Appendix A

Subdifferentials and Subgradients

Let $\Phi(\cdot)$ be a concave criterion function defined on some set $\mathcal{M} \subset \mathbb{M}$, e.g., $\mathcal{M} = \mathbb{M}^{\geq}$. The definition of $\Phi(\cdot)$ can be extended to any $p \times p$ symmetric matrix in $\mathbb{M}$ by setting $\Phi(\mathbf{M}) = -\infty$ for $\mathbf{M} \notin \mathcal{M}$. This extension is then concave on $\mathbb{M}$; its effective domain is the set $\mathrm{dom}(\Phi) = \{\mathbf{M} \in \mathbb{M} : \Phi(\mathbf{M}) > -\infty\}$. Note that $\mathbb{M}^{\geq} \subset \mathrm{dom}(\Phi)$ when $\Phi(\cdot)$ positively homogeneous and isotonic; see Lemma 5.4-(*iii*). A concave function $\Phi(\cdot)$ is called proper when $\mathrm{dom}(\Phi) \neq \varnothing$ and $\Phi(\mathbf{M}) < \infty$ for all $\mathbf{M} \in \mathbb{M}$. As a rule all the criteria we consider are proper.

When $\Phi(\cdot) : \mathbb{M} \longrightarrow \mathbb{R}$ is non-differentiable, the notion of gradient can be generalized as follows. A matrix $\tilde{\mathbf{M}}$ is called a subgradient of $\Phi(\cdot)$ at $\mathbf{M}$ if

$$\Phi(\mathbf{A}) \leq \Phi(\mathbf{M}) + \mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M})]\,,\ \forall \mathbf{A} \in \mathbb{M}\,. \tag{A.1}$$

Here $\mathrm{trace}(\mathbf{A}, \mathbf{B})$ is the usual scalar product between $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{M}$. The set of all subgradients of $\Phi(\cdot)$ at $\mathbf{M}$ is called the subdifferential[1] of $\Phi(\cdot)$ at $\mathbf{M}$ and is denoted by $\partial\Phi(\mathbf{M})$. The fact that these notions generalize that of gradient is due to the property $\partial\Phi(\mathbf{M}) = \{\nabla_{\mathbf{M}}\Phi(\mathbf{M})\}$ when $\Phi(\cdot)$ is differentiable at $\mathbf{M}$. In other situations $\partial\Phi(\mathbf{M})$ is not reduced to that singleton; it defines a convex set, closed if bounded, empty when $\mathbf{M} \notin \mathrm{dom}(\Phi)$, and satisfies the following properties. For any $\Phi(\cdot)$ concave on $\mathbb{M}$,

$$\partial(\alpha\Phi)(\mathbf{M}) = \alpha\partial\Phi(\mathbf{M})\,,\ \forall \mathbf{M} \in \mathbb{M}\,,\ \forall \alpha > 0\,. \tag{A.2}$$

For any $\Phi(\cdot)$ and $f(\cdot)$ concave on $\mathbb{M}$

$$\partial[\Phi + f](\mathbf{M}) = \partial\Phi(\mathbf{M}) + \partial f(\mathbf{M})\,,\ \forall \mathbf{M} \in \mathbb{M}\,, \tag{A.3}$$

[1]Subgradients and subdifferentials are usually defined for convex functions. We keep the same denomination here, although supergradients and superdifferentials might be more appropriate due to the upper-bound property (A.1); see Rockafellar (1970, p. 308).

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

if there exists some $\mathbf{A} \in \mathbb{M}$ where $f(\mathbf{A})$ is finite and $\Phi(\cdot)$ is continuous; see Alexéev et al. (1987, Sect. 3). Another sufficient condition is that the effective domains of $\Phi(\cdot)$ and $f(\cdot)$ overlap sufficiently, i.e., that their relative interiors[2] have a point in common; see Rockafellar (1970, p. 223). Also, for any $\Phi_1(\cdot)$, $\Phi_2(\cdot)$ concave on $\mathbb{M}$, continuous at $\hat{\mathbf{M}}$ such that $\Phi_1(\hat{\mathbf{M}}) = \Phi_2(\hat{\mathbf{M}})$,

$$\partial[\min(\Phi_1, \Phi_2)](\hat{\mathbf{M}}) = \mathrm{conv}[\partial\Phi_1(\hat{\mathbf{M}}) \cup \partial\Phi_2(\hat{\mathbf{M}})] \tag{A.4}$$

with $\mathrm{conv}(\mathcal{S})$ the convex hull of the set $\mathcal{S}$; see Alexéev et al. (1987, Sect. 3). For a continuous version of this property, consider a set of proper criteria functions $\Phi_\gamma(\cdot)$ from $\mathbb{M}$ to $\mathbb{R}$ (i.e., such that $\Phi_\gamma(\mathbf{M}) > -\infty$ for some $\mathbf{M}$ and $\Phi_\gamma(\mathbf{M}) < \infty$ for all $\mathbf{M} \in \mathbb{M}$) with $\gamma \in \Gamma$, a compact subset of $\mathbb{R}$, such that $\Phi_\gamma(\cdot)$ is concave and upper semicontinuous for all $\gamma \in \Gamma$ and the function $\gamma \longrightarrow \Phi_\gamma(\mathbf{M})$ is lower semicontinuous in γ for all $\mathbf{M}$. Suppose that $\Phi_\gamma(\cdot)$ is continuous at $\hat{\mathbf{M}}$ for all $\gamma \in \Gamma$ and define $\Phi^*(\mathbf{M}) = \min_{\gamma\in\Gamma} \Phi_\gamma(\mathbf{M})$ and $\Gamma^*(\mathbf{M}) = \{\gamma \in \Gamma : \Phi_\gamma(\mathbf{M}) = \Phi^*(\mathbf{M})\}$. Then, $\Phi^*(\cdot)$ is concave, and any element $\tilde{\mathbf{M}}$ of its subdifferential $\partial\Phi^*(\hat{\mathbf{M}})$ at $\hat{\mathbf{M}}$ can be written as

$$\tilde{\mathbf{M}} = \sum_{i=1}^{r} \alpha_i \tilde{\mathbf{M}}_i \tag{A.5}$$

with $r \leq p(p+1)/2 + 1$, $\sum_{i=1}^r \alpha_i = 1$, $\alpha_i > 0$, and $\tilde{\mathbf{M}}_i \in \partial\Phi_{\gamma_i}(\hat{\mathbf{M}})$ for some $\gamma_i \in \Gamma^*(\hat{\mathbf{M}})$, $i = 1, \ldots, r$; see Alexéev et al. (1987, p. 67).

Subgradients can also be defined for indicator functions. Let $\mathcal{M}$ be a convex subset of $\mathbb{M}$ and define

$$\hat{\mathbb{I}}_{\mathcal{M}}(\mathbf{M}) = \begin{cases} 0 \text{ if } \mathbf{M} \in \mathcal{M} \\ -\infty \text{ otherwise.} \end{cases}$$

Then, $\tilde{\mathbf{M}} \in \partial\hat{\mathbb{I}}_{\mathcal{M}}(\mathbf{M})$ if and only if $\hat{\mathbb{I}}_{\mathcal{M}}(\mathbf{A}) \leq \hat{\mathbb{I}}_{\mathcal{M}}(\mathbf{M}) + \mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M})]$ for all $\mathbf{A} \in \mathbb{M}$, see (A.1), and therefore $\mathbf{A} \in \mathcal{M}$ implies $\mathbf{M} \in \mathcal{M}$ and $\mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M})] \geq 0$, which means that $-\tilde{\mathbf{M}}$ is normal to $\mathcal{M}$ at $\mathbf{M}$; see Rockafellar (1970, p. 215).

With the notions of subgradients and subdifferentials a large part of the results of differential calculus remain valid for non-differentiable functions. In particular, a necessary-and-sufficient condition for a concave criterion $\Phi(\cdot)$ to reach its maximum value on $\mathbb{M}$ at $\mathbf{M}^*$ is that $\mathbf{O} \in \partial\Phi(\mathbf{M}^*)$, with $\mathbf{O}$ the null matrix; see Rockafellar (1970, p. 264). From this we directly obtain the following; see Pukelsheim (1993, p. 162).

[2] The relative interior of a convex set $\mathcal{S}$ is the interior of $\mathcal{S}$ regarded as a subset of the smallest affine set containing $\mathcal{S}$ (i.e., the affine hull of $\mathcal{S}$).

Theorem A.1. *Let $\Phi(\cdot)$ be a concave criterion taking finite values on $\mathbb{M}^>$ and let $\mathcal{M}$ be a convex subset of $\mathbb{M}^\geq$ that intersects $\mathbb{M}^>$. Then $\mathbf{M}^*$ maximizes $\Phi(\cdot)$ over $\mathcal{M}$ if and only if there exists $\tilde{\mathbf{M}} \in \partial\Phi(\mathbf{M}^*)$ such that*

$$\mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M}^*)] \leq 0 \,, \forall \mathbf{A} \in \mathcal{M} \,. \tag{A.6}$$

Indeed, using (A.3) the necessary-and-sufficient condition $\mathbf{O} \in \partial[\Phi + \hat{\mathbb{I}}_{\mathcal{M}}](\mathbf{M}^*)$ becomes: there exists $\tilde{\mathbf{M}} \in \partial\Phi(\mathbf{M}^*)$ such that $-\tilde{\mathbf{M}} \in \partial\hat{\mathbb{I}}_{\mathcal{M}}(\mathbf{M}^*)$, which gives (A.6).

In the particular case where $\Phi(\cdot)$ is differentiable with $\mathcal{M} = \mathcal{M}_\theta(\Xi)$, Theorem A.1 says that $\mathbf{M}^*$ is Φ-optimal on $\mathcal{M}_\theta(\Xi)$ if and only if $F_\Phi(\mathbf{M}^*, \mathbf{A}) = \mathrm{trace}[\nabla_{\mathbf{M}}\Phi(\mathbf{M}^*)(\mathbf{A} - \mathbf{M}^*)] \leq 0$ for all $\mathbf{A} \in \mathcal{M}_\theta(\Xi)$. Writing $\mathbf{M}^* = \mathbf{M}(\xi^*)$ and $\mathbf{A} = \mathbf{M}(\nu)$ for some ξ^* and ν in Ξ, we obtain that ξ^* is ϕ-optimal on Ξ if and only if $F_\phi(\xi^*; \nu) \leq 0$ for all $\nu \in \Xi$, see (5.34), which corresponds to the equivalence theorem 5.21 (note that $F_\phi(\xi; \xi) = 0$ for all ξ).

More generally, consider the case where $\Phi(\cdot)$ is not differentiable everywhere. Then, the one-sided directional derivative $\Phi'(\mathbf{M}^*, \mathbf{A})$ defined by (5.30) is given by

$$\Phi'(\mathbf{M}^*, \mathbf{A}) = \inf\{\mathrm{trace}(\tilde{\mathbf{M}}\mathbf{A}) : \tilde{\mathbf{M}} \in \partial\Phi(\mathbf{M}^*)\} \,, \tag{A.7}$$

see Rockafellar (1970, pp. 216–217), and the subgradient theorem says that $\mathbf{M}^*$ is Φ-optimal on $\mathcal{M}_\theta(\Xi)$ if and only if

$$\begin{aligned} F_\Phi(\mathbf{M}^*, \mathbf{A}) &= \Phi'(\mathbf{M}^*, \mathbf{A} - \mathbf{M}^*) \\ &= \inf_{\tilde{\mathbf{M}} \in \partial\Phi(\mathbf{M}^*)} \mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M}^*)] \leq 0 \,, \ \forall \mathbf{A} \in \mathcal{M}_\theta(\Xi) \,, \end{aligned} \tag{A.8}$$

which again corresponds to the equivalence theorem. Since the subdifferential $\partial\Phi(\mathbf{M}^*)$ is convex, the minimax theorem applies (Dem'yanov and Malozemov 1974, Theorem 5.2, p. 218). The necessary-and-sufficient condition (A.8) for the Φ-optimality of $\mathbf{M}^*$ on $\mathcal{M}_\theta(\Xi)$ can be expressed as the existence of $\tilde{\mathbf{M}} \in \partial\Phi(\mathbf{M}^*)$ such that $\mathrm{trace}[\tilde{\mathbf{M}}(\mathbf{A} - \mathbf{M}^*)] \leq 0$ for all $\mathbf{A} \in \mathcal{M}_\theta(\Xi)$. This type of condition has been used, for instance, in Theorem 5.38.

Finally, notice that the expressions for the directional derivatives of non-differentiable criteria obtained in Sect. 5.2.1 by using Lemmas 5.17 and 5.18 are direct consequences of (A.7) and (A.4), (A.5).

Appendix B

Computation of Derivatives Through Sensitivity Functions

The computation of the derivatives $\partial\eta(x,\theta)/\partial\theta_i$, $i = 1, \ldots, p$, of the model response $\eta(x,\theta)$ with respect to the model parameters θ is a mandatory step for most of the developments presented throughout the book: they are required, for instance, to evaluate the information matrix, the curvatures of the model, etc. However, in many circumstances the analytic expression of $\eta(x,\theta)$ is unknown, and its derivatives can only be obtained numerically. This appendix shows that this does not raise any particular difficulty, apart perhaps the computational time required by numerical calculations performed on a computer.

Consider the case, often met in practical applications, when $\eta(x,\theta)$ is the solution of a differential equation (similar developments can be made for recurrence equations).[1] Then, the derivatives $\partial\eta(x,\theta)/\partial\theta_i$, also called *sensitivity functions*, are solutions of other differential equations, which can easily be derived from the original one; see, e.g., Rabitz et al. (1983) and Walter and Pronzato (1997, Chap. 4). Only first-order derivatives are considered hereafter, but the developments easily extend to higher-order derivation. One can refer to classical textbooks on numerical analysis for methods to solve initial-value problems; see, e.g., Stoer and Bulirsch (1993).

Consider, for instance, the following state-space representation for the equations that give the response $\eta(x,\theta)$:

$$\dot{\mathbf{v}}(x,\theta,t) = \frac{\mathrm{d}\mathbf{v}(x,\theta,t)}{\mathrm{d}t} = \mathbf{F}[\mathbf{v}(x,\theta,t),\theta], \quad \mathbf{v}(x,\theta,0) = \mathbf{v}_0(x,\theta)\,, \tag{B.1}$$

$$\eta(x,\theta,t) = \mathbf{H}[\mathbf{v}(x,\theta,t),\theta]\,. \tag{B.2}$$

Here t denotes the time and $\mathbf{v}(x,\theta,t)$ the vector of state variables at time t for experimental conditions x and model parameters θ. The dependence of the

[1]Other methods for exact differentiation, based on adjoint state or adjoint code approaches, are available for more general situations where $\eta(x,\theta)$ is not given by a differential or recurrence equation, see, e.g., Walter and Pronzato (1997, Chap. 4).

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

right-hand side in some input signal $u(t)$ is omitted for the sake of simplicity of notations. Also, the (matrix) functions $\mathbf{F}$ and $\mathbf{H}$ might depend explicitly on t, which would correspond to a nonstationary system. The notation $\mathbf{v}_0(x,\theta)$ is to stress the fact that the initial conditions may be part of the unknown parameters to be estimated.

We wish to determine the values of $\partial\eta(x,\theta,t)/\partial\theta_i$, $i=1,\ldots,p$, at some particular values of t given by the sampling times $t_1,t_2,\ldots,t_N$ at which the observations are performed. Note that, although we write $\eta(x,\theta,t)$, these sampling times may be part of the design variables x. Also, x may include some control variables that influence the input signal $u(t)$, in which case we would write $u(t)=u(x,t)$.

The derivation of (B.2) with respect to θ_i gives

$$\frac{\partial\eta(x,\theta,t)}{\partial\theta_i} = \left.\frac{\partial\mathbf{H}(\mathbf{v},\theta)}{\partial\mathbf{v}^\top}\right|_{\mathbf{v}(x,\theta,t)} \frac{\partial\mathbf{v}(x,\theta,t)}{\partial\theta_i} + \left.\frac{\partial\mathbf{H}(\mathbf{v},\theta)}{\partial\theta_i}\right|_{\mathbf{v}(x,\theta,t)} \tag{B.3}$$

which requires the evaluation of the derivative $\partial\mathbf{v}(x,\theta,t)/\partial\theta_i$. It is obtained by differentiating the evolution equations (B.1) of the system,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathbf{v}(x,\theta,t)}{\partial\theta_i} = \left.\frac{\partial\mathbf{F}(\mathbf{v},\theta)}{\partial\mathbf{v}^\top}\right|_{\mathbf{v}(x,\theta,t)} \frac{\partial\mathbf{v}(x,\theta,t)}{\partial\theta_i} + \left.\frac{\partial\mathbf{F}(\mathbf{v},\theta)}{\partial\theta_i}\right|_{\mathbf{v}(x,\theta,t)}, \tag{B.4}$$

and the initial conditions

$$\frac{\partial\mathbf{v}(x,\theta,0)}{\partial\theta_i} = \frac{\partial\mathbf{v}_0(x,\theta)}{\partial\theta_i}\,.$$

Therefore, the solution of the initial-value problem (B.1) gives $\eta(x,\theta,t)$, and the solution of p initial-value problems similar to (B.4) gives the sensitivity functions $\partial\eta(x,\theta,t)/\partial\theta_i$, $i=1,\ldots,p$, through (B.3); see Valko and Vajda (1984) and Bilardello et al. (1993) for details. Notice that the differential equations (B.4) corresponding to θ_i and θ_j with $i\neq j$ are independent, i.e., the solutions can be obtained independently once the trajectory of $\mathbf{v}(x,\theta,t)$ has been obtained. Also note that (B.4) is linear in $\partial\mathbf{v}(x,\theta,t)/\partial\theta_i$ (but nonstationary since $\partial\mathbf{F}(\mathbf{v},\theta)/\partial\mathbf{v}^\top\big|_{\mathbf{v}(x,\theta,t)}$ depends on t), and only the driving term $\partial\mathbf{F}(\mathbf{v},\theta)/\partial\theta_i\big|_{\mathbf{v}(x,\theta,t)}$ and initial conditions $\partial\mathbf{v}_0(x,\theta)/\partial\theta_i$ depend on i.

On the other hand, although approximating the derivatives $\partial\eta(x,\theta,t)/\partial\theta_i$ by finite differences might seem simpler, it would require the solutions of $p+1$ initial-value problems of the type (B.1) and would thus only produce approximate results for similar efforts. The situation is even more favorable to exact calculations when the state-space representation (B.1) is linear, i.e., when the differential equation that gives $\eta(x,\theta,t)$ is linear, with known initial conditions. Then, if $\eta(x,\theta,t)$ is solution of an m-th-order differential equation, $\eta(x,\theta,t)$ and its derivatives $\partial\eta(x,\theta,t)/\partial\theta_i$, $i=1,\ldots,p$, can be obtained by solving an initial-value problem for a differential equation of order $2m$ only, whatever the number p of parameters. Indeed, consider the following m-th-order differential equation

$$\eta^{(m)}(x,\theta,t) + \sum_{i=0}^{m-1} \theta_i \, \eta^{(i)}(x,\theta,t) = \sum_{i=m}^{m+q} \theta_i \, u^{(i-m)}(t)\,, \tag{B.5}$$

where $\eta^{(k)}(x,\theta,t)$ and $u^{(k)}(t)$, $k \geq 0$, respectively denote the k-th-order derivatives of $\eta(x,\theta,t)$ and $u(t)$ with respect to t (with $\eta^{(0)}(x,\theta,t) = \eta(x,\theta,t)$ and $u^{(0)}(t) = u(t)$) and where the initial conditions $\eta^{(i)}(x,\theta,0) = \alpha_i$, $i = 0,\ldots,m-1$, are known. Denote by $s_j(x,\theta,t)$ the sensitivity functions

$$s_j(x,\theta,t) = \frac{\partial \eta(x,\theta,t)}{\partial \theta_j}\,,\ j = 0,\ldots,m+q\,.$$

They are solutions of differential equations of order m, obtained by differentiating (B.5) with respect to the $m+q+1$ parameters θ_i,

$$s_j^{(m)}(x,\theta,t) + \sum_{i=0}^{m-1} \theta_i \, s_j^{(i)}(x,\theta,t) = u^{(j-m)}(t)\,,\ j = m,\ldots,m+q\,, \tag{B.6}$$

$$s_j^{(m)}(x,\theta,t) + \sum_{i=0}^{m-1} \theta_i \, s_j^{(i)}(x,\theta,t) = -\eta^{(j)}(x,\theta,t)\,,\ j = 0,\ldots,m-1\,, \tag{B.7}$$

with zero initial conditions since the α_i are known. The computation of $\eta(x,\theta,t)$ and its derivatives $s_j(x,\theta,t)$ then seems to require the solution of an initial-value problem for $m+q+2$ differential equations of order m. However, one may notice that all these differential equations have the same homogeneous part (left-hand side) and only differ by their driving terms. The computations can thus be simplified as follows. First solve (B.6) for $j = m$ to obtain $s_m(x,\theta,t)$. Then, by linearity, we have $s_{m+k}(x,\theta,t) = \dot{s}_{m+k-1}(x,\theta,t)$ for $k = 1,\ldots,q$. Assume for the moment that the initial conditions α_i equal zero. Then, by linearity again, $\eta(x,\theta,t) = \sum_{j=m}^{m+q} \theta_j s_j(x,\theta,t)$. The solution of (B.7) for $j = 0$ gives $s_0(x,\theta,t)$, and by differentiation with respect to t we get $s_k(x,\theta,t) = \dot{s}_{k-1}(x,\theta,t)$ for $k = 1,\ldots,m-1$. The response $\eta(x,\theta,t)$ and the $m+q+1$ sensitivity functions are thus obtained by solving two initial-value problems for a differential equation of order m. When the α_i are not zero, the solution $\eta(x,\theta,t)$ must be corrected to take those initial conditions into account. This can be done through a state-space representation. Define the vector of state variables at time t by $\mathbf{w}(x,\theta,t) = [\eta^{(m-1)}(x,\theta,t), \eta^{(m-2)}(x,\theta,t), \ldots, \eta^{(0)}(x,\theta,t)]^\top$. It satisfies the differential equation

$$\dot{\mathbf{w}}(x,\theta,t) = \mathbf{A}(\theta)\mathbf{w}(x,\theta,t) + \sum_{i=m}^{m+q} \theta_i \, u^{(i-m)}(t)\ \mathbf{e}_1$$

with $\mathbf{e}_1 = (1,0,\ldots,0)^\top$ the first basis vector of $\mathbb{R}^m$ and $\mathbf{A} = \mathbf{A}(\theta)$ the $m \times m$ matrix

$$\mathbf{A} = \begin{pmatrix} -\theta_{m-1} & -\theta_{m-2} & \cdots & -\theta_1 & -\theta_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} .$$

The response $\eta(x, \theta, t)$ is then given by

$$\eta(x, \theta, t) = \sum_{j=m}^{m+q} \theta_j s_j(x, \theta, t) + \sum_{i=1}^{n_\lambda} \sum_{j=1}^{n_{\lambda_i}} c_{i,j}\, t^{j-1} \exp(\lambda_i t) ,$$

where n_λ denotes the number of distinct eigenvalues of $\mathbf{A}$, the eigenvalue λ_i having the multiplicity n_{λ_i}. The m constants $c_{i,j}$ are determined from the initial conditions $\eta^{(i)}(x, \theta, 0) = \alpha_i$, $i = 0, \ldots, m-1$.

Appendix C

Proofs

Lemma 2.5. *Let $\{x_i\}$ be an asymptotically discrete design with measure ξ. Assume that $a(x, \theta)$ is a bounded function on $\mathscr{X} \times \Theta$ and that to every $x \in \mathscr{X}$ we can associate a random variable $\varepsilon(x)$. Let $\{\varepsilon_i\}$ be a sequence of independent random variables, with ε_i distributed like $\varepsilon(x_i)$, and assume that for all $x \in \mathscr{X}$*

$$\mathbb{E}\{b[\varepsilon(x)]\} = m(x)\,, \ |m(x)| < \bar{m} < \infty\,,$$
$$\text{var}\{b[\varepsilon(x)]\} = V(x) < \bar{V} < \infty\,,$$

with $b(\cdot)$ a Borel function on $\mathbb{R}$. Then we have

$$\frac{1}{N}\sum_{k=1}^{N} a(x_k, \theta) b(\varepsilon_k) \overset{\theta}{\rightsquigarrow} \sum_{x \in S_\xi} a(x, \theta) m(x) \xi(x)$$

as N tends to ∞, where $\overset{\theta}{\rightsquigarrow}$ means uniform convergence with respect to $\theta \in \Theta$, and the convergence is almost sure (a.s.), i.e., with probability one, with respect to the random sequence $\{\varepsilon_i\}$.

Proof. For any θ, we can write

$$\begin{aligned}
&\left| \frac{1}{N}\sum_{k=1}^{N} a(x_k, \theta) b(\varepsilon_k) - \sum_{x \in S_\xi} a(x, \theta) m(x) \xi(x) \right| \\
&\leq \left| \frac{1}{N} \sum_{k=1,\, x_k \notin \mathcal{S}_\xi}^{N} a(x_k, \theta) b(\varepsilon_k) \right| \\
&\quad + \left| \frac{1}{N} \sum_{k=1,\, x_k \in \mathcal{S}_\xi}^{N} a(x_k, \theta) b(\varepsilon_k) - \sum_{x \in S_\xi} a(x, \theta) m(x) \xi(x) \right| . \qquad \text{(C.1)}
\end{aligned}$$

Let $N(x)/N$ be the relative frequency of the point x in the sequence $x_1, \ldots, x_N$. The second term is bounded by

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

$$\sum_{x\in\mathcal{S}_\xi} \sup_{\theta\in\Theta} |a(x,\theta)| \left| \frac{N(x)}{N} \frac{1}{N(x)} \sum_{k=1}^{N(x)} b(\varepsilon_k) - m(x)\xi(x) \right| ;$$

since $\sum_{k=1}^{N(x)} b(\varepsilon_k)/N(x)$ converges a.s. to $m(x)$ (SLLN) and $N(x)/N - \xi(x)$ tends to zero, this term tends a.s. to zero, uniformly in θ. Let A_N denote the first term on the right-hand side of (C.1) and $N(\mathscr{X} \setminus \mathcal{S}_\xi)$ denote the number of points among $x_1, \ldots, x_N$ that belong to the set $\mathscr{X} \setminus \mathcal{S}_\xi$. If $N(\mathscr{X} \setminus \mathcal{S}_\xi)$ is finite, the lemma is proved. Otherwise, A_N satisfies

$$|A_N| = \left| \frac{1}{N} \sum_{k=1,\, x_k\notin\mathcal{S}_\xi}^{N} a(x_k,\theta)b(\varepsilon_k) \right| \leq$$
$$\sup_{x\in\mathscr{X},\,\theta\in\Theta} |a(x,\theta)| \frac{N(\mathscr{X} \setminus \mathcal{S}_\xi)}{N} \frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k\notin\mathcal{S}_\xi}^{N} |b(\varepsilon_k)| .$$

Now, the SLLN applied to the independent sequence of random variables $|b(\varepsilon_k)|$ gives

$$\frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k\notin\mathcal{S}_\xi}^{N} |b(\varepsilon_k)| - \frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k\notin\mathcal{S}_\xi}^{N} \mathbb{E}\{|b[\varepsilon(x_k)]|\} \stackrel{\text{a.s.}}{\to} 0$$

as $N \to \infty$. Moreover,

$$\frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k\notin\mathcal{S}_\xi}^{N} \mathbb{E}\{|b[\varepsilon(x_k)]|\} \leq \sup_{x\in\mathscr{X}} \mathbb{E}\{|b[\varepsilon(x)]|\}$$
$$\leq \sup_{x\in\mathscr{X}} \sqrt{V(x) + m^2(x)} < \infty .$$

Since $N(\mathscr{X} \setminus \mathcal{S}_\xi)/N \to 0$, A_N tends to zero a.s. and uniformly in θ, which completes the proof. ■

Lemma 2.6. *Let $\{z_i\}$ be a sequence of i.i.d. random vectors of $\mathbb{R}^r$ and $a(z,\theta)$ be a Borel measurable real function on $\mathbb{R}^r \times \Theta$, continuous in $\theta \in \Theta$ for any z, with Θ a compact subset of $\mathbb{R}^p$. Assume that*

$$\mathbb{E}\{\max_{\theta\in\Theta} |a(z_1,\theta)|\} < \infty , \tag{C.2}$$

then $\mathbb{E}\{a(z_1,\theta)\}$ is continuous in $\theta \in \Theta$ and

$$\frac{1}{N} \sum_{i=1}^{N} a(z_i,\theta) \overset{\theta}{\rightsquigarrow} \mathbb{E}[a(z_1,\theta)] \text{ a.s. when } N \to \infty .$$

Proof. We use a construction similar to that in (Bierens, 1994, p. 43). Take some fixed $\theta^1 \in \Theta$ and consider the set

$$\mathscr{B}(\theta^1, \delta) = \left\{\theta \in \Theta : \ \left\|\theta - \theta^1\right\| \leq \delta\right\} .$$

Define $\bar{a}_\delta(z)$ and $\underline{a}_\delta(z)$ as the maximum and the minimum of $a(z, \theta)$ over the set $\mathscr{B}(\theta^1, \delta)$, which are properly defined random variables from Lemma 2.9. The expectations $\mathbb{E}\{|\underline{a}_\delta(z_1)|\}$ and $\mathbb{E}\{|\bar{a}_\delta(z_1)|\}$ are bounded by

$$\mathbb{E}\{\max_{\theta \in \Theta} |a(z_1, \theta)|\} < \infty .$$

Also, $\bar{a}_\delta(z) - \underline{a}_\delta(z)$ is an increasing function of δ. Hence, we can interchange the order of the limit and expectation in the following expression:

$$\lim_{\delta \searrow 0} \left[\mathbb{E}\{\bar{a}_\delta(z_1)\} - \mathbb{E}\{\underline{a}_\delta(z_1)\}\right] = \mathbb{E}\left\{\lim_{\delta \searrow 0} \left[\bar{a}_\delta(z_1) - \underline{a}_\delta(z_1)\right]\right\} = 0 ,$$

which proves the continuity of $\mathbb{E}\{a(z_1, \theta)\}$ at θ^1 and implies

$$\forall \beta > 0 , \ \exists \delta(\beta) > 0 \ \text{ such that } \ \left|\mathbb{E}\{\bar{a}_{\delta(\beta)}(z_1)\} - \mathbb{E}\{\underline{a}_{\delta(\beta)}(z_1)\}\right| < \frac{\beta}{2} .$$

Hence we can write for every $\theta \in \mathscr{B}(\theta^1, \delta(\beta))$

$$\begin{aligned}
\frac{1}{N}\sum_k \underline{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\underline{a}_{\delta(\beta)}(z_1)\} - \frac{\beta}{2} &\leq \frac{1}{N}\sum_k \underline{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\bar{a}_{\delta(\beta)}(z_1)\} \\
&\leq \frac{1}{N}\sum_k a(z_k, \theta) - \mathbb{E}\{a(z_1, \theta)\} \\
&\leq \frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\underline{a}_{\delta(\beta)}(z_1)\} \\
&\leq \frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\bar{a}_{\delta(\beta)}(z_1)\} + \frac{\beta}{2} .
\end{aligned}$$

From the SLLN, we have that $\forall \gamma > 0$, $\exists N_1(\beta, \gamma)$ such that

$$\text{Prob}\left\{\forall N > N_1(\beta, \gamma) , \ \left|\frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\bar{a}_{\delta(\beta)}(z_1)\}\right| < \frac{\beta}{2}\right\} > 1 - \frac{\gamma}{2} ,$$

$$\text{Prob}\left\{\forall N > N_1(\beta, \gamma) , \ \left|\frac{1}{N}\sum_k \underline{a}_{\delta(\beta)}(z_k) - \mathbb{E}\{\underline{a}_{\delta(\beta)}(z_1)\}\right| < \frac{\beta}{2}\right\} > 1 - \frac{\gamma}{2} .$$

Combining with previous inequalities, we obtain

$$\text{Prob}\left\{\forall N > N_1(\beta, \gamma), \max_{\theta \in \mathscr{B}(\theta^1, \delta(\beta))} \left|\frac{1}{N}\sum_k a(z_k, \theta) - \mathbb{E}\{a(z_1, \theta)\}\right| < \beta\right\} > 1 - \gamma .$$

It only remains to cover Θ with a finite numbers of sets $\mathscr{B}(\theta^i, \delta(\beta))$, $i = 1, \ldots, n(\beta)$, which is always possible from the compactness assumption. For any $\alpha > 0, \beta > 0$, take $\gamma = \alpha/n(\beta)$, $N(\beta) = \max_i N_i(\beta, \gamma)$. We obtain

$$\text{Prob}\left\{\forall N > N(\beta)\,,\ \max_{\theta\in\Theta}\left|\frac{1}{N}\sum_k a(z_k,\theta) - \mathbb{E}\{a(z_1,\theta)\}\right| < \beta\right\} > 1-\alpha\,,$$

which completes the proof. ■

Lemma 2.7. *Let $\{z_i\}$, θ, Θ and $a(z,\theta)$ be defined as in Lemma 2.6. Assume that*

$$\sup_{\theta\in\Theta} \mathbb{E}\{|a(z_1,\theta)|\} < \infty$$

and that $a(z,\theta)$ is continuous in $\theta \in \Theta$ uniformly in z. Then the conclusions of Lemma 2.6 apply.

Proof. We only need to prove (C.2). The continuity of $a(z,\theta)$ with respect to θ being uniform in z, we have: $\forall\theta^1 \in \Theta\,,\ \forall\epsilon > 0\,,\ \exists\delta(\epsilon) > 0$ such that

$$\forall\theta \in \mathcal{C}(\theta^1,\delta(\epsilon)) = \mathscr{B}(\theta^1,\delta(\epsilon)) \cap \Theta\,,\quad \sup_z |a(z,\theta) - a(z,\theta^1)| < \epsilon\,.$$

This implies that for all $\theta \in \mathcal{C}(\theta^1,\delta(\epsilon))$, $|a(z,\theta)| < |a(z,\theta^1)| + \epsilon\ \forall z$; that is,

$$\bar{a}^1_{\delta(\epsilon)}(z) = \max_{\theta\in\mathcal{C}(\theta^1,\delta(\epsilon))} |a(z,\theta)| < |a(z,\theta^1)| + \epsilon\ \forall z\,,$$

with $\mathbb{E}\{|a(z_1,\theta^1)|\} < \infty$ by assumption. Therefore, $\mathbb{E}\{|\bar{a}^1_{\delta(\epsilon)}(z_1)|\} < \infty$. Now, we can cover Θ by a finite number of balls $\mathscr{B}(\theta^k, \delta_k(\epsilon))$, $k = 1, \ldots, n(\epsilon)$ and

$$\max_{\theta\in\Theta} |a(z,\theta)| = \max_{k=1,\ldots,n(\epsilon)} \bar{a}^1_{\delta_k(\epsilon)}(z)$$

which implies (C.2). ■

Lemma 2.8. *Let $\{x_i\}$ be an asymptotically discrete design with measure ξ. Assume that to every $x \in \mathscr{X}$ we can associate a random variable $\varepsilon(x)$. Let $\{\varepsilon_i\}$ be a sequence of independent random variables, with ε_i distributed like $\varepsilon(x_i)$. Let $a(x,\varepsilon,\theta)$ be a Borel measurable function of ε for any $(x,\theta) \in \mathscr{X}\times\Theta$, continuous in $\theta \in \Theta$ for any x and ε, with Θ a compact subset of $\mathbb{R}^p$. Assume that*

$$\forall x \in \mathcal{S}_\xi\,,\quad \mathbb{E}\{\max_{\theta\in\Theta} |a[x,\epsilon(x),\theta]|\} < \infty\,, \tag{C.3}$$

$$\forall x \in \mathscr{X} \setminus \mathcal{S}_\xi\,,\quad \mathbb{E}\{\max_{\theta\in\Theta} |a[x,\epsilon(x),\theta]|^2\} < \infty\,. \tag{C.4}$$

Then we have

$$\frac{1}{N}\sum_{k=1}^{N} a(x_k, \epsilon_k, \theta) \overset{\theta}{\rightsquigarrow} \sum_{x \in S_\xi} \mathbb{E}\{a[x, \epsilon(x), \theta]\}\, \xi(x) \;\; a.s. \; when \; N \to \infty \,,$$

where the function on the right-hand side is continuous in θ on Θ.

Proof. We have

$$\left| \frac{1}{N}\sum_{k=1}^{N} a(x_k, \epsilon_k, \theta) - \sum_{x \in S_\xi} \mathbb{E}\{a[x, \epsilon(x), \theta]\}\xi(x) \right| \leq A_N + B_N$$

with

$$A_N = \left| \frac{1}{N} \sum_{k=1,\, x_k \notin \mathcal{S}_\xi}^{N} a(x_k, \epsilon_k, \theta) \right|$$

and

$$B_N = \left| \frac{1}{N} \sum_{k=1,\, x_k \in \mathcal{S}_\xi}^{N} a(x_k, \epsilon_k, \theta) - \sum_{x \in S_\xi} \mathbb{E}\{a[x, \epsilon(x), \theta]\}\xi(x) \right| .$$

Then

$$B_N \leq \sum_{x \in S_\xi} \left| \left[\frac{N(x)}{N} \frac{1}{N(x)} \sum_{k=1,\, x_k = x}^{N} a(x, \epsilon_k, \theta) \right] - \mathbb{E}\{a[x, \epsilon(x), \theta]\}\xi(x) \right|$$

where, for each $x \in \mathcal{S}_\xi$, the $a(x, \epsilon_k, \theta)$ are i.i.d. random variables satisfying (C.3). Lemma 2.6 thus applies, and, since $N(x)/N$ tends to $\xi(x)$, $B_N \overset{\theta}{\rightsquigarrow} 0$ a.s. when $N \to \infty$. Also, from the same lemma, $\mathbb{E}\{a[x, \epsilon(x), \theta]\}$ is a continuous function of θ for any x.

$$A_N \leq \bar{A}_N = \frac{N(\mathscr{X} \setminus \mathcal{S}_\xi)}{N} \frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k \notin \mathcal{S}_\xi}^{N} \max_{\theta \in \Theta} |a(x_k, \epsilon_k, \theta)| \,,$$

where $N(\mathscr{X} \setminus \mathcal{S}_\xi)$ denotes the number of points among $x_1, \ldots, x_N$ that belong to the set $\mathscr{X} \setminus \mathcal{S}_\xi$. If $N(\mathscr{X} \setminus \mathcal{S}_\xi) < \infty$, $A_N \overset{\theta}{\rightsquigarrow} 0$ a.s. when $N \to \infty$. Otherwise, the independent random variables $\max_{\theta \in \Theta} |a(x_k, \epsilon_k, \theta)|$ satisfy (C.4), and the SLLN then implies

$$\frac{1}{N(\mathscr{X} \setminus \mathcal{S}_\xi)} \sum_{k=1,\, x_k \notin \mathcal{S}_\xi}^{N} \left(\max_{\theta \in \Theta} |a(x_k, \epsilon_k, \theta)| - \mathbb{E}\{\max_{\theta \in \Theta} |a[x_k, \epsilon(x_k), \theta]|\} \right) \overset{\text{a.s.}}{\to} 0$$

when $N \to \infty$, which implies $\bar{A}_N \overset{\text{a.s.}}{\to} 0$ since $N(\mathscr{X} \setminus \mathcal{S}_\xi)/N \to 0$, and therefore $A_N \overset{\theta}{\rightsquigarrow} 0$ a.s. when $N \to \infty$. ■

Lemma 2.9 (Jennrich 1969). *Let Θ be a compact subset of $\mathbb{R}^p$, $\mathscr{Z}$ be a measurable subset of $\mathbb{R}^m$ and $J(z,\theta)$ be a Borel measurable real function on $\mathscr{Z} \times \Theta$, continuous in $\theta \in \Theta$ for any $z \in \mathscr{Z}$. Then there exists a mapping $\hat{\theta}$ from $\mathscr{Z}$ into Θ with Borel measurable components such that $J[z, \hat{\theta}(z)] = \min_{\theta\in\Theta} J(z,\theta)$, which therefore is also Borel measurable. If, moreover, $J(z,\theta)$ is continuous on $\mathscr{Z} \times \Theta$, then $\min_{\theta\in\Theta} J(z,\theta)$ is a continuous function on $\mathscr{Z}$.*

Proof. $J(z,\theta)$ is a measurable function of z for any $\theta \in \Theta$ and a continuous function of θ for any $z \in \mathscr{Z}$. Let $\{\Theta_k\}$ be an increasing sequence of finite subsets of Θ whose limit is dense in Θ. For any k, there exists a measurable function $\tilde{\theta}^k$ from $\mathscr{Z}$ into Θ_k such that

$$\forall z \in \mathscr{Z}\,,\ J(z, \tilde{\theta}^k) = \min_{\theta\in\Theta_k} J(z,\theta)\,.$$

Define $\hat{\theta}_1 = \hat{\theta}_1(z) = \liminf_{k\to\infty} \tilde{\theta}_1^k(z)$ (with $\tilde{\theta}_1^k$ the first component of $\tilde{\theta}^k$), and notice that $\hat{\theta}_1$ is measurable. For any $z \in \mathscr{Z}$, there exists a subsequence $\{\tilde{\theta}^{k_i}(z)\}$ of $\{\tilde{\theta}^k(z)\}$ that converges to a point $\tilde{\theta} \in \Theta$ such that

$$\tilde{\theta} = \tilde{\theta}(z) = (\hat{\theta}_1(z), \tilde{\theta}_2, \ldots, \tilde{\theta}_p)\,.$$

Now,

$$\begin{aligned}\min_{(\theta_2,\ldots,\theta_p);(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p)\in\Theta} J[z,(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p)] &\le J[z,(\hat{\theta}_1(z),\tilde{\theta}_2,\ldots,\tilde{\theta}_p)]\\ &= J(z,\tilde{\theta})\\ &= \lim_{i\to\infty} J[z,\tilde{\theta}^{k_i}(z)]\\ &= \lim_{i\to\infty} \min_{\theta\in\Theta_{k_i}} J(z,\theta)\\ &= \min_{\theta\in\Theta} J(z,\theta)\end{aligned}$$

where the last equality follows from the fact that $\lim_{k\to\infty} \Theta_k$ is dense in Θ. Therefore, for any $z \in \mathscr{Z}$,

$$\min_{(\theta_2,\ldots,\theta_p);(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p)\in\Theta} J[z,(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p)] = \min_{\theta\in\Theta} J(z,\theta)\,.$$

Define $J_1[z,(\theta_2,\ldots,\theta_p)] = J[z,(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p)]$. It is a continuous function of $(\theta_2,\ldots,\theta_p)$ for all $z \in \mathscr{Z}$ and a measurable function of z for all $(\theta_2,\ldots,\theta_p)$ such that $(\hat{\theta}_1(z),\theta_2,\ldots,\theta_p) \in \Theta$. Apply the same arguments to J_1 to obtain a measurable function $\hat{\theta}_2$ such that, for any $z \in \mathscr{Z}$,

$$\min_{(\theta_3,\ldots,\theta_p);(\hat{\theta}_1(z),\hat{\theta}_2(z),\theta_3,\ldots,\theta_p)\in\Theta} J[z,(\hat{\theta}_1(z),\hat{\theta}_2(z),\theta_3,\ldots,\theta_p)] = \min_{\theta\in\Theta} J(z,\theta)\,.$$

Continuing in this manner, we construct real-valued functions $\hat{\theta}_1,\ldots,\hat{\theta}_p$ such that, for any $z \in \mathscr{Z}$,

$$J[z, (\hat{\theta}_1(z), \dots, \hat{\theta}_p(z))] = \min_{\theta \in \Theta} J(z, \theta) .$$

Hence, $\hat{\theta} = (\hat{\theta}_1(z), \dots, \hat{\theta}_p(z))$ is a measurable function from $\mathscr{Z}$ into Θ with the desirable property.

We show now that the continuity of $J(z, \theta)$ on $\mathscr{Z} \times \Theta$, Θ compact, implies that $\min_{\theta \in \Theta} J(z, \theta)$ is continuous in z.

$J(\cdot, \cdot)$ is uniformly continuous on compact subsets of $\mathscr{Z} \times \Theta \subset \mathbb{R}^m \times \mathbb{R}^p$. Therefore, $\forall z_0 \in \mathscr{Z}$, $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$\forall \theta \in \Theta\,,\ \forall z \in \mathscr{B}(z_0, \delta)\,,\ J(z_0, \theta) - \epsilon < J(z, \theta) < J(z_0, \theta) + \epsilon\,,$$

where $\mathscr{B}(z_0, \delta) = \{z \in \mathbb{R}^m : \|z - z_0\| \leq \delta\}$. This implies

$$\forall z \in \mathscr{B}(z_0, \delta)\,,\ \min_{\theta \in \Theta} J(z_0, \theta) - \epsilon \leq \min_{\theta \in \Theta} J(z, \theta) \leq \min_{\theta \in \Theta} J(z_0, \theta) + \epsilon$$

and $\min_{\theta \in \Theta} J(z, \theta)$ is thus continuous at z_0. Since z_0 is arbitrary, it is continuous for all $z \in \mathscr{Z}$. ■

Lemma 2.10. *Assume that the sequence of functions $\{J_N(\theta)\}$ converges uniformly on Θ to the function $J_{\bar{\theta}}(\theta)$, with $J_N(\theta)$ continuous with respect to $\theta \in \Theta$ for any N, Θ a compact subset of $\mathbb{R}^p$, and $J_{\bar{\theta}}(\theta)$ such that*

$$\forall \theta \in \Theta\,,\ \theta \neq \bar{\theta} \Longrightarrow J_{\bar{\theta}}(\theta) > J_{\bar{\theta}}(\bar{\theta})\,.$$

Then $\lim_{N \to \infty} \hat{\theta}^N = \bar{\theta}$, where $\hat{\theta}^N \in \arg\min_{\theta \in \Theta} J_N(\theta)$. When the functions $J_N(\cdot)$ are random, and the uniform convergence to $J_{\bar{\theta}}(\cdot)$ is almost sure, the convergence of $\hat{\theta}^N$ to $\bar{\theta}$ is also almost sure.

Proof. The function $J_{\bar{\theta}}(\cdot)$ is continuous, and therefore, $\forall \beta > 0$, $\exists \epsilon > 0$ such that $J_{\bar{\theta}}(\theta) < J_{\bar{\theta}}(\bar{\theta}) + \epsilon$ implies $\|\theta - \bar{\theta}\| < \beta$. Indeed, for any $\beta > 0$ define

$$\underline{J}(\beta) = \min_{\{\theta \in \Theta:\, \|\theta - \bar{\theta}\| \geq \beta\}} J_{\bar{\theta}}(\theta)\,,\ \epsilon = \epsilon(\beta) = \frac{\underline{J}(\beta) - J_{\bar{\theta}}(\bar{\theta})}{2}\,.$$

We have $\underline{J}(\beta) > J_{\bar{\theta}}(\bar{\theta})$ and thus $\epsilon(\beta) > 0$. Assume that $J_{\bar{\theta}}(\theta) < J_{\bar{\theta}}(\bar{\theta}) + \epsilon = [\underline{J}(\beta) + J_{\bar{\theta}}(\bar{\theta})]/2$. It implies $J_{\bar{\theta}}(\theta) < \underline{J}(\beta)$ and thus $\|\theta - \bar{\theta}\| < \beta$.

Now, the uniform convergence of $J_N(\cdot)$ implies that there exists N_0 such that $\forall N > N_0$ and $\forall \theta \in \Theta$, $|J_N(\theta) - J_{\bar{\theta}}(\theta)| < \epsilon/2$. Therefore, $|J_N(\bar{\theta}) - J_{\bar{\theta}}(\bar{\theta})| < \epsilon/2$, $|J_N(\hat{\theta}^N) - J_{\bar{\theta}}(\hat{\theta}^N)| < \epsilon/2$; hence, $J_{\bar{\theta}}(\hat{\theta}^N) < J_N(\hat{\theta}^N) + \epsilon/2 \leq J_N(\bar{\theta}) + \epsilon/2 < J_{\bar{\theta}}(\bar{\theta}) + \epsilon$, and thus $\|\hat{\theta}^N - \bar{\theta}\| < \beta$. Almost sure statements follow immediately. ■

Lemma 2.11. *Assume that the sequence of functions $\{J_N(\theta)\}$ converges uniformly on Θ to the function $J_{\bar{\theta}}(\theta)$, with $J_N(\theta)$ continuous with respect to*

$\theta \in \Theta$ for any N, Θ a compact subset of $\mathbb{R}^p$. Let $\Theta^\# = \arg\min_{\theta\in\Theta} J_{\bar\theta}(\theta)$ denote the set of minimizers of $J_{\bar\theta}(\theta)$. Then $\lim_{N\to\infty} d(\hat\theta^N, \Theta^\#) = 0$, where $\hat\theta^N \in \arg\min_{\theta\in\Theta} J_N(\theta)$. When the functions $J_N(\cdot)$ are random and the uniform convergence to $J_{\bar\theta}(\cdot)$ is almost sure, the convergence of $d(\hat\theta^N, \Theta^\#)$ to 0 is also almost sure.

Proof. The proof is similar to that of Lemma 2.10, we simply change $J_{\bar\theta}(\bar\theta)$ into $J_{\bar\theta}(\Theta^\#) = \min_{\theta\in\Theta} J_{\bar\theta}(\theta)$ and define

$$\underline{J}(\beta) = \min_{\{\theta\in\Theta:\, d(\theta,\Theta^\#)\geq\beta\}} J_{\bar\theta}(\theta)$$

with $d(\theta, \Theta^\#) = \min_{\theta'\in\Theta^\#} \|\theta - \theta'\|$. ■

Lemma 2.12 (Jennrich 1969). *Let Θ be a convex compact subset of $\mathbb{R}^p$, $\mathscr{Z}$ be a measurable subset of $\mathbb{R}^m$ and $J(z,\theta)$ be a Borel measurable real function on $\mathscr{Z}\times\Theta$, continuously differentiable in $\theta \in \mathrm{int}(\Theta)$ for any $z \in \mathscr{Z}$. Let $\theta^1(z)$ and $\theta^2(z)$ be measurable functions from $\mathscr{Z}$ into Θ. There exists a measurable function $\tilde\theta$ from $\mathscr{Z}$ into $\mathrm{int}(\Theta)$ such that for all $z \in \mathscr{Z}$ $\tilde\theta(z)$ lies on the segment joining $\theta^1(z)$ and $\theta^2(z)$ and*

$$J[z,\theta^1(z)] - J[z,\theta^2(z)] = \left.\frac{\partial J(z,\theta)}{\partial\theta^\top}\right|_{\tilde\theta(z)} [\theta^1(z) - \theta^2(z)]\,.$$

Proof. Let $d(z,\theta)$ denote the Euclidian distance from θ to the segment joining $\theta^1(z)$ and $\theta^2(z)$ and define

$$D(z,\theta) = \left| J[z,\theta^1(z)] - J[z,\theta^2(z)] - \frac{\partial J(z,\theta)}{\partial\theta^\top}[\theta^1(z) - \theta^2(z)] \right| + d(z,\theta)\,,$$

which is a measurable function of z for any $\theta \in \Theta$ and is continuous in θ for any $z \in \mathscr{Z}$. We can then apply Lemma 2.9: there exists a measurable function $\tilde\theta(z)$ from $\mathscr{Z}$ into Θ such that for any $z \in \mathscr{Z}$, $\tilde\theta(z)$ minimizes $D(z,\theta)$ with respect to $\theta \in \Theta$. From the (Taylor) mean value theorem, this $\tilde\theta(z)$ has the property that for any $z \in \mathscr{Z}$, $D[z,\tilde\theta(z)] = 0$, which completes the proof. ■

Lemma 3.4 (Wu 1981). *If for any $\delta > 0$*

$$\liminf_{N\to\infty} \inf_{\|\theta-\bar\theta\|\geq\delta} [S_N(\theta) - S_N(\bar\theta)] > 0 \text{ a.s.} \tag{C.5}$$

then $\hat\theta^N_{LS} \stackrel{\text{a.s.}}{\to} \bar\theta$ as $N\to\infty$. If for any $\delta > 0$

$$\mathrm{Prob}\left\{ \inf_{\|\theta-\bar\theta\|\geq\delta} [S_N(\theta) - S_N(\bar\theta)] > 0 \right\} \to 1\,,\ N\to\infty\,, \tag{C.6}$$

then $\hat\theta^N_{LS} \stackrel{\text{p}}{\to} \bar\theta$ as $N\to\infty$.

Proof. If $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$ is not true, there exists some $\delta > 0$ such that

$$\text{Prob}(\limsup_{N\to\infty} \|\hat{\theta}^N_{LS} - \bar{\theta}\| \geq \delta) > 0\,.$$

Now, $S_N(\hat{\theta}^N_{LS}) - S_N(\bar{\theta}) \leq 0$ from the definition of $\hat{\theta}^N_{LS}$, and $\|\hat{\theta}^N_{LS} - \bar{\theta}\| \geq \delta$ implies $\inf_{\|\theta-\bar{\theta}\|\geq\delta}[S_N(\theta) - S_N(\bar{\theta})] \leq 0$. Therefore, $\text{Prob}[\liminf_{N\to\infty} \inf_{\|\theta-\bar{\theta}\|\geq\delta} [S_N(\theta) - S_N(\bar{\theta})] \leq 0] > 0$, which contradicts (C.5).

When $\inf_{\|\theta-\bar{\theta}\|\geq\delta}[S_N(\theta)-S_N(\bar{\theta})] > 0$, then $\|\hat{\theta}^N_{LS}-\bar{\theta}\| < \delta$. Therefore, when (C.6) is satisfied, for any $\delta > 0$ and $\epsilon > 0$ there exists N_0 such that for all $N > N_0$ $\text{Prob}\{\|\hat{\theta}^N_{LS} - \bar{\theta}\| < \delta\} \geq 1 - \epsilon$, that is, $\hat{\theta}^N_{LS} \overset{\text{p}}{\to} \bar{\theta}$. ∎

Lemma 3.7. *Let* $\mathbf{u}, \mathbf{v}$ *be two random vectors of* $\mathbb{R}^r$ *and* $\mathbb{R}^s$ *respectively defined on a probability space with measure* μ, *with* $\mathbb{E}(\|\mathbf{u}\|^2) < \infty$ *and* $\mathbb{E}(\|\mathbf{v}\|^2) < \infty$. *We have*

$$\mathbb{E}(\mathbf{u}\mathbf{u}^\top) \succeq \mathbb{E}(\mathbf{u}\mathbf{v}^\top)[\mathbb{E}(\mathbf{v}\mathbf{v}^\top)]^{+}\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\,, \tag{C.7}$$

where $\mathbf{M}^+$ *denotes the Moore–Penrose g-inverse of* $\mathbf{M}$ *and* $\mathbf{A} \succeq \mathbf{B}$ *means that* $\mathbf{A} - \mathbf{B}$ *is nonnegative definite. Moreover, the equality is obtained in (C.7) if and only if* $\mathbf{u} = \mathbf{A}\mathbf{v}$ μ*-a.s. for some nonrandom matrix* $\mathbf{A}$.

Proof. Since $\mathbb{E}(\{\mathbf{u}\}_i^2) < \infty$, $i = 1, \ldots, r$ and $\mathbb{E}(\{\mathbf{v}\}_i^2) < \infty$, $i = 1, \ldots, s$, Cauchy–Schwarz inequality gives

$$\left\{\mathbb{E}\left[\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix}(\mathbf{u}^\top\ \mathbf{v}^\top)\right]\right\}_{ij} < \infty$$

for any $i, j = 1, \ldots, r + s$, so that $\mathbb{E}(\mathbf{u}\mathbf{u}^\top)$, $\mathbb{E}(\mathbf{u}\mathbf{v}^\top)$, and $\mathbb{E}(\mathbf{v}\mathbf{v}^\top)$ are well defined. Consider $\mathbb{E}[(\mathbf{x}^\top\mathbf{u} + \mathbf{y}^\top\mathbf{v})^2]$ for some nonrandom $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^r \times \mathbb{R}^s$. By direct expansion, we obtain

$$\mathbb{E}[(\mathbf{x}^\top\mathbf{u})^2] + 2\mathbf{x}^\top\mathbb{E}(\mathbf{u}\mathbf{v}^\top)\mathbf{y} + \mathbf{y}^\top\mathbb{E}(\mathbf{v}\mathbf{v}^\top)\mathbf{y} \geq 0 \tag{C.8}$$

which reaches its minimum value with respect to $\mathbf{y}$ when

$$\mathbb{E}(\mathbf{v}\mathbf{v}^\top)\mathbf{y} = -\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\mathbf{x}\,.$$

This system is compatible, and thus consistent; see Harville (1997, p. 73). Indeed,

$$\begin{aligned}\mathbf{z}^\top\mathbb{E}(\mathbf{v}\mathbf{v}^\top) = \mathbf{0}^\top &\Longrightarrow \mathbb{E}(\mathbf{z}^\top\mathbf{v}\mathbf{v}^\top\mathbf{z}) = 0 \Longrightarrow \mathbf{z}^\top\mathbf{v} = 0\ \mu\text{-a.s.}\\ &\Longrightarrow \mathbf{z}^\top\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\mathbf{x} = \mathbb{E}(\mathbf{z}^\top\mathbf{v}\mathbf{u}^\top\mathbf{x}) = 0\,.\end{aligned}$$

Therefore, the solution $\mathbf{y}^*$ is given by

$$\mathbf{y}^* = -[\mathbb{E}(\mathbf{v}\mathbf{v}^\top)]^{-}\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\mathbf{x}$$

for any g-inverse of $\mathbb{E}(\mathbf{v}\mathbf{v}^\top)$; see Harville (1997, p. 108). Take

$$\mathbf{y}^* = -[\mathbb{E}(\mathbf{v}\mathbf{v}^\top)]^+\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\mathbf{x}$$

with $\mathbf{M}^+$ the Moore–Penrose g-inverse of $\mathbf{M}$, see Harville (1997, p. 493), and substitute $\mathbf{y}^*$ for $\mathbf{y}$ in (C.8). We obtain

$$\mathbf{x}^\top\mathbb{E}(\mathbf{u}\mathbf{u}^\top)\mathbf{x} \geq \mathbf{x}^\top\mathbb{E}(\mathbf{u}\mathbf{v}^\top)[\mathbb{E}(\mathbf{v}\mathbf{v}^\top)]^+\mathbb{E}(\mathbf{v}\mathbf{u}^\top)\mathbf{x}$$

for any nonrandom vector $\mathbf{x} \in \mathbb{R}^r$, i.e., (C.7).

Assume that equality is attained. Taking $\mathbf{A} = \mathbb{E}(\mathbf{u}\mathbf{v}^\top)[\mathbb{E}(\mathbf{v}\mathbf{v}^\top)]^+$ we obtain $\mathbb{E}[(\mathbf{u}-\mathbf{A}\mathbf{v})(\mathbf{u}-\mathbf{A}\mathbf{v})^\top] = \mathbf{O}$ and thus $\mathbf{u} = \mathbf{A}\mathbf{v}$, μ-a.s. ∎

Lemma 5.1. *Let* $\mathbf{A}$ *be a* $p\times p$ *positive-definite matrix and let* $\mathcal{E}_A = \{\mathbf{t} \in \mathbb{R}^p : \mathbf{t}^\top\mathbf{A}\mathbf{t} \leq 1\}$*. Then:*

(i) $\mathrm{vol}(\mathcal{E}_A) = V_p \det^{-1/2}\mathbf{A}$*, with* $V_p = \pi^{p/2}/\Gamma(p/2+1) = \mathrm{vol}[\mathscr{B}(\mathbf{0},1)]$*, the volume of the unit ball* $\mathscr{B}(\mathbf{0},1)$ *in* $\mathbb{R}^p$*.*

(ii) For any vector $\mathbf{c} \in \mathbb{R}^p$ *we have*

$$\max_{\mathbf{t}\in\mathcal{E}_A}(\mathbf{c}^\top\mathbf{t})^2 = \mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c}\,;$$

in particular, when $\|\mathbf{c}\| = 1$*, then* $\mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c}$ *is the squared half-length of the orthogonal projection of* $\mathcal{E}_A$ *onto the straight line defined by* $\mathbf{c}$*.*

(iii) $\max_{\|\mathbf{c}\|=1}\mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c} = 1/\lambda_{\min}(\mathbf{A}) = R^2(\mathcal{E}_A)$*, with* $\lambda_{\min}(\mathbf{A})$ *the minimum eigenvalue of* $\mathbf{A}$ *and* $R(\mathcal{E}_A)$ *the radius of the smallest ball containing* $\mathcal{E}_A$*; the length of a principal axis of* $\mathcal{E}_A$ *equals* $2/\sqrt{\lambda_i(\mathbf{A})}$ *with* $\lambda_i(\mathbf{A})$ *an eigenvalue of* $\mathbf{A}$*.*

(iv) The squared length of the half-diagonal of the parallelepiped containing $\mathcal{E}_A$ *and parallel to the coordinate axes of the Euclidean space* $\mathbb{R}^p$ *equals the sum of the squared half-lengths of the principal axes of* $\mathcal{E}_A$ *and is given by* $\mathrm{trace}(\mathbf{A}^{-1})$*.*

(v) Let $\mathcal{E}_B$ *be defined similarly to* $\mathcal{E}_A$ *but for the* $p\times p$ *positive-definite matrix* $\mathbf{B}$*, then the following statements are equivalent:*

(a) $\mathcal{E}_A \subseteq \mathcal{E}_B$*.*

(b) $\mathbf{A} \succeq \mathbf{B}$*, i.e., the matrix* $\mathbf{A}-\mathbf{B}$ *is nonnegative definite.*

(c) For any $\mathbf{c} \in \mathbb{R}^p$*,* $\mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c} \leq \mathbf{c}^\top\mathbf{B}^{-1}\mathbf{c}$*, i.e.,* $\mathbf{B}^{-1} \succeq \mathbf{A}^{-1}$*.*

Proof.

(*i*) We can write

$$\begin{aligned}\mathrm{vol}(\mathcal{E}_A) &= \int_{\{\mathbf{t}\in\mathbb{R}^p:\mathbf{t}^\top\mathbf{A}\mathbf{t}\leq 1\}} d\mathbf{t} = \int_{\{\mathbf{u}\in\mathbb{R}^p:\mathbf{u}^\top\mathbf{u}\leq 1\}} \det{}^{-1/2}(\mathbf{A})\,d\mathbf{u}\\ &= [\det{}^{-1/2}\mathbf{A}]\,\mathrm{vol}[\mathscr{B}(\mathbf{0},1)]\,.\end{aligned}$$

(*ii*) From Cauchy–Schwarz inequality we have

$$\forall \mathbf{t} \in \mathbb{R}^p\,, \quad (\mathbf{c}^\top \mathbf{t})^2 = [(\mathbf{A}^{-1/2}\mathbf{c})^\top (\mathbf{A}^{1/2}\mathbf{t})]^2 \leq (\mathbf{c}^\top \mathbf{A}^{-1}\mathbf{c})(\mathbf{t}^\top \mathbf{A}\mathbf{t})\,.$$

Therefore,

$$\mathbf{c}^\top \mathbf{A}^{-1}\mathbf{c} \geq \sup_{\mathbf{t}\neq \mathbf{0}} \frac{(\mathbf{c}^\top \mathbf{t})^2}{\mathbf{t}\mathbf{A}\mathbf{t}}\,.$$

For $\mathbf{t} = \mathbf{A}^{-1}\mathbf{c}$, the ratio on right-hand side equals $\mathbf{c}^\top \mathbf{A}^{-1}\mathbf{c}$, so that

$$\mathbf{c}^\top \mathbf{A}^{-1}\mathbf{c} = \max_{\mathbf{t}\neq \mathbf{0}} \frac{(\mathbf{c}^\top \mathbf{t})^2}{\mathbf{t}\mathbf{A}\mathbf{t}} = \max_{\{\mathbf{t}\in\mathbb{R}^p:\mathbf{t}^\top\mathbf{A}\mathbf{t}\leq 1\,,\ \mathbf{t}\neq\mathbf{0}\}} \frac{(\mathbf{c}^\top \mathbf{t})^2}{\mathbf{t}\mathbf{A}\mathbf{t}} = \max_{\mathbf{t}\in\mathcal{E}_A} (\mathbf{c}^\top \mathbf{t})^2\,.$$

When $\|\mathbf{c}\| = 1$, then $\mathbf{c}\mathbf{c}^\top\mathbf{t}$ is the orthogonal projection of $\mathbf{t}$ onto the straight line defined by $\mathbf{c}$ and its squared length equals $\|\mathbf{c}\mathbf{c}^\top\mathbf{t}\|^2 = (\mathbf{c}^\top\mathbf{t})^2$.

(*iii*) The largest orthogonal projection of $\mathcal{E}_A$ onto the straight line defined by $\mathbf{c}$ is obtained when $\mathbf{c}$ goes in the direction of the main axis of $\mathcal{E}_A$. Let $\lambda_1 = \lambda_{\min}(\mathbf{A}) \leq \lambda_2 \leq \cdots \leq \lambda_p$ denote the eigenvalues of $\mathbf{A}$. In a basis of associated eigenvectors, $\mathcal{E}_A$ is defined by $\{\mathbf{y} \in \mathbb{R}^p : \sum_{i=1}^p y_i^2\,\lambda_i \leq 1\}$, with y_i the i-th component of $\mathbf{y}$. The half-length of the longest principal axis of $\mathcal{E}_A$ is thus $R(\mathcal{E}_A) = 1/\sqrt{\lambda_1}$. The length of the i-th principal axis is $2/\sqrt{\lambda_i}$.

(*iv*) From the same arguments as above, the sum of the squared half-lengths of the principal axes of $\mathcal{E}_A$ is $\sum_{i=1}^p \lambda_i^{-1} = \mathrm{trace}(\mathbf{A}^{-1})$. Let $\mathbf{e}_k$ denote the k-th basis vector of $\mathbb{R}^p$; then $\{\mathbf{A}^{-1}\}_{kk} = \mathbf{e}_k^\top\mathbf{A}^{-1}\mathbf{e}_k$ is the squared half-length of the orthogonal projection of $\mathcal{E}_A$ onto the k-th coordinate axis. By the Pythagorean relation in $\mathbb{R}^p$, we obtain that the squared length of the half-diagonal of the parallelepiped containing $\mathcal{E}_A$ and parallel to the coordinate axes equals $\sum_{k=1}^p \{\mathbf{A}^{-1}\}_{kk} = \mathrm{trace}(\mathbf{A}^{-1})$.

(*v*) The implication (b) $\Longrightarrow$ (a) is a direct consequence of the definitions of $\mathcal{E}_A$ and $\mathcal{E}_B$. The implication (a) $\Longrightarrow$ (c) follows from (*ii*). Suppose that (c) holds. Take any vector $\mathbf{v} \in \mathbb{R}^p$ and denote $\mathbf{s} = \mathbf{A}\mathbf{v}$, $\mathbf{z} = \mathbf{B}\mathbf{v}$. We have

$$\begin{aligned} 0 &\leq (\mathbf{s}-\mathbf{z})^\top\mathbf{A}^{-1}(\mathbf{s}-\mathbf{z}) = \mathbf{s}^\top\mathbf{A}^{-1}\mathbf{s} + \mathbf{z}^\top\mathbf{A}^{-1}\mathbf{z} - 2\mathbf{s}^\top\mathbf{A}^{-1}\mathbf{z} \\ &\leq \mathbf{s}^\top\mathbf{A}^{-1}\mathbf{s} + \mathbf{z}^\top\mathbf{B}^{-1}\mathbf{z} - 2\mathbf{v}^\top\mathbf{z} = \mathbf{s}^\top\mathbf{A}^{-1}\mathbf{s} + \mathbf{z}^\top\mathbf{B}^{-1}\mathbf{z} - 2\mathbf{z}^\top\mathbf{B}^{-1}\mathbf{z} \\ &= \mathbf{v}^\top\mathbf{A}\mathbf{v} - \mathbf{v}^\top\mathbf{B}\mathbf{v}\,; \end{aligned}$$

that is, $\mathbf{A} \succeq \mathbf{B}$. ■

Lemma 5.2. *Suppose that the estimator* $\hat\theta^N$ *in the regression model (3.2) satisfies* $\sqrt{N}(\hat\theta^N - \bar\theta) \xrightarrow{\mathrm{d}} \mathbf{z} \sim \mathscr{N}(\mathbf{0}, \mathbf{M}^{-1}(\xi,\bar\theta))$ *as* $N \to \infty$*. Then, for* N *large we have approximately*

$$\text{Prob}\left\{y(x_1),\ldots,y(x_N) : \forall x \in \mathscr{X}\,, |\eta(x,\hat{\theta}^N) - \eta(x,\bar{\theta})| \leq \frac{1}{\sqrt{N}} \left[\chi_p^2(1-\alpha)\frac{\partial \eta(x,\theta)}{\partial \theta^\top}\bigg|_{\bar{\theta}} \mathbf{M}^{-1}(\xi,\bar{\theta})\frac{\partial \eta(x,\theta)}{\partial \theta}\bigg|_{\bar{\theta}}\right]^{1/2}\right\} \geq 1-\alpha$$

where $\chi_p^2(1-\alpha)$ *is the* $(1-\alpha)$ *quantile of the* χ_p^2 *distribution.*

Proof. Since $\hat{\theta}^N - \bar{\theta}$ is approximately normal $\mathscr{N}(\mathbf{0}, \mathbf{M}^{-1}(\xi,\bar{\theta})/N)$, the quantity $N(\hat{\theta}^N - \bar{\theta})^\top \mathbf{M}(\xi,\bar{\theta})(\hat{\theta}^N - \bar{\theta})$ follows approximately the χ_p^2 distribution. Hence, for N large

$$\text{Prob}\left\{\mathbf{y} : (\hat{\theta}^N - \bar{\theta})^\top \mathbf{H}(\hat{\theta}^N - \bar{\theta}) \leq 1\right\} \simeq 1-\alpha$$

where $\mathbf{H} = N\mathbf{M}(\xi,\bar{\theta})/\chi_p^2(1-\alpha)$ and $\mathbf{y} = [y(x_1),\ldots,y(x_N)]^\top$. Since $\mathbf{u}^\top\mathbf{H}\mathbf{u} \leq 1$ is equivalent to $(\mathbf{v}^\top\mathbf{u})^2 \leq \mathbf{v}^\top\mathbf{H}^{-1}\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^p$ (from Cauchy–Schwarz inequality), for large N we have

$$\begin{aligned}
&\text{Prob}\left\{\mathbf{y} : \forall x \in \mathscr{X}\,, |\eta(x,\hat{\theta}^N) - \eta(x,\bar{\theta})|^2 \leq \frac{1}{N}\,\chi_p^2(1-\alpha)\frac{\partial \eta(x,\theta)}{\partial \theta^\top}\bigg|_{\bar{\theta}} \mathbf{M}^{-1}(\xi,\bar{\theta})\frac{\partial \eta(x,\theta)}{\partial \theta}\bigg|_{\bar{\theta}}\right\} \\
&\simeq \text{Prob}\left\{\mathbf{y} : \forall x \in \mathscr{X}\,, \left|\frac{\partial \eta(x,\theta)}{\partial \theta^\top}\bigg|_{\bar{\theta}}(\hat{\theta}^N - \bar{\theta})\right|^2 \leq \frac{\partial \eta(x,\theta)}{\partial \theta^\top}\bigg|_{\bar{\theta}} \mathbf{H}^{-1}\frac{\partial \eta(x,\theta)}{\partial \theta}\bigg|_{\bar{\theta}}\right\} \\
&\geq \text{Prob}\left\{\mathbf{y} : \forall \mathbf{v} \in \mathbb{R}^p\,, \left|\mathbf{v}^\top(\hat{\theta}^N - \bar{\theta})\right|^2 \leq \mathbf{v}^\top\mathbf{H}^{-1}\mathbf{v}\right\} \\
&= \text{Prob}\left\{\mathbf{y} : (\hat{\theta}^N - \bar{\theta})^\top \mathbf{H}(\hat{\theta}^N - \bar{\theta}) \leq 1\right\} \simeq 1-\alpha\,.
\end{aligned}$$

■

Lemma 5.4 (Pukelsheim 1993, Sects. 5.2, 5.4). *Let* $\Phi(\cdot)$ *be a function from* $\mathbb{M}^{\geq}$ *to* $\mathbb{R}$. *Then,*

(i) When $\Phi(\cdot)$ *is positively homogeneous, it is concave if and only if it is superadditive, i.e.,* $\Phi(\mathbf{M}_1+\mathbf{M}_2) \geq \Phi(\mathbf{M}_1)+\Phi(\mathbf{M}_2)$ *for all* $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{M}^{\geq}$.
(ii) When $\Phi(\cdot)$ *is superadditive, nonnegativity implies isotonicity.*
(iii) When $\Phi(\cdot)$ *is positively homogeneous, isotonicity implies nonnegativity (i.e.,* $\Phi(\mathbf{M}) \geq 0$ *for all* $\mathbf{M}$ *in* $\mathbb{M}^{\geq}$*); moreover, either* Φ *is identically zero or* $\Phi(.)$ *is strictly positive on the open set* $\mathbb{M}^{>}$.

Proof.

(i) Take any $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{M}^{\geq}$, any $\alpha \in (0,1)$. Superadditivity gives $\Phi[(1-\alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] \geq \Phi[(1-\alpha)\mathbf{M}_1] + \Phi(\alpha\mathbf{M}_2) = (1-\alpha)\Phi(\mathbf{M}_1) + \alpha\Phi(\mathbf{M}_2)$ and thus implies concavity. Conversely, concavity implies $\Phi(\mathbf{M}_1+\mathbf{M}_2) = \Phi[(2\mathbf{M}_1 + 2\mathbf{M}_2)/2] \geq (1/2)\Phi(2\mathbf{M}_1) + (1/2)\Phi(2\mathbf{M}_2) = \Phi(\mathbf{M}_1) + \Phi(\mathbf{M}_2)$.

(*ii*) Take any $\mathbf{M}_1 \succeq \mathbf{M}_2 \in \mathbb{M}^{\geq}$. Superadditivity and nonnegativity imply $\Phi(\mathbf{M}_1) - \Phi(\mathbf{M}_2) = \Phi(\mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_2) - \Phi(\mathbf{M}_2) \geq \Phi(\mathbf{M}_1 - \mathbf{M}_2) \geq 0$ so that $\Phi(\cdot)$ is isotonic.

(*iii*) Isotonicity implies $\Phi(\mathbf{M}) \geq \Phi(\mathbf{O})$ for any $\mathbf{M} \in \mathbb{M}^{\geq}$, and positive homogeneity gives $\Phi(\mathbf{O}) = \Phi(0\mathbf{M}) = 0$, so that $\Phi(\cdot)$ is nonnegative. If Φ is non identically zero, there exists some $\mathbf{M}^*$ in $\mathbb{M}^{\geq}$ such that $\Phi(\mathbf{M}^*) > 0$. Then, for any $\mathbf{M} \in \mathbb{M}^{>}$, there exists $\alpha > 0$ such that $\alpha\mathbf{M} - \mathbf{M}^* \succeq \mathbf{O}$ and isotonicity with positive homogeneity imply $\Phi(\mathbf{M}) = \Phi(\alpha\mathbf{M})/\alpha \geq \Phi(\mathbf{M}^*)/\alpha > 0$. ■

Lemma 5.5. *For any $p \times p$ matrix $\mathbf{M}$ in $\mathbb{M}^{\geq}$ and any $\mathbf{c} \in \mathcal{M}(\mathbf{M})$ (i.e., such that $\mathbf{c} = \mathbf{M}\mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^p$) we have*

$$\Phi_c(\mathbf{M}) = -\mathbf{c}^\top \mathbf{M}^- \mathbf{c} = \min_{\mathbf{z} \in \mathbb{R}^p} [\mathbf{z}^\top \mathbf{M} \mathbf{z} - 2\mathbf{z}^\top \mathbf{c}] \, .$$

When $\mathbf{c} \notin \mathcal{M}(\mathbf{M})$, the right-hand side equals $-\infty$.

Proof. When $\mathbf{c} \in \mathcal{M}(\mathbf{M})$, we can write $[\mathbf{z}^\top \mathbf{M} \mathbf{z} - 2\mathbf{z}^\top \mathbf{c}] - \Phi_c(\mathbf{M}) = (\mathbf{M}^-\mathbf{c} - \mathbf{z})^\top \mathbf{M} (\mathbf{M}^-\mathbf{c} - \mathbf{z}) \geq 0$, with $\mathbf{M}^-$ any g-inverse of $\mathbf{M}$. When $\mathbf{c} \notin \mathcal{M}(\mathbf{M})$, take $\mathbf{z} = \gamma\mathbf{u}$ with $\gamma > 0$ and $\mathbf{u}$ any element of $\mathcal{N}(\mathbf{M}) = \{\mathbf{u} \in \mathbb{R}^p : \mathbf{M}\mathbf{u} = \mathbf{0}\}$ such that $\mathbf{c}^\top \mathbf{u} = s > 0$. Then, $\mathbf{z}^\top \mathbf{M} \mathbf{z} = 0$ and $\mathbf{z}^\top \mathbf{c} = \gamma s$ which can be made arbitrarily large. ■

Lemma 5.6. *For any $p \times p$ matrix $\mathbf{M}$ in $\mathbb{M}^{\geq}$ and any $\mathbf{c} \in \mathcal{M}(\mathbf{M})$, we have*

$$\Phi_c^+(\mathbf{M}) = (\mathbf{c}^\top \mathbf{c})(\mathbf{c}^\top \mathbf{M}^- \mathbf{c})^{-1} = (\mathbf{c}^\top \mathbf{c}) \min_{\mathbf{z}^\top \mathbf{c} = 1} \mathbf{z}^\top \mathbf{M} \mathbf{z} \, .$$

When $\mathbf{c} \notin \mathcal{M}(\mathbf{M})$, the minimum on the right-hand side equals 0.

Proof. When $\mathbf{c} \in \mathcal{M}(\mathbf{M})$, Cauchy–Schwarz inequality gives

$$\begin{aligned}(\mathbf{c}^\top \mathbf{M}^- \mathbf{c})(\mathbf{z}^\top \mathbf{M} \mathbf{z}) &= (\mathbf{c}^\top \mathbf{M}^- \mathbf{M} \mathbf{M}^- \mathbf{c})(\mathbf{z}^\top \mathbf{M} \mathbf{z}) \\ &\geq (\mathbf{c}^\top \mathbf{M}^- \mathbf{M} \mathbf{z})^2 = (\mathbf{c}^\top \mathbf{z})^2\end{aligned}$$

for any $\mathbf{z} \in \mathbb{R}^p$. Therefore,

$$\mathbf{c}^\top \mathbf{M}^- \mathbf{c} \geq \sup_{\mathbf{z}^\top \mathbf{M} \mathbf{z} \neq 0} \frac{(\mathbf{c}^\top \mathbf{z})^2}{\mathbf{z}^\top \mathbf{M} \mathbf{z}} \, .$$

Taking $\mathbf{z} = \mathbf{M}^- \mathbf{c}$ gives equality since then $\mathbf{z}^\top \mathbf{M} \mathbf{z} = \mathbf{c}^\top \mathbf{M}^- \mathbf{c} > 0$. We can thus write

$$(\mathbf{c}^\top \mathbf{M}^- \mathbf{c}) = \sup_{\mathbf{z}^\top \mathbf{M} \mathbf{z} \neq 0} \frac{(\mathbf{c}^\top \mathbf{z})^2}{\mathbf{z}^\top \mathbf{M} \mathbf{z}} = \sup_{\mathbf{z}^\top \mathbf{c} = 1} \frac{1}{\mathbf{z}^\top \mathbf{M} \mathbf{z}} \, .$$

When $\mathbf{c} \notin \mathcal{M}(\mathbf{M})$, take $\mathbf{z}$ as any element of $\mathcal{N}(\mathbf{M})$ such that $\mathbf{c}^\top \mathbf{z} \neq 0$. Then, $\mathbf{z}^\top \mathbf{M} \mathbf{z} = 0$, and the supremum in the equation above is infinite. ■

Lemma 5.7. *For any $p \times p$ matrix $\mathbf{M}$ in $\mathbb{M}^{\geq}$ partitioned as*

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}$$

with $\mathbf{M}_{11}$ of dimension $s \times s$, we have

$$\log\det(\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-}\mathbf{M}_{21}) \leq \log\det(\mathbf{M}_{11} + \mathbf{D}^\top\mathbf{M}_{22}\mathbf{D} - \mathbf{M}_{12}\mathbf{D} - \mathbf{D}^\top\mathbf{M}_{21})$$

for any $\mathbf{D} \in \mathbb{R}^{(p-s)\times s}$, with equality if and only if $\mathbf{M}_{22}\mathbf{D} = \mathbf{M}_{12}$.

Proof. Take any matrix $\mathbf{C}$ solution of $\mathbf{M}_{22}\mathbf{C} = \mathbf{M}_{21}$ (which is equivalent to $\mathbf{C} = \mathbf{M}_{22}^{-}\mathbf{M}_{21}$ for some g-inverse $\mathbf{M}_{22}^{-}$ of $\mathbf{M}_{22}$), and denote $\mathbf{M}^* = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-}\mathbf{M}_{21}$. Then, for any matrix $\mathbf{D}$ in $\mathbb{R}^{(p-s)\times s}$,

$$\mathbf{M}_{11} + \mathbf{D}^\top\mathbf{M}_{22}\mathbf{D} - \mathbf{M}_{12}\mathbf{D} - \mathbf{D}^\top\mathbf{M}_{21} - \mathbf{M}^* = (\mathbf{C} - \mathbf{D})^\top\mathbf{M}_{22}(\mathbf{C} - \mathbf{D}),$$

which is positive definite unless $\mathbf{M}_{22}\mathbf{D} = \mathbf{M}_{22}\mathbf{C} = \mathbf{M}_{21}$. When $\mathbf{M}^*$ is nonsingular, the strict isotonicity of the function $\log\det(\cdot)$ on $\mathbb{M}^{>}$ (see Sect. 5.1.5) concludes the proof. When $\mathbf{M}^*$ is singular, $\log\det(\mathbf{M}^*) = -\infty$, we also have $\log\det[\mathbf{M}_{11} + \mathbf{D}^\top\mathbf{M}_{22}\mathbf{D} - \mathbf{M}_{12}\mathbf{D} - \mathbf{D}^\top\mathbf{M}_{21}] = -\infty$ when $\mathbf{M}_{22}\mathbf{D} = \mathbf{M}_{21}$ since then $\mathbf{M}_{11} + \mathbf{D}^\top\mathbf{M}_{22}\mathbf{D} - \mathbf{M}_{12}\mathbf{D} - \mathbf{D}^\top\mathbf{M}_{21} = \mathbf{M}^*$. ■

Lemma 5.11. *The criterion $\phi_c(\cdot) = \Phi_c[\mathbf{M}(\cdot)]$, with $\Phi_c(\mathbf{M})$ given by (5.9), is upper semicontinuous at any $\xi_* \in \Xi_c = \{\xi \in \Xi : \mathbf{c} \in \mathcal{M}[\mathbf{M}(\xi)]\}$.*

Proof. Take $\xi_* \in \Xi_c$, and consider any sequence $\{\xi_n\}$ of measures in Ξ converging weakly to ξ_*. We have $\phi_c(\xi_n) = -\infty$ if $\xi_n \in \Xi \setminus \Xi_c$ and, from Lemma 5.5, $\phi_c(\xi_n) \leq \mathbf{z}^\top\mathbf{M}(\xi_n)\mathbf{z} - 2\mathbf{z}^\top\mathbf{c}$ for any $\mathbf{z} \in \mathbb{R}^p$ otherwise. Therefore, for any $\mathbf{z} \in \mathbb{R}^p$,

$$\limsup_{n\to\infty} \phi_c(\xi_n) \leq \limsup_{n\to\infty} [\mathbf{z}^\top\mathbf{M}(\xi_n)\mathbf{z} - 2\mathbf{z}^\top\mathbf{c}] = \mathbf{z}^\top\mathbf{M}(\xi_*)\mathbf{z} - 2\mathbf{z}^\top\mathbf{c};$$

that is,

$$\limsup_{n\to\infty} \phi_c(\xi_n) \leq \min_{\mathbf{z}\in\mathbb{R}^p} [\mathbf{z}^\top\mathbf{M}(\xi_*)\mathbf{z} - 2\mathbf{z}^\top\mathbf{c}] = \phi_c(\xi_*),$$

so that $\phi_c(\cdot)$ is upper semicontinuous at ξ_*. ■

Lemma 5.12. *Consider a sequence of matrices satisfying (5.19) and suppose that $\mathbf{c} \in \mathcal{M}(\mathbf{M}_0)$. Then, under the conditions*
C1: $\|\mathbf{R}_t\| = [\mathrm{trace}(\mathbf{R}_t^\top\mathbf{R}_t)]^{1/2} = o(t^\alpha)$ as $t \to 0^+$,
and
C2: $\mathbf{M}_0 + t^\alpha\mathbf{M}_\alpha \in \mathbb{M}^{>}$ for arbitrary small $t > 0$,
we have

$$\lim_{t\to 0^+} \Phi_c[\mathbf{M}(t)] = \Phi_c(\mathbf{M}_0). \tag{C.9}$$

Proof. Define $r_t = \|\mathbf{R}_t\|/t^\alpha$. We first show that $\mathbf{M}(t) - (1-\sqrt{r_t})(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha) \in \mathbb{M}^>$ for t small enough. Take any $\mathbf{z} \in \mathbb{R}^p$, $\mathbf{z} \neq \mathbf{0}$. We have

$$\begin{aligned}\mathbf{z}^\top \mathbf{M}(t)\mathbf{z} - (1-\sqrt{r_t})\,\mathbf{z}^\top(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha)\mathbf{z} &= \sqrt{r_t}\,\mathbf{z}^\top(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha)\mathbf{z} + \mathbf{z}^\top \mathbf{R}_t \mathbf{z} \\ &\geq \sqrt{r_t}\,\mathbf{z}^\top(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha)\mathbf{z} - \|\mathbf{z}\|^2\|\mathbf{R}_t\| \\ &= \sqrt{r_t}\,t^\alpha\,\|\mathbf{z}\|^2 \left(\frac{\mathbf{z}^\top(\mathbf{M}_0/t^\alpha + \mathbf{M}_\alpha)\mathbf{z}}{\|\mathbf{z}\|^2} - \sqrt{r_t}\right) \\ &\geq \sqrt{r_t}\,t^\alpha\,\|\mathbf{z}\|^2 \left(\frac{\mathbf{z}^\top(\mathbf{M}_0 + t_0^\alpha \mathbf{M}_\alpha)\mathbf{z}}{t_0^\alpha\|\mathbf{z}\|^2} - \sqrt{r_t}\right)\end{aligned}$$

for all $t < t_0$. $\mathbf{z}^\top(\mathbf{M}_0 + t_0^\alpha \mathbf{M}_\alpha)\mathbf{z}/[t_0^\alpha\|\mathbf{z}\|^2] > 0$ from C2, while $\sqrt{r_t}$ tends to zero as $t \to 0$ from C1. Therefore, there exists t_1 such that $\mathbf{M}(t) - (1-\sqrt{r_t})(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha) \in \mathbb{M}^>$ for $0 < t < t_1$. We thus obtain $\Phi_c[\mathbf{M}(t)] \geq (1-\sqrt{r_t})^{-1}\Phi_c(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha)$ for $0 < t < t_1$.

Next, we write $\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha$ as

$$\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha = (1-\gamma_t)\mathbf{M}_0 + \gamma_t \mathbf{M}_{0,\alpha}$$

with $\mathbf{M}_{0,\alpha} = \mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha$ and $\gamma_t = (t/t_0)^\alpha$. Then, from the concavity of $\Phi_c(\cdot)$, $\Phi_c(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha) \geq (1-\gamma_t)\Phi_c(\mathbf{M}_0) + \gamma_t \Phi_c(\mathbf{M}_{0,\alpha})$, which implies

$$\liminf_{t\to 0^+} \Phi_c[\mathbf{M}(t)] \geq \lim_{t\to 0^+} \frac{1}{1-\sqrt{r_t}}\,[(1-\gamma_t)\Phi_c(\mathbf{M}_0) + \gamma_t \Phi_c(\mathbf{M}_{0,\alpha})] = \Phi_c(\mathbf{M}_0)\,.$$

Finally, from the upper semicontinuity of $\Phi_c(\cdot)$ we have $\limsup_{t\to 0^+} \Phi_c[\mathbf{M}(t)] \leq \Phi_c(\mathbf{M}_0)$, which implies (C.9). ■

Corollary 5.13. *For a sequence of matrices $\mathbf{M}(t) \in \mathbb{M}^\geq$ satisfying (5.19) with $\mathbf{c} \in \mathcal{M}(\mathbf{M}_0)$ and the condition C1, either the continuity property (C.9) is satisfied or the convergence of $\mathbf{M}(t)$ to $\mathbf{M}_0$ is along a hyperplane tangent to the cone $\mathbb{M}^\geq$ at $\mathbf{M}_0$, i.e., $\mathbf{M}_\alpha$ belongs to a supporting hyperplane to $\mathbb{M}^\geq$ at $\mathbf{M}_0$.*

Proof. Let $\mathbf{A} \in \mathbb{M}$ define a supporting hyperplane $\mathcal{H}_A$ to the cone $\mathbb{M}^\geq$ at $\mathbf{M}_0$; it satisfies $\text{trace}(\mathbf{A}\mathbf{M}_0) = 0$ and $\text{trace}(\mathbf{A}\mathbf{M}) \geq 0$ for any $\mathbf{M} \in \mathbb{M}^\geq$ ($\mathbf{A}$ is thus normal to $\mathbb{M}^\geq$ at $\mathbf{M}_0$ and $\mathbf{A} \in \mathbb{M}^\geq$). We have $\text{trace}[\mathbf{A}\mathbf{M}(t)] = t^\alpha\,\text{trace}(\mathbf{A}\mathbf{M}_\alpha) + \text{trace}(\mathbf{A}\mathbf{R}_t) \geq 0$ (since $\mathbf{M}(t) \in \mathbb{M}^\geq$), and thus $\text{trace}(\mathbf{A}\mathbf{M}_\alpha) \geq -\|\mathbf{A}\|\,\|\mathbf{R}_t\|/t^\alpha$, which tends to zero from C1. This implies $\text{trace}(\mathbf{A}\mathbf{M}_\alpha) \geq 0$; that is, $\mathbf{M}_\alpha$ is on the same side of $\mathcal{H}_A$ as $\mathbb{M}^\geq$.

There are two alternatives. Either C2 is satisfied and Lemma 5.12 implies (C.9), or C2 in not satisfied. In the latter case, for any $t > 0$ there exists $\mathbf{z}_t$ with $\|\mathbf{z}_t\| = 1$ such that $\mathbf{z}_t^\top(\mathbf{M}_0 + t^\alpha \mathbf{M}_\alpha)\mathbf{z}_t \leq 0$. From any such sequence $\{\mathbf{z}_t\}$ we extract a subsequence converging to some $\mathbf{z}_*$, which thus satisfies $\mathbf{z}_*^\top \mathbf{M}_0 \mathbf{z}_* \leq 0$, and therefore $\mathbf{z}_*^\top \mathbf{M}_0 \mathbf{z}_* = 0$ since $\mathbf{M}_0 \in \mathbb{M}^\geq$. Also, $\mathbf{z}_t^\top \mathbf{M}_\alpha \mathbf{z}_t \leq 0$ (since $\mathbf{z}_t^\top \mathbf{M}_0 \mathbf{z}_t \geq 0$) and thus $\mathbf{z}_*^\top \mathbf{M}_\alpha \mathbf{z}_* \leq 0$. Take $\mathbf{A} = \mathbf{z}_* \mathbf{z}_*^\top$; it defines a supporting hyperplane $\mathcal{H}_A$ to $\mathbb{M}^{\geq}$ at $\mathbf{M}_0$. From the developments above we obtain $\text{trace}(\mathbf{A}\mathbf{M}_\alpha) = \mathbf{z}_*^\top \mathbf{M}_\alpha \mathbf{z}_* \geq 0$ and thus $\text{trace}(\mathbf{A}\mathbf{M}_\alpha) = 0$; that is, $\mathbf{M}_\alpha$ belongs to $\mathcal{H}_A$. ■

Remark.

(*i*) When the sequence of matrices $\mathbf{M}(t)$ satisfies (5.19) and C1 with $\mathbf{c} \in \mathcal{M}(\mathbf{M}_0)$, if $\lim_{t\to 0^+} \Phi_c[\mathbf{M}(t)] \neq \Phi_c(\mathbf{M}_0)$, it means that C2 is not satisfied and, from the proof of Corollary 5.13, that $\mathbf{M}_\alpha$ belongs to a supporting hyperplane to $\mathbb{M}^{\geq}$ at $\mathbf{M}_0$. Conversely, if $\mathbf{M}_\alpha$ does not belong to such a tangent hyperplane, C2 and thus (C.9) are satisfied.

(*ii*) The condition C1 in Corollary 5.13 can be replaced by $\mathbf{R}_t \in \mathbb{M}^{\geq}$. Indeed, in that case $\mathbf{M}(t) - (\mathbf{M}_0 + t^\alpha \mathbf{R}_t) \in \mathbb{M}^{\geq}$, $\Phi_c[\mathbf{M}(t)] \geq \Phi_c(\mathbf{M}_0 + t^\alpha \mathbf{R}_t)$, and the rest of the proof is similar to that of Corollary 5.13. □

Lemma 5.28. *When the design criterion $\Phi(\cdot)$ is isotonic, an optimal design is supported at values of x such that $\mathbf{g}_\theta(x)$ is on the boundary of the Elfving's set $\mathscr{F}_\theta$.*

Proof. Suppose that $\mathbf{M}(\xi,\theta) = \sum_{i=1}^m \xi_i \, \mathbf{g}_\theta(x^{(i)}) \mathbf{g}_\theta^\top(x^{(i)})$, $m \leq p(p+1)/2+1$, see Sect. 5.2.3, with $x^{(1)}$ such that $\mathbf{g}_\theta(x^{(1)})$ lies in the interior of $\mathscr{F}_\theta$. We can then decompose $\mathbf{g}_\theta(x^{(1)})$ into

$$\mathbf{g}_\theta(x^{(1)}) = \sum_{j=1}^{p+1} \alpha_j \mathbf{g}_j \,,$$

with $\alpha_j \geq 0$, $\sum_{j=1}^{p+1} \alpha_j = 1$, and $\mathbf{g}_j = \pm \mathbf{g}_\theta(x^{(j)})$ for some $x^{(j)}$ with $\mathbf{g}_j$ belonging to the boundary of $\mathscr{F}_\theta$. For any $\mathbf{u} \in \mathbb{R}^p$, consider $\Delta(\mathbf{u}) = \mathbf{u}^\top \mathbf{M}'(\xi,\theta)\mathbf{u} - \mathbf{u}^\top \mathbf{M}(\xi,\theta)\mathbf{u}$, where $\mathbf{M}'(\xi,\theta)$ is obtained by substituting $\sum_{j=1}^{p+1} \alpha_j \mathbf{g}_j \mathbf{g}_j^\top$ for $\mathbf{g}_\theta(x^{(1)})\mathbf{g}_\theta^\top(x^{(1)})$ in $\mathbf{M}(\xi,\theta)$. We have

$$\begin{aligned}
\Delta(\mathbf{u}) &= \xi_1 \left\{ \sum_{j=1}^{p+1} \alpha_j \, [\mathbf{u}^\top \mathbf{g}_\theta(x^{(j)})]^2 - [\mathbf{u}^\top \mathbf{g}_\theta(x^{(1)})]^2 \right\} \\
&= \xi_1 \left\{ \sum_{j=1}^{p+1} \alpha_j \, [\mathbf{u}^\top \mathbf{g}_j]^2 - \left[\sum_{j=1}^{p+1} \alpha_j (\mathbf{u}^\top \mathbf{g}_j) \right]^2 \right\}
\end{aligned}$$

and $\Delta(\mathbf{u}) \geq 0$ from Cauchy–Schwarz inequality. Therefore $\mathbf{M}'(\xi,\theta) \succeq \mathbf{M}(\xi,\theta)$ and $\Phi[\mathbf{M}'(\xi,\theta)] \geq \Phi[\mathbf{M}(\xi,\theta)]$. ■

Lemma 7.9. *Assume that $\eta(\theta)$ is continuous for $\theta \in \Theta$, a compact subset of $\mathbb{R}^p$. We have:*

(i) *For any* $\theta, \theta' \in \Theta_\eta(t)$, $E_\eta(\|\theta - \theta'\|^2) < 4t$, *and the maximum diameter* $\overline{D}(t)$ *of any connected part of* $\Theta_\eta(t)$ *satisfies* $\overline{D}^2(t) \leq \inf\{\delta : E_\eta(\delta) \geq 4t\}$.

(ii) *Suppose that the probability measure of the observations* $\mathbf{y}$ *has a density with respect to the Lebesgue measure in* $\mathbb{R}^N$. *If there exists* $\delta' < \delta$ *such that* $E_\eta(\delta) < t < E_\eta(\delta')$, *then the probability that the set* $\Theta_\eta(t)$ *is not connected is strictly positive.*

Proof.

(i) For any $\theta, \theta' \in \Theta_\eta(t)$, $\|\eta(\theta) - \eta(\theta')\| \leq \|\eta(\theta) - \mathbf{y}\| + \|\mathbf{y} - \eta(\theta')\| < 2\sqrt{t}$, therefore $E_\eta(\|\theta - \theta'\|^2) < 4t$. Let $\mathcal{C}(t)$ denote any connected part of $\Theta_\eta(t)$; if θ and θ' are in $\mathcal{C}(t)$, for any $\delta \in [0, \|\theta - \theta'\|^2]$, there exists $\theta'' \in \mathscr{B}(\theta, \sqrt{\delta})$ such that $\theta'' \in \mathcal{C}(t)$, with $\mathscr{B}(\theta, \sqrt{\delta})$ the closed ball of center θ and radius $\sqrt{\delta}$. This implies $E_\eta(\delta) < 4t$ for any $\delta \in [0, \mathrm{diam}^2[\mathcal{C}(t)]]$, and therefore $\mathrm{diam}^2[\mathcal{C}(t)] \leq \inf\{\delta : E_\eta(\delta) \geq 4t\}$.

(ii) Define $\alpha = (1/2)\min\{E_\eta(\delta') - t, t - E_\eta(\delta)\}$, take θ_1 and θ_2 in Θ such that $\|\theta_1 - \theta_2\|^2 = \delta$ and $\|\eta(\theta_1) - \eta(\theta_2)\|^2 = E_\eta(\delta)$. Consider the set $\mathcal{A}_{\delta'} = \Theta \cap \{\theta : \|\theta - \theta_1\|^2 = \delta'\}$.
Suppose first that $\mathcal{A}_{\delta'}$ is empty. Suppose that $\|\mathbf{y} - \eta(\theta_1)\| \leq \sqrt{E_\eta(\delta)}$ and $\|\mathbf{y} - \eta(\theta_2)\| \leq \sqrt{E_\eta(\delta)}$, which happens with a strictly positive probability. Then $E_\eta(\delta) < t$ implies that $\theta_1 \in \Theta_\eta(t)$, $\theta_2 \in \Theta_\eta(t)$, and $\Theta_\eta(t)$ is not connected.
Suppose now that $\mathcal{A}_{\delta'}$ is not empty and that $\mathbf{y}$ satisfies $\|\mathbf{y} - \eta(\theta_2)\| \leq \sqrt{E_\eta(\delta)}$ and $\|\mathbf{y} - \eta(\theta_1)\| < \sqrt{t + \alpha} - \sqrt{t}$, which again happens with strictly positive probability. Since $\alpha < t$, $\sqrt{t+\alpha} - \sqrt{t} < \sqrt{t}$ and $\theta_1 \in \Theta_\eta(t)$. Also, $E_\eta(\delta) < t$ implies $\theta_2 \in \Theta_\eta(t)$. Any θ in $\mathcal{A}_{\delta'}$ satisfies $\|\eta(\theta) - \eta(\theta_1)\|^2 \geq E_\eta(\delta') > t + \alpha$; therefore,

$$
\begin{aligned}
\|\mathbf{y} - \eta(\theta)\| &\geq \|\eta(\theta) - \eta(\theta_1)\| - \|\mathbf{y} - \eta(\theta_1)\| \\
&> \sqrt{t+\alpha} - \|\eta(\theta_1) - \mathbf{y}\| > \sqrt{t}
\end{aligned}
$$

and $\theta \notin \Theta_\eta(t)$, which implies that $\Theta_\eta(t)$ is not connected. ■

Symbols and Notation

$\Rightarrow$	Convergence in general or weak convergence (of probability measures or distribution functions)
$\stackrel{\mathrm{d}}{\to}$	Convergence in distribution
$\stackrel{\mathrm{p}}{\to}$	Convergence in probability
$\stackrel{\mathrm{a.s.}}{\to}$	Almost sure convergence
$\stackrel{\theta}{\rightsquigarrow}$	Uniform convergence with respect to θ
$\sim$	Distributed
a, A	Scalars
$\mathcal{A}$, $\mathscr{A}$, $\mathbb{A}$	Sets
$\mathbf{a}$	Column vector
α	Scalar or column vector
$\mathbf{A}$	Matrix
$\mathbf{a}^\top$, $\mathbf{A}^\top$	Transposed of $\mathbf{a}$ and $\mathbf{A}$
$\mathbf{A}^-$	A generalized inverse of $\mathbf{A}$ (i.e., $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$)
$\mathbf{A}^+$	The Moore–Penrose g-inverse of $\mathbf{A}$ (i.e., $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$, $(\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+$ and $(\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A}$)
$\\|\mathbf{a}\\| = \\|\mathbf{a}\\|_2$	Euclidian norm of $\mathbf{a} = (a_1, \ldots, a_d)^\top \in \mathbb{R}^d$, $\\|\mathbf{a}\\| = \left(\sum_{i=1}^d a_i^2\right)^{1/2}$
$\\|\mathbf{a}\\|_1$	$\mathscr{L}_1$ norm of $\mathbf{a} = (a_1, \ldots, a_d)^\top \in \mathbb{R}^d$, $\\|\mathbf{a}\\|_1 = \sum_{i=1}^d \|a_i\|$
$\\|\mathbf{a}\\|_{\boldsymbol{\Omega}}$	$(\mathbf{a}^\top \boldsymbol{\Omega} \mathbf{a})^{1/2}$, for some $\boldsymbol{\Omega} \in \mathbb{M}^{\geq}$
$\\|\cdot\\|_\xi$	Norm in $\mathscr{L}_2(\xi)$, $\\|\phi\\|_\xi = \left[\int_{\mathscr{X}} \phi^2(x)\, \xi(\mathrm{d}x)\right]^{1/2}$, $\phi \in \mathscr{L}_2(\xi)$
$\langle \cdot, \cdot \rangle_\xi$	Inner product in $\mathscr{L}_2(\xi)$, $\langle \phi, \psi \rangle_\xi = \int_{\mathscr{X}} \phi(x)\, \psi(x)\, \xi(\mathrm{d}x)$, $\phi, \psi \in \mathscr{L}_2(\xi)$
$\stackrel{\xi}{\equiv}$	Parameter equivalence for the design ξ in a regression model, $\theta \stackrel{\xi}{\equiv} \theta^*$ when $\\|\eta(\cdot, \theta) - \eta(\cdot, \theta^*)\\|_\xi = 0$
$\{\mathbf{a}_i\}_j$	j-th component of $\mathbf{a}_i$

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

$\{\mathbf{A}\}_{ij}$	(i,j)-th entry of $\mathbf{A}$
$\{\mathbf{A}\}_{i,\cdot}$	i-th row of $\mathbf{A}$
$\mathbf{A} \succeq \mathbf{B}$	$\mathbf{A} - \mathbf{B} \in \mathbb{M}^{\geq}$ (is nonnegative definite), $\mathbf{A}, \mathbf{B} \in \mathbb{M}$
$\mathbf{A} \succ \mathbf{B}$	$\mathbf{A} - \mathbf{B} \in \mathbb{M}^{>}$ (is positive definite), $\mathbf{A}, \mathbf{B} \in \mathbb{M}$
$\nabla f(\cdot)$	Gradient vector of $f(\cdot)$, $\{\nabla f(\alpha)\}_i = \partial f(\theta)/\partial\theta_i\big\|_{\theta=\alpha}$
$\nabla^2 f(\cdot)$	Hessian matrix of $f(\cdot)$, $\left\{\nabla^2 f(\alpha)\right\}_{ij} = \partial^2 f(\theta)/\partial\theta_i\partial\theta_j\big\|_{\theta=\alpha}$
$\tilde{\nabla} f(\cdot)$	Subgradient of $f(\cdot)$
$\partial f(\cdot)$	Subdifferential of $f(\cdot)$
$x'(\cdot)$	Derivative of $x(\cdot)$
$x''(\cdot)$	Second-order derivative of $x(\cdot)$
$\mathbf{0}$	Null vector, $\{\mathbf{0}\}_i = 0$ for all i
$\mathbf{O}$	Null matrix, $\{\mathbf{O}\}_{i,j} = 0$ for all i, j
$\mathbf{1}$	Vector of ones, $\{\mathbf{1}\}_i = 1$ for all i
a.s.	Almost sure(ly)
$\mathscr{B}(\mathbf{c}, r)$	Closed ball $\{\mathbf{x} \in \mathbb{R}^d : \\|\mathbf{x} - \mathbf{c}\\| \leq r\}$
$C_{int}(\xi, \theta)$	Intrinsic curvature of a regression model at θ for the design measure ξ
$C_{int}(X, \theta)$	Intrinsic curvature of a regression model at θ for the exact design X
$C_{par}(\xi, \theta)$	Parametric curvature at θ for ξ
$C_{par}(X, \theta)$	Parametric curvature at θ for X
$C_{tot}(\xi, \theta)$	Total curvature at θ for ξ
diag($\mathbf{a}$)	Diagonal matrix with vector $\mathbf{a}$ on its diagonal
d.f.	(Cumulative) distribution function
$\mathbf{e}_i$	i-th basis vector
$\mathscr{E}_\phi(\cdot)$	Efficiency criterion associated with $\phi(\cdot)$, $\mathscr{E}_\phi(\xi) = \frac{\phi^+(\xi)}{\phi^+(\xi^*)}$ with ξ^* optimal for $\phi(\cdot)$
$\mathbb{E}(\cdot)$	Expectation
$\mathbb{E}_\mu(\cdot)$	Expectation for the probability measure μ
$\mathbb{E}_\pi(\cdot)$	Expectation for the p.d.f. $\pi(\cdot)$
$\mathbb{E}_x(\cdot)$	Conditional expectation for a given x, $\mathbb{E}_x(\omega) = \mathbb{E}(\omega\\|x)$
$\mathbb{F}(\cdot)$	Distribution function (d.f.)
$\mathbf{f}(\cdot)$	Regressor in a linear regression model, $\eta(x,\theta) = \mathbf{f}^\top(x)\theta$
$\mathbf{f}_\theta(\cdot)$	Derivative in a nonlinear regression model, $\mathbf{f}_\theta(x) = \partial\eta(x,\theta)/\partial\theta$
$F_\phi(\xi; \nu)$	Directional derivative of $\phi(\cdot)$ at ξ in the direction ν
$F_\phi(\xi, x)$	Directional derivative of $\phi(\cdot)$ at ξ in the direction δ_x
$\mathscr{F}_\theta$	Elfving's set, convex closure of the set $\{\mathbf{f}_\theta(x) : x \in \mathscr{X}\} \cup \{-\mathbf{f}_\theta(x) : x \in \mathscr{X}\}$
$\mathbf{I}_q$	q-dimensional identity matrix
i.i.d.	Independently and identically distributed
$\mathbb{I}_\mathcal{A}(\cdot)$	Indicator function of the set $\mathcal{A}$
int($\mathcal{A}$)	Interior of the set $\mathcal{A}$

$J_N(\cdot)$	Estimation criterion
$J_{\bar{\theta}}(\cdot)$	Limiting value of $J_N(\cdot)$ as $N \to \infty$ (under uniform convergence conditions)
ℓ	Number of elements in the finite design space $\mathscr{X}_\ell$
LP	Linear programming
LS	Least squares
$\mathrm{L}_{X,\mathbf{y}}(\theta)$	Likelihood of parameters θ for the design X at observations $\mathbf{y}$
$\mathscr{L}_2(\xi)$	Hilbert space of square-integrable real-valued functions ϕ, $\mathscr{L}_2(\xi) = \{\phi(\cdot) : \mathscr{X} \longrightarrow \mathbb{R}, \int_{\mathscr{X}} \phi^2(x)\, \xi(\mathrm{d}x) < \infty\}$
ML	Maximum likelihood
MSE	Mean-squared error
$\mathbf{M}_X(\theta)$	Information matrix for the (exact) design X
$\mathbf{M}(X,\theta)$	Normalized information matrix for the (exact) design X, $\mathbf{M}(X,\theta) = \mathbf{M}_X(\theta)/N$
$\mathbf{M}(\xi,\theta)$	Normalized information matrix for the design measure ξ
$\mathbf{M}(\xi)$	Normalized information matrix $\mathbf{M}(\xi,\theta^0)$ (local design)
$\mathbf{M}_\theta(x)$	Normalized information matrix $\mathbf{M}(\delta_x,\theta)$
$\mathbb{M}$	Set of symmetric $p \times p$ matrices
$\mathbb{M}^{\geq}$	Subset of $\mathbb{M}$ formed by nonnegative-definite matrices
$\mathbb{M}^{>}$	Subset of $\mathbb{M}$ formed by positive-definite matrices
$\mathcal{M}(\mathbf{M})$	Column space of the matrix $\mathbf{M}$, $\mathcal{M}(\mathbf{M}) = \{\mathbf{M}\mathbf{u} : \mathbf{u} \in \mathbb{R}^p\}$
$\mathcal{M}_\theta(\mathscr{X})$	$\{\mathbf{M}_\theta(x) : x \in \mathscr{X}\}$
$\mathcal{M}_\theta(\Xi)$	$\{\mathbf{M}(\xi,\theta) : \xi \in \Xi\}$
$\mathscr{M}$	Set of probability measures
N	Number of observations
$\mathcal{N}(\mathbf{M})$	Null space of the matrix $\mathbf{M}$, $\mathcal{N}(\mathbf{M}) = \{\mathbf{u} \in \mathbb{R}^p : \mathbf{M}\mathbf{u} = \mathbf{0}\}$
$\mathscr{N}(\mathbf{a},\mathbf{V})$	Normal distribution (mean $\mathbf{a}$, variance–covariance matrix $\mathbf{V}$)
$o_{\mathrm{p}}(\cdot)$	$\alpha_n = o_{\mathrm{p}}(\beta_n)$ if $\{\alpha_n/\beta_n\} \xrightarrow{\mathrm{p}} 0$, $n \to \infty$
$\mathcal{O}_{\mathrm{p}}(\cdot)$	$\alpha_n = \mathcal{O}_{\mathrm{p}}(\beta_n)$ if $\{\alpha_n/\beta_n\}$ is bounded in probability, $n \to \infty$
p	Dimension of the parameter vector θ
p.d.f.	Probability density function
P_θ, $\mathbf{P}_\theta$	Projectors
$\mathscr{P}_{\ell-1}$	Probability simplex $\{\mathbf{w} \in \mathbb{R}^\ell : w_i \geq 0\,,\ \sum_{i=1}^{\ell} w_i = 1\}$
QP	Quadratic programming
$\mathbb{R}^p$	p-dimensional Euclidian space of real column vectors
$\mathbf{R}(\theta)$	Riemannian curvature tensor
$s(x)$	Skewness of the p.d.f. $\bar{\varphi}_x(\cdot)$, $s(x) = \mathbb{E}_x\{\varepsilon^3(x)\}\sigma^{-3}(x)$
$\mathbb{S}_\eta$	Expectation surface, $\mathbb{S}_\eta = \{\eta(\theta) :\ \theta \in \Theta\}$
$\mathcal{S}_\xi$	Support of the design measure ξ
SLLN	Strong law of large numbers
TSLS	Two-stage least squares
$\mathrm{var}(\cdot)$	Variance
$\mathrm{var}_x(\cdot)$	Conditional variance for a given x

$\mathrm{Var}(\cdot)$	Variance–covariance matrix
w.p.1	With probability one
$\mathscr{W}_{\ell-1}$	$\{\underline{\mathbf{w}} \in \mathbb{R}^{\ell-1} : \underline{w}_i \geq 0\,,\ \sum_{i=1}^{\ell-1} \underline{w}_i \leq 1\}$
WLS	Weighted least squares
x_i	i-th design point, experimental variables for the i-th trial
$x^{(i)}$	i-th element in a finite design space $\mathscr{X}_\ell$
X	Exact design with fixed size N, $X = (x_1, \ldots, x_N)$
$\mathscr{X}$	Design space (in general a compact subset of $\mathbb{R}^d$)
$\mathscr{X}_\ell$	Finite design space with ℓ elements
$y(x)$	Observation (random variable) at x
$\mathbf{y}$	Vector of observations, $\mathbf{y} = [y(x_1), \ldots, y(x_N)]^\top$
δ_x	Delta measure with mass 1 at x
$\varepsilon_i = \varepsilon(x_i)$	Measurement error (with zero mean, $\mathbb{E}_x\{\varepsilon(x)\} = 0$)
$\eta(x, \theta)$	Mean (or expected) response at $x \in \mathscr{X}$ for parameters θ in a regression model
$\eta(\theta)$	Vector of responses, $\eta(\theta) = \eta_X(\theta) = [\eta(x_1, \theta), \ldots, \eta(x_N, \theta)]^\top$
θ	Vector of parameters $(\theta_1, \ldots, \theta_p)^\top \in \Theta \subset \mathbb{R}^p$
$\hat{\theta}^N$	Estimator of θ for N observations
$\bar{\theta}$	True value of θ
Θ	Parameter space, a subset of $\mathbb{R}^p$
$\overline{\Theta}$	Closure of Θ
$\partial\Theta$	Boundary of Θ
$\Theta^\#$	Set of global minimizers of $J_{\bar{\theta}}(\cdot)$
$\kappa(x)$	Kurtosis of the p.d.f. $\bar{\varphi}_x(\cdot)$, $\kappa(x) = \mathbb{E}_x\{\varepsilon^4(x)\}\sigma^{-4}(x) - 3$
$\lambda(x, \bar{\theta})$	Parameterized variance function $\mathbb{E}_x\{\varepsilon^2(x)\}$ in a (mixed) regression model
$\lambda_{\min}(\mathbf{A})$	Minimum eigenvalue of $\mathbf{A}$
$\lambda_{\max}(\mathbf{A})$	Maximum eigenvalue of $\mathbf{A}$
μ	Probability measure (e.g., prior measure for θ)
ξ	Design measure (a probability measure on $\mathscr{X}$)
ξ^*	Optimum design measure
Ξ	Set of design measures on $\mathscr{X}$
$\pi(\cdot)$	Prior p.d.f. for θ
$\pi_{X,\mathbf{y}}(\cdot)$	Posterior p.d.f. for θ given $\mathbf{y}$ for the design X
$\varpi_X(\cdot,\cdot)$	p.d.f. of the joint distribution of θ and $\mathbf{y}$ for the design X
$\sigma^2(x)$	Variance of the error $\varepsilon(x)$
$\bar{\varphi}(\cdot)$	p.d.f. of the errors ε (regression model with i.i.d. errors)
$\bar{\varphi}_x(\cdot)$	p.d.f. of the errors $\varepsilon(x)$ (regression model)
$\varphi_{x,\theta}(\cdot)$	p.d.f. of the observations $y(x)$ (e.g., exponential family with parameters θ)
$\varphi_{X,\theta}(\cdot)$	p.d.f. of $\mathbf{y}$ given θ for the design X
$\varphi_X^*(\cdot)$	p.d.f. of the marginal distribution of $\mathbf{y}$ for the design X
$\phi(\cdot)$	Design criterion, function of a design measure ξ
$\phi^+(\cdot)$	Positively homogenous form of $\phi(\cdot)$

ϕ^*	Optimum (i.e., maximum) value of $\phi(\xi)$, $\xi \in \Xi$
$\Phi(\cdot)$	Design criterion, function of an information matrix
$\Phi^+(\cdot)$	Positively homogenous form of $\Phi(\cdot)$

List of Labeled Assumptions

H$_\Theta$, page 22: Θ is a compact subset of $\mathbb{R}^p$ such that $\Theta \subset \overline{\text{int}(\Theta)}$.

H1$_\eta$, page 22: $\eta(x,\theta)$ is bounded on $\mathscr{X} \times \Theta$ and $\eta(x,\theta)$ is continuous on Θ, $\forall x \in \mathscr{X}$.

H2$_\eta$, page 22: $\bar{\theta} \in \text{int}(\Theta)$ and, $\forall x \in \mathscr{X}$, $\eta(x,\theta)$ is twice continuously differentiable with respect to $\theta \in \text{int}(\Theta)$, and its first two derivatives are bounded on $\mathscr{X} \times \text{int}(\Theta)$.

H1$_h$, page 36: The function $h(\cdot) : \Theta \longrightarrow \mathbb{R}$ is continuous and has continuous second-order derivatives in $\text{int}(\Theta)$.

H3$_\eta$, page 43: Let $\mathscr{S}_\epsilon$ denote the set $\left\{\theta \in \text{int}(\Theta) : \left\|\eta(\cdot,\theta) - \eta(\cdot,\bar{\theta})\right\|_\xi^2 < \epsilon\right\}$, then there exists $\epsilon > 0$ such that for every $\theta^{\#}$ and θ^* in $\mathscr{S}_\epsilon$ we have

$$\left[\frac{\partial}{\partial\theta}\left\|\eta(\cdot,\theta) - \eta(\cdot,\theta^{\#})\right\|_\xi^2\right]_{\theta=\theta^*} = \mathbf{0} \implies \theta^{\#} \overset{\xi}{\equiv} \theta^* .$$

H4$_\eta$, page 43: For any point $\theta^* \overset{\xi}{\equiv} \bar{\theta}$ there exists a neighborhood $\mathcal{V}(\theta^*)$ such that

$$\forall \theta \in \mathcal{V}(\theta^*) \, , \; \text{rank}[\mathbf{M}(\xi,\theta)] = \text{rank}[\mathbf{M}(\xi,\theta^*)] \, .$$

H2$_h$, page 43: The function $h(\cdot)$ is defined and has a continuous nonzero vector of derivatives $\partial h(\theta)/\partial\theta$ on $\text{int}(\Theta)$. Moreover, for any $\theta \overset{\xi}{\equiv} \bar{\theta}$, there exists a linear mapping A_θ from $\mathscr{L}_2(\xi)$ to $\mathbb{R}$ (a continuous linear functional on $\mathscr{L}_2(\xi)$), such that $A_\theta = A_{\bar{\theta}}$ and that

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

$$\frac{\partial h(\theta)}{\partial \theta_i} = A_\theta\left[\{\mathbf{f}_\theta\}_i\right],\ i = 1, \ldots, p,$$

where $\{\mathbf{f}_\theta\}_i$ is defined by (3.42).

H2$'_h$, page 44: There exists a function $\Psi(\cdot)$, with continuous gradient, such that $h(\theta) = \Psi[\eta(\theta)]$, with $\eta(\theta) = (\eta(x^{(1)}, \theta), \ldots, \eta(x^{(k)}, \theta))^\top$.

H2$''_h$, page 44: $h(\theta) = \Psi[h_1(\theta), \ldots, h_k(\theta)]$ with $\Psi(\cdot)$ a continuously differentiable function of k variables and with

$$h_i(\theta) = \int_{\mathscr{X}} g_i[\eta(x, \theta), x]\, \xi(\mathrm{d}x),\ i = 1, \ldots, k,$$

for some functions $g_i(t, x)$ differentiable with respect to t for any x in the support of ξ.

H3$_h$, page 47: The vector function $\mathbf{h}(\theta)$ has a continuous Jacobian $\partial \mathbf{h}(\theta)/\partial \theta^\top$ on $\mathrm{int}(\Theta)$. Moreover, for each $\theta \overset{\xi}{\equiv} \bar{\theta}$ there exists a continuous linear mapping B_θ from $\mathscr{L}_2(\xi)$ to $\mathbb{R}^q$ such that $B_\theta = B_{\bar{\theta}}$ and that

$$\frac{\partial \mathbf{h}(\theta)}{\partial \theta_i} = B_\theta\left[\{\mathbf{f}_\theta\}_i\right],\ i = 1, \ldots, p,$$

where $\{\mathbf{f}_\theta\}_i$ is given by (3.42).

H1$_\lambda$, page 48: $\lambda(x, \bar{\theta})$ is bounded and strictly positive on $\mathscr{X}$, $\lambda^{-1}(x, \theta)$ is bounded on $\mathscr{X} \times \Theta$, and $\lambda(x, \theta)$ is continuous on Θ for all $x \in \mathscr{X}$.

H2$_\lambda$, page 48: For all $x \in \mathscr{X}$, $\lambda(x, \theta)$ is twice continuously differentiable with respect to $\theta \in \mathrm{int}(\Theta)$, and its first two derivatives are bounded on $\mathscr{X} \times \mathrm{int}(\Theta)$.

H$_\mathcal{S}$, page 172: There exists $r > 0$ such that:

(a) $\mathrm{Prob}_{\bar{\theta}}[\mathcal{G}(r)] = \mathrm{Prob}(\|\mathbf{y} - \eta(\theta)\| < r) \geq 1 - \epsilon$.
(b) Every $\mathbf{y} \in \mathcal{T}(r)$ has one r-projection only.

H$_\mathscr{X}$-(i), page 273: $\inf_{\theta \in \Theta} \lambda_{\min}\left[\sum_{i=1}^{\ell} \mathbf{f}_\theta(x^{(i)}) \mathbf{f}_\theta^\top(x^{(i)})\right] > \gamma > 0$.

H$_\mathscr{X}$-(ii), page 273: For all $\delta > 0$ there exists $\epsilon(\delta) > 0$ such that for any subset $\{i_1, \ldots, i_p\}$ of distinct elements of $\{1, \ldots, \ell\}$, $\inf_{\|\theta - \bar{\theta}\| \geq \delta} \sum_{j=1}^{p} [\eta(x^{(i_j)}, \theta) - \eta(x^{(i_j)}, \bar{\theta})]^2 > \epsilon(\delta)$.

H$_\mathscr{X}$-(iii), page 273: $\lambda_{\min}\left[\sum_{j=1}^{p} \mathbf{f}_{\bar{\theta}}(x^{(i_j)}) \mathbf{f}_{\bar{\theta}}^\top(x^{(i_j)})\right] \geq \bar{\gamma} > 0$ for any subset $\{i_1, \ldots, i_p\}$ of distinct elements of $\{1, \ldots, \ell\}$.

References

Ahipasaoglu, S., P. Sun, and M. Todd (2008). Linear convergence of a modified Frank-Wolfe algorithm for computing minimum volume enclosing ellipsoids. *Optim. Methods Softw. 23*, 5–19.

Alexéev, V., E. Galéev, and V. Tikhomirov (1987). *Recueil de Problèmes d'Optimisation.* Moscou: MIR.

Andrews, D. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica 55*(6), 1465–1471.

Andrews, D. (1992). Generic uniform convergence. *Econometric Theory 8*, 241–257.

Atkinson, A. (2003). Transforming both sides and optimum experimental design for a nonlinear model arising from second-order chemical kinetics. In *Tatra Mountains Math. Pub.*, Volume 26, pp. 29–39.

Atkinson, A. (2004). Some Bayesian optimum designs for response transformation in nonlinear models with nonconstant variance. In A. Di Bucchianico, H. Läuter, and H. Wynn (Eds.), *mODa'7 – Advances in Model–Oriented Design and Analysis, Proc. 7th Int. Workshop, Heeze (Netherlands)*, Heidelberg, pp. 13–21. Physica Verlag.

Atkinson, A., K. Chaloner, A. Herzberg, and J. Juritz (1993). Optimal experimental designs for properties of a compartmental model. *Biometrics 49*, 325–337.

Atkinson, A. and R. Cook (1996). Designing for a response transformation parameter. *J. Roy. Statist. Soc. B59*, 111–124.

Atkinson, A. and D. Cox (1974). Planning experiments for discriminating between models (with discussion). *J. Roy. Statist. Soc. B36*, 321–348.

Atkinson, A. and A. Donev (1992). *Optimum Experimental Design.* Oxford Univ. Press.

Atkinson, A. and V. Fedorov (1975a). The design of experiments for discriminating between two rival models. *Biometrika 62*(1), 57–70.

Atkinson, A. and V. Fedorov (1975b). Optimal design: experiments for discriminating between several models. *Biometrika 62*(2), 289–303.

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

Atwood, C. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist. 40*(5), 1570–1602.

Atwood, C. (1973). Sequences converging to D-optimal designs of experiments. *Ann. Statist. 1*(2), 342–352.

Atwood, C. (1976). Convergent design sequences for sufficiently regular optimality criteria. *Ann. Statist. 4*(6), 1124–1138.

Atwood, C. (1980). Convergent design sequences for sufficiently regular optimality criteria, II singular case. *Ann. Statist. 8*(4), 894–912.

Avriel, M. (2003). *Nonlinear Programming. Analysis and Methods.* New York: Dover. [Originally published by Prentice Hall, 1976].

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Chichester: Wiley.

Bartlett, P. (2003). Prediction algorithms: complexity, concentration and convexity. In *Prep. 13th IFAC Symp. on System Identification, Rotterdam*, pp. 1507–1517.

Bates, D. and D. Watts (1980). Relative curvature measures of nonlinearity. *J. Roy. Statist. Soc. B42*, 1–25.

Bates, D. and D. Watts (Eds.) (1988). *Nonlinear regression Analysis and its Applications.* New York: Wiley.

Ben-Tal, A. and A. Nemirovskii (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* Philadelphia: MPS/SIAM Series on Optim. 2.

Beran, R. (1974). Asymptotically efficient rank estimates in location models. *Ann. Statist. 2*, 63–74.

Berlinet, A. and C. Thomas-Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics.* Boston: Kluwer.

Bickel, P. (1982). On adaptive estimation. *Ann. Statist. 10*, 647–671.

Bickel, P., C. Klassen, Y. Ritov, and J. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins Univ. Press.

Biedermann, S. and H. Dette (2001). Minimax optimal designs for nonparametric regression — a further optimality property of the uniform distribution. In A. Atkinson, P. Hackl, and W. Müller (Eds.), *mODa'6 – Advances in Model–Oriented Design and Analysis, Proc. 6th Int. Workshop, Puchberg/Schneberg (Austria)*, Heidelberg, pp. 13–20. Physica Verlag.

Bierens, H. (1994). *Topics in Advanced Econometrics.* Cambridge: Cambridge Univ. Press.

Bilardello, P., X. Joulia, J.-M. Le Lann, H. Delmas, and B. Koehret (1993). A general strategy for parameter estimation in differential-algebraic systems. *Comput. Chem. Eng. 17*, 517–525.

Billingsley, P. (1995). *Probability and Measure.* New York: Wiley. [3rd ed.].

Birgin, E., J. Martínez, and M. Raydan (2000). Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim. 10*(4), 1196–1211.

Bland, R., D. Goldfarb, and M. Todd (1981). The ellipsoid method: a survey. *Oper. Res. 29*(6), 1039–1091.

Bloomfield, P. and G. Watson (1975). The inefficiency of least squares. *Biometrika 62*(1), 121–128.

Bohachevsky, I., M. Johnson, and M. Stein (1986). Generalized simulated annealing for function optimization. *Technometrics 28*(3), 209–217.

Böhning, D. (1985). Numerical estimation of a probability measure. *J. Stat. Plann. Inference 11*, 57–69.

Böhning, D. (1986). A vertex-exchange-method in D-optimal design theory. *Metrika 33*, 337–347.

Boissonnat, J.-D. and M. Yvinec (1998). *Algorithmic Geometry.* Cambridge Univ. Press.

Bonnans, J., J. Gilbert, C. Lemaréchal, and C. Sagastizábal (2006). *Numerical Optimization. Theoretical and Practical Aspects.* Heidelberg: Springer. [2nd ed.].

Bornkamp, B. (2011). Functional uniform priors for nonlinear modelling. Technical Report arXiv:1110.4400v1.

Box, G. and D. Cox (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. B26*, 211–252.

Box, G. and H. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika 46*, 77–90.

Box, M. (1971). Bias in nonlinear estimation. *J. Roy. Statist. Soc. B33*, 171–201.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization.* Cambridge: Cambridge Univ. Press.

Braess, D. and H. Dette (2007). On the number of support points of maximin and Bayesian optimal designs. *Ann. Statist. 35*(2), 772–792.

Caines, P. (1988). *Linear Stochastic Systems.* New York: Wiley.

Calamai, P. and J. Moré (1987). Projected gradient methods for linearly constrained problems. *Math. Programming 39*, 93–116.

Carroll, R. and D. Ruppert (1982). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *J. Amer. Statist. Assoc. 77*(380), 878–882.

Casella, G. (2008). *Statistical Design.* New York: Springer.

Chaloner, K. and K. Larntz (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Stat. Plann. Inference 21*, 191–208.

Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: a review. *Statist. Sci. 10*(3), 273–304.

Chaudhuri, P. and P. Mykland (1993). Nonlinear experiments: optimal design and inference based likelihood. *J. Amer. Statist. Assoc. 88*(422), 538–546.

Chaudhuri, P. and P. Mykland (1995). On efficiently designing of nonlinear experiments. *Statistica Sinica 5*, 421–440.

Chavent, G. (1983). Local stability of the output least square parameter estimation technique. *Matematicada Applicada e Computacional 2*(1), 3–22.

Chavent, G. (1987). Identifiability of parameters in the output least square formulation. In E. Walter (Ed.), *Identifiability of Parametric Models*, Chapter 6, pp. 67–74. Oxford: Pergamon.

Chavent, G. (1990). A new sufficient condition for the wellposedness of non-linear least-square problems arising in identification and control. In A. Bensoussan and J. Lions (Eds.), *Analysis and Optimization of Systems,* Lecture Notes in Control and Inform. Sci., vol. 144, pp. 452–463. Springer.

Chavent, G. (1991). New size × curvature conditions for strict quasiconvexity of sets. *SIAM J. Control Optim. 29*(6), 1348–1372.

Cheng, M.-Y., P. Hall, and D. Titterington (1998). Optimal design for curve estimation by local linear smoothing. *Bernoulli 4*(1), 3–14.

Clarke, B. and A. Barron (1994). Jeffrey's prior is asymptotically less favorable under entropy risk. *J. Stat. Plann. Inference 41*, 37–60.

Clarke, G. (1980). Moments of the least-squares estimators in a non-linear regression model. *J. Roy. Statist. Soc. B42*, 227–237.

Clyde, M. and K. Chaloner (2002). Constrained design strategies for improving normal approximations in nonlinear regression problems. *J. Stat. Plann. Inference 104*, 175–196.

Cochran, W. and G. Cox (1957). *Experimental Designs.* New York: Wiley. [2nd ed., 2005].

Cook, D. and V. Fedorov (1995). Constrained optimization of experimental design (invited discussion paper). *Statistics 26*, 129–178.

Cook, D. and W. Wong (1994). On the equivalence between constrained and compound optimal designs. *J. Amer. Statist. Assoc. 89*(426), 687–692.

Cook, R., D. Hawkins, and S. Weisberg (1993). Exact iterative computation of the robust multivariate minimum volume ellipsoid estimator. *Statist. Probab. Lett. 16*, 213–218.

Cook, R. and C. Nachtsheim (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics 22*(3), 315–324.

Cormen, T., C. Leiserson, R. Rivest, and C. Stein (2001). *Introduction to Algorithms.* MIT Press and McGraw-Hill. [2nd ed.].

Correa, R. and C. Lemaréchal (1993). Convergence of some algorithms for convex minimization. *Math. Programming 62*, 261–275.

Cox, D. and D. Hinkley (1974). *Theoretical Statistics.* London: Chapman & Hall.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton, N.J.: Princeton University Press.

Cucker, F. and S. Smale (2001). On the mathematical foundations of learning. *Bull. AMS 39*(1), 1–49.

Dai, Y.-H. and R. Fletcher (2006). New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Math. Programming A 106*, 403–421.

D'Argenio, D. (1990). Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Math. Biosci. 99*, 105–118.

Davies, L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist. 20*(4), 1828–1843.

de la Garza, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Statist. 25*, 123–130.

del Pino, G. (1989). The unifying role of iterative generalized least squares in statistical algorithms (with discussion). *Statist. Sci. 4*(4), 394–408.

Dembski, W. (1990). Uniform probability. *J. Theoret. Probab. 3*(4), 611–626.

Demidenko, E. (1989). *Optimization and Regression.* Moscow: Nauka. [In Russian].

Demidenko, E. (2000). Is this the least squares estimate? *Biometrika 87*(2), 437–452.

Dem'yanov, V. and V. Malozemov (1974). *Introduction to Minimax.* New York: Dover.

den Boeff, E. and D. den Hertog (2007). Efficient line search methods for convex functions. *SIAM J. Optim. 18*(1), 338–363.

den Hertog, D. (1994). *Interior Point Approach to Linear, Quadratic and Convex Programming.* Dordrecht: Kluwer.

Dette, H. (1993). Elfving's theorem for D-optimality. *Ann. Statist. 21*(2), 753–766.

Dette, H. and V. Melas (2011). A note on de la Garza phenomenon for locally optimal designs. *Ann. Statist. 39*(2), 1266–1281.

Dette, H., A. Pepelyshev, and A. Zhigljavsky (2008). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *J. Stat. Plann. Inference 53*, 312–320.

Dette, H. and W. Studden (1993). Geometry of E-optimality. *Ann. Statist. 21*(1), 416–433.

Downing, D., V. Fedorov, and S. Leonov (2001). Extracting information from the variance function: optimal design. In A. Atkinson, P. Hackl, and W. Müller (Eds.), *mODa'6 – Advances in Model–Oriented Design and Analysis, Proc. 6th Int. Workshop, Puchberg/Schneberg (Austria)*, pp. 45–52. Heidelberg: Physica Verlag.

Dragalin, V. and V. Fedorov (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *J. Stat. Plann. Inference 136*, 1800–1823.

Eaton, M., A. Giovagnoli, and P. Sebastiani (1996). A predictive approach to the Bayesian design problem with application to normal regression models. *Biometrika 83*(1), 111–125.

Ecker, J. and M. Kupferschmid (1983). An ellipsoid algorithm for nonlinear programming. *Math. Programming 27*, 83–106.

Ecker, J. and M. Kupferschmid (1985). A computational comparison of the ellipsoidal algorithm with several nonlinear programming algorithms. *SIAM J. Control Optim. 23*(5), 657–674.

Efron, B. and D. Hinkley (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika 65*(3), 457–487.

Eisenhart, L. (1960). *Riemannian Geometry.* Princeton, N.J.: Princeton Univ. Press.

Elfving, G. (1952). Optimum allocation in linear regression. *Ann. Math. Statist. 23*, 255–262.

Ermoliev, Y. and R.-B. Wets (Eds.) (1988). *Numerical Techniques for Stochastic Optimization Problems.* Berlin: Springer.

Fedorov, V. (1971). The design of experiments in the multiresponse case. *Theory Probab. Appl. 16*(2), 323–332.

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

Fedorov, V. (1980). Convex design theory. *Math. Operationsforsch. Statist. Ser. Statist. 11*(3), 403–413.

Fedorov, V. (1989). Optimal design with bounded density: optimization algorithms of the exchange type. *J. Stat. Plann. Inference 22*, 1–13.

Fedorov, V. and P. Hackl (1997). *Model-Oriented Design of Experiments.* Berlin: Springer.

Fedorov, V. and A. Pázman (1968). Design of physical experiments. *Fortschritte der Physik 16*, 325–355.

Fellman, J. (1974). On the allocation of linear observations. *Comment. Phys. Math. 44*, 27–78.

Fellman, J. (1989). An empirical study of a class of iterative searches for optimal designs. *J. Stat. Plann. Inference 21*, 85–92.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

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

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

Fourgeaud, C. and A. Fuchs (1967). *Statistique.* Paris: Dunod.

Frank, M. and P. Wolfe (1956). An algorithm for quadratic programming. *Naval Res. Logist. Quart. 3*, 95–110.

Galil, Z. and J. Kiefer (1977). Comparison of simplex designs for quadratic mixture models. *Technometrics 19*(4), 445–453.

Galil, Z. and J. Kiefer (1980). Time- and space-saving computer methods, related to Mitchell's DETMAX, for finding D-optimum designs. *Technometrics 22*(3), 301–313.

Gallant, A. (1987). *Nonlinear Statistical Models.* New York: Wiley.

Gauchi, J.-P. and A. Pázman (2006). Designs in nonlinear regression by stochastic minimization of functionals of the mean square error matrix. *J. Stat. Plann. Inference 136*, 1135–1152.

Gautier, R. and L. Pronzato (1998). Sequential design and active control. In N. Flournoy, W. Rosenberger, and W. Wong (Eds.), *New Developments and Applications in Experimental Design*, Lecture Notes — Monograph Series, vol. 34, pp. 138–151. Hayward: IMS.

Gautier, R. and L. Pronzato (1999). Some results on optimal allocation in two–step sequential design. *Biometrical Lett. 36*(1), 15–30.

Gautier, R. and L. Pronzato (2000). Adaptive control for sequential design. *Discuss. Math. Probab. Stat. 20*(1), 97–114.

Gilmour, S. and L. Trinca (2012). Optimum design of experiments for statistical inference (with discussion). *J. Roy. Statist. Soc. C61*(3), 1–25.

Goffin, J.-L. and K. Kiwiel (1999). Convergence of a simple subgradient level method. *Math. Programming 85*, 207–211.

Goodwin, G. and R. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic Press.

Gorman, J. and A. Hero (1990). Lower bounds for parametric estimation with constraints. *IEEE Trans. Information Theory 26*, 1285–1301.

Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). *J. Roy. Statist. Soc. B-46*(2), 149–192.

Grötschel, M., L. Lovász, and A. Schrijver (1980). *Geometric Algorithms and Combinatorial Optimization*. Berlin: Springer.

Hadjihassan, S., L. Pronzato, E. Walter, and I. Vuchkov (1997). Robust design for quality improvement by ellipsoidal bounding. In G. Yin and Q. Zhang (Eds.), *Proc. 1996 AMS–SIAM Summer Seminar, Math. of Stochastic Manufacturing Systems*, Lectures in Applied Math., vol. 33, Providence, Rhode Island, pp. 127–138. American Math. Soc.

Halperin, M. (1963). Confidence interval estimation in non-linear regression. *J. Roy. Statist. Soc. B25*, 330–333.

Hamilton, D. and D. Watts (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics 27*, 241–250.

Hamilton, D., D. Watts, and D. Bates (1982). Accounting for intrinsic nonlinearities in nonlinear regression parameter inference regions. *Ann. Statist. 10*, 386–393.

Harman, R. (2004a). Lower bounds on efficiency ratios based on ϕ_p-optimal designs. In A. Di Bucchianico, H. Läuter, and H. Wynn (Eds.), *mODa'7 – Advances in Model–Oriented Design and Analysis, Proc. 7th Int. Workshop, Heeze (Netherlands)*, Heidelberg, pp. 89–96. Physica Verlag.

Harman, R. (2004b). Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika 60*, 137–153.

Harman, R. and T. Jurík (2008). Computing c-optimal experimental designs using the simplex method of linear programming. *Comput. Statist. Data Anal. 53*, 247–254.

Harman, R. and L. Pronzato (2007). Improvements on removing non-optimal support points in D-optimum design algorithms. *Statist. Probab. Lett. 77*, 90–94.

Harman, R. and M. Trnovská (2009). Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices. *Math. Slovaca 59*(5), 693–704.

Hartley, H. (1964). Exact confidence regions for parameters in non-linear regression laws. *Biometrika 51*(3&4), 347–353.

Harville, D. (1997). *Matrix Algebra from a Statistician's Perspective.* Heidelberg: Springer.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning. Data Mining, Inference and Prediction.* Heidelberg: Springer.

Hearn, D. and S. Lawphongpanich (1989). Lagrangian dual ascent by generalized linear programming. *Oper. Research Lett. 8*, 189–196.

Heyde, C. (1997). *Quasi-likelihood and its Application. A General Approach to Optimal Parameter Estimation.* New York: Springer.

Hill, P. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics 20*, 15–21.

Hiriart-Urruty, J. and C. Lemaréchal (1993). *Convex Analysis and Minimization Algorithms, part 1 and 2.* Berlin: Springer.

Hougaard, P. (1985). Saddlepoint approximations for curved exponential families. *Statist. Probab. Lett. 3*, 161–166.

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

Huber, P. (1981). *Robust Statistics.* New York: John Wiley.

Ibragimov, I. and R. Has'minskii (1981). *Statistical Estimation. Asymptotic Theory.* Heidelberg: Springer.

Ivanov, A. (1997). *Asymptotic Theory of Nonlinear Regression.* Dordrecht: Kluwer.

Jennrich, R. (1969). Asymptotic properties of nonlinear least squares estimation. *Ann. Math. Statist. 40*, 633–643.

Jobson, J. and W. Fuller (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Amer. Statist. Assoc. 75*(369), 176–181.

Johnson, M. and C. Nachtsheim (1983). Some guidelines for constructing exact D-optimal designs on convex design spaces. *Technometrics 25*, 271–277.

Jørgensen, B. (1997). Exponential dispersion models. *J. Roy. Statist. Soc. B49*, 127–167.

Karlin, S. and W. Studden (1966). Optimal experimental designs. *Ann. Math. Statist. 37*, 783–815.

Kelley, J. (1960). The cutting plane method for solving convex programs. *SIAM J. 8*, 703–712.

Khachiyan, L. (1979). A polynomial algorithm in linear programming. *Doklady Akademïa Nauk SSSR 244*, 1093–1096. [English transl. *Soviet Math. Doklady*, 20, 191–194].

Khachiyan, L. (1996). Rounding of polytopes in the real number model of computation. *Math. Oper. Res. 21*(2), 307–320.

Khachiyan, L. and M. Todd (1993). On the complexity of approximating the maximal inscribed ellipsoid for a polytope. *Math. Programming A61*(2), 137–159.

Kiefer, J. (1962). Two more criteria equivalent to D-optimality of designs. *Ann. Math. Statist. 33*(2), 792–796.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist. 2*(5), 849–879.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canad. J. Math. 12*, 363–366.

Kieffer, M. and E. Walter (1998). Interval analysis for guaranteed nonlinear parameter estimation. In A. Atkinson, L. Pronzato, and H. Wynn (Eds.), *MODA5 – Advances in Model-Oriented Data Analysis and Experimental Design, Marseilles (France), June 1998*, pp. 115–125. Heidelberg: Physica Verlag.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *Ann. Statist. 18*(1), 191–219.

Koopmans, T. and O. Reiersøl (1950). The identification of structural characteristics. *Ann. Math. Statist. 21*(2), 163–181.

Kushner, H. and D. Clark (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Heidelberg: Springer.

Kushner, H. and J. Yang (1993). Stochastic approximation with averaging of the iterates: optimal asymptotic rate of convergence for general processes. *SIAM J. Control Optim. 31*(4), 1045–1062.

Kushner, H. and G. Yin (1997). *Stochastic Approximation Algorithms and Applications.* Heidelberg: Springer.

Lai, T. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist. 22*(4), 1917–1930.

Lai, T., H. Robbins, and C. Wei (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. USA 75*(7), 3034–3036.

Lai, T., H. Robbins, and C. Wei (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal. 9*, 343–361.

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

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Pub. in Stat. 1*, 277–230.

Le Cam, L. (1960). Local asymptotically normal families of distributions. *Univ. California Pub. in Stat. 3*, 37–98.

Lehmann, E. and G. Casella (1998). *Theory of Point Estimation.* Heidelberg: Springer.

Lemaréchal, C., A. Nemirovskii, and Y. Nesterov (1995). New variants of bundle methods. *Math. Programming 69*(1), 111–147.

Levin, A. (1965). On an algorithm for the minimization of convex functions. *Soviet Math. Dokl. 6*, 286–290.

Li, X.-S. and S.-C. Fang (1997). On the entropic regularization method for solving min-max problems with applications. *Math. Methods Oper. Res. 46*, 119–130.

Liang, K.-Y. and S. Zeger (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statist. Sci. 10*(2), 158–173.

Lindley, D. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist. 27*, 986–1005.

Lindsay, B. and B. Li (1997). On the second-order optimality of the observed Fisher information. *Ann. Statist. 25*(5), 2172–2199.

Liu, S. and H. Neudecker (1997). Kantorovich inequalities and efficiency comparisons for several classes of estimators in linear models. *Statistica Neerlandica 51*(3), 345–355.

Ljung, L. (1987). *System Identification, Theory for the User.* Englewood Cliffs: Prentice Hall.

López-Fidalgo, J. and J. Rodríguez-Díaz (1998). Characteristic polynomial criteria in optimal experimental design. In A. Atkinson, L. Pronzato, and H. Wynn (Eds.), *Advances in Model–Oriented Data Analysis and Experimental Design, Proc. MODA'5, Marseilles, June 22–26, 1998*, pp. 31–38. Heidelberg: Physica Verlag.

López-Fidalgo, J., B. Torsney, and R. Ardanuy (1998). MV-optimisation in weighted linear regression. In A. Atkinson, L. Pronzato, and H. Wynn (Eds.), *Advances in Model–Oriented Data Analysis and Experimental Design, Proc. MODA'5, Marseilles, June 22–26, 1998*, pp. 39–50. Heidelberg: Physica Verlag.

Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus, with Applications in Statistics and Econometrics.* New York: Wiley.

Manski, C. (1984). Adaptive estimation of nonlinear regression models. *Econometric Rev. 3*(2), 145–194.

McCormick, G. and R. Tapia (1972). The gradient projection method under mild differentiability conditions. *SIAM J. Control 10*(1), 93–98.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models.* London: Chapman & Hall. [2nd ed.].

Mikulecká, J. (1983). On a hybrid experimental design. *Kybernetika 19*(1), 1–14.

Minoux, M. (1983). *Programmation Mathématique, Théorie et Algorithmes, vol. 1 & 2.* Paris: Dunod.

Mitchell, T. (1974). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics 16*, 203–210.

Molchanov, I. and S. Zuyev (2001). Variational calculus in the space of measures and optimal design. In A. Atkinson, B. Bogacka, and A. Zhigljavsky (Eds.), *Optimum Design 2000*, Chapter 8, pp. 79–90. Dordrecht: Kluwer.

Molchanov, I. and S. Zuyev (2002). Steepest descent algorithm in a space of measures. *Stat. Comput. 12*, 115–123.

Montgomery, D. (1976). *Design and Analysis of Experiments.* New York: Wiley. [6th ed., 2005].

Morris, M. and T. Mitchell (1995). Exploratory designs for computational experiments. *J. Stat. Plann. Inference 43*, 381–402.

Müller, C. and A. Pázman (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika 48*, 1–19.

Müller, H.-G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Lett. 2*, 285–290.

Müller, W. (2007). *Collecting Spatial Data. Optimum Design of Experiments for Random Fields.* Heidelberg: Physica-Verlag. [3rd ed.].

Müller, W. and A. Pázman (2003). Measures for designs in experiments with correlated errors. *Biometrika 90*(2), 423–434.

Müller, W. and B. Pötscher (1992). Batch sequential design for a nonlinear estimation problem. In V. Fedorov, W. Müller, and I. Vuchkov (Eds.), *Model-Oriented Data Analysis II, Proc. 2nd IIASA Workshop, St Kyrik (Bulgaria), May 1990*, pp. 77–87. Heidelberg: Physica Verlag.

Nadaraya, E. (1964). On estimating regression. *Theory Probab. Appl. 9*, 141–142.

Nesterov, Y. (1995). Complexity estimates of some cutting plane methods based on the analytic center. *Math. Programming 69*, 149–176.

Nesterov, Y. (2004). *Introductory Lectures to Convex Optimization: A Basic Course.* Dordrecht: Kluwer.

Nesterov, Y. and A. Nemirovskii (1994). *Interior-Point Polynomial Algorithms in Convex Programming.* Philadelphia: SIAM.

Newey, W. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica 9*(4), 1161–1167.

Pan, S., S. He, and X. Li (2007). Smoothing method for minimizing the sum of the r largest functions. *Optim. Methods Softw. 22*, 267–277.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist. 35*, 1065–1076.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc. 74*(365), 105–121.

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

Pázman, A. (1984a). Nonlinear least squares – uniqueness versus ambiguity. *Math. Operationsforsch. Statist. Ser. Statist. 15*, 323–336.

Pázman, A. (1984b). Probability distribution of the multivariate nonlinear least-squares estimates. *Kybernetika 20*, 209–230.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Dordrecht (co-pub. VEDA, Bratislava): Reidel (Kluwer group).

Pázman, A. (1990). Small-sample distributional properties of nonlinear regression estimators. a geometric approach (with discussion). *Statistics 21*(3), 323–367.

Pázman, A. (1992a). A classification of nonlinear regression models and parameter confidence regions. *Kybernetika 28*(6), 444–453.

Pázman, A. (1992b). Geometry of the nonlinear regression with prior. *Acta Math. Univ. Comenianae LXI*, 263–276.

Pázman, A. (1993a). Higher dimensional nonlinear regression — a statistical use of the Riemannian curvature tensor. *Statistics 25*, 17–28.

Pázman, A. (1993b). *Nonlinear Statistical Models.* Dordrecht: Kluwer.

Pázman, A. (2001). Concentration sets, Elfving sets and norms in optimum design. In A. Atkinson, B. Bogacka, and A. Zhigljavsky (Eds.), *Optimum Design 2000*, Chapter 10, pp. 101–112. Dordrecht: Kluwer.

Pázman, A. (2002a). Optimal design of nonlinear experiments with parameter constraints. *Metrika 56*, 113–130.

Pázman, A. (2002b). Results on nonlinear least squares estimators under nonlinear equality constraints. *J. Stat. Plann. Inference 103*, 401–420.

Pázman, A., S. Bilenkaya, J. Bystrický, Z. Janout, and F. Lehar (1969). The problem of un-ambiguity of phase-shift analysis I: Description of a powerful discriminating criterion. *Czech J. Physics B19*, 882–890.

Pázman, A. and W. Müller (2001). Optimum design of experiments subject to correlated errors. *Statist. Probab. Lett. 52*, 29–34.

Pázman, A. and W. Müller (2010). A note on the relationship between two approaches to optimal design under correlation. In A. Giovagnoli, A. Atkinson, and B. Torsney (Eds.), *mODa9 — Advances in Model-Oriented Design and Analysis, Proc. 9th Int. Workshop in Model-Oriented Design and Analysis, Bertinoro, Italy, June 14–18, 2010*, pp. 145–148. Heidelberg: Physica Verlag.

Pázman, A. and L. Pronzato (1992). Nonlinear experimental design based on the distribution of estimators. *J. Stat. Plann. Inference 33*, 385–402.

Pázman, A. and L. Pronzato (1996). A Dirac function method for densities of nonlinear statistics and for marginal densities in nonlinear regression. *Statist. Probab. Lett. 26*, 159–167.

Pázman, A. and L. Pronzato (1998). Approximate densities of two bias–corrected nonlinear LS estimators. In A. Atkinson, L. Pronzato, and H. Wynn (Eds.), *MODA'5 – Advances in Model–Oriented Data Analysis and Experimental Design, Proc. 5th Int. Workshop, Marseille, 22–26 juin 1998*, pp. 145–152. Heidelberg: Physica Verlag.

Pázman, A. and L. Pronzato (2004). Simultaneous choice of design and estimator in nonlinear regression with parameterized variance. In A. Di Bucchianico, H. Läuter, and H. Wynn (Eds.), *mODa'7 – Advances in Model–Oriented Design and Analysis, Proc. 7th Int. Workshop, Heeze (Netherlands)*, Heidelberg, pp. 117–124. Physica Verlag.

Pázman, A. and L. Pronzato (2006a). Asymptotic criteria for designs in nonlinear regression with model errors. *Math. Slovaca 56*(5), 543–553.

Pázman, A. and L. Pronzato (2006b). On the irregular behavior of LS estimators for asymptotically singular designs. *Statist. Probab. Lett. 76*, 1089–1096.

Pázman, A. and L. Pronzato (2009). Asymptotic normality of nonlinear least squares under singular experimental designs. In L. Pronzato and A. Zhigljavsky (Eds.), *Optimal Design and Related Areas in Optimization and Statistics*, Chapter 8, pp. 167–191. Springer.

Pečarić, J., S. Puntanen, and G. Styan (1996). Some further matrix extensions of the Cauchy-Schwarz and Kantorovich inequalities, with some statistical applications. *Linear Algebra Appl. 237/238*, 455–476.

Phillips, R. (2002). Least absolute deviations estimation via the EM algorithm. *Stat. Comput. 12*, 281–285.

Pilz, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models*, Volume 55. Leipzig: Teubner-Texte zur Mathematik. [Also published by Wiley, New York, 1991].

Polak, E. (1971). *Computational Methods in Optimization, a Unified Approach.* New York: Academic Press.

Polak, E. (1987). *Optimization. Algorithms and Consistent Approximations.* New York: Springer.

Polyak, B. (1987). *Introduction to Optimization.* New York: Optimization Software.

Polyak, B. (1990). New stochastic approximation type procedures. *Automat. i Telemekh. 7*, 98–107.

Polyak, B. and A. Juditsky (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim. 30*, 838–855.

Pronzato, L. (2004). A minimax equivalence theorem for optimum bounded design measures. *Statist. Probab. Lett. 68*, 325–331.

Pronzato, L. (2006). On the sequential construction of optimum bounded designs. *J. Stat. Plann. Inference 136*, 2783–2804.

Pronzato, L. (2008). Optimal experimental design and some related control problems. *Automatica 44*, 303–325.

Pronzato, L. (2009a). Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statist. Probab. Lett. 79*, 2307–2313.

Pronzato, L. (2009b). On the regularization of singular c-optimal designs. *Math. Slovaca 59*(5), 611–626.

Pronzato, L. (2010a). One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. *Metrika 71*(2), 219–238.

Pronzato, L. (2010b). Penalized optimal designs for dose-finding. *J. Stat. Plann. Inference 140*, 283–296.

Pronzato, L., C.-Y. Huang, and E. Walter (1991). Nonsequential T-optimal design for model discrimination: new algorithms. In A. Pázman and J. Volaufová (Eds.), *Proc. ProbaStat'91*, Bratislava, pp. 130–136.

Pronzato, L. and W. Müller (2012). Design of computer experiments: space filling and beyond. *Stat. & Comput. 22*(3), 681–701.

Pronzato, L. and A. Pázman (1994a). Bias correction in nonlinear regression via two-stages least-squares estimation. In M. Blanke and T. Söderström (Eds.), *Prep. 10th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, Volume 1, Copenhagen, pp. 137–142. Danish Automation Society.

Pronzato, L. and A. Pázman (1994b). Second-order approximation of the entropy in nonlinear least-squares estimation. *Kybernetika 30*(2), 187–198. [*Erratum* 32(1):104, 1996].

Pronzato, L. and A. Pázman (2001). Using densities of estimators to compare pharmacokinetic experiments. *Comput. Biol. Med. 31*(3), 179–195.

Pronzato, L. and A. Pázman (2004). Recursively re-weighted least-squares estimation in regression models with parameterized variance. In *Proc. EUSIPCO'2004, Vienna, Austria*, pp. 621–624.

Pronzato, L. and E. Walter (1985). Robust experiment design via stochastic approximation. *Math. Biosci. 75*, 103–120.

Pronzato, L. and E. Walter (1988). Robust experiment design via maximin optimization. *Math. Biosci. 89*, 161–176.

Pronzato, L. and E. Walter (1994). Minimal-volume ellipsoids. *Internat. J. Adapt. Control Signal Process. 8*, 15–30.

Pronzato, L. and E. Walter (1996). Volume-optimal inner and outer ellipsoids. In M. Milanese, J.-P. Norton, H. Piet-Lahanier, and E. Walter (Eds.), *Bounding Approaches to System Identification*, pp. 119–138. London: Plenum.

Pronzato, L. and E. Walter (2001). Eliminating suboptimal local minimizers in nonlinear parameter estimation. *Technometrics 43*(4), 434–442.

Pronzato, L., H. Wynn, and A. Zhigljavsky (2000). *Dynamical Search.* Boca Raton: Chapman & Hall/CRC.

Pronzato, L., H. Wynn, and A. Zhigljavsky (2005). Kantorovich-type inequalities for operators via D-optimal design theory. *Linear Algebra Appl. 410*, 160–169.

Pukelsheim, F. (1993). *Optimal Experimental Design.* New York: Wiley.

Pukelsheim, F. and S. Reider (1992). Efficient rounding of approximate designs. *Biometrika 79*(4), 763–770.

Rabitz, H., M. Kramer, and D. Dacol (1983). Sensitivity analysis in chemical kinetics. *Ann. Rev. Phys. Chem. 34*, 414–461.

Raynaud, H.-F., L. Pronzato, and E. Walter (2000). Robust identification and control based on ellipsoidal parametric uncertainty descriptions. *Eur. J. Control 6*(3), 245–257.

Rényi, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, pp. 547–561.

Robertazzi, T. and S. Schwartz (1989). An accelerated sequential algorithm for producing D-optimal designs. *SIAM J. Sci. Statist. Comput. 10*(2), 341–358.

Rockafellar, R. (1970). *Convex Analysis.* Princeton: Princeton Univ. Press.

Rojas, C., J. Welsh, G. Goodwin, and A. Feuer (2007). Robust optimal experiment design for system identification. *Automatica 43*, 993–1008.

Ross, G. (1990). *Nonlinear estimation.* New York: Springer.

Rousseeuw, P. (1984). Least median of squares regression. *J. Amer. Statist. Assoc. 79*, 871–880.

Rousseeuw, P. and A. Leroy (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

Rudin, W. (1987). *Real and Complex Analysis.* New York: McGraw-Hill. [3rd ed.].

Sahm, M. and R. Schwabe (2001). A note on optimal bounded designs. In A. Atkinson, B. Bogacka, and A. Zhigljavsky (Eds.), *Optimum Design 2000*, Chapter 13, pp. 131–140. Dordrecht: Kluwer.

Schaback, R. (2003). Mathematical results concerning kernel techniques. In *Prep. 13th IFAC Symp. on System Identification, Rotterdam*, pp. 1814–1819.

Schlossmacher, E. (1973). An iterative technique for absolute deviations curve fitting. *J. Amer. Statist. Assoc. 68*(344), 857–859.

Schwabe, R. (1995). Designing experiments for additive nonlinear models. In C. Kitsos and W. Müller (Eds.), *MODA4 – Advances in Model-Oriented Data Analysis, Spetses (Greece), june 1995*, pp. 77–85. Heidelberg: Physica Verlag.

Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models.* New York: Springer.

Schwabe, R. (1997). Maximin efficient designs. Another view at D-optimality. *Statist. Probab. Lett. 35*, 109–114.

Schwabe, R. and W. Wong (1999). Efficiency bounds for product designs in linear models. *Ann. Inst. Statist. Math. 51*, 723–730.

Scott, D. (1992). *Multivariate Density Estimation.* New York: Wiley.

Shah, S., J. Mitchell, and M. Kupferschmid (2000). An ellipsoid algorithm for equality-constrained nonlinear programs. *Comput. Oper. Res. 28*(1), 85–92.

Shimizu, K. and E. Aiyoshi (1980). Necessary conditions for min-max problems and algorithm by a relaxation procedure. *IEEE Trans. Automatic Control 25*, 62–66.

Shiryaev, A. (1996). *Probability.* Berlin: Springer.

Shor, N. (1977). Cut-off method with space extension in convex programming problems. *Kibernetika 13*(1), 94–95. [English Trans. *Cybernetics Syst. Anal.*, **13**(1), 94–96].

Shor, N. (1985). *Minimization Methods for Non-Differentiable Functions.* Berlin: Springer.

Shor, N. and O. Berezovski (1992). New algorithms for constructing optimal circumscribed and inscribed ellipsoids. *Optim. Methods Softw. 1*, 283–299.

Sibson, R. (1972). Discussion on a paper by H.P. Wynn. *J. Roy. Statist. Soc. B34*, 181–183.

Sibson, R. (1974). D_A-optimality and duality. *Progress in Stat. Colloq. Math. Soc. Janos Bolyai 9*, 677–692.

Silvey, S. (1980). *Optimal Design.* London: Chapman & Hall.

Silvey, S. and D. Titterington (1973). A geometric approach to optimal design theory. *Biometrika 60*(1), 21–32.

Silvey, S., D. Titterington, and B. Torsney (1978). An algorithm for optimal designs on a finite design space. *Comm. Statist. – Theory Methods A7*(14), 1379–1389.

Söderström, T. and P. Stoica (1981). Comparison of some instrumental variable methods—consistency and accuracy aspects. *Automatica 17*(1), 101–115.

Söderström, T. and P. Stoica (1983). *Instrumental Variable Methods for System Identification.* New York: Springer.

Söderström, T. and P. Stoica (1989). *System Identification.* New York: Prentice Hall.

Spokoinyi, V. (1992). On asymptotically optimal sequential experimental design. *Advances in Soviet Math. 12*, 135–150.

St. John, R. and N. Draper (1975). D-optimality for regression designs: a review. *Technometrics 17*(1), 15–23.

Stein, C. (1956). Efficient nonparametric testing and estimation. In *Proc. 3rd Berkeley Symp. Math. Stat. Prob.*, Volume 1, pp. 187–196. Berkeley: Univ. of California Press.

Stoer, J. and R. Bulirsch (1993). *Introduction to Numerical Analysis (2nd Edition).* Heidelberg: Springer.

Stoica, P. (1998). On the Cramér-Rao bound under parametric constraints. *IEEE Signal Proc. Lett. 5*(7), 177–179.

Stoica, P. (2001). Parameter estimation problems with singular information matrices. *IEEE Trans. Signal Proc. 49*, 87–90.

Stone, C. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist. 3*(2), 267–284.

Studden, W. (1971). Elfving's theorem and optimal designs for quadratic loss. *Ann. Math. Statist. 42*, 1613–1621.

Sun, P. and R. Freund (2004). Computation of minimum-volume covering ellipsoids. *Oper. Res. 52*(5), 690–706.

Sundaraij, N. (1978). A method for confidence regions for nonlinear models. *Austral. J. Statist. 20*(3), 270–274.

Tarasov, S., L. Khachiyan, and I. Erlich (1988). The method of inscribed ellipsoids. *Soviet Math. Dokl. 37*(1), 226–230.

Titterington, D. (1975). Optimal design: some geometrical aspects of D-optimality. *Biometrika 62*(2), 313–320.

Titterington, D. (1976). Algorithms for computing D-optimal designs on a finite design space. In *Proc. 1976 Conf. on Information Science and Systems*, Baltimore, pp. 213–216. Dept. of Electronic Engineering, John Hopkins Univ.

Titterington, D. (1978). Estimation of correlation coefficients by ellipsoidal trimming. *J. Roy. Statist. Soc. C27*(3), 227–234.

Todd, M. (1982). On minimum volume ellipsoids containing part of a given ellipsoid. *Math. Oper. Res. 7*(2), 253–261.

Todd, M. and E. Yildirim (2007). On Khachiyan's algorithm for the computation of minimum volume enclosing ellipsoids. *Discrete Appl. Math. 155*, 1731–1744.

Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. In K. Kortanek and A. Fiacco (Eds.), *Proc. Int. Symp. on Semi-infinite Programming and Applications*, Heidelberg, pp. 249–260. Springer.

Torsney, B. (1986). Moment inequalities via optimal design theory. *Linear Algebra Appl. 82*, 237–253.

Torsney, B. (2009). W-iterations and ripples therefrom. In L. Pronzato and A. Zhigljavsky (Eds.), *Optimal Design and Related Areas in Optimization and Statistics*, Chapter 1, pp. 1–12. Springer.

Tsybakov, A. (2004). *Introduction à l'estimation non-paramétrique.* Heidelberg: Springer.

Uciński, D. and M. Patan (1982). *D*-optimal design of a monitoring network for parameter estimation of distributed systems. *J. Global Optim. 39*, 291–322.

Valko, P. and S. Vajda (1984). An extended ODE solver for sensitivity calculations. *Comput. Chem. 8*(4), 255–271.

van de Geer, S. (2000). *Empirical Processes in M-estimation.* Cambridge: Cambridge Univ. Press.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge: Cambridge Univ. Press.

van der Vaart, A. (2002). The statistical work of Lucien Le Cam. *Ann. Statist. 30*(3), 631–682.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes, with Applications to Statistics.* New York: Springer.

Vandenberghe, L., S. Boyd, and S.-P. Wu (1998). Determinant maximisation with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl. 19*(2), 499–533.

Vapnik, V. (1998). *Statistical Learning Theory.* New York: Wiley.

Vapnik, V. (2000). *The Nature of Statistical Learning Theory.* New York: Springer. [2nd ed.].

Vazquez, E. (2005). Modélisation comportementale de systèmes nonlinéaires multivariables par méthodes à noyaux et applications. Ph.D. Thesis, Université Paris XI, Orsay, France.

Vazquez, E. and E. Walter (2003). Multi-output support vector regression. In *Prep. 13th IFAC Symp. on System Identification, Rotterdam*, pp. 1820–1825.

Veinott, A. (1967). The supporting hyperplane method for unimodal programming. *Oper. Research 15*, 147–152.

Vila, J.-P. (1990). Exact experimental designs via stochastic optimization for nonlinear regression models. In *Proc. Compstat, Int. Assoc. for Statistical Computing*, pp. 291–296. Heidelberg: Physica Verlag.

Vila, J.-P. and J.-P. Gauchi (2007). Optimal designs based on exact confidence regions for parameter estimation of a nonlinear regression model. *J. Stat. Plann. Inference 137*, 2935–2953.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist. 20*, 595–601.

Walter, E. (1982). *Identifiability of State-Space Models.* Berlin: Springer Verlag.

Walter, E. (Ed.) (1987). *Identifiability of Parametric Models.* Oxford: Pergamon Press.

Walter, E. and L. Pronzato (1995). Identifiabilities and nonlinearities. In A. Fossard and D. Normand-Cyrot (Eds.), *Nonlinear Systems. Modeling and Estimation*, Chapter 3, pp. 111–143. London: Chapman & Hall.

Walter, E. and L. Pronzato (1997). *Identification of Parametric Models from Experimental Data*. Heidelberg: Springer.

Watson, G. (1964). Smooth regression analysis. *Sankhya, Series A 26*, 359–372.

Welch, W. (1982). Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics 24*(1), 41–28.

Welzl, E. (1991). Smallest enclosing disks (balls and ellipsoids). In *New Results and New Trends in Computer Science*, Lecture Notes in Comput. Sci., vol. 555, pp. 359–370. Heidelberg: Springer.

Wolfe, P. (1970). Convergence theory in nonlinear programming. In J. Abadie (Ed.), *Integer and Nonlinear Programming*, pp. 1–36. Amsterdam: North Holland.

Wright, M. (1998). The interior-point revolution in constrained optimization. Technical Report 98–4–09, Computing Sciences Research Center, Bell Laboratories, Murray Hill, New Jersey 07974.

Wu, C. (1978a). Some algorithmic aspects of the theory of optimal designs. *Ann. Statist. 6*(6), 1286–1301.

Wu, C. (1978b). Some iterative procedures for generating nonsingular optimal designs. *Comm. Statist. – Theory Methods A7*(14), 1399–1412.

Wu, C. (1980). Characterizing the consistent directions of least squares estimates. *Ann. Statist. 8*(4), 789–801.

Wu, C. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist. 9*(3), 501–513.

Wu, C. (1983). Further results on the consistent directions of least squares estimators. *Ann. Statist. 11*(4), 1257–1262.

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

Wu, C. and H. Wynn (1978). The convergence of general step–length algorithms for regular optimum design criteria. *Ann. Statist. 6*(6), 1273–1285.

Wynn, H. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Statist. 41*, 1655–1664.

Wynn, H. (1972). Results in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. B34*, 133–147.

Wynn, H. (1977). Optimum designs for finite populations sampling. In S. Gupta and D. Moore (Eds.), *Statistical Decision Theory and Related Topics II*, pp. 471–478. New York: Academic Press.

Wynn, H. (1982). Optimum submeasures with applications to finite population sampling. In S. Gupta and J. Berger (Eds.), *Statistical Decision Theory and Related Topics III. Proc. 3rd Purdue Symp., vol. 2*, pp. 485–495. New York: Academic Press.

Wynn, H. (2004). Maximum entropy sampling and general equivalence theory. In A. Di Bucchianico, H. Läuter, and H. Wynn (Eds.), *mODa'7 – Advances in Model–Oriented Design and Analysis, Proc. 7th Int. Workshop, Heeze (Netherlands)*, pp. 211–218. Heidelberg: Physica Verlag.

Wynn, H. (2007). Bayesian information-based learning and majorization. Technical report, http://mucm.group.shef.ac.uk/index.html.

Yang, M. (2010). On de la Garza phenomenon. *Ann. Statist. 38*(4), 2499–2524.

Yang, M. and J. Stufken (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist. 37*(1), 518–541.

Yang, S.-S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc. 80*(392), 1004–1011.

Ye, Y. (1997). *Interior-Point Algorithms: Theory and Analysis.* Chichester: Wiley.

Yu, Y. (2010a). Monotonic convergence of a general algorithm for computing optimal designs. *Ann. Statist. 38*(3), 1593–1606.

Yu, Y. (2010b). Strict monotonicity and convergence rate of Titterington's algorithm for computing D-optimal designs. *Comput. Statist. Data Anal. 54*, 1419–1425.

Yu, Y. (2011). D-optimal designs via a cocktail algorithm. *Stat. Comput. 21*, 475–481.

Zarrop, M. (1979). *Optimal Experiment Design for Dynamic System Identification.* Heidelberg: Springer.

Zhang, Y. and L. Gao (2003). On numerical solution of the maximum volume ellipsoid problem. *SIAM J. Optim. 14*(1), 53–76.

Zhu, Z. and H. Zhang (2006). Spatial sampling design under the infill asymptotic framework. *Environmetrics 17*(4), 323–337.

Author Index

Ahipasaoglu, S.D. 296
Aiyoshi, E. 311, 326
Alexéev, V. 124, 135, 336
Andrews, D.W.K. 20
Ardanuy, R. 112
Atkinson, A.C. 3, 70, 229, 233
Atwood, C.L. 110, 138, 139, 279, 281, 283, 284, 291
Avriel, M. 115, 331

Barndorff-Nielsen, O.E. 89
Barron, A.R. 241
Bartlett, P.L. 3
Bates, D.M. 169, 170, 185, 193, 223
Ben-Tal, A. 277
Beran, R. 97
Berezovski, O.A. 161, 295
Berlinet, A. 3
Bickel, P.J. 97, 102
Biedermann, S. 4
Bierens, H.J. 18, 19, 42, 97, 345
Bilardello, P. 340
Bilenkaya, S. 215
Billingsley, P. 11, 12
Birgin, E.G. 287
Bland, R.G. 296, 316, 318
Bloomfield, P. 154
Bohachevsky, I.O. 299
Böhning, D. 284, 285
Boissonnat, J.-D. 278
Bonnans, J.F. 291, 302, 304, 306, 307, 331
Bornkamp, B. 242
Box, G.E.P. 69, 133
Box, M.J. 183
Boyd, S. 161, 277, 315
Braess, D. 237, 245, 247
Bulirsch, R. 57, 339
Bystrický, J. 215

Caines, P.E. 104
Calamai, P.H. 287
Carroll, R.J. 58
Casella, G. 6, 36, 83, 95, 96, 99, 102
Chaloner, K. 163, 170, 215, 229, 237, 240, 241
Chaudhuri, P. 269, 272
Chavent, G. 191
Cheng, M.-Y 4
Clark, D.N. 179, 312
Clarke, B.S. 241
Clarke, G.P.Y. 183
Clyde, M. 170, 215
Cochran, W.G. 6
Cook, D. 124, 135, 140
Cook, R.D. 70, 161, 299
Cormen, T.H. 278
Correa, R. 309
Cox, D.R. 69, 83, 233, 240
Cox, G.M. 6
Cramér, H. 108
Cucker, F. 3

Dacol, D. 339
Dai, Y.-H. 287
D'Argenio, D.Z. 240

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

Davies, L. 161
de la Garza, A. 139
del Pino, G. 56, 93
Delmas, H. 340
Dembski, W.A. 242
Demidenko, E.Z. 191–193
Dem'yanov, V.F. 127, 138, 149, 302, 307, 337
den Boeff, E. 279, 303
den Hertog, D. 277, 279, 303, 315
Dette, H. 4, 139, 142, 146, 237, 245, 247, 292
Donev, A.N. 3
Downing, D. 50, 57, 93
Dragalin, V. 125
Draper, N.R. 284

Eaton, M.L. 241
Ecker, J.G. 314, 316
Efron, B. 170
Eisenhart, L.P. 171
Elfving, G. 39, 142
Erlich, I.I. 161, 315

Fang, S.-C. 254, 255
Fedorov, V.V. 2, 3, 50, 57, 93, 108, 111, 124, 125, 133, 135, 136, 140, 153, 161–164, 233, 236, 259, 281, 297
Fellman, J. 140, 292
Feuer, A. 245, 246
Firth, D. 183
Fletcher, R. 287
Ford, I. 272, 274
Fourgeaud, C. 17
Frank, M. 279
Freund, R.M. 140, 161
Friedman, J. 3
Fuchs, A. 17
Fuller, W.A. 64

Galéev, E. 124, 135, 336
Galil, Z. 123, 299
Gallant, A.R. 7
Gao, L. 161
Gauchi, J.-P. 8, 176, 179, 185, 314
Gautier, R. 267, 269
Gilbert, J.C. 291, 302, 304, 306, 307, 331
Gilmour, S.G. 186
Giovagnoli, A. 241
Goffin, J.-L. 309
Goldfarb, D. 296, 316, 318
Goodwin, G.C. 2, 103, 245, 246, 297
Gorman, J.D. 104
Green, P.J. 56, 68, 93
Grötschel, M. 316, 318

Hackl, P. 2, 3, 124, 125, 136, 153, 164
Hadjihassan, S. 161
Hall, P. 4
Halperin, M. 185
Hamilton, D.C. 8, 185
Harman, R. 113, 121, 123, 145, 162, 295
Hartley, H.O. 185
Harville, D.A. 128, 130, 148, 149, 162, 351, 352
Has'minskii, R.Z. 102
Hastie, T. 3
Hawkins, D.M. 161
He, S. 255
Hearn, D.W. 330
Hero, A.O. 104
Herzberg, A.M. 229
Heyde, C.C. 102
Hill, P.D.H. 233
Hinkley, D.V. 83, 170, 240
Hiriart-Urruty, J.B. 277
Hougaard, P. 176
Hu, I. 268, 272, 274
Huang, C.-Y. 137
Huber, P.J. 79

Ibragimov, I.A. 102
Ivanov, A.V. 7

Janout, Z. 215
Jennrich, R.I. 7, 14, 18, 19, 25, 348, 350
Jobson, J.D. 64
Johnson, M.E. 299
Jørgensen, B. 89
Joulia, X. 340
Juditsky, A.B. 312, 313
Jurík, T. 145
Juritz, J. 229

Karlin, S. 110, 138
Kelley, J.E. 315
Khachiyan, L.G. 161, 286, 295, 296, 315, 316

Kiefer, J. 107, 112, 117, 123, 132, 147, 148, 299
Kieffer, M. 222, 224
Kim, J. 77
Kiwiel, K.C. 309
Klassen, C.A.J. 97
Koehret, B. 340
Koopmans, T.C. 188
Kramer, M. 339
Kupferschmid, M. 314, 316, 317, 322
Kushner, H.J. 179, 312, 313

Lai, T.L. 2, 25, 28, 29, 34, 272, 274
Larntz, K. 237
Lawphongpanich, S. 330
Le Cam, L. 102
Le Lann, J.-M. 340
Lehar, F. 215
Lehmann, E.L. 36, 83, 95, 96, 99, 102
Leiserson, C.E. 278
Lemaréchal, C. 291, 302, 304, 306, 307, 309, 331
Leonov, S. 50, 57, 93
Leroy, A.M. 77
Levin, A.Ju. 315
Li, B. 170
Li, X. 255
Li, X.-S. 254, 255
Liang, K.-Y. 102
Lindley, D.V. 239
Lindsay, B.G. 170
Liu, S. 154
Ljung, L. 2, 103, 104, 297
López-Fidalgo, J. 112, 113
Lovász, L. 316, 318
Lucas, H.L. 133

Magnus, J.R. 117
Malozemov, V.N. 127, 138, 149, 302, 307, 337
Manski, C.F. 97, 102
Martínez, J.M. 287
McCormick, G.P. 287
McCullagh, P. 89
Melas, V.B. 139
Mikulecká, J. 124, 136
Minoux, M. 301, 302
Mitchell, J.E. 317, 322
Mitchell, T.J. 4
Molchanov, I. 284
Montgomery, D.C. 6
Moré, J.J. 287
Morris, M.D. 4
Müller, Ch.H. 137, 148
Müller, H.-G. 4
Müller, W.G. 2, 4, 272
Mykland, P.A. 269, 272

Nachtsheim, Ch.J. 299
Nelder, J.A. 89
Nemirovskii, A. 306, 307, 331
Nemirovskii, A.S. 277, 315
Nesterov, Yu. 277, 306, 307, 315, 331, 332
Neudecker, H. 117, 154
Newey, W.K. 20

Pan, S. 255
Parzen, E. 76, 264
Patan, M. 301
Payne, R.L. 2, 103, 297
Pázman, A. 2, 3, 8, 15, 38, 44, 45, 55, 67, 72–74, 104, 108, 111, 114, 117–119, 137, 140, 145, 148, 152, 162, 168, 169, 171, 175–179, 183, 184, 189, 215, 281, 283, 291, 314
Pepelyshev, A. 292
Pečarić, J.E. 154
Phillips, R.F. 57
Pilz, J. 163, 241
Polak, E. 138, 149, 288
Pollard, D. 77
Polyak, B.T. 286, 291, 312, 313
Pötscher, B.M. 272
Pronzato, L. 2, 4, 6, 8, 28, 34, 44, 45, 55, 67, 73, 103, 125, 134, 137, 140, 153, 161, 165, 177–179, 183, 184, 211, 222, 227, 228, 233, 267, 269, 273, 274, 292, 295, 297, 311, 313, 314, 339
Pukelsheim, F. 3, 107, 111, 113, 114, 116–119, 133, 141, 149, 163, 297, 336, 354
Puntanen, S. 154

Rabitz, H. 339
Raydan, M. 287
Raynaud, H.-F. 161

Reider, S. 297
Reiersøl, O. 188
Rényi, A. 109
Ritov, Y. 97
Rivest, R.L. 278
Robbins, H. 25
Robertazzi, T.G. 293, 294
Rockafellar, R.T. 118, 126, 335–337
Rodríguez-Díaz, J.M. 113
Rojas, C.R. 245, 246
Ross, G.J.S. 169, 211
Rousseeuw, P.J. 77
Rudin, W. 251
Ruppert, D. 58

Sagastizábal, C. 291, 302, 304, 306, 307, 331
Sahm, M. 164
Schaback, R. 3
Schlossmacher, E.J. 57
Schrijver, A. 316, 318
Schwabe, R. 134, 135, 148, 164
Schwartz, S.C. 293, 294
Scott, D.W. 266
Sebastiani, P. 241
Shah, S. 317, 322
Shimizu, K. 311, 326
Shiryaev, A.N. 11, 12, 15, 27
Shor, N.Z. 161, 295, 302, 316
Sibson, R. 130, 141
Silvey, S.D. 3, 38, 126, 139–142, 272, 274, 292
Smale, S. 3
Söderström, T. 102, 104
Spokoinyi, V.G. 268
St. John, R.C. 284
Stein, C. 97, 278
Stein, M.L. 299
Stoer, J. 57, 339
Stoica, P. 102, 104
Stone, C.J. 97
Studden, W.J. 110, 138, 142, 146
Stufken, J. 139
Styan, G.P.H. 154
Sun, P. 140, 161, 296
Sundaraij, N. 185

Tapia, R.A. 287
Tarasov, S.P. 161, 315
Thomas-Agnan, C. 3
Tibshirani, R. 3
Tikhomirov, V. 124, 135, 336
Titterington, D.M. 4, 141, 160, 161, 274, 291, 292
Todd, M.J. 161, 295, 296, 315, 316, 318
Torsney, B. 112, 134, 292
Trinca, L.A. 186
Trnovská, M. 162
Tsybakov, A.B. 3

Uciński, D. 301

Vajda, S. 340
Valko, P. 340
van de Geer, S. 20, 79
van der Vaart, A.W. 20, 77, 79, 97, 100, 102
Vandenberghe, L. 161, 277, 315
Vapnik, V.N. 3
Vazquez, E. 3
Veinott, A.F. 331
Verdinelli, I. 163, 240, 241
Vila, J.-P. 185
Vuchkov, I.N. 161

Wald, A. 25
Walter, E. 2, 3, 6, 103, 137, 140, 161, 179, 188, 211, 222, 224, 227, 228, 233, 297, 311, 313, 314, 339
Watson, G.S. 154
Watts, D.G. 8, 169, 170, 185, 193, 223
Wei, C.Z. 2, 25, 28, 34, 274
Weisberg, S. 161
Welch, W.J. 299, 301, 302
Wellner, J.A. 20, 97
Welsh, J.S. 245, 246
Welzl, E. 140, 161
Wolfe, P. 279, 284
Wolfowitz, J. 107, 132
Wong, W.K. 124, 134
Wright, M.H. 315
Wu, C.F.J. 7, 14, 25, 28, 34, 38, 274, 277, 281–283, 288, 289, 310, 350
Wu, S.-P. 161
Wynn, H.P. 45, 134, 161, 163, 164, 239, 281, 283, 292, 310

Yang, J. 313

Yang, M. 139
Yang, S.-S. 264
Ye, Y. 315
Yildirim, E.A. 161, 296
Yin, G.G. 179, 312, 313
Yu, Y. 292, 293
Yvinec, M. 278
Zarrop, M.B. 2, 103, 297
Zeger, S.L. 102
Zhang, H. 2
Zhang, Y. 161
Zhigljavsky, A.A. 134, 161, 292
Zhu, Z. 2
Zuyev, S. 284

Subject Index

A-optimality, 110
for the MSE matrix, 178
adaptive design, 267
algorithm
black-box method, 307
branch-and-bound, 299
bundle method, 306, 331
conditional gradient, 286
constrained gradient, 286
cutting-plane, 219, 326
DETMAX, 298
ellipsoid, 315
exchange, 298
interior-point, 315
level method, 331
multi-vertex direction, 303
multiplicative, 291
Newton, 291
projected gradient, 286
projected subgradient, 306
simulated annealing, 299
vertex-direction, 137, 272, 279
vertex-exchange, 285
approximate bias, 182
approximate design, 2
asymptotic normality, 29
local, 102
of a function of the LS estimator, 36
of the M estimator, 81
of the ML estimator, 88
in regression, 84
of the penalized WLS estimator, 51
of the TSLS estimator, 54
of the WLS estimator, 29
regular, *see* regular asymptotic normality
average efficiency, 236
average-optimum design, 236
algorithm, 312
equivalence theorem, 237
average-optimum exact design, 313

batch-sequential design, 267
Bayesian design, 163, 238
Bayesian estimation, 98
Bernoulli experiments, 67, 90
bias-corrected LS estimator, 183
bounded design measure, 163
bundle method, 306, 331

c-maximin efficiency, 147
c-maximin optimality, 112
derivative, 129
c-optimality, 38, 42, 111, 142, 203
algorithm, 145, 328
derivative, 129
duality, 148
equivalence theorem, 149
extended, 219
for nonlinear models, 149
for the MSE matrix, 179
semicontinuity, 15, 119
Caratheodory's theorem, 139, 143
central cut, 316
compound optimum design, 123
derivative, 127
equivalence theorem, 135

L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*,
Lecture Notes in Statistics 212, DOI 10.1007/978-1-4614-6363-4,

concavity, 114, 116
condition number, 113
confidence
  ellipsoid, 109
  region, 109, 184
    approximate volume, 185
consistency
  of a function of the LS estimator, 24
  of the LS estimator, 22
    with finite $\mathscr{X}$, 26
  of the M estimator, 79
  of the ML estimator, 86
    in regression, 84
  of the penalized WLS estimator, 50
  of the TSLS estimator, 53
  of the WLS estimator, 24
  strong, 12, 15
constrained optimum design, 123, 320, 331
  equivalence theorem, 135
continuous design, 2
convergence
  rate, 28, 76
  uniform, 16, 20
  weak, 11
Cramér–Rao inequality, 85, 89, 94
  Bayesian form, 99
  with constraints, 104
curvature
  intrinsic, 71, 169
  parametric, 170, 198, 203
  total, 223
cutting plane, 326
cutting planes, 219

D-optimality, 109, 185, 271
  algorithms, 293
  and c-maximin efficiency, 147
  average, 240, 241
  Bayesian, 240
  derivative, 128
  duality, 140
  equivalence theorem, 132
  for the MSE matrix, 178
D_A-optimality
  derivative, 130
D_s-optimality, 110
  derivative, 130
  duality, 141
  equivalence theorem, 138
deep cut, 316, 318
delta method, 36
dependent observations, 2, 25, 28, 102
design
  asymptotically discrete, 12
  product, 134
  randomized, 12
  sequential, 267
design criterion
  global, 118
  local, 106
  partial, 118
  positively homogenous, 115
  regularized, 152
  singular, 118
  with constraints, 123
design measure, 2
  bounded, 163
differential equation, 339
directional derivative, 125
discrimination, 5, 137, 233
distinguishability, 233
duality properties, 140

E-optimality, 111, 197, 215, 255
  algorithm, 327
  derivative, 129
  extended, 215, 231, 245
E_k-optimality, 113, 255
efficiency, 94
  asymptotic, 96
  super, 102
efficiency criterion, 121
  lower bounds, 121
Elfving's set, 140, 278
Elfving's theorem, 142
ellipsoid
  confidence, 109
  method, 296, 315
  of concentration, 108
  of maximum volume, 161, 315
  of minimum volume, 140, 160, 295
entropy, 109
  approximation, 180, 183
equality constraints, 74, 104
equivalence theorem, 132, 337
estimability, 6, 189, 231
  function, 196

estimating functions and equations, 101
estimation criterion, 16
estimation with constraints, 74, 162
estimator
  adaptive, 76, 96
  Bayesian, 98
  L_1, 57
  LS, *see* LS estimator
  M, *see* M estimator
  ML, *see* ML estimator
  one-(Newton) step, 97
  (data-) recursive, 55, 103
  robust, 77
  TSLS, *see* TSLS estimator
  unbiased, 94
  WLS, *see* WLS estimator
exact design, 1, 296
expectation surface, 43, 168, 190
  overlapping, 168, 172, 196, 208
exponential family, 4, 89, 95

Fisher information
  for location, 28, 82, 84
  matrix, 85, 88, 91, 92, 94
Frobenius norm, 113
full sequential design, 271

g-inverse, 14
  Moore–Penrose, 29
G-optimality, 112
  algorithm, 328
  extended, 220
generalized linear model, 5, 89
global criterion, 118
gradient, 126, 286

heteroscedasticity, 21, 167
homoscedasticity, 5, 21

identifiability, 6, 188
implicit function theorem, 74, 181
information
  function, 119
  mapping, 113
information matrix, 2, 38, 105, 268
  observed, 170
instrumental variables, 102
intrinsic curvature, 71, 169
isotonicity, 114, 118
iteratively reweighted LS, 56, 92

Jeffrey's prior, 241

Kullback–Leibler divergence, 84, 233, 239
kurtosis, 52

L-optimality, 111, 112
  average, 240
  Bayesian, 241
level method, 331
likelihood function, 83
linear programming, 145, 218, 220, 296, 303, 316
locally optimum design, 106, 216
Loewner's ordering, 114
LS estimability, 22, 189, 196, 212
LS estimator, 21
  approximate bias, 182
  asymptotic normality, 29, 274
  bias-corrected, 183
  consistency, 22
  iteratively reweighted, 56
  marginal densities, 177
  probability density, 173
  with parameterized variance, 48

M estimator, 79, 159, 233
  asymptotic normality, 81
  consistency, 79
marginal densities, 177, 178
maximin efficiency, 245
  bounds, 250
maximin-optimum design, 124, 244
  algorithm, 302
  derivative, 127
  equivalence theorem, 136, 245
maximin-optimum exact design, 311
maximum a posteriori estimator, 98, 152
  probability density, 177
maximum-entropy regularization, 221, 254
mean-squared error matrix, 6, 178, 183
ML estimator, 49, 83, 233
  asymptotic normality, 84, 88
  consistency, 84, 86

model
  flat, 171, 175
  generalized linear, 5, 89
  intrinsically linear, 169, 174, 212
  linear, 170
  mixed regression, 4, 48
  nonparametric, 3
  parametrically linear, 170
  product-type, 134
  regular, 35, 168
  variance-component, 4, 48
modeling error, 70, 157
monotonicity, 114
MV-optimality, 111, 255
  derivative, 128

non-differentiable optimization, 302
non-informative prior, 242
nuisance parameter, 50, 61, 70, 232

orthogonal invariance, 113

partial criterion, 118
penalized WLS estimator, 49, 60, 93, 156, 232
  asymptotic normality, 51
  consistency, 50
polarity, 141
positive homogeneity, 114
prediction-error method, 103
probability density
  of the LS estimator, 173
  of the maximum a posteriori estimator, 177
  of the penalized LS estimator, 176
probability level criterion, 259
product design, 134
projection, 287
projector, 71, 168

quadratic programming, 287, 307, 331
quantile criterion, 259
quasi-likelihood, 102

radius of curvature, 192
regression model, 21
  classification, 172
  geometry, 168
  linear, 4
  logistic, 90
  mixed, 4, 48
  multidimensional, 161
regular asymptotic normality, 37, 151
  for finite $\mathscr{X}$, 41
  of a multidimensional function, 47
regularization, 152
  via $\mathscr{L}_q$ norms, 248
  via maximum-entropy, 221, 254
reparameterization, 110, 169
repetitions, 1, 12, 169, 173, 185, 201, 211
replications, *see* repetitions
response function, 3
Riemannian curvature tensor, 171
rounding, 296

saddle-point approximation, 176
Schur's ordering, 114
score function, 101, 183
semicontinuity, 15, 118, 152, 198, 336
sensitivity functions, 339
sequential design, 267
singular criterion, 118
singular design, 34, 204
  in nonlinear models, 38, 149
skewness, 52
small sample asymptotics, 176
space filling design, 4
state-space representation, 339
steepest ascent, 284, 303, 307
stochastic approximation, 180, 267, 312
stochastic dynamic programming, 267
stochastic equicontinuity, 20
subdifferential, 306, 335
subgradient, 335
  algorithm, 307
  inequality, 309
  projection, 306
submodularity, 293
sufficient statistic
  multidimensional, 92
  one-dimensional, 90
superadditivity, 114, 354
support points, 1, 295
  number, 139, 145, 245

T-optimality, 137, 233
thinnest cylinder, 141

TSLS estimator, 53, 62, 91
  asymptotic normality, 54
  consistency, 53
two-stage allocation, 268

uniform SLLN, 16, 19
universal optimality, 123

variance
  misspecification, 57
  parameterized, 48, 93
  stabilization, 69

WLS estimator, 21, 154, 232
  asymptotic normality, 29
  consistency, 24