# Asymptotic criteria for designs in nonlinear regression with model errors

Andrej Pazman, Luc Pronzato

**To cite this version:**

Andrej Pazman, Luc Pronzato. Asymptotic criteria for designs in nonlinear regression with model errors. Mathematica Slovaca, 2006, 56 (5), pp.543-553. hal-00158853

**HAL Id: hal-00158853**

**https://hal.science/hal-00158853**

Submitted on 28 Feb 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Asymptotic criteria for designs in nonlinear regression with model errors**

Andrej Pázman and Luc Pronzato

*Department of Applied Mathematics and Statistics*
*Faculty of Mathematics, Physics and Informatics*
*Comenius University*
*Mlynská dolina 84248 Bratislava, Slovakia*
*pazman@fmph.uniba.sk*
*and*
*Laboratoire I3S, CNRS/UNSA, Bât. Euclide, Les Algorithmes,*
*2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France*
*pronzato@i3s.unice.fr*

*The research of the first author supported by the Slovak VEGA-grants No. 1/0264/03 and 1/3016/06*

*The work of the second author was partially supported by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors's view.*

**AMS subject classification**: 62K05, 62J02
**Keywords and phrases:** nonlinear regression, least squares, optimum design, asymptotic normality, measures of nonlinearity, information matrix

Abstract

*We derive bounds for the design optimality criteria under the assumption that the supposed regression model $y(x_k) = \eta(x_k, \theta) + \varepsilon_k$ ; $\quad k = 1, 2, \ldots$ does not correspond to the true one. The investigation is based on the asymptotic properties of the LSE of $\theta$, and full proofs of these properties are presented under the assumption that the sequence of design points $\{x_k\}_{k=1}^{\infty}$ is randomly sampled according to a design measure $\xi$. The bounds and the asymptotic properties are related to the intrinsic measure of nonlinearity of the model.*

## 1. Introduction.

In optimal experimental design one often assumes that there are no errors in the considered model, a condition rarely satisfied in practice. Papers considering robustness to model errors are usually dealing with linear models, and the main stress is on bias issues, that is on the term $\|\nu(.) - \eta(., \theta)\|_\xi$ in Theorem 1, below (cf. [6,9,11]). Here we want to show that the situation is different in a *nonlinear* regression model

$$y(x_k) = \eta(x_k, \theta) + \varepsilon_k \ ; \quad \theta \in \Theta \tag{1}$$

with $\varepsilon_k$ i.i.d. random variables, $E(\varepsilon_k) = 0, Var(\varepsilon_k) = \sigma^2$ ; $k = 1, 2, ....$ The model error consists in that the true description of the data is

$$y(x_k) = \nu(x_k) + \varepsilon_k \tag{2}$$

where $\nu(x_k)$ are unknown, and are outside the model (1), i.e. there is no $\theta \in \Theta$ such that $\nu(x_k) = \eta(x_k, \theta)$ ; $k = 1, 2, ....$ We shall show that changes of the asymptotic variance of the LS estimators are modified by this misspecification of the model. The main question to be answered here is "how far" from the model can be the values $\nu(x_k)$, so as not to influence too much the usual criteria for the design of the experiment.

*Assumptions on models:* We take assumptions usual in experimental design theory. The set $\Theta$ is a compact subset of $\mathbb{R}^p$ having no isolated points, i.e. $\overline{\text{int}(\Theta)} = \Theta$. Each $x_k$ is an element of $\mathcal{X}$, the design space, which is a compact subset of an Euclidean space. The function $\eta(x, \theta)$ is continuous on $\mathcal{X} \times \Theta$, and its first and second order derivatives with respect to the components of $\theta$ exist, and are continuous on $\mathcal{X} \times$ int$(\Theta)$. The function $\nu(x)$ is continuous on $\mathcal{X}$.

## 2. The asymptotic properties of the LSE.

The estimator considered is the LSE

$$\hat{\theta}^{(N)} = \arg \min_{\theta \in \Theta} \left\| y^{(N)} - \eta^{(N)}(\theta) \right\|^2$$

where $y^{(N)} = (y(x_1), ..., y(x_N))^T$, $\eta^{(N)}(\theta) = (\eta(x_1, \theta), ..., \eta(x_N, \theta))^T$ .The sequence of estimators $\left\{ \hat{\theta}^{(N)} \right\}_{N=1}^{\infty}$ has no stabilized asymptotic properties, even when model (1) is true, unless the sequences $\{\eta(x_k, \theta)\}_{k=1}^{\infty}$, $\left\{ \frac{\partial \eta(x_k, \theta)}{\partial \theta_i} \right\}_{k=1}^{\infty}$, $\left\{ \frac{\partial^2 \eta(x_k, \theta)}{\partial \theta_i \partial \theta_j} \right\}_{k=1}^{\infty}$ have some regularity properties. There are different ways how to formulate these properties. The classical approach in [5] requires that these sequences has "finite tail products", the book [4] requires assumptions formulated as suprema and infima of very complex terms , the same is true for [2]. Here we start from the point of view that in the context of experimental design the stress should be on formulating the assumptions in terms of the design measure, which by definition is a probability measure $\xi$ on $\mathcal{X}$. The standard interpretation of $\xi$ is that the sequence of frequencies of the design points $\{x_k\}_{k=1}^{\infty}$ must approach $\xi$ (in a certain sense). Here we restrict our attention to the particular case that $\{x_k\}_{k=1}^{\infty}$ is obtained by independent random sampling from $\xi$. This allows to obtain complete proofs of asymptotic properties of $\hat{\theta}^{(N)}$ in a much simpler way than in classical papers on nonlinear regression, maintaining the main feature of a general approach.

The proofs of the following statements are presented in Section 4.

**Lemma 1**. *If $\{x_k\}_{k=1}^{\infty}$ is randomly sampled from the probability distribution $\xi$, then, with probability equal to one, for any real function $a(x, \varepsilon, \theta)$ defined and continuous on $\mathcal{X} \times \mathbb{R} \times \Theta$, and such that*

$$\int_{\mathbb{R}} \left[ \int_{\mathcal{X}} \max_{\theta \in \Theta} |a(x, \varepsilon, \theta)| \, \xi(dx) \right] P(d\varepsilon) < \infty$$

*the uniform strong law of large numbers (USLLN) holds, i.e.*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}a\left(x_k,\varepsilon_k,\theta\right)=\int_{\mathbb{R}}\left[\int_{\mathcal{X}}a\left(x,\varepsilon,\theta\right)\xi\left(dx\right)\right]P\left(d\varepsilon\right)$$

*uniformly over $\theta\in\Theta$, and a.s. with respect to $P$* and *$\xi$.*

Let us denote

$$\left\|\nu\left(.\right)-\eta\left(.,\theta\right)\right\|_{\xi}^{2}=\int_{\mathcal{X}}\left[\nu\left(x\right)-\eta\left(x,\theta\right)\right]^{2}\xi\left(dx\right)$$

**Theorem 1.** *Suppose that the sequence $\{x_k\}_{k=1}^{\infty}$ is randomly sampled from $\xi$. Then under the assumptions on the model (2), the sequence of estimators $\left\{\hat{\theta}^{(N)}\right\}_{N=1}^{\infty}$ converges a.s. to the set*

$$\tilde{\Theta}_{\xi}=\arg\min_{\theta\in\Theta}\left\|\nu\left(.\right)-\eta\left(.,\theta\right)\right\|_{\xi}^{2}$$

*Moreover the usual estimator of $\sigma^2$*

$$\left[s^{2}\right]^{(N)}=\frac{1}{N-p}\sum_{k=1}^{N}\left[y\left(x_k\right)-\eta\left(x_k,\hat{\theta}^{(N)}\right)\right]^{2}$$

*converges a.s. to $\sigma^2+\min_{\theta\in\Theta}\left\|\nu\left(.\right)-\eta\left(.,\theta\right)\right\|_{\xi}^{2}$.*

**Remark**: According to [5], the estimator $\hat{\theta}^{(N)}$ is a measurable function of $x_k,\varepsilon_k;k=1,...,N$, if it is defined uniquely, otherwise, there is a measurable choice for $\hat{\theta}^{(N)}$.

**Theorem 2**. *Suppose that the sequence $\{x_k\}_{k=1}^{\infty}$ is randomly sampled from $\xi$. Suppose that $\tilde{\Theta}_{\xi}=\left\{\tilde{\theta}\right\}$ is a one-point set, and that $\tilde{\theta}\in$*int$(\Theta)$. *Suppose that $M\left(\xi,\tilde{\theta}\right)$ is nonsingular, where*

$$M\left(\xi,\theta\right)=\int_{\mathcal{X}}\frac{\partial\eta\left(x,\theta\right)}{\partial\theta}\frac{\partial\eta\left(x,\theta\right)}{\partial\theta^{T}}\xi\left(dx\right)$$

*and that*

$$C_{int}\left(\xi,\tilde{\theta}\right)\left\|\nu\left(.\right)-\eta\left(.,\tilde{\theta}\right)\right\|_{\xi}<1 \tag{3}$$

*where*

$$C_{int}\left(\xi,\theta\right)=\max_{u\in\mathbb{R}^{p},u\neq0}\frac{\left\|\left[I-P^{\theta}\right]\left[\sum_{i,j}u_i\frac{\partial^2\eta(.,\theta)}{\partial\theta_i\partial\theta_j}u_j\right]\right\|_{\xi}}{u^{T}M\left(\xi,\theta\right)u}$$

*is the generalized intrinsic measure of nonlinearity (curvature) of Bates and Watts (cf. [1]) in model (1), at the point $\theta$, and under the design $\xi$. Denote*

$$\begin{aligned}D_{\nu}\left(\xi,\theta\right)&=\int_{\mathcal{X}}\left[\eta\left(x,\theta\right)-\nu\left(x\right)\right]\frac{\partial^{2}\eta\left(x,\theta\right)}{\partial\theta\partial\theta^{T}}\xi\left(dx\right)\\M_{\nu}\left(\xi,\theta\right)&=\int_{\mathcal{X}}\left[\eta\left(x,\theta\right)-\nu\left(x\right)\right]^{2}\frac{\partial\eta\left(x,\theta\right)}{\partial\theta}\frac{\partial\eta\left(x,\theta\right)}{\partial\theta^{T}}\xi\left(dx\right)\end{aligned}$$

*Then* $M\left(\xi,\tilde{\theta}\right)+D_\nu\left(\xi,\tilde{\theta}\right)$, $M\left(\xi,\tilde{\theta}\right)+\sigma^{-2}M_\nu\left(\xi,\tilde{\theta}\right)$ *are nonsingular matrices, and the sequence* $\left\{\sqrt{N}\left(\hat{\theta}^{(N)}-\tilde{\theta}\right)\right\}_{N=1}^{\infty}$ *converges in distribution to a random vector distributed normally,* $N\left(0,\sigma^2 J_\nu^{-1}\left(\xi,\tilde{\theta}\right)\right)$*, where*

$$J_\nu\left(\xi,\tilde{\theta}\right) = \left[M\left(\xi,\tilde{\theta}\right)+D_\nu\left(\xi,\tilde{\theta}\right)\right] \times\left[M\left(\xi,\tilde{\theta}\right)+\sigma^{-2}M_\nu\left(\xi,\tilde{\theta}\right)\right]^{-1}\left[M\left(\xi,\tilde{\theta}\right)+D_\nu\left(\xi,\tilde{\theta}\right)\right]$$

**Remarks:** Cf. [7] for a detailed derivation of the intrinsic curvature for the case of a finite number of design points; the generalization to $L_2(\xi)$ is straightforward. We denoted by $I$ the identity operator in $L_2(\xi)$, and $P^\theta$ is an orthogonal projector in $L_2(\xi)$ defined by

$$\left(P^\theta\phi\right)(x)=\frac{\partial\eta(x,\theta)}{\partial\theta^T}M^{-1}(\xi,\theta)\int_{\mathcal{X}}\frac{\partial\eta(x^*,\theta)}{\partial\theta}\phi(x^*)\,\xi(dx^*)$$

for any $\phi(.)\in L_2(\xi)$. Notice also that without the assumption of nonsingularity of $M\left(\xi,\tilde{\theta}\right)$ we can have irregularities of the distribution of $\hat{\theta}^{(N)}$, even when model (1) is correct (cf. [8]).

**3. Bounds for the criteria of optimality.**

When model (1) is correct, and $\tilde{\theta}$ is the true parameter value, then $\sigma^2\left[M\left(\xi,\tilde{\theta}\right)\right]^{-1}$ is the asymptotic variance matrix of $\hat{\theta}^{(N)}$. The experimenter designing the experiment does not know that model (1) is wrong, and even if he suspects that, he does not know the true mean function $\nu(.)$. The usual approach consists in taking a point $\theta^{(o)}\in\mathrm{int}(\Theta)$, which is supposed to be close to the hypothetical true value of $\theta$, and applying an optimality criterion $\Phi$ on the information matrix $M\left(\xi,\theta^{(o)}\right)$. One then computes the optimum design in a standard way, still supposing that model (1) is true. The question is how efficient is the resulting design in model (2)

Any optimality criterion should respect the ordering of information matrices, hence a standard assumption on optimality criteria is the monotonicity

$$M\le M^*\Rightarrow\Phi[M]\le\Phi[M^*]$$

where $M\le M^*$ means $u^TMu\le u^TM^*u$ for every $u\in\mathbb{R}^p$. A good design $\xi$ should thus give a large value of $\Phi\left[M\left(\xi,\theta^{(o)}\right)\right]$. This ordering is opposite to the standard one in most books and papers on optimum experimental design, but it is recommended in [10] using information arguments: the criterion should measure the amount of information contained in the experiment. From this interpretation other important properties also follow: the positivity of $\Phi$ ($\Phi[M]>0$), and the homogeneity of $\Phi$

$$\Phi[kM]=k\Phi[M]\ ;\quad k>0$$

For example $\Phi[M]=\det^{1/p}[M]$ is a positive homogeneous monotone form of the criterion of D-optimality.

A design $\xi^*$ is called locally $\Phi$-optimal (in the locality of $\theta^{(o)}$) when $\Phi\left[M\left(\xi^*,\theta^{(o)}\right)\right] = \max_\xi \Phi\left[M\left(\xi,\theta^{(o)}\right)\right]$ On the other hand, when model (2) holds, the criterion function $\Phi$ should be applied to $J_\nu\left(\xi,\tilde{\theta}\right)$. Suppose $\tilde{\theta}=\theta^{(o)}$ and compare both criteria. The following theorem gives limits for $\Phi\left[J_\nu\left(\xi,\tilde{\theta}\right)\right]$ in terms of $\Phi\left[M\left(\xi,\tilde{\theta}\right)\right]$

**Theorem 3.** *When $\Phi$ is positive, monotone and homogeneous, and when the assumptions of Theorem 2 hold, then we have the following bounds for $\Phi\left[J_\nu\left(\xi,\tilde{\theta}\right)\right]$*

$$\frac{\delta^2}{b}\Phi\left[M\left(\xi,\tilde{\theta}\right)\right] \le \Phi\left[J_\nu\left(\xi,\tilde{\theta}\right)\right] \le \Delta^2\Phi\left[M\left(\xi,\tilde{\theta}\right)\right]$$

*where*

$$\begin{aligned}
\delta &= 1 - C_{int}\left(\xi,\tilde{\theta}\right)\left\|\nu(.)-\eta\left(.,\tilde{\theta}\right)\right\|_\xi \\
\Delta &= 1 + C_{int}\left(\xi,\tilde{\theta}\right)\left\|\nu(.)-\eta\left(.,\tilde{\theta}\right)\right\|_\xi \\
b &= 1 + \frac{\max_{x\in\mathcal{X}}\left[\nu(x)-\eta\left(x,\tilde{\theta}\right)\right]^2}{\sigma^2}
\end{aligned}$$

**Proof.** Denote $M = M\left(\xi,\tilde{\theta}\right)$, $M_\nu = M_\nu\left(\xi,\tilde{\theta}\right)$, $D_\nu = D_\nu\left(\xi,\tilde{\theta}\right)$, $J = J_\nu\left(\xi,\tilde{\theta}\right)$, $J^* = \left[M\left(\xi,\tilde{\theta}\right)+D_\nu\left(\xi,\tilde{\theta}\right)\right]\left[M\left(\xi,\tilde{\theta}\right)\right]^{-1}\left[M\left(\xi,\tilde{\theta}\right)+D_\nu\left(\xi,\tilde{\theta}\right)\right]$, $\nu = \nu(.)$, $\tilde{\eta} = \eta\left(.,\tilde{\theta}\right)$, $C = C_{int}\left(\xi,\tilde{\theta}\right)$, $P = P^{\tilde{\theta}}$. From the definition of $\tilde{\theta}$ we have $\nu-\tilde{\eta} = (I-P)(\nu-\tilde{\eta})$. So for every $u\in\mathbb{R}^p$ we have

$$u^T(M+D_\nu)u = u^TMu\left[1+\frac{<\tilde{\eta}-\nu,(I-P)\,u^T\frac{\partial^2\eta(.,\theta)}{\partial\theta\partial\theta^T}\mid_{\tilde{\theta}} u>_\xi}{u^TMu}\right]$$

where $<a,b>_\xi = \int_{\mathcal{X}} a(x)\,b(x)\,\xi(dx)$. From the Schwarz inequality

$$\begin{aligned}
u^T(M+D_\nu)u &\le \left(1+\|\nu-\tilde{\eta}\|_\xi C\right)u^TMu = \Delta\ u^TMu \\
u^T(M+D_\nu)u &\ge \left(1-\|\nu-\tilde{\eta}\|_\xi C\right)u^TMu = \delta\ u^TMu
\end{aligned} \tag{4}$$

Hence, since $u^TM^{-1}u = \max_\alpha\left[\left(u^T\alpha\right)^2/\alpha^TM\alpha\right]$ (cf. [3]), we have

$$\frac{1}{\Delta}u^TM^{-1}u \le u^T(M+D_\nu)^{-1}u \le \frac{1}{\delta}u^TM^{-1}u$$

Set $u = (M+D_\nu)v$, $\quad v\in\mathbb{R}^p$, to obtain

$$\frac{1}{\Delta}v^TJ^*v \le v^T(M+D_\nu)v \le \frac{1}{\delta}v^TJ^*v$$

which together with (4) gives

$$\delta^2v^TMv \le v^TJ^*v \le \Delta^2v^TMv$$

Since $M_\nu$ is p.s.d. we have

$$v^TJv \le v^TJ^*v \le bv^TJv$$

so that

$$\begin{aligned} \delta^2 v^T M v &\leq b v^T J v \\ \Delta^2 v^T M v &\geq v^T J v \end{aligned}$$

Since $\Phi$ is monotone and homogeneous it follows that

$$\frac{\delta^2}{b}\Phi\left[M\left(\xi,\tilde{\theta}\right)\right] \leq \Phi\left[J_\nu\left(\xi,\tilde{\theta}\right)\right] \leq \Delta^2\Phi\left[M\left(\xi,\tilde{\theta}\right)\right]$$

$\square$

## 4. Proofs of asymptotic properties.

**Proof of Lemma 1.** Denote $\mathcal{Z} = \mathbb{R}\times\mathcal{X}$, and $z = (\varepsilon, x)^T$ the random vector distributed $P \times \xi$ on $\mathcal{Z}$. Take some fixed $\theta^1 \in \Theta$, and consider the set

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

Define $\bar{a}_\delta(z)$ and $\underset{-\delta}{a}(z)$ as the maximum and the minimum of $a(z,\theta)$ over the set $\mathcal{B}\left(\theta^1, \delta\right)$. We have that $E\left\{\left|\underset{-\delta}{a}(z)\right|\right\}$ and $E\{|\bar{a}_\delta(z)|\}$ are bounded by $E\{\max_{\theta\in\Theta}|a(z,\theta)|\} < \infty$. Moreover, $\bar{a}_\delta(z) - \underset{-\delta}{a}(z)$ is an increasing function of $\delta$. Hence we can change the order of the limit and the mean to prove that

$$\lim_{\delta\searrow 0}\left[E\{\bar{a}_\delta(z)\} - E\left\{\underset{-\delta}{a}(z)\right\}\right] = E\left\{\lim_{\delta\searrow 0}\left[\bar{a}_\delta(z) - \underset{-\delta}{a}(z)\right]\right\} = 0$$

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

$$\forall_{\beta>0}\exists_{\delta(\beta)>0} \;:\; \left|E\left\{\bar{a}_{\delta(\beta)}(z)\right\} - E\left\{\underset{-\delta(\beta)}{a}(z)\right\}\right| < \frac{\beta}{2}$$

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

$$\begin{aligned} \frac{1}{N}\sum_k \underset{-\delta(\beta)}{a}(z_k) - E\left\{\underset{-\delta(\beta)}{a}(z)\right\} - \frac{\beta}{2} &\leq \frac{1}{N}\sum_k \underset{-\delta(\beta)}{a}(z_k) - E\left\{\bar{a}_{\delta(\beta)}(z)\right\} \\ &\leq \frac{1}{N}\sum_k a(z_k,\theta) - E\{a(z,\theta)\} \\ &\leq \frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - E\left\{\underset{-\delta(\beta)}{a}(z)\right\} \\ &\leq \frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - E\left\{\bar{a}_{\delta(\beta)}(z)\right\} + \frac{\beta}{2} \end{aligned}$$

By the strong law of large numbers we have that $\forall_{\gamma>0}\exists_{N(\beta,\gamma)}$

$$\begin{aligned} \text{Prob}\left\{\forall_{N>N(\beta,\gamma)}\left|\frac{1}{N}\sum_k \bar{a}_{\delta(\beta)}(z_k) - E\left\{\bar{a}_{\delta(\beta)}(z)\right\}\right| < \frac{\beta}{2}\right\} &> 1 - \frac{\gamma}{2} \\ \text{Prob}\left\{\forall_{N>N(\beta,\gamma)}\left|\frac{1}{N}\sum_k \underset{-\delta(\beta)}{a}(z_k) - E\left\{\underset{-\delta(\beta)}{a}(z)\right\}\right| < \frac{\beta}{2}\right\} &> 1 - \frac{\gamma}{2} \end{aligned}$$

Combining with the previous inequality we obtain: $\forall_{\gamma>0}\exists_{N(\beta,\gamma)}$

$$\text{Prob}\left\{\forall_{N>N(\beta,\gamma)} \max_{\theta\in\mathcal{B}(\theta^1,\delta(\beta))}\left|\frac{1}{N}\sum_k a(z_k,\theta)-E\{a(z,\theta)\}\right|<\beta\right\}>1-\gamma$$

It only remains to cover the compact set $\Theta$ by a finite numbers of open sets $\mathcal{B}\left(\theta^i,\delta(\beta)\right)$ ; $i=1,...,n(\beta)$. For any $\alpha>0$ take $\gamma=\alpha/n(\beta)$ , $N(\beta)=\max_i N_i(\beta,\gamma)$. We obtain

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

which completes the proof. □

**Proof of Theorem 1.** For any $\theta\in\Theta$

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}[y(x_k)-\eta(x_k,\theta)]^2 &= \frac{1}{N}\sum_{k=1}^{N}\varepsilon_k^2+\frac{2}{N}\sum_{k=1}^{N}[\nu(x_k)-\eta(x_k,\theta)]\varepsilon_k \\ &\quad +\frac{1}{N}\sum_{k=1}^{N}[\nu(x_k)-\eta(x_k,\theta)]^2\end{aligned}$$

since $\varepsilon_k=y(x_k)-\nu(x_k)$. The function $(x,\theta,\varepsilon)\to\varepsilon^2+2[\nu(x)-\eta(x,\theta)]\varepsilon+[\nu(x)-\eta(x,\theta)]^2$ is continuous on $\mathcal{X}\times\Theta\times\mathbb{R}$, so the USLLN holds, and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}[y(x_k)-\eta(x_k,\theta)]^2=\sigma^2+\|\nu(.)-\eta(.,\theta)\|_\xi^2 \tag{5}$$

a.s., and uniformly over $\Theta$, since $E(\varepsilon)=0$. Denote $z_k=(\varepsilon_k,x_k)$ and take a fixed sequence $z=\{z_k\}_{k=1}^\infty$ such that this limit holds, and denote by $\theta^{\#}(z)$ a limit point of $\left\{\hat\theta^{(N)}(z)\right\}_{N=1}^\infty$. It exists, since $\Theta$ is compact, and there is a subsequence $\left\{\hat\theta^{(N_k)}(z)\right\}_{k=1}^\infty$ converging to it. Take any $\tilde\theta\in\tilde\Theta$. By the definition of the LSE we have

$$\frac{1}{N_t}\sum_{k=1}^{N_t}\left[y(x_k)-\eta\left(x_k,\tilde\theta\right)\right]^2\geq\frac{1}{N_t}\sum_{k=1}^{N_t}\left[y(x_k)-\eta\left(x_k,\hat\theta^{(N_t)}(z)\right)\right]^2$$

According to (5), the left hand side converges to $\sigma^2+\left\|\nu(.)-\eta\left(.,\tilde\theta\right)\right\|_\xi^2$ while the right hand side to $\sigma^2+\left\|\nu(.)-\eta\left(.,\theta^{\#}(z)\right)\right\|_\xi^2$, consequently $\theta^{\#}(z)\in\tilde\Theta$. Hence all limit points of $\left\{\hat\theta^{(N)}(z)\right\}_{N=1}^\infty$, for any sequence $z=\{z_k\}_{k=1}^\infty$ which satisfies (5), are in $\tilde\Theta$.

The limit value of $\left[s^2\right]^{(N)}$ then follows from (5). □

**Proof of Theorem 2.**

For any $v \in \mathbb{R}^p$ we have by the Schwarz inequality

$$
\begin{aligned}
& v^T \left[ M\left(\xi, \tilde{\theta}\right) + D_\nu \left(\xi, \tilde{\theta}\right) \right] v \\
= \quad & \left[ v^T M\left(\xi, \tilde{\theta}\right) v \right] \left[ 1 + \frac{\int_{\mathcal{X}} \left[ \eta\left(x, \tilde{\theta}\right) - \nu(x) \right]^T \left[ I - P\left(\tilde{\theta}\right) \right] \left[ v^T \frac{\partial^2 \eta(x,\theta)}{\partial\theta\partial\theta^T} |_{\tilde{\theta}} \, v \right] \xi(dx)}{v^T M\left(\xi, \tilde{\theta}\right) v} \right] \\
\geq \quad & \left[ v^T M\left(\xi, \tilde{\theta}\right) v \right] \left[ 1 - \left\| \eta\left(., \tilde{\theta}\right) - \nu(.) \right\|_\xi C_{int}\left(\xi, \tilde{\theta}\right) \right] \geq 0
\end{aligned}
$$

hence $M\left(\xi, \tilde{\theta}\right) + D_\nu\left(\xi, \tilde{\theta}\right)$ is p.d. Evidently $M_\nu\left(\xi, \tilde{\theta}\right)$ is p.s.d. hence $M\left(\xi, \tilde{\theta}\right) + \sigma^{-2} M_\nu\left(\xi, \tilde{\theta}\right)$ is also p.d.

Take a sequence $z = \{z_k\}_{k=1}^\infty$ such that $\left\{\hat{\theta}^{(N)}(z)\right\}_{N=1}^\infty$ converges to $\tilde{\theta}$. According to Theorem 1, the probability of sampling such a sequence is equal to one. Denote $J_N(\theta, z) = \frac{1}{N}\sum_{k=1}^N \left[y(x_k) - \eta(x_k, \theta)\right]^2$. By the Taylor formula we have

(6)

$$
0 = \left[ \frac{\partial J_N(\theta, z)}{\partial\theta} \right]_{\hat{\theta}^{(N)}(z)} = \left[ \frac{\partial J_N(\theta, z)}{\partial\theta} \right]_{\tilde{\theta}} + \left[ \frac{\partial^2 J_N(\theta, z)}{\partial\theta\partial\theta^T} \right]_{\beta^{(N)}} \left[ \hat{\theta}^{(N)}(z) - \tilde{\theta} \right]
$$

with $\beta^{(N)}$ a point between $\hat{\theta}^{(N)}(z)$ and $\tilde{\theta}$. We have

$$
\begin{aligned}
-\sqrt{N} \left[ \frac{1}{2} \frac{\partial J_N(\theta, z)}{\partial\theta} \right]_{\tilde{\theta}} &= \frac{1}{\sqrt{N}} \sum_{k=1}^N \left[ y(x_k) - \eta\left(x_k, \tilde{\theta}\right) \right] \left[ \frac{\partial \eta(x_k, \theta)}{\partial\theta} \right]_{\tilde{\theta}} \\
&= \frac{1}{\sqrt{N}} \sum_{k=1}^N \left[ \varepsilon_k + \left( \nu(x_k) - \eta\left(x_k, \tilde{\theta}\right) \right) \right] \left[ \frac{\partial \eta(x_k, \theta)}{\partial\theta} \right]_{\tilde{\theta}}
\end{aligned}
$$

According to the central limit theorem, the last term converges in distribution to a random vector distributed $N\left(0, \sigma^2 M\left(\xi, \tilde{\theta}\right) + M_\nu\left(\xi, \tilde{\theta}\right)\right)$ since

$$
E\left\{ \left[ \varepsilon + \left( \nu(x) - \eta\left(x, \tilde{\theta}\right) \right) \right] \left[ \frac{\partial \eta(x, \theta)}{\partial\theta} \right]_{\tilde{\theta}} \right\} = 0
$$

and

$$
\begin{aligned}
& Var\left\{ \left[ \varepsilon + \left( \nu(x) - \eta\left(x, \tilde{\theta}\right) \right) \right] \left[ \frac{\partial \eta(x, \theta)}{\partial\theta} \right]_{\tilde{\theta}} \right\} \\
= \quad & E\left\{ \left[ \varepsilon + \left( \nu(x) - \eta\left(x, \tilde{\theta}\right) \right) \right]^2 \left[ \frac{\partial \eta(x, \theta)}{\partial\theta} \frac{\partial \eta(x, \theta)}{\partial\theta^T} \right]_{\tilde{\theta}} \right\} \\
= \quad & \sigma^2 M\left(\xi, \tilde{\theta}\right) + M_\nu\left(\xi, \tilde{\theta}\right)
\end{aligned}
$$

Further we have

$$
\begin{aligned}
\frac{\partial^2 J_N(\theta, z)}{\partial\theta\partial\theta^T} &= \frac{1}{N} \sum_{k=1}^N \left[ \nu(x_k) + \varepsilon_k - \eta(x_k, \theta) \right] \left[ \frac{\partial^2 \eta(x_k, \theta)}{\partial\theta\partial\theta^T} \right] \\
&+ \frac{1}{N} \sum_{k=1}^N \left[ \frac{\partial \eta(x_k, \theta)}{\partial\theta} \frac{\partial \eta(x_k, \theta)}{\partial\theta^T} \right]
\end{aligned}
$$

and according to Lemma 1 this expression converges a.s. with respect to $x_k$ and $\varepsilon_k$, and uniformly with respect to $\theta$ to

$$\begin{aligned} & E\left\{\left[\nu\left(x\right)+\varepsilon-\eta\left(x,\theta\right)\right]\left[\frac{\partial^2\eta\left(x,\theta\right)}{\partial\theta\partial\theta^T}\right]\right\}+E\left\{\frac{\partial\eta\left(x,\theta\right)}{\partial\theta}\frac{\partial\eta\left(x,\theta\right)}{\partial\theta^T}\right\} \\ = \quad & D_\nu\left(\xi,\theta\right)+M\left(\xi,\theta\right) \end{aligned}$$

Consequently $\left[\frac{\partial^2 J_N(\theta,z)}{\partial\theta\partial\theta^T}\right]_{\beta^{(N)}}$ converges a.s. to $D_\nu\left(\xi,\tilde{\theta}\right)+M\left(\xi,\tilde{\theta}\right)$ since $\beta^{(N)}$ converges a.s. to $\tilde{\theta}$. The required result now follows from (6) $\square$

**Acknowledgment.** The authors thank the referee for a very careful revision of the paper.

**References**

[1] Bates, D.M., Watts, D.G.: Nonlinear Regression Analysis and its Applications. New York, Wiley 1988.

[2] Gallant, A. R.: Nonlinear Statistical Models. New York, Wiley 1987.

[3] Harville, D.A.: Matrix Algebra from a Statistician's Perspective. New York, Springer 1997.

[4] Ivanov, A.V.: Asymptotic Theory of Nonlinear Regression. Dordrecht, Kluwer, 1997.

[5] Jennrich, R. L.: Asymptotic properties of nonlinear least squares estimation. Ann. Math. Statist., **40** (1969), p. 633-643.

[6] Kiefer, J.: Optimum design for fitting biased multivariate surfaces. In: P.R.Krishnaian, editor: Multivariate analysis p. 287-297. North Holland, Amsterdam 1973.

[7] Pázman A.: Nonlinear Statistical Models. Dordrecht, Kluwer, 1993.

[8] Pázman A., Pronzato, L.: On the irregular behavior of LS estimators for asymptotically singular designs. Lab. I3S, CNRS Sophia Antipolis. Rapport de recherche ISRN I3S/RR-2005-12-FR, Juin 2005 (To appear in Stat. & Prob. Letters)

[9] Pesotchinski, L.: Optimal robust designs: linear regression in $R^k$. Annals of Statistics **10** (1982), p. 511-525.

[10] Pukelsheim, F.: Optimal Experimental Design. New York, Wiley 1993.

[11] Welch, W. J.: A mean squared error criterion for the design of experiments. Biometrika **70** (1983), p. 205-213.