On the irregular behavior of LS estimators for asymptotically singular designs ⋆

Andrej Pázman [a] and Luc Pronzato [b,*]

[a] *Department of Applied Mathematics and Statistics*
Mlynská Dolina, 84248 Bratislava, Slovakia

[b] *Laboratoire I3S, CNRS/UNSA, Bât. Euclide, Les Algorithmes,*
2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France

Abstract

Optimum design theory sometimes yields singular designs. An example with a linear regression model often mentioned in the literature is used to illustrate the difficulties induced by such designs. The estimation of the model parameters θ, or of a function of interest $h(\theta)$, may be impossible with the singular design ξ^*. Depending on how ξ^* is approached by the empirical measure ξ^n of the design points, with n the number of observations, consistency is achieved but the speed of convergence may depend on ξ^n and on the value of θ. Even in situations where convergence is in $1/\sqrt{n}$ and the asymptotic distribution of the estimator of θ or $h(\theta)$ is normal, the asymptotic variance may still differ from that obtained from ξ^*.

Key words: singular design, optimum design, asymptotic normality, consistency, LS estimation
PACS: 62K05, 62E20

⋆ The research of the 1st author has been supported by the VEGA-grant nb. 1/0264/03. The research of the 2nd author has been supported in part by the IST Programme of the European Community under the PASCAL network of Excellence IST2002506778. This publication only reflects the authors views.
* author for correspondence
Email addresses: pazman@fmph.uniba.sk (Andrej Pázman), pronzato@i3s.unice.fr (Luc Pronzato).

1 Introduction

We consider the following linear regression model

$$\eta(x,\theta) = \theta_1 x + \theta_2 x^2 = \mathbf{f}^\top(x)\theta \tag{1}$$

with $\mathbf{f}(x) = (x,\ x^2)^\top$, $x \in \mathcal{X} = [-1,1]$ and observations $y_k = \eta(x_k,\bar{\theta}) + \varepsilon_k$ where $\bar{\theta}$ is the (unknown) true value of the model parameters and the errors ε_k are i.i.d. with zero mean and variance σ^2. We shall take $\sigma = 1$ throughout the paper. We shall denote $\hat{\theta}^n$ the LS estimator of θ obtained from the observations $y_1, y_2, \ldots, y_n$; ξ_n will denote the empirical measure of the associated design points $x_1, x_2, \ldots, x_n$. We shall also denote

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}^\top(x)\,\xi(dx)$$

the information matrix for a design measure ξ.

We assume that $\bar{\theta}_1 \geq 0$ and $\bar{\theta}_2 < 0$ and are interested in the estimation of

$$h(\theta) = -\frac{\theta_1}{2\theta_2}\,, \tag{2}$$

the value of x where $\eta(x,\theta)$ is maximal. When the design space is $\mathcal{X} = [-1,1]$, the optimum design measure $\xi^* = \xi^*_\theta$ for the estimation of $h(\theta)$ (c-optimality) has its support included in $\{-1,1\}$, the weight of each point depending on the value of $h(\theta)$ (a standard situation in nonlinear problems). One can show, see, e.g., Silvey (1980, p. 57), that

$$\xi^*(1) = \begin{cases} \frac{1}{2} + \frac{1}{4h} & \text{if } h \geq \frac{1}{2}\,, \\ \frac{1}{2} + h & \text{if } 0 \leq h \leq \frac{1}{2}\,, \end{cases} \tag{3}$$

so that when $h(\theta) = 1/2$ the optimum design is singular with $\xi^*(1) = 1$ (and $\xi^*(-1) = 0$). Therefore, if we know *a priori* that $h(\bar{\theta})$ is close to 1/2 we should put the design points, or the majority of them, close to 1. It is the purpose of this paper to show that depending how this is realized, the asymptotic behavior of $\hat{\theta}^n$, or of $h(\hat{\theta}^n)$, may have some unexpected features. Note that $h(\theta)$ is not estimable when ξ^* is singular. Therefore, when $h(\hat{\theta}^n)$ obtained with the design ξ_n converges to $h(\bar{\theta})$, $n \to \infty$, it means that the sequence $x_1, x_2, \ldots$ itself, not the limiting design ξ^*, is responsible for consistency. In particular, it thus seems legitimate to question the adjective "optimum" for ξ^*.

The example considered is extremely simple but the conclusions are of general consequences: we show that it is only in very particular circumstances that approaching a singular "optimum" design conveys some optimal properties to the nonlinear LS estimation for which it was designed. The example has been chosen due to its frequent use in the optimum-design literature, see e.g. Silvey (1980); Ford and Silvey (1980); Ford et al. (1985); Wu (1985). Here, the singularity of ξ^* is obtained for a particular value of $h(\theta)$. This should not give the reader the impression that singular "optimum" designs are exceptional. Indeed, consider the situation where $\eta(x,\theta) = \theta_1 + \theta_2 x$ and one wishes to estimate $h'(\theta) = -\theta_1/\theta_2$, the value of x for which $\eta(x,\theta) = 0$. Here the "optimum" design ξ^* for the LS estimation of h' is singular for any θ, with $\xi^*[h'(\theta)] = 1$. Also, the estimation of $h(\theta) = -\theta_1/(2\theta_2)$ in the full quadratic regression model $\eta(x,\theta) = \theta_0 + \theta_1 x + \theta_2 x^2$ yields singular "optimum" designs for a full range of values of h, see Chaloner (1989); Fedorov and Müller (1997). See also Buonaccorsi and Iyer (1986) for the estimation of ratios of linear combinations of the parameters.

Sections 2 and 3 concern the situation where we know *a priori* that $h(\bar{\theta})$ is close to 1/2 and non-sequential designs approaching the singular measure ξ^* are used: ξ_n converges weakly to ξ^* in Section 2 whereas strong convergence is considered in Section 3. The sequential construction of the design is briefly discussed in Section 4. Throughout the paper we denote

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

2 ξ_n converges weakly to ξ^*

By weak convergence we mean convergence in distribution, which we denote $\xrightarrow{\text{w}}$. Let ξ^* be the singular measure that puts weight 1 at $x = 1$. Throughout the section we use the design measure ξ_n constructed from

$$x_i = \begin{cases} 1 & \text{if } i = 2k-1\,, \\ 1 + (1/k)^{1/4} & \text{if } i = 2k\,, \end{cases}$$

for $k = 1, 2, \ldots$ so that $\xi_n \xrightarrow{\text{w}} \xi^*$, $n \to \infty$.

2.1 Consistency.

From Corollary 1 of (Wu, 1980), $\mathbf{u}^\top\hat{\theta}^n \stackrel{\text{a.s.}}{\rightarrow} \mathbf{u}^\top\bar{\theta}$ for any $\mathbf{u} \in \mathbb{R}^2$ when $S^\infty(\mathbf{w}) = \sum_{i=1}^\infty [\mathbf{w}^\top \mathbf{f}(x_i)]^2 = \infty$ for all $\mathbf{w} = (w_1\ w_2)^\top \neq \mathbf{0}$. Here we obtain

$$S^\infty(\mathbf{w}) = \sum_{k=1}^\infty (w_1+w_2)^2 + \sum_{k=1}^\infty \left\{ w_1[1+1/k^{1/4}] + w_2[1+1/k^{1/4}]^2 \right\}^2$$

so that $S^\infty(\mathbf{w}) = \infty$ when $w_1 + w_2 \neq 0$. For $w_1 + w_2 = 0$ (and $w_1 \neq 0$ since $\mathbf{w} \neq \mathbf{0}$) we have

$$S^\infty(\mathbf{w}) = w_1^2 \sum_{k=1}^\infty [1/k^{1/2} + 1/k^{1/4}]^2 > w_1^2 \sum_{k=1}^\infty 1/k = \infty \,.$$

Therefore $\mathbf{u}^\top\hat{\theta}^n \stackrel{\text{a.s.}}{\rightarrow} \mathbf{u}^\top\bar{\theta}$ for any $\mathbf{u} \in \mathbb{R}^2$ so that $\hat{\theta}^n \stackrel{\text{a.s.}}{\rightarrow} \bar{\theta}$ and $h(\hat{\theta}^n) \stackrel{\text{a.s.}}{\rightarrow} h(\bar{\theta})$, $n \rightarrow \infty$.

2.2 Asymptotic normality of $\mathbf{u}^\top\hat{\theta}^n$.

This paragraph is auxiliary to the investigation of the asymptotic distribution of $h(\hat{\theta}^n)$.

Consider the case $\mathbf{u} = \mathbf{1}$. When the design ξ^* is used, all design points $x_i = 1$, but $\mathbf{1}^\top\hat{\theta}^n$ is estimable in spite of the singularity of ξ^* since $\mathbf{1}$ is in the range of

$$\mathbf{M}(\xi^*) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} .$$

The variance of $\mathbf{1}^\top\hat{\theta}^n$, which we denote $\text{var}_{\xi^*}(\mathbf{1}^\top\hat{\theta}^n)$, then satisfies

$$n \, \text{var}_{\xi^*}(\mathbf{1}^\top\hat{\theta}^n) = \mathbf{1}^\top \mathbf{M}^-(\xi^*)\mathbf{1} = 1$$

with $\mathbf{M}^-$ any g-inverse of $\mathbf{M}$. On the other hand, the variance of $\mathbf{1}^\top\hat{\theta}^n$ for the design ξ_n satisfies

$$\lim_{n\rightarrow\infty} n \, \text{var}_{\xi_n}(\mathbf{1}^\top\hat{\theta}^n) = 9/5 \neq \mathbf{1}^\top \mathbf{M}^-(\xi^*)\mathbf{1} \,, \tag{4}$$

where $n \, \text{var}_{\xi_n}(\mathbf{1}^\top\hat{\theta}^n) = \mathbf{1}^\top \mathbf{M}^{-1}(\xi_n)\mathbf{1}$. Indeed, take $n = 2m$, then

$$\mathbf{M}(\xi_n) = \begin{pmatrix} \mu_2(n) & \mu_3(n) \\ \mu_4(n) & \mu_5(n) \end{pmatrix}$$

with $\mu_i(n) = (1/n)[m + \sum_{k=1}^{m}(1+k^{-1/4})^i]$. We then obtain (4) by direct calculations. The difference between $\lim_{n\to\infty} n \,\mathrm{var}_{\xi^*}(\mathbf{1}^\top\hat{\theta}^n)$ and $\lim_{n\to\infty} n \,\mathrm{var}_{\xi_n}(\mathbf{1}^\top\hat{\theta}^n)$ is due to the discontinuity of the function $\mathbf{M}(\xi) \mapsto n \,\mathrm{var}_{\xi}(\mathbf{1}^\top\hat{\theta}^n)$, see Pázman (1980).

Next, following the same lines as in (Huber, 1973), we can show that Lindeberg's condition is satisfied, and for any direction $\mathbf{u} \neq \mathbf{0}$

$$\sqrt{n}\frac{\mathbf{u}^\top(\hat{\theta}^n - \bar{\theta})}{\sqrt{\mathbf{u}^\top\mathbf{M}^{-1}(\xi_n)\mathbf{u}}} \stackrel{\mathrm{d}}{\to} \zeta_u \sim \mathcal{N}(0,1)\,.$$

For $\mathbf{u} = \mathbf{1}$ it gives

$$\sqrt{n}\,\mathbf{1}^\top(\hat{\theta}^n - \bar{\theta}) \stackrel{\mathrm{d}}{\to} \zeta_1 \sim \mathcal{N}(0, 9/5)\,,$$

but for a direction $\mathbf{u}$ such that $(\mathbf{u}^\top\mathbf{1})^2 \neq 2\|\mathbf{u}\|^2$ (i.e., not parallel to $\mathbf{1}$) the convergence of $\mathbf{u}^\top(\hat{\theta}^n - \bar{\theta})$ is in $n^{-1/4}$ since $\mathbf{u}^\top\mathbf{M}^{-1}(\xi_n)\mathbf{u}$ grows as $\sqrt{n}$ (note that $\mathbf{u}^\top\theta$ is not estimable from the limiting design ξ^*). In particular, one can check that

$$n^{1/4}\,\mathbf{u}^\top(\hat{\theta}^n - \bar{\theta}) \stackrel{\mathrm{d}}{\to} \zeta_1 \sim \mathcal{N}(0, 9/5)$$

for $\mathbf{u} = (0,\ 1)^\top$ or $(1,\ 0)^\top$.

Hence, when $\mathbf{u}^\top\theta$ is estimable under the limiting design ξ^*, $\mathbf{u}^\top\hat{\theta}^n$ converges as $1/\sqrt{n}$ but the limiting variance differs from $\mathbf{u}^\top\mathbf{M}^-(\xi^*)\mathbf{u}$; when $\mathbf{u}^\top\theta$ is not estimable under ξ^* ($\mathbf{u}$ is not in the range of $\mathbf{M}(\xi^*)$), then $\mathbf{u}^\top\hat{\theta}^n$ converges as $n^{-1/4}$.

2.3 Asymptotic normality of $h(\hat{\theta}^n)$.

Consider now the estimation of $h(\theta)$ given by (2).

When $\bar{\theta}_1 + \bar{\theta}_2 \neq 0$, we have

$$h(\hat{\theta}^n) = h(\bar{\theta}) + (\hat{\theta}^n - \bar{\theta})^\top \left[\frac{\partial h(\theta)}{\partial\theta}_{|\bar{\theta}} + o_p(1)\right]$$

where $\partial h(\theta)/\partial\theta = -1/(2\theta_2)[1,\ 2h(\theta)]^\top$, so that $\partial h(\theta)/\partial\theta_{|\bar{\theta}}$ is not parallel to $\mathbf{1}$. Therefore, $n^{1/4}[h(\hat{\theta}^n) - h(\bar{\theta})]$ has the same limiting distribution as

$$-\frac{1}{2\bar{\theta}_2} n^{1/4}[1,\ 2h(\bar{\theta})](\hat{\theta}^n - \bar{\theta}) \stackrel{\mathrm{d}}{\to} \zeta_2 \sim \mathcal{N}(0, v_{\bar{\theta}}) \qquad (5)$$

with

$$v_{\bar\theta} = 1/(4\bar\theta_2^2) \lim_{n\to\infty} (1/\sqrt{n})[1,\ 2h(\bar\theta)]\mathbf{M}^{-1}(\xi_n)[1,\ 2h(\bar\theta)]^\top = \frac{9}{5}\,\frac{[2h(\bar\theta)-1]^2}{4\bar\theta_2^2}\,.$$

$h(\hat\theta^n)$ is thus asymptotically normal, but converges as $n^{-1/4}$.

When $\bar\theta_1 + \bar\theta_2 = 0$, we write

$$h(\hat\theta^n) = h(\bar\theta) + (\hat\theta^n - \bar\theta)^\top \frac{\partial h(\theta)}{\partial \theta}_{|\bar\theta} + \frac{1}{2}(\hat\theta^n - \bar\theta)^\top \left[\frac{\partial^2 h(\theta)}{\partial\theta\partial\theta^\top}_{|\bar\theta} + o_p(1)\right](\hat\theta^n - \bar\theta)$$

with

$$\frac{\partial h(\theta)}{\partial \theta}_{|\bar\theta} = -\frac{1}{2\bar\theta_2}\mathbf{1} \text{ and } \frac{\partial^2 h(\theta)}{\partial\theta\partial\theta^\top}_{|\bar\theta} = \frac{1}{2\bar\theta_2^2}\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

Direct calculations give

$$\sqrt{n}(\hat\theta^n - \bar\theta)^\top \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} (\hat\theta^n - \bar\theta) = o_p(1)$$

so that

$$\sqrt{n}[h(\hat\theta^n) - h(\bar\theta)] \xrightarrow{\text{d}} \zeta_3 \sim \mathcal{N}(0, \frac{9}{5}\,\frac{1}{4\bar\theta_2^2})\,. \tag{6}$$

$h(\hat\theta^n)$ is thus asymptotically normal and converges as $1/\sqrt{n}$, but its asymptotic variance differs from the value $1/(4\bar\theta_2^2) = \partial h(\theta)/\partial\theta_{|\bar\theta}^\top \mathbf{M}^-(\xi^*)\partial h(\theta)/\partial\theta_{|\bar\theta}$.

3 ξ_n converges strongly to ξ^*

By strong convergence we mean that $\lim_{n\to\infty} \xi_n(x) = \xi^*(x)$ for all $x \in \mathcal{X}$, ξ^* being the limiting *discrete* design. In this section we consider different simple examples of strongly converging ξ_n and study the asymptotic properties of estimators. The first example corresponds to a design generated by an optimisation algorithm.

3.1 Steepest descent algorithm.

Consider the steepest descent algorithm (Wynn, 1972) for the construction of an optimum design for the estimation of $\mathbf{1}^\top\theta$ in the model (1). The optimum design ξ^* on $\mathscr{X} = [-1, 1]$ is singular with $\xi^*(1) = 1$ (and $\mathbf{1}^\top\theta$ is estimable for ξ^*). It is well known that the algorithm generates a sequence of points such that ξ_n converges to the optimum, in the sense that $\lim_{n\to\infty} \mathbf{1}^\top\mathbf{M}^{-1}(\xi_n)\mathbf{1} = \mathbf{1}^\top\mathbf{M}^{-}(\xi^*)\mathbf{1}$. We show by elementary calculus that ξ_n converges strongly to ξ^*, in contrast with the situation considered in Section 2.

Take x_1, x_2 such that $\mathbf{M}(\xi_2)$ is non singular. By construction, $\mathbf{M}(\xi_k)$ is then non singular for all k and the design sequence is such that

$$x_{k+1} = \arg\max_{x\in[-1,1]} \left[\mathbf{1}^\top\mathbf{M}^{-1}(\xi_k)\begin{pmatrix} x \\ x^2 \end{pmatrix}\right]^2 , \tag{7}$$

see eq. 4.1 in (Wynn, 1972). Straightforward calculation shows that x_{k+1} maximizes

$$x^2\sum_{i=1}^{k}(x_i^2 - x_i^3) + x\sum_{i=1}^{k}(x_i^4 - x_i^3) .$$

This quadratic function is minimum at $x_k^* = -S_k/(2S'_k)$, with $S_k = \sum_{i=1}^k x_i^3(x_i - 1)$ and $S'_k = \sum_{i=1}^k x_i^2(1 - x_i)$. Note that $S'_k > 0$. It thus reaches its maximum at $x = \pm 1$ and

$$x_{k+1} = \begin{cases} 1 & \text{if } S_k > 0 , \\ -1 & \text{otherwise} . \end{cases}$$

When $x_{k+1} = -1$, $S_{k+1} = 2 + S_k$ so that S_k ultimately becomes positive and x_{j+1} equals 1 for some j. When this happens, $S_{j+1} = S_j$ and $x_i = 1$ for all subsequent i, $i = j+1, j+2, \ldots$ The number of observations at $x \neq 1$ is thus finite. The design measure ξ_n converges strongly to ξ^* and $\lim_{n\to\infty} n\text{var}(\mathbf{1}^\top\hat{\theta}^n) = \mathbf{1}^\top\mathbf{M}^{-}(\xi^*)\mathbf{1} = 1$. Notice the difference with (4).

The method of steepest-descent for designing an optimal experiment for the estimation of $h(\theta)$ in the model (1) minimizes $\partial h(\theta)/\partial\theta^\top_{|\bar{\theta}}\mathbf{M}^{-}(\xi)\partial h(\theta)/\partial\theta_{|\bar{\theta}}$ and is based on the iterations

$$x_{k+1} = \arg\max_{x\in[-1,1]} \left[\frac{\partial h(\theta)}{\partial\theta^\top}_{|\bar{\theta}}\mathbf{M}^{-1}(\xi_k)\begin{pmatrix} x \\ x^2 \end{pmatrix}\right]^2 . \tag{8}$$

For $h(\theta)$ given by (2), when $\bar{\theta}_1 + \bar{\theta}_2 \neq 0$ the limiting optimum design is non singular and there are no difficulties. When $\bar{\theta}_1 + \bar{\theta}_2 = 0$, the iterations are given by (7) and ξ_n converges strongly to ξ^* which is singular. Moreover, from the results above the number of observations at $x \neq 1$ is finite. It is this type of situation that we investigate below in more details.

In the rest of the section we consider the estimation of $h(\theta)$ for different cases of measures that converge strongly to ξ^*. Suppose that m observations are performed at $x = z$ for some $z \in [-1, 1]$, $z \neq 1$, $z \neq 0$, and $n - m$ at $x = 1$. The LS estimator of θ is then given by

$$\hat{\theta}^n = \bar{\theta} + \frac{1}{z - z^2} \left[\frac{\delta_m}{\sqrt{m}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \frac{\gamma_{n-m}}{\sqrt{n-m}} \begin{pmatrix} -z^2 \\ z \end{pmatrix} \right] \tag{9}$$

where $\delta_m = \sum_{x_i=z} \varepsilon_i / \sqrt{m}$ and $\gamma_{n-m} = \sum_{x_i=1} \varepsilon_i / \sqrt{n-m}$. They are independent and both tend to be distributed $\mathcal{N}(0, 1)$.

3.2 Consistency.

We have $\hat{\theta}^n \overset{\text{a.s.}}{\to} \bar{\theta}$ (and $h(\hat{\theta}^n) \overset{\text{a.s.}}{\to} h(\bar{\theta})$) as soon as $m \to \infty$ and $n \to \infty$. However, when $n \to \infty$ with m fixed, then

$$\hat{\theta}^n \overset{\text{a.s.}}{\to} \hat{\theta}^\# = \bar{\theta} + \frac{1}{z - z^2} \frac{\delta_m}{\sqrt{m}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{10}$$

and $\hat{\theta}^n$ is not consistent. $h(\hat{\theta}^n)$ is then *not consistent*, except when $\bar{\theta}_1 + \bar{\theta}_2 = 0$. Indeed, in that case we obtain $\hat{\theta}_1^\# + \hat{\theta}_2^\# = 0$ so that $h(\hat{\theta}^\#) = h(\bar{\theta}) = 1/2$. Only this situation is investigated further when m is finite.

3.3 Asymptotic distribution of $h(\hat{\theta}^n)$.

case a) m is fixed and $\bar{\theta}_1 + \bar{\theta}_2 = 0$.

We can write

$$\sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] = \sqrt{n}[h(\hat{\theta}^n) - h(\hat{\theta}^\#)] = \sqrt{n}(\hat{\theta}^n - \hat{\theta}^\#)^\top \left[\frac{\partial h(\theta)}{\partial \theta}_{|\hat{\theta}^\#} + o_p(1) \right],$$

with $\partial h(\theta)/\partial \theta_{|\hat{\theta}^\#} = -\mathbf{1}/(2\hat{\theta}_2^\#)$. From (9) and (10),

$$\sqrt{n}(\hat{\theta}^n - \hat{\theta}^\#) = \frac{\sqrt{n}}{z - z^2} \frac{\gamma_{n-m}}{\sqrt{n-m}} \begin{pmatrix} -z^2 \\ z \end{pmatrix} \overset{\text{d}}{\to} \zeta_4 \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} z^4 & -z^3 \\ -z^3 & z^2 \end{pmatrix}),$$

which gives

$$\sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] \overset{\text{d}}{\to} \frac{1}{2} \frac{\nu}{\zeta} \tag{11}$$

where $\nu \sim \mathcal{N}(0,1)$ and $\zeta \sim \mathcal{N}(\bar{\theta}_2, 1/[m(z-z^2)^2])$ are independent. $h(\hat{\theta}^n)$ thus converges as $1/\sqrt{n}$ but its limiting distribution is not normal and depends on the choice of z.

case b) $m \to \infty$ and $m/n \to 0$, $n \to \infty$.

Suppose first that $\bar{\theta}_1 + \bar{\theta}_2 \neq 0$. We can write

$$\sqrt{m}[h(\hat{\theta}^n) - h(\bar{\theta})] = \sqrt{m}(\hat{\theta}^n - \bar{\theta})^\top \left[\frac{\partial h(\theta)}{\partial \theta}_{|\bar{\theta}} + o_p(1) \right] ,$$

with $\partial h(\theta)/\partial\theta_{|\bar{\theta}} = -1/(2\bar{\theta}_2)[1,\ 2h(\bar{\theta})]^\top$. From (9), since $m/n \to 0$,

$$\sqrt{m}(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\text{d}} \zeta_5 \sim \mathcal{N}(\mathbf{0}, \frac{1}{(z-z^2)^2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix})$$

which gives

$$\sqrt{m}[h(\hat{\theta}^n) - h(\bar{\theta})] \xrightarrow{\text{d}} \zeta_6 \sim \mathcal{N}(0, \frac{(\bar{\theta}_1 + \bar{\theta}_2)^2}{4\bar{\theta}_2^4 (z-z^2)^2}) \,. \tag{12}$$

$h(\hat{\theta}^n)$ is thus asymptotically normal and converges as $1/\sqrt{m}$. The limiting variance depends on z.

Suppose now that $\bar{\theta}_1 + \bar{\theta}_2 = 0$. We obtain from (9),

$$\begin{aligned} \sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] &= \sqrt{n} \left(-\frac{\hat{\theta}_1^n}{2\hat{\theta}_2^n} - \frac{1}{2} \right) \\ &= -\frac{\gamma_{n-m}}{2} \sqrt{\frac{n}{n-m}} \left[\bar{\theta}_2 - \frac{\delta_m}{\sqrt{m}} \frac{1}{z-z^2} + \frac{\gamma_{n-m}}{\sqrt{n-m}} \frac{z}{z-z^2} \right]^{-1} \end{aligned}$$

so that

$$\sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] \xrightarrow{\text{d}} \zeta_7 \sim \mathcal{N}(0, \frac{1}{4\bar{\theta}_2^2}) \,. \tag{13}$$

In contrast with (5, 6, 11) and (12), this is the unique case that leads to the expression used by Silvey (1980), with a speed of convergence that coincides with that obtained for non singular designs.

4 Discussion

Suppose that one knows *a priori* that $\bar{\theta}_1 + \bar{\theta}_2$ is close to 0, and designs an experiment that tries to approach ξ^* which puts weight 1 at $x = 1$, for estimating $h(\theta)$ given by (2) in the model (1). The justification for this choice lies in the asymptotic result (13): $h(\hat{\theta}^n)$ converges in $1/\sqrt{n}$, the asymptotic variance $1/(4\bar{\theta}_2^2)$ is the minimum over all possible designs when $\bar{\theta}_1 + \bar{\theta}_2 = 0$, that is, when $h(\bar{\theta}) = 1/2$. However, the results of Sections 2 and 3 evidence the risk of using a design approaching ξ^*.

- When $h(\bar{\theta}) = 1/2$ but ξ_n converges weakly to ξ^*, the limiting variance of $h(\hat{\theta}^n)$ is larger than $1/(4\bar{\theta}_2^2)$, see (4).
- When $h(\bar{\theta}) = 1/2$ but the number of observations at $x \neq 1$ is finite, as is the case of a design generated by the steepest descent algorithm, the limiting distribution of $h(\hat{\theta}^n)$ is not normal, see (11).
- When $h(\bar{\theta}) \neq 1/2$, although close to 1/2 (and one cannot be sure that $h(\bar{\theta}) = 1/2$, otherwise no experiment would be needed), the speed of convergence of $h(\hat{\theta}^n)$ is slower than $\sqrt{n}$, see (5, 12), and $h(\hat{\theta}^n)$ may even be not consistent, see (10).

A first possibility to avoid these difficulties is to use a non singular design, at the cost of a possible loss of efficiency. For instance, a design ξ_α that puts weight α at $x = 1$ and $1 - \alpha$ at -1, $0 < \alpha < 1$, ensures $\sqrt{n}$-convergence of $h(\hat{\theta}^n)$, and

$$\sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] \xrightarrow{\text{d}} \zeta_8 \sim \mathcal{N}(0, \frac{1}{4\bar{\theta}_2^2}[1,\ 2h(\bar{\theta})]\mathbf{M}^{-1}(\xi_\alpha)[1,\ 2h(\bar{\theta})]^\top)$$

as $n \to \infty$. When one knows that $\bar{\theta}_1 + \bar{\theta}_2$ is close to 0, one may then use ξ_α with α close to 1. Its efficiency is given by

$$\text{eff}(\alpha, h) = \frac{[1,\ 2h]\mathbf{M}^{-1}(\xi_{\alpha^*})[1,\ 2h]^\top}{[1,\ 2h]\mathbf{M}^{-1}(\xi_\alpha)[1,\ 2h]^\top}$$

where $\alpha^* = \alpha^*(h)$ corresponds to $\xi^*(1)$ in (3). $\text{eff}(\alpha, h)$ is plotted in Figure 1 for $\alpha \in [0.5, 1)$, $h \in [0.25, 0.75]$. Although $\text{eff}(\alpha, 1/2)$ quickly decreases when α moves away from 1, the loss of efficiency remains reasonable for small departures. In particular, $\xi_{3/4}$ is maximin-efficient, see Silvey (1980, p. 59): it guarantees $\text{eff}(3/4, h) \geq 0.75$ for any $h \geq 0$, the minimum efficiency being obtained for $h = 0$ and $h = 1/2$. (Note the difference with (Schwabe, 1997) where θ_1 is not restricted to be positive. The maximin-efficient design is then $\xi_{1/2}$, it is D-optimal and its minimum efficiency is 0.5.)

Another option consists in designing ξ_n sequentially, that is, using the algorithm (8) with $\partial h(\theta)/\partial\theta_{|\hat{\theta}^k}$ substituted for $\partial h(\theta)/\partial\theta_{|\bar{\theta}}$ in the determination of

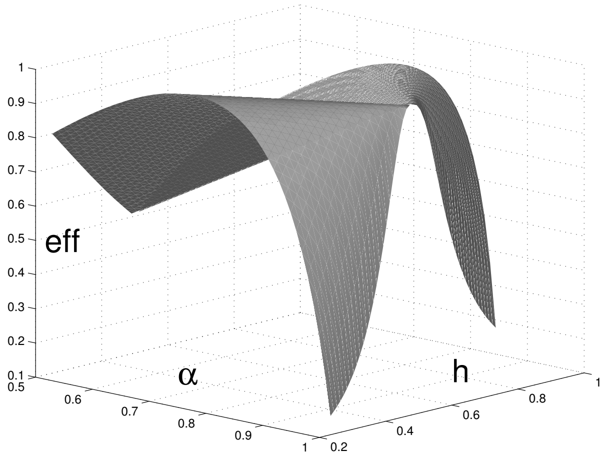

Fig. 1. Efficiency $\mathrm{eff}(\alpha, h)$.

x_{k+1}. Strong consistency of $\hat{\theta}^n$ is proved in (Ford and Silvey, 1980), ξ_n converges to the optimum design $\xi^*_{\bar{\theta}}$ for $\bar{\theta}$, and the asymptotic normality

$$\sqrt{n}[h(\hat{\theta}^n) - h(\bar{\theta})] \xrightarrow{\mathrm{d}} \zeta_9 \sim \mathcal{N}(0, \frac{1}{4\bar{\theta}_2^2}[1,\ 2h(\bar{\theta})]\mathbf{M}^-(\xi^*_{\bar{\theta}})[1,\ 2h(\bar{\theta})]^\top) \qquad (14)$$

$n \to \infty$, is proved in (Wu, 1985). The asymptotic efficiency thus equals one. In particular, (14) remains valid when $\bar{\theta}_1 + \bar{\theta}_2 = 0$, and then coincides with (13). When feasible, sequential design thus appears as the natural remedy to the issues raised in Sections 2 and 3. However, some difficulties should not be underestimated. The proof in (Wu, 1985) of the asymptotic result (14) under a sequential design is very much problem specific. Strong consistency of the LS estimator in the linear model under a sequential design requires stronger conditions than $\mathbf{M}^{-1}(\xi_n)/n \to 0$, see Lai and Wei (1982). Bayesian imbedding permits to weaken those conditions (Sternby, 1977) (at the expense of obtaining strong consistency of the estimator for *almost all* values of $\bar{\theta}$ with respect to some prior distribution), but its application to the sequential design of experiments (Hu, 1998) prohibits singular designs.

We hope we have convinced the reader of the richness of possible asymptotic behaviors of estimators under asymptotically singular designs. Combining this with a sequential construction of the design raises many challenging issues.

References

Buonaccorsi, J., Iyer, H., 1986. Optimal designs for ratios of linear combinations in the general linear model. Journal of Statistical Planning and Inference 13, 345–356.

Chaloner, K., 1989. Bayesian design for estimating the turning point of a quadratic regression. Commun. Statist.-Theory Meth. 18 (4), 1385–1400.

Fedorov, V., Müller, W., 1997. Another view on optimal design for estimating the point of extremum in quadratic regression. Metrika 46, 147–157.

Ford, I., Silvey, S., 1980. A sequentially constructed design for estimating a nonlinear parametric function. Biometrika 67 (2), 381–388.

Ford, I., Titterington, D., Wu, C., 1985. Inference and sequential design. Biometrika 72 (3), 545–551.

Hu, I., 1998. On sequential designs in nonlinear problems. Biometrika 85 (2), 496–503.

Huber, P., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Annals of Statistics 1 (5), 799–821.

Lai, T., Wei, C., 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Annals of Statistics 10 (1), 154–166.

Pázman, A., 1980. Singular experimental designs. Math. Operationsforsch. Statist., Ser. Statistics 16, 137–149.

Schwabe, R., 1997. Maximin efficient designs. Another view at D-optimality. Statistics & Probability Letters 35, 109–114.

Silvey, S., 1980. Optimal Design. Chapman & Hall, London.

Sternby, J., 1977. On consistency for the method of least squares using martingale theory. IEEE Transactions on Automatic Control 22 (3), 346–352.

Wu, C., 1985. Asymptotic inference from sequential design in a nonlinear situation. Biometrika 72 (3), 553–558.

Wu, C.-F., 1980. Characterizing the consistent directions of least squares estimates. Annals of Statistics 8 (4), 789–801.

Wynn, H., 1972. Results in the theory and construction of D-optimum experimental designs. Journal of Royal Statistical Society B34, 133–147.