Journal of Statistical Planning and Inference 40 (1994) 135–160
North-Holland

Robust designs for approximately linear regression: M-estimated parameters

Douglas P. Wiens*

Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

Received 15 February 1993; revised manuscript recieved 20 July 1993

Abstract

We obtain designs, to be used for investigations of response surfaces by regression techniques, when
(i) the fitted, linear (in the parameters) response is incorrect and
(ii) the parameters are to be estimated robustly.
Minimax designs are determined for 'small' departures from the fitted response. We specialize to the case in which the experimenter fits a plane, when in fact the true response contains quadratic and interaction terms. In this case, minimax rotatable designs are derived, subject to a lower bound on the power of a robust test of model adequacy. The optimal designs place their mass at the centre of the design space, and on a sphere interior to the design space.

AMS Subject Classification: Primary 62K05; secondary 62F35, 62J05.

Key words: Robustness; regression designs; M-estimation; response surfaces.

1. Introduction and summary

Consider the fixed carriers linear regression model

$$y_i=\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x}_i)+\varepsilon_i,\quad i=1,2,\ldots,n, \tag{1.1}$$

$$\boldsymbol{\theta}:r\times 1;\quad \varepsilon_1,\ldots,\varepsilon_n\quad \text{uncorrelated,}$$

$$\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\in S\subseteq\mathbb{R}^p,\quad \int_S \mathrm{d}\boldsymbol{x}=1.$$

When the response function $E[y(\boldsymbol{x})]=\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})$ is exactly correct, and the errors ε_i are normally distributed, then the least-squares estimator (LSE) of $\boldsymbol{\theta}$ is unbiased and — for a fixed design — has minimum variance among unbiased estimators.

Correspondence to: Dr. Douglas P. Wiens, Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, Canada T6G 2G1.
*Research supported by the Natural Sciences and Engineering Research Council of Canada.

SSDI 0378-3758(93)E0055-L

Various methods have been proposed, and successfully implemented, to protect *separately* against violations of the assumptions of normality and linearity in $\boldsymbol{\theta}$. Protection against a broad class of nonnormal error distributions is afforded by the use of M-estimation, whereby the regression/scale estimate $(\hat{\boldsymbol{\theta}}, \hat{\sigma})$ is a solution to

$$n^{-1} \sum_{i=1}^{n} \psi\left(\frac{y_i - \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{u}(\boldsymbol{x}_i)}{\sigma}\right) \boldsymbol{u}(\boldsymbol{x}_i) = \mathbf{0}, \tag{1.2}$$

$$n^{-1} \sum_{i=1}^{n} \chi\left(\frac{y_i - \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{u}(\boldsymbol{x}_i)}{\sigma}\right) = 0 \tag{1.3}$$

for appropriately chosen functions ψ and χ. See Huber (1981).

To formalize departures from the fitted response, suppose that, rather than (1.1), the true model is

$$y_i = \boldsymbol{\theta}_0^{\mathrm{T}} \boldsymbol{u}(\boldsymbol{x}_i) + f(\boldsymbol{x}_i) + \varepsilon_i \tag{1.4}$$

for an unknown function f constrained by

$$\int \boldsymbol{u}(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \mathbf{0}, \tag{1.5}$$

$$\int f^2(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \leqslant \eta^2, \quad \eta^2 \text{ fixed.} \tag{1.6}$$

(All integrals in this paper are over the design space S, unless it is explicitly indicated otherwise.) Condition (1.5) is imposed without loss of generality, since any function f which satisfies (1.6) may be modified, by subtracting its L_2-projection on $\boldsymbol{u}$, to satisfy (1.5) as well. In turn, (1.5) guarantees that $E[y(\boldsymbol{x})]$ is uniquely parametrized.

Under (1.4), the LSE $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ may be biased, although the bias may be made small by an appropriate choice of design. If the $\boldsymbol{x}_i$ are governed by a design measure ξ, i.e. a probability measure on S given by

$$\xi(S') = \frac{\text{No. of } \boldsymbol{x}_i \text{ in } S'}{n}$$

for $S' \subseteq S$, then the necessary and sufficient condition for unbiasedness of $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ is

$$\int \boldsymbol{u}(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\xi(\boldsymbol{x}) = \mathbf{0}. \tag{1.7}$$

If this is to hold for all f satisfying (1.5) and (1.6), then it is required that ξ be the (continuous) uniform d.f. $\lambda(\boldsymbol{x})(\mathrm{d}\lambda(\boldsymbol{x}) = \mathrm{d}\boldsymbol{x})$ on S.

Uniformity of ξ is of course attainable only asymptotically. Even if (1.7) holds only approximately, the corresponding design will likely be quite inefficient for model (1.1). There is a rich literature on the construction of designs for (1.1), concentrating on the minimization of some scalar-valued function of the covariance matrix of $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$, see e.g. Fedorov (1972). These variance-optimal designs are typically supported on a small number of extreme points of S, and so differ markedly from any design which comes close to satisfying (1.7) for any broad class of functions f.

There is also a body of work concerned with the choice of appropriate design measures for (1.4), or submodels thereof, when $\hat{\theta}$ is the LSE and loss incorporates both bias and variance. Box and Draper (1959) studied designs for the submodel in which $\theta_0^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})$ and $f(\boldsymbol{x})$ are multinomials in $\boldsymbol{x}$ of degrees $d_1 < d_2$, respectively. Taking loss as the integrated mean squared error of $\hat{y}(\boldsymbol{x})$,

$$\mathscr{L}(f,\xi)=\int E[\{\hat{y}(\boldsymbol{x})-(\theta_0^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})+f(\boldsymbol{x}))\}^2]\,\mathrm{d}\boldsymbol{x}, \tag{1.8}$$

they obtained designs to minimize $\mathscr{L}$, *averaged* over f. Huber (1975) obtained designs to minimize the *maximum* of $\mathscr{L}(f,\xi)$ for f ranging over the entire class defined by (1.5) and (1.6). This was extended to multiple linear regression in Wiens (1990) and also to a variety of other loss functions in Wiens (1992b, 1993); see Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984), Notz (1989) and Wiens (1991) for other approaches to robust designs for least-squares estimation problems.

A more complete robustification of version (1.4) of the regression problem requires the construction of designs which are appropriate for use when the parameters are to be estimated robustly. In this paper we obtain such designs when $\hat{\theta}$ and $\hat{\sigma}$ are to be obtained from (1.2) and (1.3). As loss functions, we consider both the integrated, asymptotic squared bias and a finite sample analogue of the integrated mean squared error of $\hat{y}(\boldsymbol{x})$. The required asymptotic theory is presented in Section 2. In Section 3 we specialize to functions $f(\boldsymbol{x})$ of the form

$$f_{\alpha}(\boldsymbol{x})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{v}(\boldsymbol{x}), \tag{1.9}$$

where $\boldsymbol{v}(\boldsymbol{x}): q\times 1$ is fixed and $\boldsymbol{\alpha}$ is constrained by (1.5) and (1.6). We obtain expansions of the loss functions in terms of $\boldsymbol{\alpha}$, with terms which are $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$ neglected.

For $f=f_{\alpha}$, condition (1.7) is attainable. It is the necessary and sufficient condition for a design to minimize the integrated squared bias of $\hat{y}(\boldsymbol{x})$, uniformly in $\boldsymbol{\alpha}$, to order $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$. Bias-optimal designs to order 6 and higher in $\|\boldsymbol{\alpha}\|$ are also considered. The conditions for their existence are found to be increasingly, and unrealistically, restrictive.

In Section 4 we investigate designs to minimize the maximum value of the aforementioned analogue of the integrated mean squared error of $\hat{y}(\boldsymbol{x})$. We specialize to the case in which one fits a plane, with constant term, in input variables $X_1,\ldots,X_p$. The true response contains as well all quadratic and interaction terms $X_iX_j(i\geqslant j)$. The experimenter wishes to retain the high efficiency of a design optimal for a linear response, while guarding against biases incurred under the second-order alternative. We obtain minimax designs, subject to a lower bound on the asymptotic power of a robust test of the hypothesis that $\boldsymbol{\alpha}=\mathbf{0}$. We restrict the search to designs which satisfy those moment conditions, in least-squares design theory, for *rotatability*. These designs have a property that, ignoring terms of $O(\|\boldsymbol{\alpha}\|^4)$, the asymptotic variance of $\sqrt{n}\,\hat{y}(\boldsymbol{x}_0)$ depends on $\boldsymbol{x}_0$ and $\boldsymbol{\alpha}$ in a fairly simple manner. Special cases are considered, asymptotically and numerically, in Section 4.3. All proofs for Section 4 are deferred to Section 4.4.

2. Asymptotic theory

Let $y_1, \ldots, y_n$ be independent observations following model (1.4), and let $\xi_n(\boldsymbol{x})$ be the design measure of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Rather than (1.2) and (1.3) we work with

$$E_H\left[\psi\left(\frac{y-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})}{\sigma}\right)\boldsymbol{u}(\boldsymbol{x})\right]=\mathbf{0}, \tag{2.1}$$

$$E_H\left[\chi\left(\frac{y-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})}{\sigma}\right)\right]=0. \tag{2.2}$$

When $H=H_n$, the empirical distribution function of $(\boldsymbol{x}_i, y_i)$, then (2.1) and (2.2) become (1.2) and (1.3). We assume that ξ_n has a weak limit ξ. Let G be the distribution function of ε. If H is the limiting, true distribution function of $(y, \boldsymbol{x})$, given by

$$H(y,\boldsymbol{x})=\int\cdots\int_{\bigcap_{i=1}^{p}\{z_i\leqslant x_i\}} G(y-\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{z})-f(\boldsymbol{z}))\,\mathrm{d}\xi(\boldsymbol{z}), \tag{2.3}$$

then (2.1) and (2.2) define the asymptotic values of the estimates.

We shall take $\psi(x)=\rho'(x)$, $\chi(x)=x\psi(x)-\rho(x)$ for a convex function ρ, with $\rho(0)>0$. A particular case is Huber's proposal 2:

$$\psi(x)=\psi_c(x):=\operatorname{sign}(x)\cdot\min(|x|, c), \tag{2.4}$$

$$\chi(x)=\chi_c(x):=\psi_c^2(x)+\chi_c(0), \tag{2.5}$$

where $c>0$ and $\chi_c(0)$ is adjusted for Fisher consistency at a particular distribution when $f\equiv 0$, e.g. $\chi(0)=-E_G[\psi_c^2(\varepsilon)]$ gives consistency when $\varepsilon\sim G$. As in Huber (1981, Ch. 7), the solutions $(\boldsymbol{\theta}, \sigma)$ to (2.1), (2.2) are unique. Under additional assumptions (see e.g. Silvapullé (1985))

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\\ \hat{\sigma}-\sigma\end{pmatrix}$$

is asymptotically normally distributed, with mean $\mathbf{0}$. The asymptotic covariance matrix is given by $D^{-1}CD^{-\mathrm{T}}$, where

$$C=E_{\xi,G}[JJ^{\mathrm{T}}], \qquad D=E_{\xi,G}\left[\frac{\partial}{\partial\boldsymbol{\theta},\sigma}J\right] \tag{2.6}$$

and

$$J=\begin{pmatrix}\psi((f(\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\varepsilon)/\sigma)\boldsymbol{u}\\ \chi((f(\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\varepsilon)/\sigma)\end{pmatrix}.$$

By redefining $(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$ as $\boldsymbol{\theta}$, we may now, without loss of generality, take $\boldsymbol{\theta}_0=\mathbf{0}$. Define functions

$$l(s,t)=E_G\left[\psi\left(\frac{\varepsilon+s}{t}\right)\right], \tag{2.7}$$

$$m(s,t)=E_G\left[\chi\left(\frac{\varepsilon+s}{t}\right)\right]. \tag{2.8}$$

From (2.1) and (2.2), the asymptotic roots $(\boldsymbol{\theta},\sigma)$ are then defined by

$$E_\xi[l(f(\boldsymbol{x})-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x}),\sigma)\boldsymbol{u}(\boldsymbol{x})]=\mathbf{0}, \tag{2.9}$$

$$E_\xi[m(f(\boldsymbol{x})-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x}),\sigma)]=0. \tag{2.10}$$

Define

$$A_1=\int \boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

and assume that A_1, C and D are nonsingular. Note that the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is $[D^{-1}CD^{-\mathrm{T}}]_{11}$. As our loss function, we shall take

$$\mathscr{L}(f,\xi)=\mathscr{V}(f,\xi)+\mathscr{B}(f,\xi)+\mathscr{Q}(f), \tag{2.11}$$

where $\boldsymbol{\theta}$ is defined by (2.9) and (2.10), f is constrained by (1.5) and (1.6), and

$$\mathscr{V}(f,\xi)=n^{-1}\operatorname{tr}A_1[D^{-1}CD^{-\mathrm{T}}]_{11}$$

$$\mathscr{B}(f,\xi)=\boldsymbol{\theta}^{\mathrm{T}}A_1\boldsymbol{\theta},\qquad \mathscr{Q}(f)=\int f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

The components of $\mathscr{L}(f,\xi)$ represent errors due to variance, bias and model misspecification, respectively. This form of $\mathscr{L}$ is motivated by a consideration of the integrated mean squared error of the fitted values,

$$\int E[\hat{y}(\boldsymbol{x})-f(\boldsymbol{x}))^2]\,\mathrm{d}\boldsymbol{x}=\operatorname{tr}A_1\operatorname{Cov}[\hat{\boldsymbol{\theta}}]+E[\hat{\boldsymbol{\theta}}]^{\mathrm{T}}A_1E[\hat{\boldsymbol{\theta}}]+\mathscr{Q}(f).$$

The mathematical formulation of the design problem is then to minimize, over ξ, the maximum, over f, of $\mathscr{L}(f,\xi)$.

3. Optimality theory: $f=f_\alpha$

Henceforth, we constrain $f(\boldsymbol{x})$ by (1.9) as well as by (1.5) and (1.6). Note that by (1.5),

$$\int \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x}=\mathrm{O}:r\times q. \tag{3.1}$$

Without loss of generality, we take

$$A_1 = I_r, \qquad \int \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x} = I_q.$$

Define

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = E_\xi \begin{bmatrix} \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} & \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\ \boldsymbol{v}\boldsymbol{u}^{\mathrm{T}} & \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} \end{bmatrix}.$$

Condition (1.6) becomes

$$\|\boldsymbol{\alpha}\| \leqslant \eta. \tag{3.2}$$

With $f = f_{\boldsymbol{\alpha}}$, (2.9) and (2.10) define functions $\boldsymbol{\theta}(\boldsymbol{\alpha})$ and $\sigma(\boldsymbol{\alpha})$. From these, we obtain Taylor expansions of $\mathscr{L}(f_{\boldsymbol{\alpha}}, \xi)$ around $\boldsymbol{\alpha} = \mathbf{0}$. These expansions require Fisher consistency of $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta} = \mathbf{0}$ when $\boldsymbol{\alpha} = \mathbf{0}$. We ensure this by assuming that the error distribution $G(\varepsilon)$ is symmetric about 0. By this, $l(0, \sigma(\mathbf{0})) = 0$, where $\sigma(\mathbf{0})$ is defined by

$$m(0, \sigma(\mathbf{0}))\left(= E_G\left[\chi\left(\frac{\varepsilon}{\sigma(\mathbf{0})}\right)\right]\right) = 0. \tag{3.3}$$

Define functions

$$c(s,t) = E_G\left[\psi^2\left(\frac{\varepsilon + s}{t}\right)\right], \qquad d(s,t) = E_G\left[\psi'\left(\frac{\varepsilon + s}{t}\right)\right]. \tag{3.4}$$

We denote differentiation with subscripts so that, for instance,

$$m_{11}(0, \sigma(\mathbf{0})) = \left.\frac{\partial^2 m(s,t)}{\partial s^2}\right|_{\substack{s=0\\ t=\sigma(\mathbf{0})}}, \qquad m_2(0, \sigma(\mathbf{0})) = \left.\frac{\partial m(s,t)}{\partial t}\right|_{\substack{s=0\\ t=\sigma(\mathbf{0})}}.$$

Assume the existence of all indicated derivatives, and that differentiation may be interchanged with E_ξ in (2.9) and (2.10).

Define constants

$$V_{\psi,G} = \sigma^2(\mathbf{0})c(0, \sigma(\mathbf{0}))/d^2(0, \sigma(\mathbf{0})),$$

$$\tau_0 = -\frac{m_{11}(0, \sigma(\mathbf{0}))}{\sigma(\mathbf{0})m_2(0, \sigma(\mathbf{0}))},$$

$$\tau_1 = \frac{c_{11}(0, \sigma(\mathbf{0}))}{2c(0, \sigma(\mathbf{0}))} - \frac{d_{11}(0, \sigma(\mathbf{0}))}{d(0, \sigma(\mathbf{0}))},$$

$$\tau_2 = \sigma(\mathbf{0})\frac{\tau_0}{\tau_1}\left(\frac{d_2(0, \sigma(\mathbf{0}))}{d(0, \sigma(\mathbf{0}))} - \frac{c_2(0, \sigma(\mathbf{0}))}{2c(0, \sigma(\mathbf{0}))} - \frac{1}{\sigma(\mathbf{0})}\right)$$

and matrices

$$\Delta = B_{11}^{-1}B_{12}, \qquad B_{22.1} = B_{22} - B_{21}B_{11}^{-1}B_{12},$$

$$P = E_\xi[(\boldsymbol{u}^{\mathrm{T}}B_{11}^{-2}\boldsymbol{u})(\boldsymbol{v} - \Delta^{\mathrm{T}}\boldsymbol{u})(\boldsymbol{v} - \Delta^{\mathrm{T}}\boldsymbol{u})^{\mathrm{T}}] - \tau_2(\operatorname{tr} B_{11}^{-1})B_{22.1}.$$

Note that

$$E_\xi[\boldsymbol{u}^{\mathrm{T}}B_{11}^{-2}\boldsymbol{u}]=\operatorname{tr}B_{11}^{-1},\qquad E_\xi[(\boldsymbol{v}-\Delta^{\mathrm{T}}\boldsymbol{u})(\boldsymbol{v}-\Delta^{\mathrm{T}}\boldsymbol{u})^{\mathrm{T}}]=B_{22.1}.$$

A straightforward expansion of the terms in (2.11) gives the following result.

Theorem 3.1. *Define* $\theta(\boldsymbol{\alpha}),\sigma(\boldsymbol{\alpha})$ *by* (2.9) *and* (2.10), *where* $f=f_\alpha$ *is given by* (1.9), (3.1) *and* (3.2). *Then*:

(i) $\theta(\boldsymbol{\alpha})=\Delta\boldsymbol{\alpha}+\mathrm{O}(\|\boldsymbol{\alpha}\|^3)$.

(ii) $\sigma(\boldsymbol{\alpha})=\sigma(\mathbf{0})+\frac{1}{2}\tau_0\sigma(\mathbf{0})\boldsymbol{\alpha}^{\mathrm{T}}B_{22.1}\boldsymbol{\alpha}+\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$.

(iii) *Ignoring terms which are* $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$,

$$[D^{-1}CD^{-\mathrm{T}}]_{12}=0.$$

Thus, $\sqrt{n}(\hat{\theta}-\theta)$ *and* $\sqrt{n}(\hat{\sigma}-\sigma)$ *are asymptotically uncorrelated and hence asymptotically independent*;

(iv) *The asymptotic covariance matrix of* $\sqrt{n}(\hat{\theta}-\theta)$ *is*

$$[D^{-1}CD^{-\mathrm{T}}]_{11}=V_{\psi,G}\{B_{11}^{-1}+\tau_1(E_\xi[(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{v}-\Delta^{\mathrm{T}}\boldsymbol{u}))^2B_{11}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}B_{11}^{-1}] \\ -\tau_2(\boldsymbol{\alpha}^{\mathrm{T}}B_{22.1}\boldsymbol{\alpha})B_{11}^{-1})\}+\mathrm{O}(\|\boldsymbol{\alpha}\|^4).$$

Thus, ignoring terms which are $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$,

$$\begin{aligned}\mathscr{V}(f_\alpha,\xi)&=n^{-1}V_{\psi,G}[\operatorname{tr}B_{11}^{-1}+\tau_1\boldsymbol{\alpha}^{\mathrm{T}}P\boldsymbol{\alpha}],\\ \mathscr{B}(f_\alpha,\xi)&=\boldsymbol{\alpha}^{\mathrm{T}}\Delta^{\mathrm{T}}\Delta\boldsymbol{\alpha},\\ \mathscr{L}(f_\alpha,\xi)&=n^{-1}V_{\psi,G}\operatorname{tr}B_{11}^{-1}+\boldsymbol{\alpha}^{\mathrm{T}}[I_q+\Delta^{\mathrm{T}}\Delta+n^{-1}V_{\psi,G}\tau_1P]\boldsymbol{\alpha}.\end{aligned}\tag{3.5}$$

Remark 3.1. Consider the special case of Huber's Proposal 2 as at (2.4) and (2.5). Let G have an absolutely continuous density g and put $k=c\sigma(\mathbf{0})$. Then

$$\tau_0=(G(k)-0.5-kg(k))\Big/\int_0^k\varepsilon^2\,\mathrm{d}G,$$

$$\tau_1=\frac{G(k)-0.5-kg(k)}{[\int_0^k\varepsilon^2\,\mathrm{d}G+k^2(1-G(k))]}-\frac{g'(k)}{G(k)-0.5},$$

$$\tau_2=\frac{\tau_0}{\tau_1}\left[\frac{kg(k)}{G(k)-0.5}+\frac{\int_0^k\varepsilon^2\,\mathrm{d}G}{\int_0^k\varepsilon^2\,\mathrm{d}G+k^2(1-G(k))}-1\right].$$

Note that $\tau_0,\tau_1>0$ if g is unimodal, since then

$$G(k)-0.5-kg(k)=\int_0^k(-\varepsilon g'(\varepsilon)\,\mathrm{d}\varepsilon>0.$$

If $G=\Phi$, the N(0, 1) d.f., then $\int_0^k \varepsilon^2\,d\Phi=\Phi(k)-0.5-k\phi(k)$ and $\phi'(k)=-k\phi(k)$, whence

$$\tau_0=1, \qquad \tau_2=1-\frac{1}{\tau_1}, \quad \tau_1\to 1 \text{ as } k\to 0,\infty.$$

We find numerically that $1\leqslant\tau_1\leqslant 1.098$, with the maximum being attained at $k=0.8$.

Remark 3.2. It is not obvious that the approximation to $[D^{-1}CD^{-\mathrm{T}}]_{11}$, in Theorem 3.1.(iv), yields a positive-definite matrix when terms of $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$ are neglected. In order that the approximation to $\boldsymbol{a}^{\mathrm{T}}[D^{-1}CD^{-\mathrm{T}}]_{11}\,\boldsymbol{a}$ be positive for all nonnull vectors $\boldsymbol{a}$, we require

$$\tau_1\tau_2<\inf_{\boldsymbol{a}}\frac{\boldsymbol{a}^{\mathrm{T}}\{B_{11}^{-1}+\tau_1 E[(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{v}-\Delta^{\mathrm{T}}\boldsymbol{\alpha}))^2 B_{11}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}B_{11}^{-1}]\}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\{(\boldsymbol{\alpha}^{\mathrm{T}}B_{22.1}\boldsymbol{\alpha})B_{11}^{-1}\}\boldsymbol{a}}$$

$$=\mathrm{ch}_{\min}\frac{\{I_p+\tau_1 E[(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{v}-\Delta^{\mathrm{T}}\boldsymbol{u}))^2 B_{11}^{-1/2}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}B_{11}^{-1/2}]\}}{\boldsymbol{\alpha}^{\mathrm{T}}B_{22.1}\boldsymbol{\alpha}}.$$

(Here $\mathrm{ch}_{\min}$ denotes the minimum characteristic root.) Hence, it suffices for positive definiteness if $\tau_1\geqslant 0$ and $\tau_1\tau_2<\inf_{\|\boldsymbol{\alpha}\|\leqslant\eta}(\boldsymbol{\alpha}^{\mathrm{T}}B_{22.1}\boldsymbol{\alpha})^{-1}$, i.e.

$$\tau_1\geqslant 0, \quad \eta^2\tau_1\tau_2\mathrm{ch}_{\max}B_{22.1}<1. \tag{3.6}$$

Remark 3.3. A possible approach to this problem is to consider arbitrary contamination within shrinking neighbourhoods. Replace η by $\eta/\sqrt{n}$ in (1.6) and (3.2). In Theorem 3.1 take $q=1$. Then, ignoring terms which are $\mathrm{o}(n^{-1})$, (3.5) gives

$$\max_{\boldsymbol{\alpha}}\mathscr{L}(f_{\boldsymbol{\alpha}},\xi)=n^{-1}\left[V_{\psi,G}\operatorname{tr}B_{11}^{-1}+\eta^2\left(1+\left\|B_{11}^{-1}\int\boldsymbol{u}(\boldsymbol{x})v(\boldsymbol{x})\,d\xi(\boldsymbol{x})\right\|^2\right)\right]$$

$$:=\mathscr{L}_0(v,\xi).$$

One can now consider the problem

$$(\mathrm{P}):\quad \min_{\xi}\max_{v}\mathscr{L}_0(v,\xi),$$

where the max is taken over all functions $v(\boldsymbol{x})$, of $L_2(S)$-norm 1, satisfying (3.1).

If least-squares estimation is employed, then the ignored $\mathrm{o}(n^{-1})$ term in fact vanishes identically, and $\mathscr{L}_0(v,\xi)$ is exact. For an arbitrary M-estimate, $\mathscr{L}_0(v,\xi)$ depends on (ψ,χ) only through $V_{\psi,G}$, which is constant with respect to v and ξ. Thus, the solutions to problem (P) for general M-estimates are identical to those in the least-squares case, except that the error variance σ^2 is replaced by $V_{\psi,G}$; see Huber (1981) and Wiens (1990, 1992b) for details and special cases.

Note that, in shrinking neighbourhoods as above, $n\mathscr{L}_0(v,\xi)$ no longer depends on n. The minimax design is then independent of the sample size. This, rather counterintuitive, conclusion is forced by the requirement that maximal bias and standard error decrease at the same rate $\mathrm{O}(n^{-1/2})$. If our primary interest is in model selection, then this requirement is quite reasonable — if maximal bias does not decrease, or decreases too slowly, then any consistent test of model adequacy will eventually reject, with power one. If however our primary interest is in prediction, then bias should be judged relative to the standard error of the predictor. This standard error contains a constant term, arising from the error variance, and a term of the order $n^{-1/2}$, arising from the standard deviation of the estimate. A practitioner would then be quite willing to tolerate biases which are O(1), as long as they are small relative to the standard deviation of the errors.

Our approach, as carried out in Section 4, attempts a compromise between these competing demands of model selection and predictive ability. It is briefly described as follows. We view η as small but fixed, while letting $n\to\infty$. We will minimize the maximum, over $\boldsymbol{\alpha}$ (subject to (3.2)) of (3.5), for particular choices of $\boldsymbol{z}$ and $\boldsymbol{v}$. It turns out that, without further restrictions, the minimax designs are such that a test of the hypothesis that $\boldsymbol{\alpha}=\mathbf{0}$ has no power. We thus minimize the maximum loss, *subject to a lower bound on the asymptotic power of this test, against the contiguous subclass of alternatives* with $\|\boldsymbol{\alpha}\|=\mathrm{O}(n^{-1/2})$.

3.1. Bias optimality

Define

$$\mathscr{B}_1(f,\xi)=\mathscr{Q}(f)+\mathscr{B}(f,\xi). \tag{3.7}$$

We say that a design ξ is *B-uniform* if $E_\xi[\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}]=E_\lambda[\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}]$, i.e. if $B_{12}=0$. Note that then $\Delta=0$. From Theorem 3.1,

$$\mathscr{B}_1(f_{\boldsymbol{\alpha}},\xi)=\boldsymbol{\alpha}^{\mathrm{T}}(I_q+\Delta^{\mathrm{T}}\Delta)\boldsymbol{\alpha}+\mathrm{O}(\|\boldsymbol{\alpha}\|^4) \tag{3.8}$$

and so any B-uniform design is bias-optimal, to this order, uniformly in $\boldsymbol{\alpha}$. When $\hat{\boldsymbol{\theta}}$ is the LSE, the remainder term in (3.8) vanishes, and the exact optimality of B-uniform designs is essentially the result in the appendix of Box and Draper (1959).

The bias optimality of B-uniform designs may be maintained even if $G(\varepsilon)$ is nonsymmetric, if at least $l(0,\sigma(\mathbf{0}))=0$ (Fisher consistency). In this case Theorem 3.1 (i) holds, but with

$$\Delta=\{B_{11}-\tau_4E_\xi[\boldsymbol{u}]E_\xi[\boldsymbol{u}^{\mathrm{T}}]\}^{-1}\{B_{12}-\tau_4E_\xi[\boldsymbol{u}]E_\xi[\boldsymbol{v}^{\mathrm{T}}]\},$$

$$\tau_4=\frac{l_2(0,\sigma(\mathbf{0}))m_1(0,\sigma(\mathbf{0}))}{l_1(0,\sigma(\mathbf{0}))m_2(0,\sigma(\mathbf{0}))}.$$

If $\boldsymbol{u}(\boldsymbol{x})$ contains a constant element, then $E_\xi[\boldsymbol{v}^{\mathrm{T}}]$ is a row of B_{12}, so that $\Delta=0$ if ξ is B-uniform. As at (3.8), bias optimality follows.

Under conditions considerably more stringent than B-uniformity, we can obtain designs which are bias-optimal to higher order in $\|\boldsymbol{\alpha}\|$. Again assuming symmetry of G, we may extend Theorem 3.1(i) to

$$\theta(\alpha)=\Delta\alpha+\tau_5 E_\xi[(\alpha^T(v-\Delta^T u))^3 B_{11}^{-1}u]/6+O(\|\alpha\|^5),$$

$$\tau_5=\ l_{111}(0,\sigma(\mathbf{0}))/\ l_1(0,\sigma(\mathbf{0})).$$

To this order, a design is bias-optimal if it is B-uniform and if, as well, $E_\xi[(\alpha^T(v-\Delta^T u))^3 u]=\mathbf{0}$ for all α, i.e.

$$E_\xi[v_i v_j v_k u_l]=0,\quad i,j,k\leqslant q,\quad l\leqslant p.$$

These conditions are attainable, but very restrictive. For instance, in the case of simple linear regression with quadratic contamination, i.e.

$$u(x)=(1,x\sqrt{12})^T,\quad v(x)=\sqrt{180}(x^2-\tfrac{1}{12}),\qquad -\tfrac{1}{2}\leqslant x\leqslant\tfrac{1}{2},\tag{3.9}$$

the requirements become

$$E_\xi[X]=\tfrac{1}{12},\qquad E_\xi[(X^2-\tfrac{1}{12})^3]=0.$$

Such a design places an inordinate amount of mass near 0.

Note that for (3.9) there is a design which is asymptotically unbiased to *all* orders in α — the design which places mass 1/2 at each of $\pm 1/\sqrt{12}$. This follows from (2.9) with $f(x)=\alpha v(x)$. Few experimenters, however, would be willing to pay the associated price in efficiency, for unbiasedness against one alternative.

4. MSE-optimality: multiple linear regression with quadratic contamination

The discussion of bias optimality in Section 3.1 suggests that bias reduction alone does not lead to practical designs, beyond the possible requirement of B-uniformity. In this section we concentrate on minimizing the maximum integrated mean squared error, as at (3.5).

We consider the case in which the fitted response contains a constant term, and linear terms $\theta_i x_i$, $i=1,\dots,p$, normalized so that $E_\lambda[uu^T]=I_{p+1}$. The true response also contains terms $\theta_{ij}x_i x_j$ $(i\geqslant j)$, translated and scaled so that $E_\lambda[vv^T]=I_q(q=p(p+1)/2)$ and $E_\lambda[uv^T]=0_{p\times q}$. As design space we take a p-dimensional sphere of unit volume:

$$S=\left\{x\,|\,\|x\|\leqslant r:=\left(\Gamma\left(\frac{p}{2}+1\right)\right)^{1/p}\Big/\sqrt{\pi}\right\}.$$

We use the notation

$$Z=\|X\|^2,\qquad \mu_k(\xi)=E_\xi[Z^k],\qquad \sigma_\xi^2=\mathrm{var}_\xi[Z],$$

$$\mu_3'(\xi)=E_\xi[(Z-\mu_1(\xi))^3],\quad \sigma_{12}(\xi)=\mathrm{Cov}_\xi[Z,Z^2].$$

Then, in particular,

$$\mu_1(\lambda)=\frac{p}{p+2}r^2, \qquad \mu_2(\lambda)=\frac{p}{p+4}r^4, \qquad \sigma_\lambda^2=\frac{4p}{(p+2)^2(p+4)}r^4. \tag{4.1}$$

For any square matrix M, let $\text{vec}_s M$ denote the vector of elements of M, on or below the main diagonal, arranged column-by-column. In this notation, we are taking

$$\boldsymbol{u}(\boldsymbol{x})=(1,\boldsymbol{x}^{\mathrm{T}}/(\mu_1(\lambda)/p)^{1/2})^{\mathrm{T}}, \qquad \boldsymbol{v}(\boldsymbol{x})=\Sigma_\lambda^{-1/2}(\underset{s}{\text{vec}}\,\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-(\mu_1(\lambda)/p)\boldsymbol{j}), \tag{4.2}$$

where

$$\Sigma_\lambda=\underset{\lambda}{\text{Cov}}[\underset{s}{\text{vec}}\,\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}], \qquad \boldsymbol{j}=\underset{s}{\text{vec}}\,I_p, \qquad (\mu_1(\lambda)/p)\boldsymbol{j}=E_\lambda[\underset{s}{\text{vec}}\,\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}].$$

It turns out that the designs which minimize the maximum of $\mathscr{L}(f_a,\xi)$ have no power to detect interaction or curvature in the response, see Section 4.3.2. We thus seek to minimax $\mathscr{L}(f_a,\xi)$, subject to a lower bound on the power of Wald's test of the hypothesis that $\boldsymbol{\alpha}=\boldsymbol{0}$. This test is carried out by fitting the full model $E[y]=\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{u}(\boldsymbol{x})+\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{v}(\boldsymbol{x})$, and then computing the test statistic

$$T=(n\hat{\boldsymbol{\alpha}}^{\mathrm{T}}B_{22.1}\hat{\boldsymbol{\alpha}})/(\hat{V}_{\psi,G}),$$

where $\hat{V}_{\psi,G}$ is a consistent estimate of $V_{\psi,G}$. When ψ depends on $\boldsymbol{x}$ only through the residual, Wald's test is identical to the τ-test of Hampel et al. (1986); see Huber (1981, Ch. 7), Street et al. (1988) and Wiens (1992a) for computational details.

For the asymptotics of the testing theory we must work in shrinking subneighbourhoods of (1.6). Under alternatives K: $\boldsymbol{\alpha}=\boldsymbol{\beta}/\sqrt{n}$, the asymptotic distribution of T is $\chi_q^2(\lambda^2)$, where the noncentrality parameter is

$$\lambda^2=(\boldsymbol{\beta}^{\mathrm{T}}B_{22.1}\boldsymbol{\beta})/(V_{\psi,G}).$$

On the sphere $\|\boldsymbol{\beta}\|=\text{constant}$, the minimum value of λ^2 is proportional to $\text{ch}_{\min} B_{22.1}$. Since

$$\max_{\|\boldsymbol{\alpha}\|\leqslant\eta}\mathscr{L}(f_a,\xi)=n^{-1}V_{\psi,G}\operatorname{tr}B_{11}^{-1} + \eta^2(1+\text{ch}_{\max}(B_{21}B_{11}^{-2}B_{12}+n^{-1}V_{\psi,G}\tau_1 P)), \tag{4.3}$$

our problem is to find ξ to minimize (4.3), subject to

$$\text{ch}_{\min} B_{22.1}\geqslant\delta_{\min}^2, \tag{4.4}$$

for fixed $\delta_{\min}^2$.

We require the following additional assumptions.

(B1) $\tau_1\geqslant 0$ and $\eta^2<16p/((p+2)^2(p+4)\tau_1\tau_2)$ if $\tau_2\geqslant 0$.

(B2) For any nonnegative integers $m_1,\dots,m_6$ with $\sum m_j\leqslant 6$, $E[\prod_{j=1}^6 X_{i_j}^{m_j}]$ is zero, if any m_j is odd, and is invariant under permutations of the i_j.

(B3) If $p=2$ then

(i) $E[X_1^4-3X_1^2X_2^2]=0$.

If $p \geqslant 3$ then, in addition,

(ii) $E[X_1^2 X_2^4 - 3X_1^2 X_2^2 X_3^2] = 0$,

(iii) $E[X_1^6 - 15X_1^2 X_2^4 + 30X_1^2 X_2^2 X_3^2] = 0$.

4.1. Discussion of assumptions

Assumption (B1) ensures that (3.6) holds, see equation (4.17). If $p=1$, then (B2) asserts only that $E_\xi[X]=0$. For $p=2$ and $p \geqslant 3$, (B2) and (B3) are the moment conditions defining *rotatable* designs of order 2 and 3, respectively, see Box and Hunter (1957). Ignoring terms of order $O(\|\boldsymbol{\alpha}\|^4)$, for such designs the asymptotic variance of $\sqrt{n}\,\hat{y}(\boldsymbol{x}_0)$ depends on $\boldsymbol{x}_0$ only through the distance $\|\boldsymbol{x}_0\|$ from the centre of the design space, and through two quadratic forms in $\boldsymbol{x}_0$, whose coefficients are linear and quadratic terms in the elements of $\boldsymbol{\alpha}$, see Theorem 4.1(iv).

The assumption of rotatability reduces the optimization problem to one of obtaining the optimal distribution of $Z = \|\boldsymbol{X}\|^2$. The minimax designs place mass $1-t$ at $Z=0$, mass t at $Z = sr^2$, $0 < s \leqslant 1$. To complete the construction, one must determine a conditional distribution of $\boldsymbol{X}$, given $\|\boldsymbol{X}\| = r\sqrt{s}$, for which the identities in (B3) hold. Such designs are not unique. They have received little attention in the literature since, for $p \geqslant 3$, the corresponding design matrices, for fitting *third-order* models by *least squares*, are deficient in rank (Draper, 1960). Nonetheless, their derivation appears to be straightforward.

Example 4.1. Let X be uniformly distributed on the $2^p \cdot p!$ values obtained by permuting the elements of each of the 2^p vectors $r\sqrt{s}(\pm x_1, \ldots, \pm x_p)^{\mathrm{T}}$ in all $p!$ possible ways. Then (B2) holds. If $p=2$, (B3) is satisfied if

$$\sum x_i^2 = 1, \qquad \sum x_i^4 = \frac{3}{p+2}.$$

We then find $x_1 = \frac{1}{2}(2+\sqrt{2})^{1/2} = 0.9239$, $x_2 = \frac{1}{2}(2-\sqrt{2})^{1/2} = 0.3827$. If $p \geqslant 3$ we require as well

$$\sum x_i^6 = \frac{5}{(p+2)(p+4)}.$$

For $p=3$ we find

$$x_1 = 0.86625, \qquad x_2 = 0.26664, \qquad x_3 = 0.42252.$$

For $p>3$, some x_i may be arbitrarily set equal to 0, or to other x_j, thus reducing the required number of points.

The optimal distributions of Z, derived below, lead to *continuous* designs, in that the product nt need not even be an integer, let along a multiple of the number of points, required to be placed on $\|\boldsymbol{x}\| = r\sqrt{s}$, for rotatability. Some approximations are therefore necessary. Perhaps the simplest, and most practical, way around the problem is to employ a *randomized* design. Note that if X is a p-vector of i.i.d. normals, then $r\sqrt{s}X/\|X\|$ has all mass on the sphere of radius $r\sqrt{s}$, and satisfies (B2) and (B3). This follows from the fact that $X/\|X\|$ is distributed independently of $\|X\|$, and X satisfies (B2) and (B3). Thus, to construct a randomized design for given n, s and t:

(i) Obtain a value n_1 of $N_1 \sim \text{bin}\,(n, t)$.

(ii) Place $n-n_1$ of the design points at $\mathbf{0}$.

(iii) Generate n_1 values x_i of $X \sim \text{N}(\mathbf{0}, I_p)$, and place the remaining n_1 design points at $r\sqrt{s}x_i/\|x_i\|$, $1 \leqslant i \leqslant n_1$.

In this case, the *expected* maximum loss is minimized by the optimal distribution of Z. For further information on rotatable designs, see Herzberg (1988) and references cited therein.

4.2. Statements of main results

In this section we state Theorem 4.1, in which the optimal designs are presented. Special (asymptotic) cases, and numerical results, are discussed in Section 4.3. Proofs are in Section 4.4. We require some preliminary definitions. We will show that for any design on S, if (4.4) holds then the first moment $\mu_1(\xi)$ of Z and $(\text{ch}_{\min} B_{22.1})^{1/2} := \delta$ are restricted to lie in the set J, where, for $p=1$,

$$J = \{(\mu_1(\xi), \delta) \mid \mu_L(\delta) \leqslant \mu_1(\xi) \leqslant \mu_U(\delta),\ \delta_{\min} \leqslant \delta \leqslant 3\sqrt{5}/4\} \tag{4.5}$$

and, for $p \geqslant 2$,

$$J = \left\{(\mu_1(\xi), \delta) \mid \frac{\delta^2 \mu_2(\lambda)}{r^2} \leqslant \mu_1(\xi) \leqslant \mu_U(\delta),\ \delta_{\min} \leqslant \delta \leqslant \frac{p+4}{p+2}\right\}. \tag{4.6}$$

Here $\mu_L(\delta), \mu_U(\delta) = (r^2/2)\ (1 \mp (1 - 4\delta^2\sigma_\lambda^2/r^2)^{1/2})$ are the smallest and largest roots of $g(\mu) = \delta^2$, where

$$g(\mu) = \mu(r^2 - \mu)/\sigma_\lambda^2. \tag{4.7}$$

The second moments of Z, under the optimal design, will be defined in terms of $\mu_1(\xi)$ and δ by

$$\begin{aligned} p=1:\quad & \sigma^2(\mu_1(\xi), \delta) = \delta^2\sigma_\lambda^2, \\ & \mu_2(\mu_1(\xi), \delta) = \delta^2\sigma_\lambda^2 + \mu_1^2(\xi), \end{aligned} \tag{4.8}$$

$$\begin{aligned} p \geqslant 2:\quad & \sigma^2(\mu_1(\xi), \delta) = \delta^2\sigma_\lambda^2 + (\delta^2\mu_1^2(\lambda) - \mu_1^2(\xi))^+, \\ & \mu_2(\mu_1(\xi), \delta) = \delta^2\mu_2(\lambda) + (\mu_1^2(\xi) - \delta^2\mu_1^2(\lambda))^+. \end{aligned} \tag{4.9}$$

Minimax loss, for fixed $(\mu_1(\xi), \delta) \in J$, will be shown to be

$$\gamma(\mu_1(\xi), \delta) = n^{-1} V_{\psi,G}\left(1 + \frac{p\mu_1(\lambda)}{\mu_1(\xi)}\right) + \eta^2(1 + \bar{v}(\mu_1(\xi), \delta)), \tag{4.10}$$

where

$$\bar{v}(\mu_1(\xi),\delta)=\begin{cases}\bar{v}_0(\mu_1(\xi),\delta), & p=1,\\ \max(\bar{v}_0(\mu_1(\xi),\delta),\bar{v}_1(\mu_1(\xi),\delta)), & p\geq 2\end{cases}$$

and

$$\bar{v}_0(\mu_1(\xi),\delta)=\frac{(\mu_1(\xi)-\mu_1(\lambda))^2}{\sigma_\lambda^2}+n^{-1}V_{\psi,G}\tau_1\frac{\sigma^2(\mu_1(\xi),\delta)}{\sigma_\lambda^2}\left\{(1-\tau_2)\left(1+\frac{p\mu_1(\lambda)}{\mu_1(\xi)}\right)\right.$$
$$\left.+\frac{p\mu_1(\lambda)}{\mu_1(\xi)}\left(\frac{\sigma^2(\mu_1(\xi),\delta)-1}{\mu_1^2(\xi)}\right)\right\},$$

$$\bar{v}_1(\mu_1(\xi),\delta)=n^{-1}V_{\psi,G}\tau_1\frac{\mu_2(\mu_1(\xi),\delta)}{\mu_2(\lambda)}\left\{(1-\tau_2)\left(1+\frac{p\mu_1(\lambda)}{\mu_1(\xi)}\right)\right.$$
$$\left.+\frac{p\mu_1(\lambda)}{\mu_1(\xi)}\left(\frac{\mu_2(\mu_1(\xi),\delta)-1}{\mu_1^2(\xi)}\right)\right\}.$$

Theorem 4.1. *Make assumptions (B1)–(B3). For the model defined by (4.2), the maximum loss (4.3) is minimized, subject to (4.4), by the design ξ_* described as follows.*

(i) *We have*

$$P_{\xi_*}(Z=0)=1-t, \qquad P_{\xi_*}\left(Z=\frac{\mu_1(\xi_*)}{t}\right)=t, \tag{4.11}$$

where

$$t=\mu_1^2(\xi_*)/\mu_2(\mu_1(\xi_*),\delta_*), \tag{4.12}$$

$$(\mu_1(\xi_*),\delta_*)=\operatorname*{argmin}_J \gamma(\mu_1(\xi),\delta). \tag{4.13}$$

(ii) *On the sphere $Z=\mu_1(\xi_*)/t$, the conditional distribution of X satisfies (B2) and (B3).*

(iii) *Furthermore, when $\bar{v}(\mu_1(\xi_*),\delta_*)=\bar{v}_0(\mu_1(\xi_*),\delta_*)$, the least favourable f_α is*

$$f_\alpha^0(\boldsymbol{x})=\frac{\eta}{\sigma_\lambda}\left(\sum_{i=1}^p X_i^2-\mu_1(\lambda)\right). \tag{4.14}$$

When $\bar{v}(\mu_1(\xi_*),\delta_*)=\bar{v}_1(\mu_1(\xi_*),\delta_*)$, a least favourable f_α, orthogonal to $f_\alpha^0(\boldsymbol{x})$, is

$$f_\alpha^1(\boldsymbol{x})=\eta\left(\frac{2(p+2)}{(p-1)\mu_2(\lambda)}\right)^{1/2}\sum\sum_{1\leq i<j\leq p} X_iX_j. \tag{4.15}$$

(iv) *For any sequence of designs satisfying the rotatability conditions (B2) and (B3), and for any $\boldsymbol{x}_0$, the asymptotic variance of $\sqrt{n}\,\hat{y}(\boldsymbol{x}_0)$ is as follows. Label the elements of $\boldsymbol{\alpha}$ as*

$$\boldsymbol{\alpha}^{\mathrm{T}}=(\alpha_{11},\alpha_{12},\ldots,\alpha_{1p},\alpha_{22},\alpha_{23},\ldots,\alpha_{2p},\ldots,\alpha_{pp})$$

and put

$$H=\begin{pmatrix} \alpha_{11} & & \frac{1}{\sqrt{2}}\alpha_{ij} \\ & \alpha_{22} & \\ & & \ddots & \\ \frac{1}{\sqrt{2}}\alpha_{ji} & & & \alpha_{pp} \end{pmatrix},$$

$$h_1=\operatorname{tr} H, \qquad h_2=\operatorname{tr} H^2=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}.$$

Then, ignoring terms of $\mathrm{O}(\|\boldsymbol{\alpha}\|^4)$, $\lim_{n\to\infty}\operatorname{var}[\sqrt{n}\,\hat{y}(\boldsymbol{x}_0)]$ *equals* $V_{\psi,G}$ *times*:

$$\left\{1+\frac{\mu_1(\lambda)}{\mu_1(\xi)}\|\boldsymbol{x}_0\|^2\right\}\left\{1+\tau_1(1-\tau_2)\left[\frac{\mu_2(\xi)}{\mu_2(\lambda)}\left(h_2-\frac{h_1^2}{p}\right)+\frac{h_1^2}{p}\frac{\sigma_\xi^2}{\sigma_\lambda^2}\right]\right\}$$

$$+\frac{\tau_1 p\mu_1(\lambda)}{\mu_1^2(\xi)}\left\{\frac{4\mu_3(\xi)}{(p+4)\mu_2(\lambda)}\boldsymbol{x}_0^{\mathrm{T}}\left[H-\frac{h_1}{p}I\right]^2\boldsymbol{x}_0\right.$$

$$+\frac{4\sigma_{12}(\xi)}{\sigma_\lambda\sqrt{2(p+2)}\mu_2(\lambda)}\left(\frac{h_1}{p}\right)\boldsymbol{x}_0^{\mathrm{T}}\left[H-\frac{h_1}{p}I\right]\boldsymbol{x}_0$$

$$\left.+\|\boldsymbol{x}_0\|^2\left[\frac{(h_2-h_1^2/p)}{p\mu_2(\lambda)}\left(\sigma_{12}(\xi)-\frac{4\mu_3(\xi)}{p+4}\right)+\left(\frac{h_1}{p}\right)^2\frac{\mu_3'(\xi)}{\sigma_\lambda^2}\right]\right\}.$$

4.3. Special cases

The minimization of $\gamma(\mu_1(\xi),\delta)$ at (4.13) is a straightforward numerical problem. Convergence can be slow, however, since the surface to be minimized is quite flat near $(\mu_1(\xi_*),\delta_*)$. In all cases which we have investigated, γ is an increasing function of δ, so that $\delta_*=\delta_{\min}$.

We have computed the relevant constants, for a variety of choices of the parameters. For $p=1$, see Figure 1. For $p=2$ some values are in Table 1. We took (ψ,χ) as in Huber's Proposal 2, with $\chi(0)=-E_\Phi[\psi_c^2(\varepsilon)]$. Only the constants corresponding to $c=1$ are presented here. Error distributions used were the Gaussian and the Cauchy.

A general feature brought out by the computations is that as η increases, there is a decrease in the amount of mass which the optimal designs place at $\|\boldsymbol{x}\|=z$, and in the value of $\mu_1(\xi_*)$. This is typically (but not always, see Figure 1 in particular) accompanied by a decrease in the value of z.

For $p=2$ the maximum characteristic root $\bar{v}$ in (4.3) tends to equal $\bar{v}_0$ for large n or small η, and then purely quadratic contamination, as at (4.14), is least favourable. For small n or large η, we have $\bar{v}=\bar{v}_1$ and contamination as at (4.15) is least favourable.

The optimal designs are quite insensitive to changes in c and to the error distribution. It is tempting then to take $c=\infty$ and use a design which is optimal for the LSE. However, even though the M-estimator tends to $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ *as* $c\to\infty$, the optimal designs do

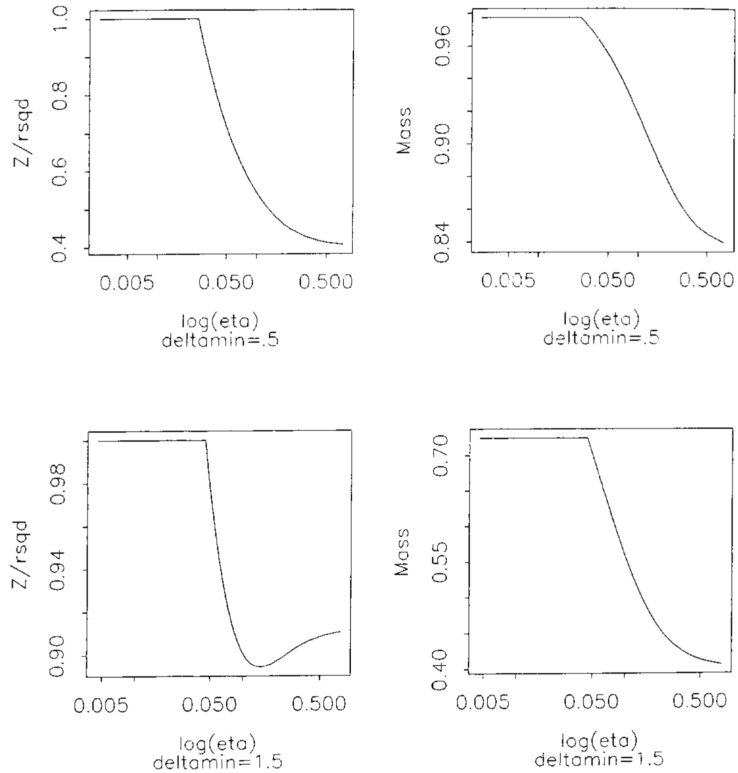

Fig. 1. Nonzero values of $Z/\text{rsq}\,d$, and corresponding mass, versus log(eta) Gaussian errors; $p=1$, $c=1$, $n=40$.

not converge in an analogous fashion. The reason is that, for all positive, *finite* values of c, $\boldsymbol{\theta}_0$ cannot be estimated in a scale equivariant manner without also estimating scale. The bias in $\sigma(\boldsymbol{\alpha})$ then contributes components to $\mathscr{L}(f_a, \xi)$, through the covariance matrix of $\hat{\boldsymbol{\theta}}$, which do not vanish as $c \to \infty$. When least-squares estimation is employed, this covariance matrix depends only on the error variance, and not on $\sigma(\boldsymbol{\alpha})$. In this sense, the LSE represents a discontinuity point in the class of M-estimators.

We now consider some special cases in more detail.

4.3.1. $n \to \infty$

As $n \to \infty$, the contribution of $\mathscr{V}(f_a, \xi)$ to $\mathscr{L}(f_a, \xi)$ vanishes, and the problem becomes one of minimaxing $\mathscr{B}_1(f_a, \xi)$, as at (3.7). Ignoring terms of order $O(\|\boldsymbol{\alpha}\|^4)$, we have

$$\max_{\|\boldsymbol{\alpha}\| \leq \eta} \mathscr{B}_1(f_a, \xi) = \eta^2 \left(1 + \frac{(\mu_1(\xi) - \mu_1(\lambda))^2}{\sigma_\lambda^2} \right).$$

Table 1
Constants for optimal designs ($p=2$, $c=1$, $\mu_1(\lambda)=0.1592$)

	$\eta=0.1$				$\eta=1$			
	Normal	Cauchy	Normal	Cauchy	Normal	Cauchy	Normal	Cauchy
$n=10$								
$\delta_{\min}=\delta_*$	0.5000	0.5000	1.2000	1.2000	0.5000	0.5000	1.2000	1.2000
nonzero z/r^2	1.0000	0.9976	1.0000	1.0000	0.4477	0.4285	0.8000	0.8000
$P_{\xi_*}(Z=z)$	0.9787	0.97861	0.8606	0.8606	0.8822	0.8695	0.7500	0.7500
$\mu_1(\xi_*)$	0.3115	0.3108	0.2739	0.2739	0.1257	0.1186	0.1910	0.1910
e_-	1.0000	0.9988	1.0000	1.0000	0.5724	0.5488	0.8108	0.8108
e_+	1.0000	1.00000	1.0000	1.0000	2.5801	2.1197	1.5286	1.4423
$\min\lambda^2$	0.0226	0.0098	0.1300	0.0566	2.2578	0.9817	13.0050	5.6549
$n=100$								
$\delta_{\min}=\delta_*$	0.5000	0.5000	1.2000	1.2000	0.5000	0.5000	1.2000	1.2000
nonzero z/r^2	0.6536	0.7452	0.8139	0.8845	0.5043	0.4882	0.9035	0.8594
$P_{\xi_*}(Z=z)$	0.9486	0.9610	0.7624	0.8108	0.9100	0.9032	0.5881	0.6499
$\mu_1(\xi_*)$	0.1974	0.2279	0.1975	0.2283	0.1461	0.1404	0.1691	0.1778
e_-	0.7738	0.8437	0.8279	0.9030	0.6359	0.6187	0.7501	0.7748
e_+	1.4725	1.1503	1.2162	1.0464	3.5673	3.3450	2.3535	2.1593
$\min\lambda^2$	0.2258	0.0982	1.3005	0.5655	22.5781	9.8175	130.0499	56.5487

Note: Gaussian errors: $\tau_0=1$, $\tau_1=1.094$, $\tau_2=0.0859$, $\sigma^2(\mathbf{0})=1$, $V_{\psi,G}=1.1073$. Cauchy errors: $\tau_0=1.2281$, $\tau_1=0.7977$, $\tau_2=-0.2784$, $\sigma^2(\mathbf{0})=1.2965$, $V_{\psi,G}=2.6578$ ($V_{\psi,G}$, when estimate is adjusted for consistency at the Cauchy, is 2.5465).

Note that this is independent of δ^2. It is minimized, subject to (4.4), by the design given by (4.11), with $\mu_1(\xi_*)=\mu_1(\lambda)$ (yielding B-uniformity) and

$$t=\frac{\mu_1^2(\lambda)}{\mu_2(\mu_1(\lambda),\delta)}=\begin{cases}\dfrac{\mu_1^2(\lambda)}{\delta^2\mu_2(\lambda)}, & \delta^2\leqslant 1,\\[2ex] \dfrac{\mu_1^2(\lambda)}{\delta^2\sigma_\lambda^2+\mu_1^2(\lambda)}, & \delta^2\leqslant 1\end{cases}$$

for chosen $\delta^2=\mathrm{ch}_{\min}\,\mathbf{B}_{22.1}$.

4.3.2. $\delta_{\min}\to 0$

We have been unable to prove, but believe it to be true, that $\gamma(\mu_1(\xi),\delta)$ is always an increasing function of δ, so that $\delta_*=\delta_{\min}$. For such cases, which include all those in Table 1, $t\to 1$ as $\delta_{\min}\to 0$ and the limiting, 'no-power' design places all mass at

$$Z=\mu_1(\xi_*)=\underset{[0,r^2]}{\operatorname{argmin}}\,\gamma(\mu_1(\xi),0).$$

4.3.3. $\delta_{\min}\to\delta_{\max}$

Recall (4.5) and (4.6), and put

$$\delta_{\max}=\begin{cases}3\sqrt{5}/4, & p=1,\\ \frac{p+4}{p+2}, & p\geqslant 2.\end{cases}$$

As $\delta_{\min} \to \delta_{\max}$, J shrinks to the point $(\frac{3}{2}\mu_1(\lambda), \delta_{\max})$ if $p=1$, and to $((p+4)/(p+2)\mu_1(\lambda), \delta_{\max})$ if $p \geqslant 2$. The first component of this point must give $\mu_1(\xi_*)$. We then find that the limiting, *maximin power* design has

$$p=1: \quad P(Z=0)=P\left(Z=\frac{1}{4}\right)=\frac{1}{2},$$

$$p \geqslant 2: \quad P(Z=0)=\frac{\sigma_\lambda^2}{\mu_2(\lambda)}=\left(\frac{2}{p+2}\right)^2=1-P(Z=r^2).$$

4.3.4. $\eta \to 0$

As $\eta \to 0$,

$$\gamma(\mu_1(\xi), \delta) \to n^{-1} V_{\psi,G}\left(1+p\,\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right).$$

This is minimized by assigning to $\mu_1(\xi)$ its largest possible value. We denote this design, which is optimal only when the fitted model is exactly correct, by ξ_0. Then $\mu_1(\xi_0)=\mu_{\mathrm{U}}(\delta_{\min})$. Note that for $p \geqslant 2$, as at (4.6),

$$\delta_{\min} \leqslant \frac{p+4}{p+2}=\frac{r^2\mu_1(\lambda)}{\mu_2(\lambda)}.$$

Then $g(\delta_{\min}\mu_1(\lambda)) \geqslant \delta_{\min}^2$, whence $\delta_{\min}\mu_1(\lambda)$ lies between the roots of $g(\mu)=\delta_{\min}^2$. In particular, $\delta_{\min}\mu_1(\lambda) \leqslant \mu_1(\xi_0)$ and so from (4.9),

$$\mu_2(\mu_1(\xi_0), \delta_{\min})=\delta_{\min}^2\sigma_\lambda^2+\mu_{\mathrm{U}}^2(\delta_{\min}). \tag{4.16}$$

For $p=1$, (4.16) follows from (4.8). Now (4.16) in (4.12) gives $t=\mu_{\mathrm{U}}(\delta_{\min})/r^2$, and so

$$P_{\xi_0}(Z=0)=1-\frac{\mu_{\mathrm{U}}(\delta_{\min})}{r^2}=1-P_{\xi_0}(Z=r^2).$$

Choosing an optimal design entails choosing values of c, η and $\delta_{\min}$. How might one make such choices? Given c and η, the choice of δ_{min} may of course be made on the basis of the desired power of the test of $\boldsymbol{\alpha}=\mathbf{0}$, discussed above. Some values of the minimum (subject to (3.2)) noncentrality parameter λ^2 are given in Table 1. Where the error distribution is Cauchy, we have adjusted $\chi(0)$ so that $\sigma(\mathbf{0})=1$. For reference, a level 0.05 test of $\boldsymbol{\alpha}=\mathbf{0}$, on 3 d.f., has a power of 0.5 when $\lambda^2=6$ and a power of 0.98 when $\lambda^2=13$.

To choose c and η, one might compare the relative efficiencies of ξ_*, optimal for fixed $\eta>0$, with the ξ_0 of Section 4.3.4. When it is in fact the case that $\eta=0$, the relative efficiency of ξ_0 is

$$e_-:=\left.\frac{\mathscr{L}(f_\alpha, \xi_0)}{\mathscr{L}(f_\alpha, \xi_*)}\right|_{\alpha=0}=\left(1+p\,\frac{\mu_1(\lambda)}{\mu_{\mathrm{U}}(\delta_{\min})}\right)\Big/\left(1+p\,\frac{\mu_1(\lambda)}{\mu_1(\xi_*)}\right)<1.$$

This may be viewed as a premium to be paid for protection against positive η. When $\eta>0$ is the true value, a measure of the protection received for this premium is the relative efficiency, computed assuming maximal contamination for both designs,

$$e_+ := \frac{\max_{\|\alpha\|\leqslant\eta} \mathscr{L}(f_\alpha, \xi_0)}{\max_{\|\alpha\|\leqslant\eta} \mathscr{L}(f_\alpha, \xi_*)} = \frac{\gamma(\mu_U(\delta_{\min}), \delta_{\min})}{\gamma(\mu_*, \delta_*)} > 1.$$

The second equality above follows from the fact brought out in the proof of Theorem 4.1, that for any 2-point mass of the form (4.11) for some $\mu_1(\xi)$ and δ, the maximum loss is $\gamma(\mu_1(\xi), \delta)$. Values of e_- and e_+ are given in Table 1.

4.4. Proofs

The proof of Theorem 4.1 employs several intermediary results. We must first obtain the characteristic roots in (4.3) and (4.4).

Lemma 4.1. (i) *The characteristic roots of* $B_{22.1}$ *are*

$$\delta_0^2 = \sigma_\xi^2/\sigma_\lambda^2 \quad \textit{with multiplicity } 1, \textit{ and}$$

$$\delta_1^2 = \mu_2(\xi)/\mu_2(\lambda) \quad \textit{with multiplicity } q-1.$$

(ii) *The characteristic roots of* $B_{21}B_{11}^{-2}B_{12} + n^{-1}V_{\psi,G}\tau_1 P$ *are*

$$v_0 = \left\{(\mu_1(\xi)-\mu_1(\lambda))^2 + n^{-1}V_{\psi,G}\tau_1((1-\tau_2)\left(1+p\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)\sigma_\xi^2 + p\frac{\mu_1(\lambda)}{\mu_1^2(\xi)}\mu_3'(\xi)\right\}\Big/\sigma_\lambda^2,$$

with multiplicity 1, *and*

$$v_1 = \left\{n^{-1}V_{\psi,G}\tau_1((1-\tau_2)\left(1+p\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)\mu_2(\xi) + p\frac{\mu_1(\lambda)}{\mu_1^2(\xi)}\sigma_{12}(\xi)\right\}\Big/\mu_2(\lambda),$$

with multiplicity $q-1$.

(iii) *Any characteristic vector corresponding to* v_0 *must be proportional to* $\Sigma_\lambda^{1/2}\mathbf{j}$. *Any characteristic vector corresponding to* v_1 *must be orthogonal to* $\Sigma_\lambda^{-1/2}\mathbf{j}$.

Remark 4.1. Note that $\sup_\xi \sigma_\xi^2 = r^4/4$, and $\sup_\xi \mu_2(\xi) = r^4$. It then follows from Lemma 4.1(i) that

$$\sup_\xi \mathrm{ch}_{\max} B_{22.1} = (p+4)(p+2)^2/16p, \tag{4.17}$$

and so assumption (B1) guarantees (3.6), for all ξ.

Proof of Lemma 4.1. We first calculate

$$B_{11}=E_\xi[\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}]=1\oplus\frac{\mu_1(\xi)}{\mu_1(\lambda)}I_p,$$

$$B_{21}=E_\xi[\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}]=\frac{\mu_1(\xi)-\mu_1(\lambda)}{p}\begin{pmatrix}\Sigma_\lambda^{-1/2}\boldsymbol{j} & \vdots & \mathrm{O}\\ q\times 1 & & q\times p\end{pmatrix},$$

$$B_{22.1}=E_\xi[(\boldsymbol{v}-B_{21}B_{11}^{-1}\boldsymbol{u})(\boldsymbol{v}-B_{21}B_{11}^{-1}\boldsymbol{u})^{\mathrm{T}}]=\Sigma_\lambda^{-1/2}\Sigma_\xi\Sigma_\lambda^{-1/2},$$

$$P=\Sigma_\lambda^{-1/2}\Bigg\{\left((1-\tau_2)\left(1+p\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)-p\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)\Sigma_\xi$$
$$+p\frac{\mu_1(\lambda)}{\mu_1^2(\xi)}E\left[(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})\left(\operatorname{vec}_s\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\frac{\mu_1(\xi)}{p}I_p\right)\right)\right.$$
$$\left.\times\left(\left(\operatorname{vec}_s\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\frac{\mu_1(\xi)}{p}I_p\right)\right)^{\mathrm{T}}\right]\Bigg\}\Sigma_\lambda^{-1/2}.$$

To evaluate $B_{22.1}$ and P, we require some special matrices. Details of the following may be found in Henderson and Searle (1979), Magnus and Neudecker (1979) and Wiens (1985).

There is a matrix $G:p^2\times q$ defined uniquely by its action

$$G\operatorname{vec}_s M=\operatorname{vec}M\quad\text{for all symmetric } M:p\times p.$$

The Moore–Penrose inverse $G^+=(G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}}$ satisfies

$$G^+\operatorname{vec}M=\operatorname{vec}_s M\quad\text{for all symmetric } M:p\times p.$$

We have the identity

$$G^{\mathrm{T}}G:=D=\operatorname{diag}(1,2,\dots,2;1,2,\dots,2;\dots;1,2;1):q\times q.$$

There is a symmetric, orthogonal, permutation matrix $K:p^2\times p^2$ defined uniquely by its action

$$K\operatorname{vec}M=\operatorname{vec}M^{\mathrm{T}}\quad\text{for all } M:p\times p,$$

with the properties

$$GG^+=\tfrac{1}{2}(I+K),$$

$$G^+(I+K)G^{+\mathrm{T}}=(\tfrac{1}{2}D)^{-1}.$$

Put

$$e=\operatorname{vec}I_p.$$

The following identities hold:

$$E_\xi[\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]=\frac{\mu_1(\xi)}{p}\,\boldsymbol{e}, \tag{4.18}$$

$$E_\xi[(\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})(\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})^{\mathrm{T}}]=\frac{\mu_2(\xi)}{p(p+2)}\,(I+K+\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}), \tag{4.19}$$

$$E_\xi[(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]=\frac{\mu_2(\xi)}{p}\,\boldsymbol{e}, \tag{4.20}$$

$$E_\xi[(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})(\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})(\operatorname{vec} \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})^{\mathrm{T}}]=\frac{\mu_3(\xi)}{p(p+2)}\,(I+K+\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}). \tag{4.21}$$

The derivation of (4.18) is straightforward. For derivations of the expectations in (4.19)–(4.21), not assuming (B3), see Wiens (1992c). The expressions above have been simplified considerably by virtue of this assumption.

We now calculate

$$\begin{aligned}\Sigma_\xi &= G^+E\left[\left(\operatorname{vec}\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\frac{\mu_1(\xi)}{p}I_p\right)\right)\left(\operatorname{vec}\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\frac{\mu_1(\xi)}{p}I_p\right)\right)^{\mathrm{T}}\right]G^{+\mathrm{T}}\\ &=\frac{\mu_2(\xi)}{p(p+2)}\left(\frac{1}{2}D\right)^{-1}+\left(\frac{\mu_2(\xi)}{p(p+2)}-\left(\frac{\mu_1(\xi)}{p}\right)^2\right)\boldsymbol{j}\boldsymbol{j}^{\mathrm{T}},\end{aligned} \tag{4.22}$$

with an analogous expression for Σ_λ. The matrix $\Sigma_\lambda^{-1/2}$ is a linear combination of $(\frac{1}{2}D)^{1/2}$ and $\boldsymbol{j}\boldsymbol{j}^{\mathrm{T}}$. Upon solving for the coefficients, we then find

$$B_{22.1}=\frac{\mu_2(\xi)}{\mu_2(\lambda)}I_p+\frac{1}{p}\left(\frac{\sigma_\xi^2}{\sigma_\lambda^2}-\frac{\mu_2(\xi)}{\mu_2(\lambda)}\right)\boldsymbol{j}\boldsymbol{j}^{\mathrm{T}},$$

from which (i) in the statement of the lemma follows easily.

We next find

$$P=\Sigma_\lambda^{-1/2}\{\kappa_1(\tfrac{1}{2}D)^{-1}+\kappa_2\boldsymbol{j}\boldsymbol{j}^{\mathrm{T}}\}\Sigma_\lambda^{-1/2},$$

where

$$\kappa_1=(1-\tau_2)\left(1+p\,\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)\frac{\mu_2(\xi)}{p(p+2)}+\frac{\mu_1(\lambda)}{(p+2)\mu_1^2(\xi)}\,\sigma_{12}(\xi),$$

$$\begin{aligned}\kappa_2&=(1-\tau_2)\left(1+p\,\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)\left(\frac{\mu_2(\xi)}{p(p+2)}-\left(\frac{\mu_1(\xi)}{p}\right)^2\right)\\&\quad+\frac{\mu_1(\lambda)}{\mu_1^2(\xi)}\left(\frac{\sigma_{12}(\xi)}{p+2}-\frac{2\mu_1(\xi)\sigma_\xi^2}{p}\right).\end{aligned}$$

The characteristic roots ν and characteristic vectors $\boldsymbol{\alpha}$ of $B_{21}B_{11}^{-2}B_{12}+n^{-1}V_{\psi,G}\tau_1 P$ are then solutions to

$$\begin{aligned}\mathbf{0}&=\Sigma_\lambda^{-1/2}\left\{\left(\frac{\mu_1(\xi)-\mu_1(\lambda)}{p}\right)^2 \boldsymbol{jj}^{\mathrm{T}}+n^{-1}V_{\psi,G}\tau_1\left(\kappa_1\left(\frac{1}{2}D\right)^{-1}\right.\right.\\&\qquad\left.\left.+\kappa_2\boldsymbol{jj}^{\mathrm{T}}\right)-\nu\Sigma_\lambda\right\}\Sigma_\lambda^{-1/2}\boldsymbol{\alpha}\\&=\Sigma_\lambda^{-1/2}\left\{\rho_1\left(\frac{1}{2}D\right)^{-1}+\rho_2\boldsymbol{jj}^{\mathrm{T}}\right\}\Sigma_\lambda^{-1/2}\boldsymbol{\alpha},\end{aligned}\tag{4.23}$$

where

$$\rho_1=n^{-1}V_{\psi,G}\tau_1\kappa_1-(\nu\mu_2(\lambda))/(p(p+2)),$$

$$\rho_2=n^{-1}V_{\psi,G}\tau_1\kappa_2+\left(\frac{\mu_1(\xi)-\mu_1(\lambda)}{p}\right)^2-\nu\left(\frac{\mu_2(\lambda)}{p(p+2)}-\left(\frac{\mu_1(\lambda)}{p}\right)^2\right).$$

The characteristic roots are obtained from

$$0=\left|\rho_1\left(\frac{1}{2}D\right)^{-1}+\rho_2\boldsymbol{jj}^{\mathrm{T}}\right|=2^p\rho_1^{q-1}\left(\rho_1+\frac{p}{2}\rho_2\right).$$

If $\rho_1+(p/2)\rho_2=0$, then $\nu=\nu_0$. If $\rho_1=0$, then $\nu=\nu_1$. This proves (ii) in the statement of the lemma. For (iii), we note that if $\nu=\nu_0$, then from (4.23),

$$\left(I_q-\frac{1}{p}\boldsymbol{jj}^{\mathrm{T}}\right)\Sigma_\lambda^{-1/2}\boldsymbol{\alpha}=\mathbf{0},$$

whence $\Sigma_\lambda^{-1/2}\boldsymbol{\alpha}$ is proportional to $\boldsymbol{j}$. If $\nu=\nu_1$, then (4.23) requires $\boldsymbol{\alpha}$ to be orthogonal to $\Sigma_\lambda^{-1/2}\boldsymbol{j}$. □

The problem is now to minimize

$$\max_{\|\boldsymbol{\alpha}\|\leqslant\eta}\mathscr{L}(f_{\boldsymbol{\alpha}},\xi)=n^{-1}V_{\psi,G}\left(1+p\frac{\mu_1(\lambda)}{\mu_1(\xi)}\right)+\eta^2(1+\max(\nu_0,\nu_1)),\tag{4.24}$$

subject to

$$\min(\delta_0^2,\delta_1^2)=\delta^2\geqslant\delta_{\min}^2.$$

We do this first for *fixed* δ^2 *and* $\mu_1(\xi)$. This fixes the first two moments of Z, under ξ. From (4.24), we are then to minimize max (ν_0,ν_1). But from Lemma 4.1(ii), both ν_0 and ν_1 are minimized, for fixed $\mu_1(\xi)$, $\mu_2(\xi)$ and $\sigma^2(\xi)$, by that ξ which has the smallest third moment.

Define

$$W=\frac{Z-\mu_1(\xi)}{r^2-\mu_1(\xi)},\qquad a=\frac{\mu_1(\xi)}{r^2-\mu_1(\xi)}.$$

Then $-a \leqslant W \leqslant 1$, $E[W]=0$, and we seek to minimize $E[W^3]$, subject to $E[W^2]$ being fixed.

Lemma 4.2. *Let W be a random variable with support in $[-a, 1]$, $a>0$.*

(i) *If $E[W]=0$, then $E[W^2] \leqslant a$.*

(ii) *If $E[W]=0$ and $E[W^2]=\sigma^2 \leqslant a$, then*

$$E[W^3] \geqslant \frac{\sigma^2}{a}(\sigma^2 - a^2), \tag{4.25}$$

with equality iff

$$P(W=-a)=\frac{\sigma^2}{\sigma^2+a^2}=1-P\left(W=\frac{\sigma^2}{a}\right). \tag{4.26}$$

Proof. (i) If $-a \leqslant W \leqslant 1$, then

$$0 \geqslant (W+a)(W-1) = W^2 + (a-1)W - a.$$

Taking expectations gives (i).

(ii) If $-a \leqslant W \leqslant 1$, then

$$0 \leqslant (W+a)\left(W-\frac{\sigma^2}{a}\right)^2 = W^3 + \left(a - \frac{2\sigma^2}{a}\right)W^2 + \left(\frac{\sigma^4}{a^2} - 2\sigma^2\right)W + \frac{\sigma^4}{a}.$$

Taking expectations gives (4.25). Equality holds in (4.25) iff $P(W \in \{-a, \sigma^2/a\})=1$, and then the condition $E[W]=0$ gives (4.26). $\square$

Corollary 4.1. *For fixed values of $\mu_1(\xi)$ and $\mu_2(\xi)$, both $\mu_3'(\xi)$ and $\sigma_{12}(\xi)$ are minimized in ξ if*

$$P_\xi(Z=0)=1-t=1-P_\xi\left(Z=\frac{\mu_1(\xi)}{t}\right),$$

where $t=\mu_1^2(\xi)/\mu_2(\xi)$.

Remark 4.2. Lemma 4.2(i) is equivalent to each of

$$\frac{\mu_1(\xi)}{t} \leqslant r^2,$$

$$\sigma_\xi^2 \leqslant \mu_1(\xi)(r^2 - \mu_1(\xi)), \tag{4.27}$$

$$\mu_2(\xi) \leqslant \mu_1(\xi) r^2. \tag{4.28}$$

Corollary 4.2. *In order that δ^2 be an attainable value of $\mathrm{ch}_{\min} B_{22.1}$ under ξ, it is necessary to have $(\mu_1(\xi), \delta) \in J$. Furthermore, for $p \geqslant 2$ and $(\mu_1(\xi), \delta) \in J$: (i) $\delta^2 = \delta_0^2$ iff $\delta \mu_1(\lambda) \leqslant \mu_1(\xi) \leqslant \mu_U(\delta)$ and (ii) $\delta^2 = \delta_1^2$ iff $\delta^2 \mu_2(\lambda)/r^2 \leqslant \mu_1(\xi) \leqslant \delta \mu_1(\lambda)$.*

Proof. If $p=1$, then by Lemma 4.1, (4.27) and (4.1) we have

$$\delta^2=\frac{\sigma_\xi^2}{\sigma_\lambda^2}\leqslant\frac{\mu_1(\xi)(r^2-\mu_1(\xi))}{\sigma_\lambda^2}=g(\mu_1(\xi))\leqslant\frac{r^4}{4\sigma_\lambda^2}=\frac{45}{16},$$

requiring $(\mu_1(\xi),\delta)\in J$.

If $p\geqslant 2$, then by Lemma 4.2(i) and (4.27), (4.28) we have

$$\delta_1^2\leqslant r^2\frac{\mu_1(\xi)}{\mu_2(\lambda)} \tag{4.29}$$

and $\delta_0^2\leqslant g(\mu_1(\xi))$, i.e.

$$\mu_L(\delta_0)\leqslant\mu_1(\xi)\leqslant\mu_U(\delta_0). \tag{4.30}$$

Note that '$\delta_0^2\leqslant\delta_1^2$' is equivalent to each inequality '$\delta_0\mu_1(\lambda)\leqslant\mu_1(\xi)$' and '$\delta_1\mu_1(\lambda)\leqslant\mu_1(\xi)$', so that

$$\delta_0^2\leqslant\delta_1^2 \quad \text{iff} \quad \delta\mu_1(\lambda)\leqslant\mu_1(\xi). \tag{4.31}$$

Now (i) follows from (4.30) and (4.31) and (ii) from (4.29) and (4.31). By (ii), if $\delta=\delta_1$ then

$$\delta\leqslant\frac{r^2\mu_1(\lambda)}{\mu_2(\lambda)}=\frac{p+4}{p+2}. \tag{4.32}$$

It remains to establish this inequality when $\delta=\delta_0$. For this, it suffices to show that, when $\delta=\delta_0$, we have

$$\mu_L(\delta)\leqslant\delta\mu_1(\lambda). \tag{4.33}$$

Then, together with (i), we will have that $\delta\mu_1(\lambda)\in[\mu_L(\delta),\mu_U(\delta)]$, so that $g(\delta\mu_1(\lambda))\geqslant\delta^2$. This is (4.32).

Suppose that (4.33) fails. Then

$$\begin{aligned}0<g'(\delta\mu_1(\lambda))&=\frac{g(\delta\mu_1(\lambda))}{\delta\mu_1(\lambda)}-\frac{\delta\mu_1(\lambda)}{\sigma_\lambda^2}\\&<\delta\left(\frac{1}{\mu_1(\lambda)}-\frac{\mu_1(\lambda)}{\sigma_\lambda^2}\right)\\&=-\frac{\delta(p+2)}{pr^2}\left(\frac{p(p+4)}{4}-1\right)\\&<0,\end{aligned}$$

which is the desired contradiction. □

Proof of Theorem 4.1. By the discussion preceding Lemma 4.2, we may restrict to measures ξ as in Corollary 4.1. For such measures,

$$\mu_3'(\xi)=\sigma_\xi^2(\sigma_\xi^2-\mu_1^2(\xi))/\mu_1(\xi), \tag{4.34}$$

$$\sigma_{12}(\xi)=\sigma_\xi^2(\sigma_\xi^2+\mu_1^2(\xi))/\mu_1(\xi). \tag{4.35}$$

Note also that by virtue of Lemma 4.1(i) for $p=1$, and Corollary 4.2(i) and (ii) for $p\geqslant 2$,

$$\mu_2(\xi)=\mu_2(\mu_1(\xi),\delta), \tag{4.36}$$

$$\sigma_\xi^2=\sigma^2(\mu_1(\xi),\delta) \tag{4.37}$$

in the notation of (4.8) and (4.9). Now substitute (4.34)–(4.37) into v_0, v_1 of Lemma 4.1(i), obtaining $\bar{v}_0(\mu_1(\xi),\delta)$ and $\bar{v}_1(\mu_1(\xi),\delta)$. As at (4.24), we now have that for fixed $\mu_1(\xi)$ and δ,

$$\min_{\xi}\max_{\|\alpha\|\leqslant\eta}\mathscr{L}(f_\alpha,\xi)=\gamma(\mu_1(\xi),\delta),$$

with $\gamma(\mu_1(\xi),\delta)$ as at (4.10). We now minimize over $(\mu_1(\xi),\delta)\in J$, obtaining (4.11)–(4.13).

To prove Theorem 4.1(iii) note that by (4.2),

$$f_\alpha(\boldsymbol{x})=\boldsymbol{\alpha}^{\mathrm{T}}\Sigma_\lambda^{-1/2}\left(\operatorname*{vec}_s \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\frac{\mu_1(\lambda)}{p}\boldsymbol{j}\right) \tag{4.38}$$

and that $\mathscr{L}(f_\alpha,\xi_*)$ is maximized by an f_α for which $\boldsymbol{\alpha}$ is a characteristic vector of $B_{21}B_{11}^{-2}B_{12}+n^{-1}V_{\psi,G}\tau_1 P$, corresponding to the maximum root $\bar{v}$ and normalized by $\int f_\alpha^2\,\mathrm{d}\boldsymbol{x}=\eta^2$. If $\bar{v}=\bar{v}_0$, then by Lemma 4.1(iii) the maximizing f_α has $\Sigma_\lambda^{-1/2}\,\boldsymbol{\alpha}$ proportional to $\boldsymbol{j}$, whence f_α is proportional to $\Sigma x_i^2-\mu_1(\lambda)$. Normalizing gives (4.14).

If $\bar{v}=\bar{v}_1$ then the maximizing f_α satisfies (4.38) and

$$\boldsymbol{\alpha}^{\mathrm{T}}\Sigma_\lambda^{-1/2}\boldsymbol{j}=0.$$

Equivalently, $f_\alpha(\boldsymbol{x})$ is a linear combination of the x_ix_j $(i\leqslant j)$, without constant terms, satisfying $E_\lambda[f_\alpha(\boldsymbol{X})]=0$. An example, orthogonal to f_α^0, is the f_α^1 of (4.15).

The proof of Theorem 4.1(iv) is omitted — it involves only a lengthy calculation, using techniques similar to those in Lemma 4.1.

References

Box, G.E.P. and N. Draper (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622–654.

Box, G.E.P. and J.S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Statist.* **28**, 195–241.

Draper, N.R. (1960). Third order rotatable designs in three dimensions. *Ann. Math. Statist.* **31**, 865–874.

Fedorov, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Hampel, F.R., R. Ronchetti, R.J. Rousseeuw and W. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

Henderson, H.V. and S.R. Searle (1979). Vec and vech operators for matrices with some uses in Jacobians and multivariate statistics, *Canad. J. Statist.* **7**, 65–81.

Herzberg, A. (1988). Rotatable designs. In: *Encyclopedia of Statistical Sciences*, Vol. 8. Wiley, New York, 187–189.

Huber, J. (1975). Robustness and designs. In: Srivastava, J.N., Ed., *A Survey of Statistical Design and Linear Models*. North-Holland, Amsterdam, 287–303.

Huber, J. (1981). *Robust Statistics*. Wiley, New York.

Li, K.C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12**, 269–282.

Li, K.C. and W. Notz (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135–151.

Magnus, J.R. and H. Neudecker (1979). The commutation matrix: some properties and applications. *Ann. Statist.* **7**, 381–394.

Marcus, M.B. and J. Sacks (1976). Robust designs for regression problems. In: S.S. Gupta and D.S. Moore, Eds., *Statistical Theory and Related Topics II*. Academic Press, New York, 245–268.

Notz, W. (1989). Optimal designs for regression models with possible bias. *J. Statist. Plann. Inference* **22**, 43–54.

Pesotchinsky, L. (1982). Optimal robust designs: linear regression in R^k. *Ann. Statist.* **10**, 511–525.

Sacks, J. and D. Ylvisaker (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6**, 1122–1137.

Silvapullé, M.J. (1985). Asymptotic behavior of robust estimates of regression and scale parameters with fixed carriers. *Ann. Statist.* **13**, 1490–1487.

Street, J.O., R.J. Carroll and D. Ruppert (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *Amer. Statist.* **42**, 152–154.

Wiens, D.P. (1985). On some pattern reduction matrices which appear in statistics. *Linear Algebra Appl.* **67**, 233–258.

Wiens, D.P. (1990). Robust minimax designs for multiple linear regression. *Linear Algebra Appl.* **127**, 327–340.

Wiens, D.P. (1991). Designs for approximately linear regression: two optimality properties of uniform designs. *Statist. Probab. Lett.* **12**, 217–221.

Wiens, D.P. (1992a). A note on the computation of robust bounded influence estimates and test statistics in regression. *Comput. Statist. Data Anal.* **13**, 211–220.

Wiens, D.P. (1992b). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference* **31**, 353–371.

Wiens, D.P. (1992c). On moments of quadratic forms in non-spherically distributed variables, *Statistics* **23**, 265–270.

Wiens, D.P. (1993). Designs for approximately linear regression: maximizing the minimum coverage probability of cofidence ellipsoids. *Canad. J. Statist.* **21**, 59–70.