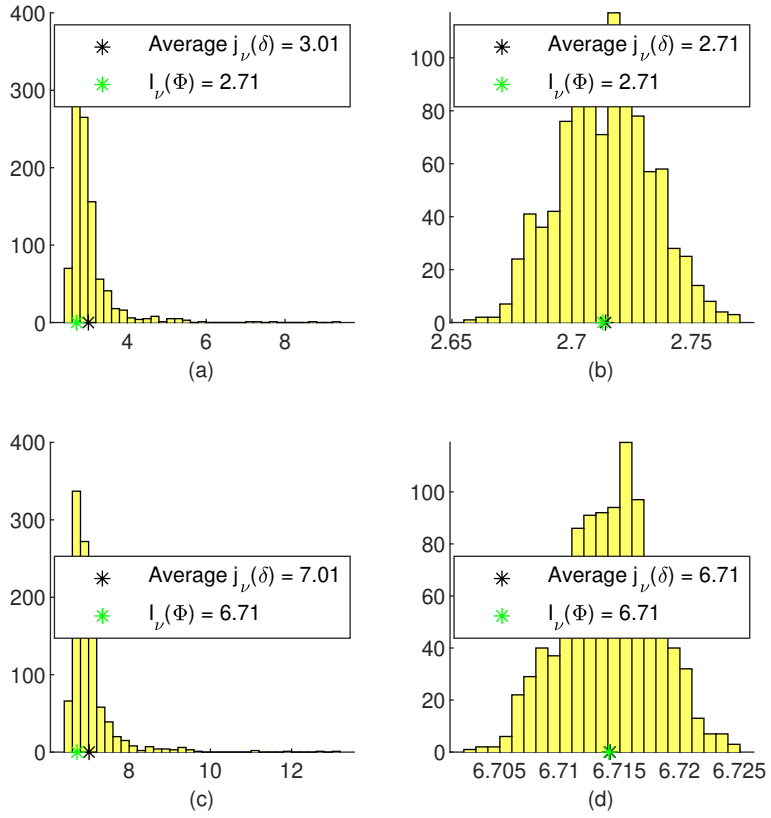

Figure 2: Values of $j_\nu(\delta)$ and their averages, estimating $J_\nu(\Phi)$, from 1000 jittered designs δ using $\nu = .5$. Top: $c = .5$; bottom: $c = .1$. (a), (c): Completely random sampling. (b), (d): Stratified random sampling.

In Figure 1(b) we plot a scaled version of $\psi_\Phi(x; c)$. The contaminant ψ_Φ has the effect of changing the uncontaminated response $E[Y(x)] = \theta_0 + \theta_1 x$ to $(\theta_0 - k) + \theta_1 x + 2kI(\phi_n(x; c) > 0)$ for (when $c = .5$) $k = \tau/\sqrt{2n}$. Thus it biases the intercept and then places contamination uniformly on the support of ϕ_n. In the parlance of game theory, it is difficult to see how Nature, knowing Φ but not δ and assumed malevolent, could respond less favourably than this.

For ease in the estimation of $J_\nu(\Phi)$ we note that $E_\Phi\left[tr\boldsymbol{A}\boldsymbol{M}_\delta^{-1}\right] = 2\left\{1 + E_\Phi\left[\left(\mu_\delta^2 + 1/3\right)/\sigma_\delta^2\right]\right\}$, where μ_δ and σ_δ^2 are the mean and variance of the design. See Figure 2 for comparative values illustrating the close agreement between $J_\nu(\Phi)$ and $I_\nu(\Phi)$.

The plots reveal that the loss associated with the design Φ decreases with c, i.e. as the design becomes closer to the uniform design on all of χ, for which the bias vanishes. This is in line with the remark of Box and Draper (1959): "The optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize bias alone*."

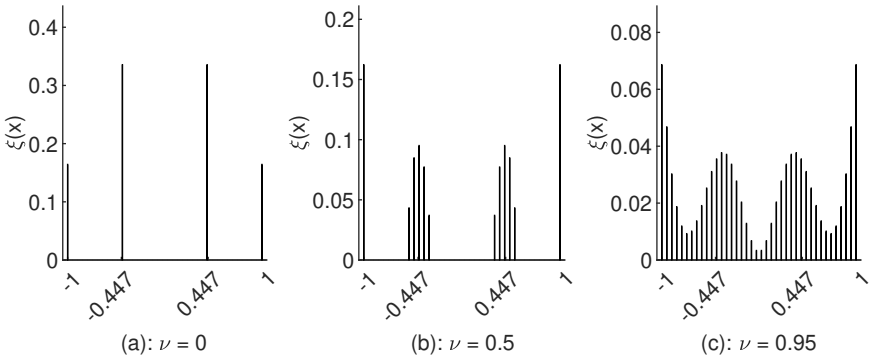

Figure 3: Minimax designs for approximate cubic regression

3.1.1. Sampling methods

We constructed 1000 completely random and stratified random designs, in order to assess their performance. A completely random design δ consisted of $n = 10$ points chosen from $\phi_n(x; c)$. The resulting values of $j_\nu(\delta)$ are plotted in Figure 2(a),(c). Stratification consisted of choosing one design point at random from each bin – Figure 2(b),(d). The sample averages of the losses from the randomized designs were smaller and closer to $I_\nu(\Phi)$ under the stratified sampling scheme, and more concentrated around their expectation of $J_\nu(\Phi)$, as exhibited by the much shorter tail in (b). In a further simulation, for which the output is not displayed here, we estimated $J_\nu(\xi)$, as at (9), by drawing 1000 samples from the minimax density $m(x)$ and averaging their IMSE. The values $\{j_\nu\}$ showed more variation than those plotted in Figure 2(a), and with an average of 2.72 – significantly larger than the target value $I_\nu(\xi) = 2.31$. From this we infer that jittering combined with stratification gives an efficient, structured implementation of the minimax solution.

Simulations using other inputs also resulted in these same conclusions – that our random design strategies typically yield designs very close to optimal with respect to our robustness and efficiency requirements, and that do not suffer from the drawback of deterministic designs of having infinite maximum loss.

4. Cluster designs in one dimension

Working in discrete design spaces, Wiens (2018) obtained minimax robust designs for a variety of approximate responses. Those shown in Figure 3 are for cubic regression. The classically I-optimal design ($\nu = 0$) minimizing integrated variance alone was derived by Studden (1977) and places masses of .1545 and .3455 at ± 1 and $\pm .4472$. The robust designs can thus be described as taking the replicates of the classical design and spreading their mass out ('clustering') over nearby regions. This same phenomenon has frequently been noticed in other situations (Fang and Wiens 2000, Heo et al. 2001 for instance).

In this section we aim to formalize this notion in order to obtain designs competing with the minimax designs, but with finite maximum loss even in continuous design spaces, and having the advantage of being much more easily derived – there is no need for the

minimax designs to be known. We consider only one-dimensional designs in this section, and will illustrate the methods in polynomial response models of degrees $p-1=1,2,3$.

Suppose that a given static design has p support points $t_1<\cdots<t_p$ in $[-1,1]$. Define midpoints $s_i=(t_i+t_{i+1})/2$, $i=1,...,p-1$. Put $s_0=\min(-1,t_1)$ and $s_p=\max\left(1,t_p\right)$. Then the p intervals $I_i=\{[s_{i-1},s_i],\ i=1,...,p\}$ cover $[-1,1]$ and have the properties that $t_i\in I_i$ and that any point in I_i is closer to t_i than to any t_j, $j\neq i$. This is then a trivial example of a *Voronoi tessellation*, to be considered when we pass to higher dimensions.

We propose designs consisting of points sampled from Beta densities on subintervals of the I_i. Specifically, for $i=1,...,p$ let $c=c(\nu)\in[0,1]$ satisfy $c(0)=0$ and $c(1)=1$. Put $J_i(c)=[t_i-c(t_i-s_{i-1}),t_i+c(s_i-t_i)]\equiv[k_i,l_i]$, with length $|J_i|=(l_i-k_i)=c(s_i-s_{i-1})=c\times|I_i|$. Let $\beta_{a,b}(x)$ be the Beta(a,b) density on $[0,1]$. Then

$$\frac{1}{|J_i|}\beta_{a,b}\left(\frac{x-k_i}{|J_i|}\right),\ x\in J_i(c) \tag{14}$$

is this density, translated and scaled to $J_i(c)$. The interpretation of 'c' is as before – it is the fraction of the design space to be sampled. Here and in the following examples we use $c=\nu^k$ where k is the dimension of $\boldsymbol{x}$, so that c varies at the same rate as the volume of χ as the dimensionality changes.

The parameters (a_i,b_i) are chosen so that the mode of (14) is at $t_i\in J_i(c)$, hence the mode $\delta_i\in[0,1]$ of $\beta_{a,b}(x)$ is given by

$$\delta_i\equiv\frac{t_i-k_i}{l_i-k_i}=\begin{cases}\frac{a_i-1}{a_i-1+b_i-1}, & a_i,b_i>1,\\ 0, & a_i\leq 1<b_i,\\ 1, & b_i\leq 1<a_i.\end{cases}$$

Then

$$(a_i-1)(1-\delta_i)=(b_i-1)\delta_i. \tag{15}$$

If $\delta_i\neq 0,1$ we determine one of (a_i,b_i) in terms of the other through (15). We define a_i in terms of b_i for $\delta_i<.5$ and b_i in terms of a_i for $\delta_i>.5$; this ensures that (16) below is symmetric. If $t_1=-1$ then $\delta_1=0$ and we set $a_1=1$. If $t_p=1$ then $\delta_p=1$ and we set $b_p=1$. In each case the remaining parameter is set equal to $1/c$, so that the density tends to a point mass at t_i as $\nu\to 0$ and to uniformity as $\nu\to 1$.

The final density $\phi(\cdot)$ from which the design points are to be sampled is a weighted average of those at (14), with weights proportional to the lengths $|I_i|$ of the I_i. Since $|J_i|=c|I_i|$ we obtain

$$\phi(x;\nu)=\frac{1}{2c}\sum_{i=1}^{p}\beta_{a_i,b_i}\left(\frac{x-k_i}{|J_i|}\right)I(x\in J_i). \tag{16}$$

Motivated by the designs of §3.1 we recommend stratified sampling, by which the sample consists of $\approx n|I_i|/2$ points drawn from (14), subject to an appropriate rounding procedure.

Table 1. Performance measures for the designs of Figure 4.

	variance		max sqd. bias		I_ν	
	$\nu = .5$	$\nu = .04$	$\nu = .5$	$\nu = .04$	$\nu = .5$	$\nu = .04$
$p = 2$	2.94	2.67	2.67	319	2.80	15.3
$p = 3$	4.65	4.27	2.62	213	3.64	12.6
$p = 4$	6.49	6.02	2.54	193	4.51	13.5

4.1. Polynomial regression

We illustrate these proposals in the context of approximate polynomial responses of degrees $p - 1 = 1, 2, 3$. As also suggested in 'Heuristic 5.1' of Waite and Woods (2022, p. 1462), t^* will consist of the support points of the classical I-optimal designs. These I-optimal designs $\xi^*$are obtained from Lemma 3.2 of Studden (1977), and are as follows.

$p = 2$: $\xi^*(\pm 1) = .5$,
$p = 3$: $\xi^*(\pm 1) = .25, \xi^*(0) = .5$,
$p = 4$: $\xi^*(\pm 1) = \frac{1}{2\left(1+\sqrt{5}\right)} \approx .1545, \xi^*\left(\pm\frac{1}{\sqrt{5}} \approx \pm.4472\right) = \frac{\sqrt{5}}{2\left(1+\sqrt{5}\right)} \approx .3455.$

Figure 4 gives the sampling densities (16), together with the subsample sizes when $n = 10$. Figure 4(a) gives output for the approximate linear model, with a maximum IMSE, as at (5), of $I_\nu(\Phi) = 2.804$. This compares very favourably with the design of Figure 1, especially given that its construction does not require the minimax design to be given. This latter point is especially germane for the design of Figure 4(b), since it is the analogue of the absolutely continuous minimax designs for approximate quadratic regression derived – with substantial theoretical and computational difficulty – by Shi et al. (2003) using methods of non-smooth optimization and by Daemi and Wiens (2013) using completely different methods.

Figure 5 gives values of $j_\nu(\delta)$ from 1000 random designs, together with their average, estimating $J_\nu(\Phi)$. On average the random designs perform almost as well against ψ_Φ – plotted in Figure 4(d)-(f) – as the continuous design Φ.

It is interesting to note – especially for the design of Figure 4(c) – the close agreement between the I-optimal design weights above, and the weights used in the computation of ϕ and detailed in the caption.

See Table 1, where the variance and maximum squared bias components of I_ν are presented for the designs of Figure 4 ($\nu = .5$) and for the corresponding designs with $\nu = .04$, very closely approximating the I-optimal design ($\nu = 0$) with maximum loss $I_0 = \infty$. That the robustness of the cluster designs is achieved for such a modest premium in terms of increased variance is both startling and encouraging.

5. Multidimensional cluster designs

See Figure 6, where a robust design, derived for fitting a full second order bivariate model – intercept, linear, quadratic and interaction terms – is depicted. It is a discrete

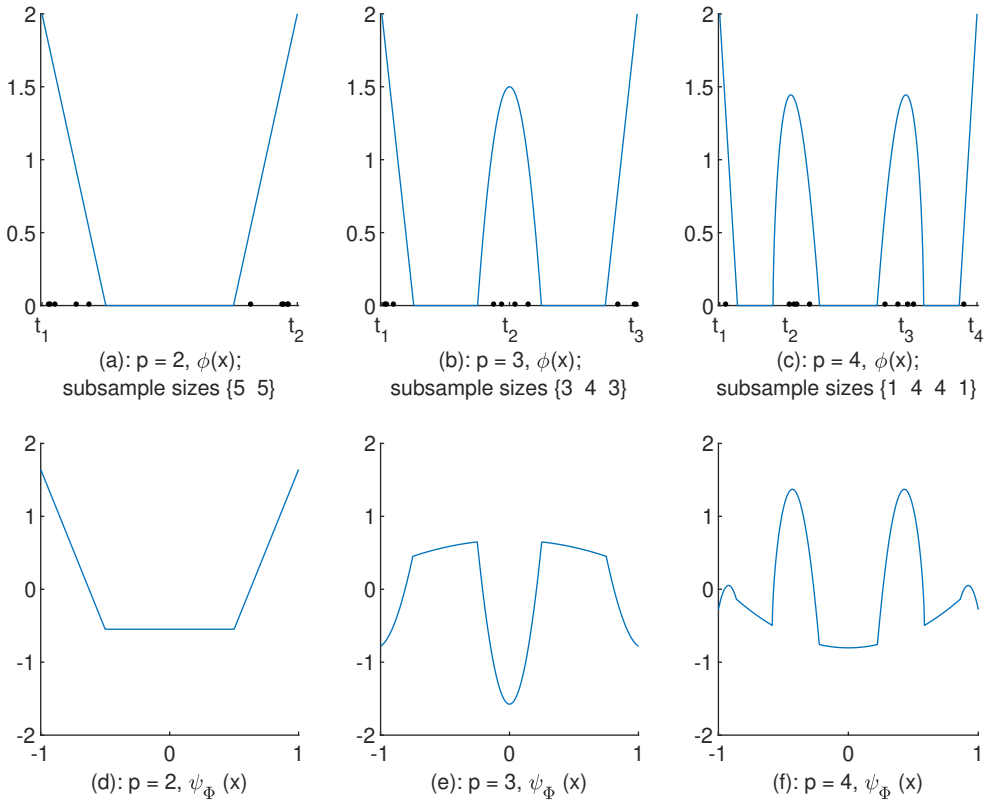

Figure 4: Top: Cluster design densities $\phi_n(x; \nu = .5)$; typical stratified samples using weights (a) {.5, .5}, (b) {.25, .5, .25}, (c) {.14, .36, .36, .14}. Bottom: Scaled least favourable contaminants $\sqrt{n}\psi_\Phi(x)/\tau$.

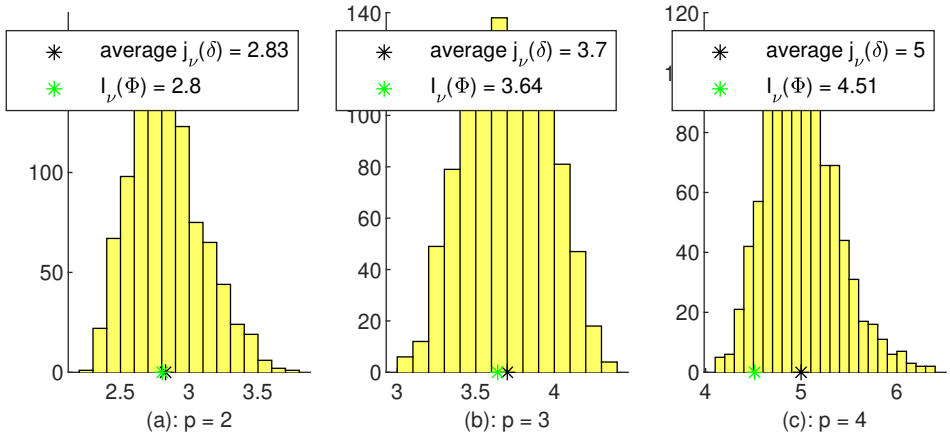

Figure 5: Values of $j_\nu(\delta)$ from 1000 cluster designs for polynomial regression ($\nu = .5$, $n = 10$) and their averages, estimating $J_\nu(\Phi)$.

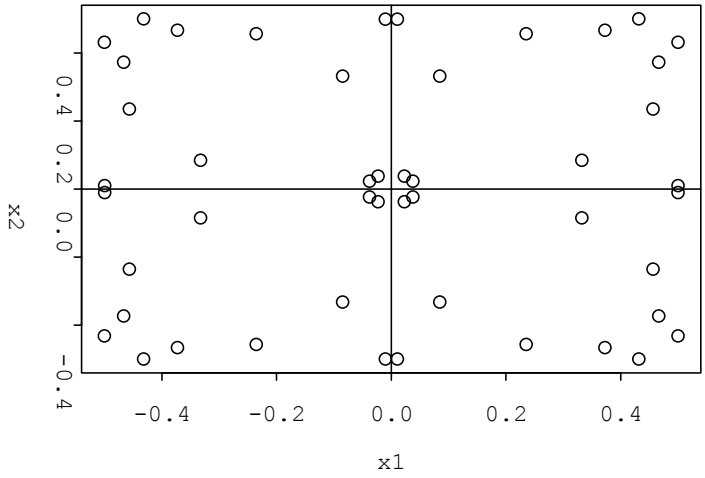

Figure 6: Design for fitting a full second order model; $n = 48$.

implementation of a design density, optimally robust against model misspecifications in a certain parametric class of densities - see Heo et al. (2001) for details. This design can roughly be described as an inscribed Central Composite Design (CCD) with 'clustering' in place of replication. It serves as motivation for the ideas of this section, which we illustrate in the context of k-dimensional, spherical CCD designs as are often used to fit second order models. Such designs utilize $2^k + 2k + 1$ points $\{\boldsymbol{t}_i\}$ consisting of 2^k corner points with $\boldsymbol{t}_{i,j} = \pm 1$ $(j = 1, ..., k)$, $2k$ axial points $\boldsymbol{t}_i = \left(0, ..., \pm\sqrt{k}, ..., 0\right)$ and a centre point $\boldsymbol{t}_i = (0, ..., 0, ..., 0)$.

In this and other multidimensional cases we propose choosing design points from spherical densities concentrated on neighbourhoods of the $\boldsymbol{t}_i$. A spherical density on a k-dimensional hypersphere

$$\mathcal{S}^{(k)}(\boldsymbol{t}, R) = \{\boldsymbol{x} \,|\, ||\boldsymbol{x} - \boldsymbol{t}|| \leq R\}$$

with centre $\boldsymbol{t}$ and radius R, in which the scaled norm $||\boldsymbol{x} - \boldsymbol{t}|| / R$ has a $Beta(k, b)$ density, is given by

$$f^{(k)}(\boldsymbol{x}; \boldsymbol{t}, R, b) = \frac{\Gamma\left(\frac{k}{2}\right)}{2\pi^{k/2} R^k \beta(k, b)} \cdot \left(1 - \frac{||\boldsymbol{x} - \boldsymbol{t}||}{R}\right)^{b-1} I\left(\boldsymbol{x} \in \mathcal{S}^{(k)}(\boldsymbol{t}, R)\right).$$

Such a density has mode $\boldsymbol{t}$ and approaches a point mass at $\boldsymbol{t}$ as $b \to \infty$, and uniformity as $b \to 1$. The choice of k as the first parameter of the beta density ensures that f is decreasing in $||\boldsymbol{x} - \boldsymbol{t}||$ and square integrable (required for the evaluation of the matrix $\boldsymbol{K}_\Phi$ as at (5)).

A sample value $\boldsymbol{x}$ from $f^{(k)}(\boldsymbol{x}; \boldsymbol{t}, R, b)$ is $\boldsymbol{x} = \boldsymbol{t} + R\boldsymbol{y}$, where $\boldsymbol{y} \sim f^{(k)}(\cdot; \boldsymbol{0}, 1, b)$ obtained by drawing a value of $\rho = ||\boldsymbol{y}|| \sim Beta(k, b)$ and, independently, drawing angles θ_i, $-\pi/2 < \theta_i \leq \pi/2$ $(i = 1, ..., k-2)$ with densities $\psi_i(\theta) = \cos^{k-i-1}\theta / \beta(1/2, (k-i)/2)$ – equivalently,

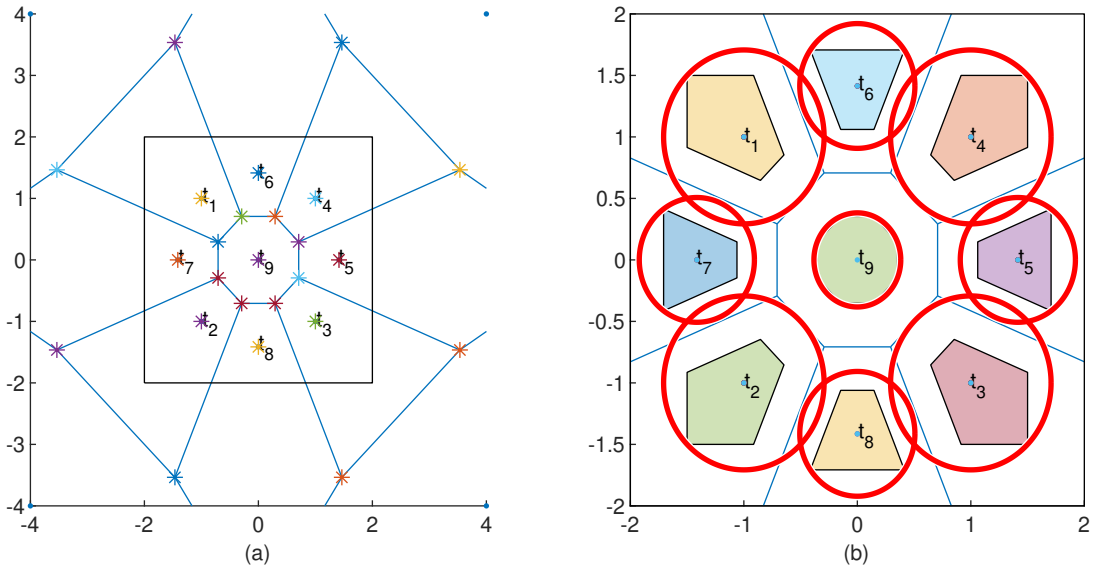

Figure 7: (a) Voronoi tessellation generated by the points $\{\boldsymbol{t}_i\}$. (b) Tessellation restricted to $\chi = [-2,2]^2$ with subtiles $\{J_i(.25)\}$ and enclosing circles $\left\{\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i)\right\}$.

$\cos^2\theta_i \sim Beta(\frac{1}{2}, \frac{k-i}{2})$ – and $\theta_{k-1} \sim Unif(-\pi, \pi)$. Then

$$\begin{aligned} y_1 &= \rho \sin\theta_1, \\ y_2 &= \rho \cos\theta_1 \sin\theta_2, \\ y_3 &= \rho \cos\theta_1 \cos\theta_2 \sin\theta_3, \\ &\cdots \\ y_{k-1} &= \rho \cos\theta_1 \cos\theta_2 \cdots \cos\theta_{k-2} \sin\theta_{k-1}, \\ y_k &= \rho \cos\theta_1 \cos\theta_2 \cdots \cos\theta_{k-2} \cos\theta_{k-1}. \end{aligned}$$

To sample θ_i for $i < k-1$ we draw $z \sim Beta(\frac{1}{2}, \frac{k-i}{2})$ and set $\theta_i = \pm \arccos\sqrt{z}$, each with probability $1/2$.

5.1. Two dimensional cluster designs on tessellations

In Figure 7(a), the nine points $\{\boldsymbol{t}_i\}$ which are displayed consist of four corner points $(-1, \pm 1)$, $(1, \pm 1)$, four axial points $\left(\pm\sqrt{2}, 0\right)$, $\left(0, \pm\sqrt{2}\right)$ and the centre point $(0,0)$. These are the generators of the *Voronoi tessellation* pictured - a tiling with the property that, within the tile T_i containing $\boldsymbol{t}_i$, all points are closer to $\boldsymbol{t}_i$ than to any $\boldsymbol{t}_j$, $j \neq i$. Figure 7(b) gives a more detailed depiction of the tessellation, restricted to the design space $\chi = [-2,2] \times [-2,2]$. Within each tile T_i, of area $|T_i|$, we have also plotted a subtile $J_i(c)$ which is a contraction of T_i with fixed point $\boldsymbol{t}_i$ and area $|J_i(c)| = c\,|T_i|$. These are then the analogues of the subintervals $J_i(c) \subseteq I_i$ from §4, and 'c' has the same interpretation – the fraction of the design space to be sampled. Surrounding each $J_i(c)$ is the smallest enclosing circle $\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i(c))$.

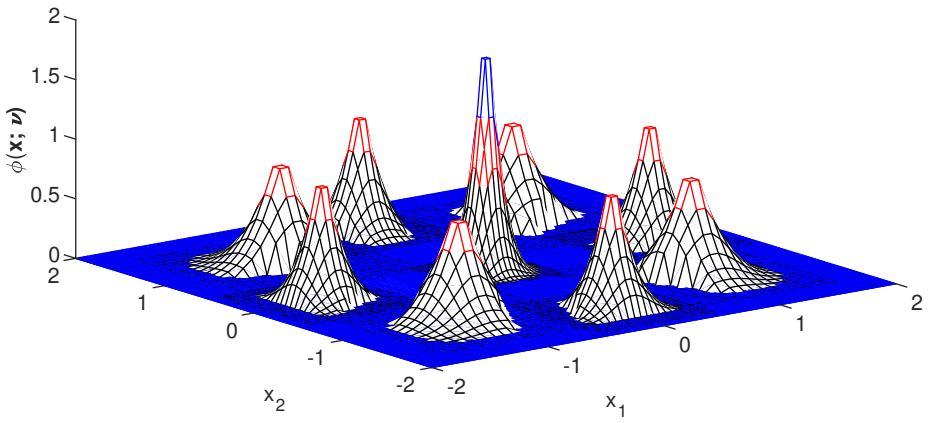

Figure 8: Sampling density $\phi(\boldsymbol{x}; \nu = .5)$ constructed for a robust, clustered CCD in two dimensions.

We sample design points from $\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i(c))$, accepting only those points which lie in $J_i(c)$. We specify $b = 1/c$ and $c = \nu^2$, then $f^{(2)}(\boldsymbol{x}; \boldsymbol{t}_i, R_i(c), b)$ approaches a point mass at $\boldsymbol{t}_i$ as $\nu \to 0$, and uniformity on $\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i(c)) \supseteq T_i$ as $\nu \to 1$. With

$$q_i(\nu) = \int_{J_i(c)} f^{(2)}(\boldsymbol{x}; \boldsymbol{t}_i, R_i(c), b)\, \mu(d\boldsymbol{x}),$$

the density of those points accepted into the design upon being drawn from $\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i(c))$ is

$$\frac{f^{(2)}(\boldsymbol{x}; \boldsymbol{t}_i, R_i(c), b)}{q_i(\nu)} I(\boldsymbol{x} \in J_i(c)).$$

We again do stratified sampling, with weights $\omega_i = |T_i| / \sum |T_i|$ proportional to the area $|T_i|$, whence the density of the design on χ is

$$\phi(\boldsymbol{x}; \nu) = \sum_{i=1}^{9} \frac{\omega_i}{q_i(\nu)} f^{(2)}(\boldsymbol{x}; \boldsymbol{t}_i, R_i(c), b)\, I(\boldsymbol{x} \in J_i(c)).$$

See Figure 8. Although we evaluate $q_i(\nu)$ by numerical integration, an estimate can be computed after the sampling is done; it is the proportion of those points which were drawn from $\mathcal{S}^{(2)}(\boldsymbol{t}_i, R_i(c))$ and then accepted into the sample. This estimate turns out to be quite accurate if an artificially large sample is simulated.

Figure 9 illustrates the results of applying the methods of the preceding discussion. We chose a total sample size of $n = 50$, $\nu = .5$, and obtained subsample sizes $n_i = n\omega_i$, rounded to $\{7, 7, 7, 7, 5, 5, 5, 5, 2\}$ with each corner point being allocated 7, each axial point being allocated 5, and the remaining 2 in the centre. The entire sample is shown in Figure 9(a), with Figure 9(b) illustrating the details for Tile 8. The required 5 points were found after 7 points were drawn from $\mathcal{S}_8$. In all, 13 points were rejected as not belonging to the appropriate subtile.

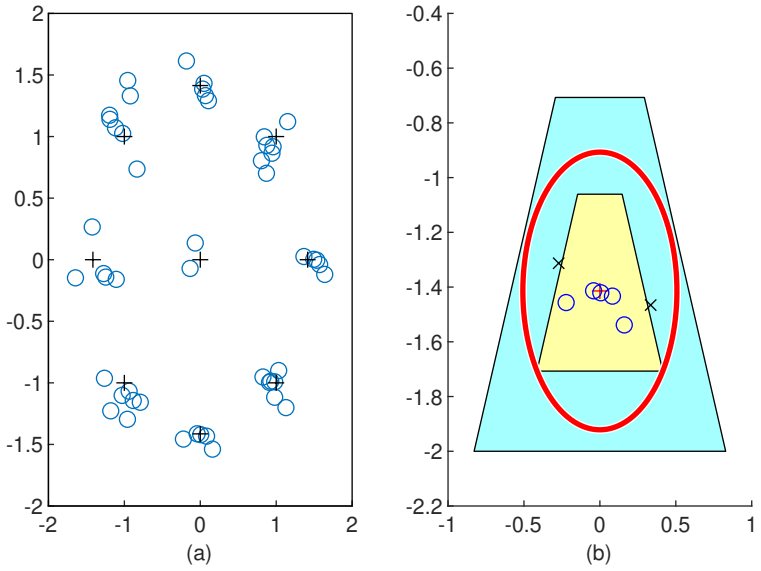

Figure 9: (a) A typical random CCD of size 50; $\nu = .5$. (b) Details of the subsample of 5 points ('o') drawn from $J_8(.25)$. Rejected points are marked as 'x'.

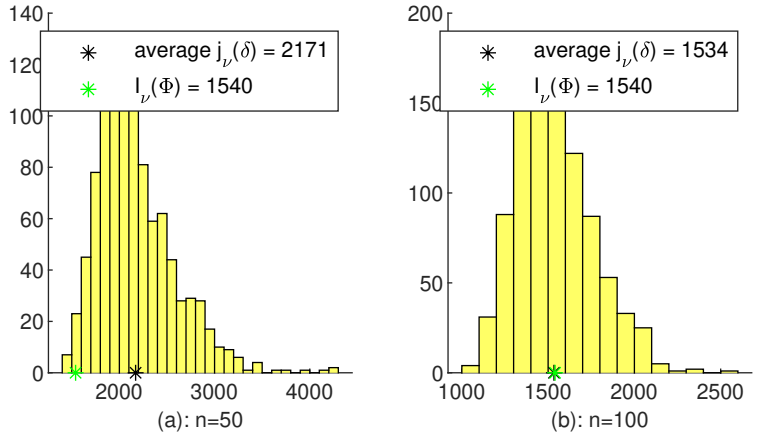

Figure 10: Values of $j_\nu(\delta)$ from 1000 stratified, clustered CCDs ($\nu = .5$) and their averages, estimating $J_\nu(\Phi)$. (a) $n = 50$, (b) $n = 100$.

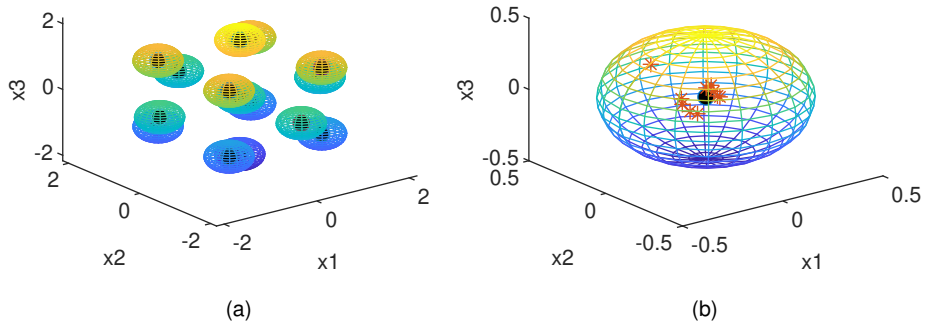

Figure 11: (a) Spheres $\mathcal{S}^{(3)}(\boldsymbol{t}_i, r_0\nu)$, $i = 1, ..., 15$ for a three dimensional spherical CCD and $\nu = .5$. (b) Centre sphere $\mathcal{S}^{(3)}(\mathbf{0}, r_0\nu)$ with sampled points.

We repeated this with a total sample size of $n = 100$. See Figure 10. With the larger sample size the random designs seem to more accurately duplicate the behaviour of the parent design Φ – a phenomenon noticed as well in the one dimensional cluster designs of the previous section.

5.2. Extensions to $k > 2$

Although the theory of §5.1 extends easily to higher dimensions, the lack of appropriate software for constructing and manipulating Voronoi tessellations becomes a severe drawback. But the general idea of sampling from spherical distributions centred on small neighbourhoods of the $\{\boldsymbol{t}_i\}$ can still be applied, albeit in a less structured manner. Let $\{\boldsymbol{t}_i\}_{i=1}^q$ be the $q = 2^k + 2k + 1$ support points of a spherical CCD in variables $\boldsymbol{x} = (x_1, ..., x_k)'$, as described at the beginning of this section. The minimum distance between these points is $\min\left(2, \sqrt{k}\right)$, and so hyperspheres $\mathcal{S}(\boldsymbol{t}_i, r_0)$ centred at the $\boldsymbol{t}_i$ and with radius $r_0 = \min\left(1, \sqrt{k}/2\right)$ are disjoint. Define subspheres

$$J_i(c) = \mathcal{S}^{(k)}\left(\boldsymbol{t}_i, r_0 c^{1/k}\right), 0 < c \leq 1.$$

Then $\int I(\boldsymbol{x} \in J_i(c))\, d\boldsymbol{x} = |J_i(c)| = c\,|\mathcal{S}(\boldsymbol{t}_i, r_0)|$. The density of $\boldsymbol{x}$ on $J_i(c)$ is $f^{(k)}\left(\cdot; \boldsymbol{t}_i, r_0 c^{1/k}, b\right)$. We again specify $b = 1/c$ and $c = \nu^k$. Then for user chosen weights $\{\omega_i\}$ the sampling density is

$$\phi(\boldsymbol{x}; \nu) = \sum_{i=1}^{q} \omega_i f^{(k)}(\boldsymbol{x}; \boldsymbol{t}_i, r_0\nu, 1/\nu)\, I\left(\boldsymbol{x} \in \mathcal{S}^{(k)}(\boldsymbol{t}_i, r_0\nu)\right).$$

See Figure 11 for an example with $k = 3$. We sampled a design of size $n = 80$ with subsamples sizes $n_i = 5$ $(i < 15)$ and $n_{15} = 10$.

Appendix

A.1. Derivations for §2.1

For an n-point design $D = \{\boldsymbol{x}_i\}_{i=1}^n$ with design measure $\delta = n^{-1}\sum \delta_{\boldsymbol{x}_i}$ define

$$\boldsymbol{F} = (\boldsymbol{f}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}(\boldsymbol{x}_n))', \boldsymbol{\psi}_\Phi = (\psi_\Phi(\boldsymbol{x}_1), ..., \psi_\Phi(\boldsymbol{x}_n))', \ \boldsymbol{D}_\phi = diag(\phi(\boldsymbol{x}_1), ..., \phi(\boldsymbol{x}_n)).$$

Then $\boldsymbol{M}_\phi = \frac{1}{n}\boldsymbol{F}'\boldsymbol{D}_\phi\boldsymbol{F}$. Define as well

$$\begin{aligned}\boldsymbol{M}_\delta &= \int_\chi \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\delta(d\boldsymbol{x}) = \frac{1}{n}\sum_{\boldsymbol{x}_i\in D}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i) = \frac{1}{n}\boldsymbol{F}'\boldsymbol{F},\\ \boldsymbol{b}_{\psi_\Phi,\delta} &= \int_\chi \boldsymbol{f}(\boldsymbol{x})\psi_\Phi(\boldsymbol{x})\delta(d\boldsymbol{x}) = \frac{1}{n}\sum_{\boldsymbol{x}_i\in D}\boldsymbol{f}(\boldsymbol{x}_i)\psi_\Phi(\boldsymbol{x}_i) = \frac{1}{n}\boldsymbol{F}'\boldsymbol{\psi}_\Phi.\end{aligned}$$

Using (7),

$$\psi_\Phi(\boldsymbol{x}_i) = \mathbf{r}_\Phi'(\boldsymbol{x}_i)\beta_\Phi = \frac{\tau}{\sqrt{n}}\left(\phi(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i) - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{A}^{-1}\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi, \tag{A.1}$$

so that

$$\boldsymbol{\psi}_\Phi = \frac{\tau}{\sqrt{n}}\left(\boldsymbol{D}_\phi\boldsymbol{F} - \boldsymbol{F}\boldsymbol{A}^{-1}\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi, \tag{A.2}$$

$$\boldsymbol{b}_{\psi_\Phi,\delta} = \frac{\tau}{\sqrt{n}}\left(\boldsymbol{M}_\phi - \boldsymbol{M}_\delta\boldsymbol{A}^{-1}\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi. \tag{A.3}$$

From (4),

$$\text{IMSE}(\delta|\psi_\Phi) = \frac{\sigma_\varepsilon^2}{n}tr\left(\boldsymbol{A}\boldsymbol{M}_\delta^{-1}\right) + \boldsymbol{b}'_{\psi_\Phi,\delta}\boldsymbol{M}_\delta^{-1}\boldsymbol{A}\boldsymbol{M}_\delta^{-1}\boldsymbol{b}_{\psi_\Phi,\delta} + \int_\chi \psi_\Phi^2(x)\mu(dx);$$

substituting (A.2) and (A.3) gives

$$\begin{aligned}\text{IMSE}(\delta|\psi_\Phi) &= \frac{\sigma_\varepsilon^2}{n}tr\left(\boldsymbol{A}\boldsymbol{M}_\delta^{-1}\right) + \frac{\tau^2}{n}\gamma_\delta, \text{ for}\\ \gamma_\delta &= \beta_\Phi'\boldsymbol{G}_\Phi^{-1/2}\left(\boldsymbol{M}_\phi\boldsymbol{M}_\delta^{-1} - \boldsymbol{M}_\Phi\boldsymbol{A}^{-1}\right)\boldsymbol{A}\left(\boldsymbol{M}_\delta^{-1}\boldsymbol{M}_\phi - \boldsymbol{A}^{-1}\ \boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi + 1.\end{aligned}$$

Now (8) and (9) are immediate.

A.2. Derivations of (12) and (13)

To evaluate (9) and establish (12) we first note that since $\phi(x_i) \equiv (2c)^{-1}$ on its support, we have that $\boldsymbol{M}_\phi\boldsymbol{M}_\delta^{-1} = (2c)^{-1}\boldsymbol{I}_2$, and then (since $\beta_\Phi'\beta_\Phi = 1$)

$$\begin{aligned}\gamma_\delta &= \beta_\Phi'\boldsymbol{G}_\Phi^{-1/2}\left(\frac{1}{2c}\boldsymbol{I}_2 - \boldsymbol{M}_\Phi\boldsymbol{A}^{-1}\right)\boldsymbol{A}\left(\frac{1}{2c}\boldsymbol{I}_2 - \boldsymbol{A}^{-1}\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi + 1\\ &= \beta_\Phi'\boldsymbol{G}_\Phi^{-1/2}\left(\frac{1}{4c^2}\boldsymbol{A} - \frac{1}{c}\boldsymbol{M}_\Phi + \boldsymbol{M}_\Phi\boldsymbol{A}^{-1}\boldsymbol{M}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi + 1\\ &= \beta_\Phi'\boldsymbol{G}_\Phi^{-1/2}\left(\frac{1}{4c^2}\boldsymbol{A} - \frac{1}{c}\boldsymbol{M}_\Phi + \boldsymbol{H}_\Phi + \boldsymbol{G}_\Phi\right)\boldsymbol{G}_\Phi^{-1/2}\beta_\Phi.\end{aligned} \tag{A.4}$$

A calculation gives

$$\boldsymbol{G}_{\Phi}^{1/2}\boldsymbol{H}_{\Phi}^{-1}\boldsymbol{G}_{\Phi}^{1/2}+\boldsymbol{I}_2=\frac{1}{c}diag\left(1,\frac{1}{3\lambda_2(c)}\right), \tag{A.5}$$

so that the maximum eigenvalue is γ_0 and

$$\beta_{\Phi}=\begin{cases}(1,0)', & \text{if } \lambda_2(c)\geq 1/3,\\ (0,1)', & \text{if } \lambda_2(c)<1/3.\end{cases}$$

The choice of β_{Φ} is somewhat arbitrary if $\lambda_2(c)=1/3$, since then (A.5) is a multiple of $\boldsymbol{I}_2$. We claim that

$$\gamma_{\delta}\equiv\gamma_0, \tag{A.6}$$

from which (12) follows, since then γ_{δ} does not depend on the design and so is non-random.

To establish (A.6), use $\boldsymbol{A}=\boldsymbol{M}_{\Phi}\boldsymbol{H}_{\Phi}^{-1}\boldsymbol{M}_{\Phi}=4c^2\boldsymbol{K}_{\Phi}\boldsymbol{H}_{\Phi}^{-1}\boldsymbol{K}_{\Phi}$ and $\boldsymbol{M}_{\Phi}=2c\boldsymbol{K}_{\Phi}$ in (A.4) to obtain

$$\gamma_{\delta}=\beta_{\Phi}'\left[\boldsymbol{G}_{\Phi}^{-1/2}\left(\boldsymbol{K}_{\Phi}\boldsymbol{H}_{\Phi}^{-1}\boldsymbol{K}_{\Phi}-\boldsymbol{K}_{\Phi}\right)\boldsymbol{G}_{\Phi}^{-1/2}\right]\beta_{\Phi}.$$

Substituting $\boldsymbol{K}_{\Phi}=\boldsymbol{G}_{\Phi}+\boldsymbol{H}_{\Phi}$, this becomes

$$\gamma_{\delta}=\beta_{\Phi}'\left[\boldsymbol{G}_{\Phi}^{1/2}\boldsymbol{H}_{\Phi}^{-1}\boldsymbol{G}_{\Phi}^{1/2}+\boldsymbol{I}_2\right]\beta_{\Phi}=\gamma_0,$$

as required.

An evaluation of (A.1), using

$$\begin{aligned}\boldsymbol{A} &= diag(2,2/3),\ \boldsymbol{M}_{\Phi}=diag(1,\lambda_2(c)),\ \text{and}\\ \boldsymbol{G}_{\Phi} &= \frac{1}{2c}diag\left((1-c),\lambda_2(c)(1-3c\lambda_2(c))\right),\end{aligned}$$

gives

$$\psi_{\Phi}(x;c)=\frac{\tau}{\sqrt{n}}\left(\frac{2c\phi_n(x;c)-c}{\sqrt{2c(1-c)}},x\frac{2c\phi_n(x;c)-3c\lambda_2(c)}{\sqrt{2c\lambda_2(c)(1-3c\lambda_2(c))}}\right)\beta_{\Phi}.$$

Using $2c\phi_n(x;c)=I_S(x;c)$, and

$$\frac{c}{\gamma_0}=\begin{cases}c, & \text{if } \lambda_2(c)\geq 1/3,\\ 3c\lambda_2(c) & \text{if } \lambda_2(c)<1/3,\end{cases}$$

this becomes

$$\psi_{\Phi}(x;c)=\frac{\tau}{\sqrt{n}}\cdot\begin{cases}\frac{I_S(x;c)-\frac{1}{\gamma_0}}{\sqrt{2c\left(1-\frac{1}{\gamma_0}\right)}}, & \text{if } \lambda_2(c)\geq 1/3,\\ \frac{x}{\sqrt{\lambda_2}}\frac{I_S(x;c)-\frac{1}{\gamma_0}}{\sqrt{2c\left(1-\frac{1}{\gamma_0}\right)}} & \text{if } \lambda_2(c)<1/3,\end{cases},$$

which is (13).

6. Acknowledgements

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada. It has benefited from conversations with Timothy Waite, University of Manchester and Xiaojian Xu, Brock University. We are grateful for the incisive and helpful comments of the reviewers.

References

Atkinson, A. C. (1996), "The Usefulness of Optimum Experimental Designs," *Journal of the Royal Statistical Society (Series B)*; 58, 59-76.

Bischoff, W. (2010), "An Improvement in the Lack-of-Fit Optimality of the (Absolutely) Continuous Uniform Design in Respect of Exact Designs," in *mODa 9 - Advances in Model-Oriented Design and Analysis*, eds. Giovagnoli, G., Atkinson, A. and Torsney, B.

Box, G.E.P., and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-654.

Daemi, M., and Wiens, D.P. (2013), "Techniques for the Construction of Robust Regression Designs," *The Canadian Journal of Statistics*, 41, 679-695.

Fang, K.T. & Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman and Hall, London and New York.

Fang, Z., and Wiens, D.P. (2000), "Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Heo, G., Schmuland, B., and Wiens, D.P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

Huber, P.J. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.

Shi, P., Ye, J., and Zhou, J. (2003), "Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 31, 397-414.

Studden, W.J. (1977), "Optimal Designs for Integrated Variance in Polynomial Regression," *Statistical Decision Theory and Related Topics II*, ed. Gupta, S.S. and Moore, D.S. New York: Academic Press, pp. 411-420.

Waite, T.W., and Woods, D.C. (2022),"Minimax Efficient Random Experimental Design Strategies With Application to Model-Robust Design for Prediction," *Journal of the American Statistical Association*, 117, 1452-1465.

Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*; 127, 327 - 340.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference,* 31, 353-371.

Wiens, D.P. (2015), "Robustness of Design", Chapter 20, *Handbook of Design and Analysis of Experiments*, Chapman & Hall/CRC.

Wiens, D.P. (2018), "I-Robust and D-Robust Designs on a Finite Design Space," *Statistics and Computing*, 28, 241-258.

Xu, X., and Yuen, W.K. (2011), "Applications and Implementations of Continuous Robust Designs," *Communications in Statistics - Theory and Methods*, 40, 969-988.