Ann. Inst. Statist. Math.
Vol. 52, No. 3, 574–597 (2000)

A UNIFIED APPROACH TO SECOND ORDER OPTIMALITY CRITERIA IN NONLINEAR DESIGN OF EXPERIMENTS

HOLGER DETTE[1] AND YURI GRIGORIEV[2*]

[1] *Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany*
[2] *Fakultät für Angewandte Mathematik, Technische Universität, 630092 Novosibirsk, Russian Federation*

(Received March 17, 1998; revised June 14, 1999)

Abstract. In the nonlinear regression model we consider the optimal design problem with a second order design D-criterion. Our purpose is to present a general approach to this problem, which includes the asymptotic second order bias and variance criterion of the least squares estimator and criteria using the volume of confidence regions based on different statistics. Under assumptions of regularity for these statistics a second order approximation of the volume of these regions is derived which is proposed as a quadratic optimality criterion. These criteria include volumes of confidence regions based on the u_n- representable statistics. An important difference between the criteria presented in this paper and the second order criteria commonly employed in the recent literature is that the former criteria are independent of the vector of residuals. Moreover, a refined version of the commonly applied criteria is obtained, which also includes effects of nonlinearity caused by third derivatives of the response function.

Key words and phrases: D-optimal design, nonlinear regression model, second order approximation, asymptotic expansion.

1. Introduction

Let

$$y = \eta(\theta) + \varepsilon = (\eta(x^1, \theta), \dots, \eta(x^n, \theta))^T + \varepsilon, \tag{1.1}$$

denote the common nonlinear regression model with the observed vector $y \in \mathbb{R}^n$, the vector of unknown parameter $\theta = (\theta^1, \dots, \theta^m)^T \in \Theta \subseteq \mathbb{R}^m$, $m \le n$, and the random vector $\varepsilon = (\varepsilon^a)_{a=1}^n$ with i.i.d. components ε^a. We suppose that ε^a has a symmetric distribution not depending on θ with expected value zero and variance σ^2. The set Θ is a convex open set and for fixed $x = (x^1, \dots, x^n)^T$ the mapping $\theta \mapsto \eta$ is supposed to be continuous with continuous third (or fourth) order derivatives (if required) such that the rank of the $n \times m$ matrix

$$F = F(\theta) = \left(\frac{\partial \eta^a}{\partial \theta^i}\right)_{i=1,\dots,m}^{a=1,\dots,n} = (F_i^a)_{i=1,\dots,m}^{a=1,\dots,n} \tag{1.2}$$

is m for all $\theta \in \Theta$. Here $\eta^a(\theta) = \eta(x^a, \theta)$ denotes a-th component of the vector $\eta(\theta)$, $x^a \in X$ is the a-th value of explanatory variable and X is the design space with sigma field containing all one point sets and containing at least m points.

* This work was done during a visit of the second author at the Department of Mathematics, Ruhr-Universität Bochum, Germany.

Following differential geometric convention we denote for a matrix $A = (A_{ij})_{i=1,\dots,n_1}^{j=1,\dots,n_2}$ with A_{ij}, A^{ij} and A_i^j simultaneously the elements of A and the matrix A itself (i.e. $A = A_{ij} = A^{ij} = A_i^j$); the specific meaning will be clear from the context. We will also make substantial use of Einstein's rule; for example $A^{ij}B_{jk} = A_{ij}B^{jk}$ denotes the matrix AB (and simultaneously the element in the position (i,k)) and $A_{ij}B^{ij} = A^{ij}B_{ij} = \text{trace}(AB^T)$.

A design ξ_n is a probability measure on X (or on its σ-field) and the matrix

$$M_{ij} = M_{ij}(\xi_n, \theta) = \int_X F_i(x,\theta)F_j(x,\theta)d\xi_n(x)$$

(where $F_i(x,\theta) = \frac{\partial}{\partial\theta^i}\eta(x,\theta)$) is proportional to the Fisher information matrix provided some conditions of regularity are satisfied (Borovkov (1998)). If ξ_n puts masses n_k/n at the points $x_{(i)}$, $i = 1,\dots,r$, $x^a \in \{x_{(1)},\dots,x_{(r)}\}$ then the experimenter takes n_i observations at each $x_{(i)}$ and the matrix M_{ij} is proportional to the inverse of the asymptotic covariance matrix of the least squares estimator $u_n = n^{1/2}(\hat\theta_n - \theta)$. This estimator of the unknown parameter $\theta \in \Theta$ is obtained from the condition

$$S(\hat\theta_n) = \inf_{\tau\in\Theta} S(\tau), \qquad S(\theta) = \|y - \eta(\theta)\|^2, \tag{1.3}$$

where $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$ and the inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^n$ is defined with respect to the matrix $n^{-1}\delta_{ab}$ (here and throughout this paper $\delta_{ab} = \delta^{ab} = \delta_a^b$ denotes Kronecker's symbol and simultaneously the identity matrix). For the sake of simplicity a function $g(\theta)$ evaluated at the least squares estimator $\theta = \hat\theta_n$ will be denoted by $\hat g$, e.g. $\hat S = S(\hat\theta_n)$ (by (1.3)) or $\hat\eta = \eta(\hat\theta_n)$ (by (1.1)).

A D-optimal design of experiment maximizes the determinant of the matrix M_{ij} and a design ξ_n^* is called locally D-optimal for given n if

$$|M(\xi_n^*, \theta)| = \max_{\xi_n} |M(\xi_n, \theta)|. \tag{1.4}$$

The statistical interpretation of the criterion (1.4) is that a D-optimal design minimizes the first order approximation of the volume of the ellipsoid of concentration

$$C_{1-\alpha} = \{\theta \in \Theta \mid M_{ij}(\xi_n, \hat\theta_n)u_n^i u_n^j \le \sigma^2\chi^2_{1-\alpha}(m)\} \tag{1.5}$$

where $\chi^2_{1-\alpha}(m)$ denotes the $(1-\alpha)$-quantile of the χ^2 distribution with m degrees of freedom. In the regular case we have

$$P\{\theta \in C_{1-\alpha}\} = 1 - \alpha + o(1), \qquad n \to \infty \tag{1.6}$$

(see Rao (1965)). Locally D-optimal designs for nonlinear models have been studied by numerous authors (see e.g. Chernoff (1953), Box and Lucas (1959), Ford *et al.* (1992)). Hamilton and Watts (1985) demonstrate that this first-order approximation can often be quite poor and introduce a second-order volume approximation. Whereas the D-optimality criterion works with a tangent plane approximation, the second order criteria take into account the curvature of the expectation surface. The theory of optimal design with respect to these criteria is not so well studied and only in development (see Hamilton and Watts (1985), Pazman (1989), O'Brien (1992), Grigoriev (1993), Pronzato and Pazman (1994), O'Brien and Rawlings (1996)).

In this paper we discuss a unified approach to a class Ψ of second order optimality criteria which could be used for the construction of optimal designs in nonlinear regression models. This class can be divided into two subclasses, say $\Psi = \Psi_1 \cup \Psi_2$. The first subclass of criteria, say Ψ_1, is motivated by the asymptotic second order expansion of the bias and covariance matrix of the estimator u_n, i.e.

$$Eu_n = b_n n^{-1/2} + o(n^{-1}), \tag{1.7}$$

$$\operatorname{Var} u_n = \sigma^2 \Lambda + \sigma^4 \Lambda_1 n^{-1} + o(n^{-1}), \tag{1.8}$$

where Λ is the inverse of the information matrix M_{ij} and the vector b_n and the matrix Λ_1 will be defined in Section 2 (see Box (1971), Clarke (1980), Ivanov (1982, 1997), Grigoriev and Ivanov (1987*a*), Grigoriev (1994), O'Brien and Rawlings (1996)). We propose the second order approximation of Wilk's generalized variance of convex combinations of the variance and the adjustment to it as a second order optimality criteria for nonlinear regression models.

The second subclass of criteria, say Ψ_2, is obtained from a second order expansion of the volume of the confidence region

$$C_{1-\alpha} = \{\theta \in \Theta \mid T_n \leq \sigma^2 \chi^2_{1-\alpha}(m)\}, \tag{1.9}$$

where the statistic T_n is representable in the form

$$T_n = T_{0n} + T_{1n} n^{-1/2} + T_{2n} n^{-1} + o_p(n^{-1}) \tag{1.10}$$

such that $T_n \xrightarrow{d} \chi^2(m)$, $n \to \infty$. Our main results give the second order expansion of the volumina of $C_{1-\alpha}$ for a broad class of statistics T_n which we call u_n-representable with c-property. This class contains the Kullback-Leibler, modified Wald and a modification of Rao's statistic. The resulting second order criteria are similar to the criterion of Hamilton and Watts (1985), who based their approach on the Neyman-Pearson statistic. The important difference is that the criteria proposed in this paper do not depend on the vector of the residuals (in contrast to the criterion of Hamilton and Watts (1985)). Moreover the methods used in the proofs of our results yield a refined second order approximation for the optimality criterion of Hamilton and Watts (1985) who obtained their expansion by ignoring the third order derivatives and by considering only the case $m = n$ (see also Seber and Wild (1989)).

The second order criteria discussed in this paper can be classified by two parameters $\beta \in [0, 1]$ and L_{2n} in the form

$$Q(\xi_n) = |\sigma^2 \Lambda| \left\{ \beta + (1-\beta) \frac{\kappa^2}{m+2} L_{2n} n^{-1} \right\}, \tag{1.11}$$

where $|\sigma^2 \Lambda|$ is the determinant of the inverse information matrix. The case $\beta = 1$ corresponds to the first order D-optimality criterion while $\beta < 1$ gives the second order criterion. For example, the criteria based on volumina of confidence regions are obtained for $\beta = 0.5$ and $\kappa^2 = \chi^2_{1-\alpha}(m)$.

The paper is organized as follows. In Section 2 the asymptotic bias and variance criteria are discussed while Section 3 introduces the second order approximation of the volume of confidence regions as optimality criteria. In Section 4 we introduce a class of generalized second order optimality criteria. Section 5 gives a detailed discussion for the one-parameter nonlinear model, for which the situation is more transparent, and illustrates the different criteria in the exponential regression model. In the same section we also discuss second order optimal designs for the two-parameter intermediate product model (see Box and Lucas (1959)). Finally, Section 6 summarizes our main results, puts the proposed criteria in larger context of experimental design and discusses possible directions for future research. The proofs are somewhat tedious and therefore deferred to the Appendix.

2. The asymptotic bias and variance criterion

The asymptotic second order expansion of the bias vector and covariance matrix of the least squares estimator $\hat{\theta}_n$ was calculated by Box (1971) and Clarke (1980). A more detailed discussion can be found in Ivanov (1982), Grigoriev and Ivanov (1987*a*, *b*), Grigoriev (1994), O'Brien and Rawlings (1996) and Ivanov (1997). Throughout this paper we use differential geometric notations for quantities, which are connected with the expectation surface

$$\mathbb{E}^m = \{\eta(\theta) \mid \theta \in \Theta\}. \tag{2.1}$$

Generalizing the notation of the information matrix we introduce

$$F_{i_1,\ldots,i_k}(x,\theta) = \frac{\partial^k}{\partial\theta^{i_1},\ldots,\partial\theta^{i_k}}\eta(x,\theta)$$
$$\Pi_{(i_1,\ldots,i_k)(j_1,\ldots,j_\ell)} = \int_X F_{i_1,\ldots,i_k}(x,\theta)F_{j_1,\ldots,j_\ell}(x,\theta)d\xi_n(x)$$

and denote the inverse of the information matrix $M_{ij} = \Pi_{(i)(j)}$ with Λ^{ij}. Following Grigoriev and Ivanov (1987*a*) the Christoffel symbols of the second kind of $\mathbb{E}^m$ are defined as

$$\Gamma^i_{jk} = \Lambda^{ir}\Pi_{(r)(jk)}, \tag{2.2}$$

and the Ricci tensor of the expectation surface $\mathbb{E}^m$ is given by

$$R_{ik} = \Lambda^{rs}(\Pi_{(ik)(rs)} - \Pi_{(ir)(ks)}) + \Lambda^{rs}\Lambda^{\alpha\beta}(\Pi_{(\alpha)(ir)}\Pi_{(\beta)(ks)} - \Pi_{(\alpha)(rs)}\Pi_{(\beta)(ik)}). \tag{2.3}$$

For the coordinates of the bias vector in (1.8) we thus obtain the following expression

$$b^k_n = -\frac{\sigma^2}{2}\Lambda^{ij}\Gamma^k_{ij} \qquad (k = 1,\ldots,m). \tag{2.4}$$

From (2.2) and (2.4) it follows that in the metric $\langle\cdot,\cdot\rangle$ with generating matrix $\sigma^{-2}\Pi_{(i)(j)}$ the norm of the bias vector is given by

$$\|b_n\|^2 = \frac{\sigma^2}{4}\Lambda^{ij}\Lambda^{k\ell}\Lambda^{rs}\Pi_{(r)(ij)}\Pi_{(s)(k\ell)}.$$

Therefore, if the bias is important, we propose the quantity

$$Q_B(\xi_n) = |\sigma^2\Lambda^{ij} + b^i_n b^j_n n^{-1}| = |\sigma^2\Lambda| \cdot (1 + B_{2n}n^{-1}) \tag{2.5}$$

with

$$B_{2n} = \|b_n\|^2 \tag{2.6}$$

as a possible design criterion of second order for minimisation. Similarly, we have the expression (see Grigoriev and Ivanov (1987*a*))

$$\operatorname{Var} u_n = D_n^{ij} + o(n^{-1}), \tag{2.7}$$

for the covariance matrix (1.8), where

$$D_n^{ij} = \sigma^2 \Lambda^{ij} + \sigma^4 \Lambda_1^{ij} n^{-1} = \sigma^2 \Lambda^{ir} (\delta_r^j + \sigma^2 \Lambda_{2,r}^j n^{-1}) \tag{2.8}$$

i.e. $\Lambda_1^{ij} = \Lambda^{ir} \Lambda_{2,r}^j$. The matrix Λ_2 with elements $\Lambda_{2,k}^j$ is defined as

$$\Lambda_{2,k}^j = -\Lambda^{jr} R_{rk} + \frac{1}{2} \Lambda^{ru} \Lambda^{sv} \Lambda^{ij} \Pi_{(k)(rs)} \Pi_{(i)(uv)} + T_{k)}^{(j} \tag{2.9}$$

where $T_{k)}^{(j} = \frac{1}{2}(T_k^j + T_j^k)$ and

$$T_k^j = \Lambda^{rs}(\Gamma_{\alpha k}^j \Gamma_{rs}^\alpha + 2\Gamma_{\alpha r}^j \Gamma_{ks}^\alpha) - \Lambda^{jr} \Lambda^{su} \Pi_{(ksu)(r)}.$$

Observing (2.9) we propose as an alternative second order criterion of D-optimality

$$Q_V(\xi_n) = |D_n^{ij}| = |\sigma^2 \Lambda| \cdot |I_m + \sigma^2 \Lambda_2 n^{-1}| = |\sigma^2 \Lambda| \cdot (1 + V_{2n} n^{-1}), \tag{2.10}$$

where

$$V_{2n} = \operatorname{tr} \sigma^2 \Lambda_2. \tag{2.11}$$

We note, that the principal drawback of the criterion (2.11) is that in general the matrix $\sigma^2 \Lambda_2$ is not positive definite and therefore the quantity V_{2n} can be negative. If bias and variance are both important we can, in addition, consider the matrix of mean squared deviation

$$\Delta_n^{ij} = D_n^{ij} + b_n^i b_n^j n^{-1}. \tag{2.12}$$

Its determinant is obviously representable in the form

$$Q_Z(\xi_n) = |\sigma^2 \Lambda| \cdot (1 + Z_{2n} n^{-1}), \tag{2.13}$$

where

$$Z_{2n} = B_{2n} + V_{2n}. \tag{2.14}$$

The values of the quantities B_{2n}, V_{2n} and Z_{2n} will be discussed in Section 4 in more detail. For regression models with a scalar parameter $\theta \in \Theta \subseteq \mathbb{R}^1$ a more transparent representation will be derived in Section 5.

3. Criteria based on the volume of confidence regions

Hamilton and Watts (1985) considered the confidence region (1.9) with the statistic of Neyman and Pearson (see Rao (1965)).

$$T_n = S(\theta) - S(\hat{\theta}_n). \tag{3.1}$$

They showed that the volume of a $100(1-\alpha)\%$ confidence region for θ is approximately equal to

$$V_n = \gamma|\sigma^2\Lambda|^{1/2}|C|^{1/2}\left\{1 + \frac{a^2}{2(m+2)}\operatorname{tr} C^{-1}Gn^{-1}\right\}, \tag{3.2}$$

where

$$a^2 = \chi^2_{1-\alpha}(m), \qquad \gamma = \frac{(a\sqrt{\pi})^m}{\Gamma\left(\dfrac{m+2}{2}\right)},$$

the matrix $G = G^{jk}$ is a function of the arrays of parameter effect curvatures (see Bates and Watts (1980)) and the matrix C is a function of the arrays of intrinsic curvatures of the expectation surface $\mathbb{E}^m$ defined in (2.1). In the following let $u : \mathbb{R}^n \to \mathbb{R}^n$ denote an orthogonal transformation of $\mathbb{R}^n$ such, that its matrix has the form

$$U = (T \vdots N),$$

where $T = FD$ (or elementwise $T^a_j = F^a_k D^k_j$) and $DD^T = \Lambda$ (or $D^k_i D^\ell_j \delta^{ij} = \Lambda^{k\ell}$). Here $\{T_i, k = 1, \ldots, m\}$ is an orthonormal basis of the tangent space

$$T^m(\hat{\theta}_n) = \{x \in \mathbb{R}^n \mid x = \hat{\eta} + \hat{F}t, t \in \mathbb{R}^m\}$$

of the expectation surface $\mathbb{E}^m$ at the point $\hat{\eta} = \eta(\hat{\theta}_n)$ and $\hat{F} = F(\hat{\theta}_n)$ is defined in (1.2). Similarly $\{N_\alpha, \alpha = 1, \ldots, n-m\}$ is an orthonormal basis of the orthogonal complement $N^{n-m}(\hat{\theta}_n)$ of the tangent space $T^m(\hat{\theta}_n)$. We also introduce the arrays of Bates and Watts (1980)

$$A_{i,rs} = \langle T_i, F_{k\ell}D^k_r D^\ell_s\rangle, \qquad i = 1, \ldots, m \tag{3.3}$$

and

$$A_{m+\alpha,rs} = \langle N_\alpha, F_{k\ell}D^k_r D^\ell_s\rangle, \qquad \alpha = 1, \ldots, n-m, \tag{3.4}$$

where

$$F_{k\ell} = \frac{\partial^2}{\partial\theta^k \partial\theta^\ell}\eta(\theta).$$

Following Hamilton and Watts (1985) we can now write

$$G_{jk} = \delta_{rj}\delta_{sk}G^{rs} = \sigma^2 \sum_{r=1}^{m}\sum_{s=1}^{m}\{A_{r,rs}A_{s,jk} + A_{r,js}A_{s,rk} + A_{r,rj}A_{s,sj}\} \tag{3.5}$$

and the elements of $C^{-1} = C_{ij}$ are given by

$$C_{jk} = \delta_{jk} - \langle\hat{e}, F_{rs}D^r_j D^s_k\rangle = \delta_{jk} - \lambda^\alpha A_{m+\alpha,jk}, \tag{3.6}$$

where λ^α are the coefficients in the expansion of the residual vector $\hat{e} = y - \hat{\eta}$ with respect to the basis $\{N_\alpha \mid \alpha = 1, \ldots, n-m\}$ of the orthogonal complement $N^{n-m}(\hat{\theta}_n)$. If $\hat{e}$ is equal zero (see Hamilton and Watts (1985)) then $C_{jk} = \delta_{jk}$ and

$$\text{tr}G = A_1 + A_2 + A_3, \tag{3.7}$$

where

$$A_1 = \sigma^2 \Lambda^{is} \Lambda^{jr} \Lambda^{kt} \Pi_{(kt)(j)} \Pi_{(ir)(s)},$$
$$A_2 = \sigma^2 \Lambda^{is} \Lambda^{jr} \Lambda^{kt} \Pi_{(ik)(j)} \Pi_{(rt)(s)},$$
$$A_3 = \sigma^2 \Lambda^{is} \Lambda^{jr} \Lambda^{kt} \Pi_{(js)(i)} \Pi_{(rt)(k)}.$$

Consequently, Hamilton and Watts (1985) proposed the functionals

$$\tilde{Q}_{HW}(\xi_n) = |\sigma^2 \Lambda|^{1/2} \left\{ 1 + \frac{a^2}{2(m+2)} (A_1 + A_2 + A_3) n^{-1} \right\} \tag{3.8}$$

or

$$Q_{HW}(\xi_n) = |\sigma^2 \Lambda| \left\{ 1 + \frac{a^2}{m+2} (A_1 + A_2 + A_3) n^{-1} \right\} \tag{3.9}$$

as a second order D-optimality criterion. A principal drawback of these criteria is that they are derived under the assumption of a vanishing vector of residuals. In general (3.2) depends on $\hat{e}$ and there is some arbitrariness in the choice $\hat{e} = 0$. As a modification of this approach O'Brien (1992, 1996) proposed to use non-informative priors for the residual vector. We additionally point out that (3.2) is only a correct second-order approximation of the volume of the confidence ellipsoid if the third order derivatives of the response function with respect to the parameters vanish (see Hamilton and Watts (1985), p. 249 and the discussion in Remark 1). Because of these difficulties we will now introduce a class of alternative confidence regions such that the second order approximations of the corresponding volumes are independent of the residuals. In Theorems 1 and 2 we will present a general formula for the second order approximation of the volume of an elliptical confidence region which also refines the formula of Hamilton and Watts (1985), because it includes the third derivatives of the response function.

Definition 1. A statistic $T_n = T_n(c)$ is called u_n-representable at the point $\hat{\theta}_n$ if there exists a vector $c = (c_i)_{i=1}^n \in \mathbb{R}^4$ such that T_n admits a stochastic asymptotic expansion of the form

$$T_n = T_n' + o_p(n^{-1}) \tag{3.10}$$

where

$$T_n' = T_{0n} + T_{1n} n^{-1/2} + T_{2n} n^{-1}$$

and

$$T_{0n} = \Pi_{(i)(j)} u_n^i u_n^j,$$
$$T_{1n} = c_1 \Pi_{(i)(jk)} u_n^i u_n^j u_n^k,$$
$$T_{2n} = (c_2 \Pi_{(ij)(k\ell)} + c_3 \Pi_{(i)(jk\ell)} + c_4 \Lambda^{rs} \Pi_{(r)(ij)} \Pi_{(s)k\ell)}) u_n^i u_n^j u_n^k u_n^\ell.$$

Table 1. c-vectors in the asymptotic expansion of the statistic $T_n^{(r)}$ $(r = 1, \ldots, 4)$ defined in (3.12)–(3.15).

Statistic	$n^{-1/2}$	n^{-1}		
	c_1	c_2	c_3	c_4
$T_n^{(1)}$, Kullback-Leibler	1	$\frac{1}{4}$	$\frac{1}{3}$	0
$T_n^{(2)}$, modified Wald	2	1	1	0
$T_n^{(3)}$, modified Rao	1	0	$\frac{1}{3}$	$\frac{1}{4}$
$T_n^{(4)}$, Pazman	0	$\frac{1}{2}$	0	$-\frac{1}{2}$

The vector $c = (c_i)_{i=1}^4 \in \mathbb{R}^4$ characterizes the u_n-representable statistic T_n.

DEFINITION 2. A u_n-representable statistic is said to have the c-property if its corresponding vector $c = (c_i)_{i=1}^n \in \mathbb{R}^4$ satisfies

$$\frac{c_1^2}{4} = c_2 + c_4. \tag{3.11}$$

The class of u_n-representable statistics with c-property is quite rich. For example, the statistics

$$T_n^{(1)} = \|\hat{\eta} - \eta\|^2, \tag{3.12}$$
$$T_n^{(2)} = \Pi_{(i)(j)}(\theta) u_n^i u_n^j, \tag{3.13}$$
$$T_n^{(3)} = \|P(\hat{\theta}_n)(\hat{\eta} - \eta)\|^2, \tag{3.14}$$
$$T_n^{(4)} = Q_{ij}(\hat{\theta}_n) u_n^i u_n^j \tag{3.15}$$

satisfy the assumptions in Definitions 1 and 2. Here $P(\hat{\theta}_n)$ is the orthogonal projector onto the tangent space $T^m(\hat{\theta}_n)$ with matrix $P^{ab} = \Lambda^{ij} F_i^a F_j^b n^{-1}$ and Q_{ij} is Pazman's matrix with elements

$$Q_{ij} = M_{ij} - \langle (I - P)(\hat{\eta} - \eta), F_{ij} \rangle,$$

(see Pazman (1992)). The validity of this proposition is easy to examine by a Taylor expansion of $T_n^{(r)}$ at the point $\hat{\theta}_n$. The c-vectors of these criteria are given in Table 1. The function $T_n^{(1)}$ is commonly called the Kullback-Leibler statistic since for a Gaussian error distribution the quantity $\sigma^{-2}\|\eta(\theta_1) - \eta(\theta_2)\|^2$ is two times the Kullback-Leibler distance between the Gaussian measures $P_{\theta_1}^n$ and $P_{\theta_2}^n$ (see Borovkov (1998)). The function $T_n^{(2)}$ is the statistic of the modified criterion of Wald (see Rao (1965)). The function $T_n^{(3)}$ is the modification of the statistic of Rao's criterion (see Rao (1965), Grigoriev (1994)). It defines so called the tangent elliptical confidence region for θ (see Pazman (1992)). Finally, we call the function $T_n^{(4)}$ Pazman's statistic since the matrix Q_{ij} is studied extensively in the numerous papers of Pazman (1989, 1990, 1992, 1993).

In Theorem 1 below we demonstrate that u_n-representability of a statistic is sufficient for a second order approximation of the volume of its corresponding confidence region (1.9) to be independent of the residuals. The important criterion of Neyman-Pearson with statistic (3.1) is not u_n-representable. The terms $T_{\nu n}$ in the stochastic expansion of Definition 1 are

$$T_{0n} = (\Pi_{(i)(j)} - b_{ij}) u_n^i u_n^j, \tag{3.16}$$

$$(3.17) \qquad T_{1n} = \left(\Pi_{(i)(jk)} - \frac{1}{3}b_{ijk}\right) u_n^i u_n^j u_n^k,$$

$$(3.18) \qquad T_{2n} = \left(\frac{1}{4}\Pi_{(i)(k\ell)} + \frac{1}{3}\Pi_{(i)(jk\ell)} - \frac{1}{12}b_{ijk\ell}\right) u_n^i u_n^j u_n^k u_n^\ell,$$

where

$$(3.19) \qquad b_{i_1,\dots,i_k} = \langle e, F_{i_1,\dots,i_k}\rangle, \qquad k = 2,3,4$$

and $e = y - \eta(\theta)$ and

$$F_{i_1,\dots,i_k} = \frac{\partial^k}{\partial\theta^{i_1},\dots,\partial\theta^{i_k}}\eta(\theta).$$

The expressions (3.16)–(3.18) contain the vectors $b_{i_1,\dots,i_k}$ and coincide with the terms $T_{\nu n}^{(1)}$ in the expansion of the Kullback-Leibler statistic if $b_{i_1,\dots,i_k} = 0$. In the general case we cannot write $T_n = T_n' + o_p(n^{-1})$ using (3.16)–(3.18) since by the central limit theorem

$$n^{1/2} b_{i_1,\dots,i_k} \xrightarrow{d} \mathcal{N}(0,\sigma^2)$$

as $n \to \infty$ and therefore

$$\begin{aligned} T_n' &= \Pi_{(i)(j)} u_n^i u_n^j + (\Pi_{(i)(jk)} u_n^i - b_{jk}) u_n^j u_n^k n^{-1/2} \\ &\quad + \left[\left(\frac{1}{4}\Pi_{(ij)(k\ell)} + \frac{1}{3}\Pi_{(i)(jk\ell)}\right) u_n^i - \frac{1}{3}b_{jk\ell}\right] u_n^j u_n^k u_n^\ell n^{-1}. \end{aligned}$$

If we introduce the following generalization of the representation (3.16)–(3.18) (which is similar to (3.10))

$$\begin{aligned} (3.20) \quad T_n' &= (\Pi_{(i)(j)} - b_{ij}) u_n^i u_n^j + (c_0 b_{ijk} + c_1 \Pi_{(i)(jk)}) u_n^i u_n^j u_n^k n^{-1/2} \\ &\quad + [c_2 \Pi_{(ij)(k\ell)} + c_3 \Pi_{(i)(jk\ell)} + c_4 \Lambda^{\alpha\beta} \Pi_{(\alpha)(ij)} \Pi_{(\beta)(k\ell)} + c_5 b_{ijk\ell}] u_n^i u_n^j u_n^k u^\ell n^{-1} \end{aligned}$$

and put

$$c_0 = -\frac{1}{3}, \quad c_1 = 1, \quad c_2 = \frac{1}{4}, \quad c_3 = \frac{1}{3}, \quad c_4 = 0, \quad c_5 = -\frac{1}{12}$$

then we obtain the statistic of Neyman-Pearson. If, in addition, $c_0 = 0$, $c_5 = 0$, we obtain the statistic of Kullback-Leibler.

For the formulation of our main result we finally define

$$\begin{aligned} A_4 &= \sigma^2 \Lambda^{is} \Lambda^{jr} \Lambda^{kt} \Pi_{(i)(jr)} \Pi_{(s)(kt)}, \\ A_5 &= \sigma^2 \Lambda^{is} \Lambda^{jr} \Lambda^{kt} \Pi_{(i)(jk)} \Pi_{(s)(rt)}, \\ A_6 &= \sigma^2 \Lambda^{is} \Lambda^{jr} \Pi_{(is)(jr)}, \\ A_7 &= \sigma^2 \Lambda^{is} \Lambda^{jr} \Pi_{(ij)(rs)}, \\ A_8 &= \sigma^2 \Lambda^{is} \Lambda^{jr} \Pi_{(i)(jrs)}. \end{aligned}$$

The quantity

$$(3.21) \qquad R = \Lambda^{ik} R_{ik} = (A_6 - A_4) - (A_7 - A_5)$$

is called the Ricci's curvature or scalar curvature (see Rashevsky (1967)) and the quantity

$$H = A_6 - A_4 \tag{3.22}$$

is called the Efron's curvature (see Efron (1975), Grigoriev and Ivanov (1987*b*)). For $m = \dim \mathbb{E}^m = 1$ and for flat models (see Pazman (1990)) Ricci's curvature R is always zero. In contrast to Ricci's curvature Efron's curvature H is not necessarily zero in the one-dimensional case. By definition Efron's curvature is non-negative, i.e. $H \geq 0$, but Ricci's curvature can have both signs. Efron's and Ricci's curvature of the expectation surface $\mathbb{E}^m$ are related by the inequality (see Grigoriev (1994), Ivanov (1997))

$$H \geq R. \tag{3.23}$$

The quantity

$$B = 3H - 2R \tag{3.24}$$

is called Beale's measure of intrinsic nonlinearity of the expectation surface $\mathbb{E}^m$ (see Beale (1960)). From (3.23) and (3.24) we obtain $B \geq 0$.

Ricci's and Efron's curvature are scalar invariants of the expectation surface $\mathbb{E}^m$ (since they are not changed under a local coordinate transformation) and together with Beale's measure of nonlinearity important objects in the theory of nonlinear regression (see Grigoriev (1994), Ivanov (1997)). In the Appendix we will give the proof of the following two assertions.

THEOREM 1. *If ξ_n is a non-degenerate design (i.e. $|M_{ij}| \neq 0$) and V_n is the volume of a confidence region* (1.9) *with u_n-representable statistic T_n, then*

$$V_n = \gamma |\sigma^2 \Lambda|^{1/2} \left\{ 1 + \frac{a^2}{2(m+2)} K_{2n} n^{-1} \right\} + o_p(n^{-1}), \tag{3.25}$$

where γ is defined in (3.2) *and*

$$K_{2n} = K_{2n}(c) = c_1^2(A_1+A_2+A_3) - c_2(A_6+2A_7) + \left(\frac{c_1^2}{4} - c_4\right)(A_4+2A_5) - 3c_3A_8. \tag{3.26}$$

Moreover, if T_n has the c-property we have

$$K_{2n} = c_1^2(A_1 + A_2 + A_3) - c_2(3H - 2R) - 3c_3A_8. \tag{3.27}$$

In order to present a similar result for statistics which are not necessarily u_n-representable we denote with S^{ij} the inverse of the matrix $\Pi_{(i)(j)} - b_{ij}$ and define the random variables

$$\begin{aligned}
B_1 &= S^{ij} S^{k\alpha} S^{\beta\gamma} \Pi_{(i)(jk)} b_{\alpha\beta\gamma}, \\
B_2 &= S^{i\alpha} S^{jk} S^{\beta\gamma} \Pi_{(i)(jk)} b_{\alpha\beta\gamma}, \\
B_3 &= S^{i\alpha} S^{j\beta} S^{k\gamma} \Pi_{(i)(jk)} b_{\alpha\beta\gamma}, \\
B_4 &= S^{ij} S^{r\alpha} S^{\beta\gamma} b_{ijk} b_{\alpha\beta\gamma}, \\
B_5 &= S^{i\alpha} S^{j\beta} S^{k\gamma} b_{ijk} b_{\alpha\beta\gamma}, \\
B_6 &= S^{ij} S^{k\ell} b_{ijk\ell},
\end{aligned}$$

where $b_{\alpha\beta\gamma}$ and $b_{ijk\ell}$ are defined in (3.19).

THEOREM 2. *If the matrix* S_{ij} *is positive definite at the point* $\hat{\theta}_n$, *then the volume* V_n *of the confidence region* (1.9) *with the statistic* (3.20) *is given by*

$$V_n = \gamma|\sigma^2\Lambda|^{1/2}\left\{1 + \frac{a^2}{2(m+2)}(K_{2n} + K'_{2n})n^{-1}\right\} + r_n, \tag{3.28}$$

where $r_n = o_p(1)$ *and* $r_n = o_p(n^{-1})$ *if* $b_{i_1,\ldots,i_k} = 0$. *Here* K_{2n} *is defined in* (3.26) *and*

$$K'_{2n} = 3c_0c_1\left(B_1 + B_2 + \frac{1}{2}B_3\right) + \frac{3c_0^2}{4}(3B_4 + 2B_5) - 3c_5B_6 \tag{3.29}$$

and everywhere in the quantities A_i *of* (3.26) *the matrix* Λ^{ij} *is replaced by the matrix* S^{ij}.

Remark 1. It is worthwhile to mention that an application of Theorem 2 gives a refined second order approximation for the volume of the confidence region based on the Neyman-Pearson statistic as considered by Hamilton and Watts (1985). To be precise we put $b_{i_1,\ldots,i_k} = 0$, $k = 2,3,4$. Then $K'_{2n} = 0$ and for Hamilton and Watts's case it follows from (3.27) with $c_1 = 1$, $c_2 = \frac{1}{4}$, $c_3 = \frac{1}{3}$ that (3.25) is valid, where

$$K_{2n} = (A_1 + A_2 + A_3) - \frac{1}{4}(3H - 2R) - A_8. \tag{3.30}$$

On the other hand Hamilton and Watts (1985) obtain (3.25) with

$$K_{2n} = A_1 + A_2 + A_3. \tag{3.31}$$

This difference can be explained by the fact that Hamilton and Watts (1985) ignore the third derivatives (see p. 249 of their paper and p. 216 of Seber and Wild (1989)). Moreover, as pointed out in their paper a rigorous derivation of (3.31) is based on the assumption $m = n$ which implies $b_{i_1,\ldots,i_k} = 0$. In this case Beale's measure of intrinsic nonlinearity vanishes and (3.30) and (3.31) coincide, except for the term A_8 which involves the third derivatives of the response function. In this sense the approximation (3.30) derived by Theorem 2 is more accurate.

Remark 2. The tangential component

$$K^T = A_1 + A_2 + A_3 \tag{3.32}$$

is changed by a reparametrisation of the expectation surface $\mathbb{E}^m$ and will be zero for the geodesic parametrisation. The normal component

$$K^N = 3H - 2R \tag{3.33}$$

is invariant with respect to reparametrisation of the expectation surface $\mathbb{E}^m$.

Remark 3. From Table 1 and (3.27) it follows that the most simple expression of the volume V_n in (3.25) is obtained for Pazman's statistic $T_n^{(4)}$. In this case we obtain

$$K_{2n} = -\frac{1}{2}(3H - 2R) \tag{3.34}$$

and as a consequence the resulting second-order optimality criterion depends only on Beale's measure of intrinsic nonlinearity.

4. The general second-order optimality criterion

In accordance with (3.25) we propose a second order criterion of D-optimality

$$Q_C(\xi_n) = |\sigma^2\Lambda| \left\{1 + \frac{a^2}{m+2} K_{2n}(c) n^{-1}\right\}. \tag{4.1}$$

Substituting quantities $A_1 - A_8$ in (2.6), (2.11) and (2.14), we obtain expressions

$$B_{2n} = \frac{1}{4} A_4, \tag{4.2}$$

$$V_{2n} = \left(A_1 + 2A_2 + \frac{1}{2}A_5\right) - R - A_8. \tag{4.3}$$

$$Z_{2n} = \left(A_1 + 2A_2 + \frac{1}{4}A_4 + \frac{1}{2}A_5\right) - R - A_8. \tag{4.4}$$

The comparison of expressions (4.1)–(4.4) now shows that all second-order optimality criteria discussed in this paper are of the form

$$Q(\xi_n) = |\sigma^2\Lambda| \left\{1 + \frac{\kappa^2}{m+2} L_{2n} n^{-1}\right\}, \tag{4.5}$$

where $K_{2n}(c)$ is defined in (3.26), $\kappa^2 \in \{m+2, \chi^2_{1-\alpha}(m)\}$ and $L_{2n} \in \Psi = \{B_{2n}, V_{2n}, Z_{2n}, K_{2n}(c)\}$ is a coefficient depending on the quantity which we wish to minimize. This second order optimiality criterion contains the terms of first and second order with equal weights, but we can write the first and quadratic design criteria in one formula if we consider the function

$$Q(\xi_n) = |\sigma^2\Lambda| \left\{\beta + (1-\beta)\frac{\kappa^2}{m+2} L_{2n} n^{-1}\right\}, \tag{4.6}$$

which puts mass β at the term of first order and mass $1-\beta$ at the term of second order.

On the other hand, we observe that in all cases the factor $L_{2n} \in \Psi$ is a linear combination of quantities $A_1 - A_8$. Therefore, the class Ψ can be extended by some criteria which have not been considered so far. For example, let $ent2$ denote the second-order approximation of the entropy in nonlinear least squares estimation (see Pronzato and Pazman (1994)). In our notation this approximation can be written in the form

$$ent2 = \text{const} + \frac{1}{2}\log|\sigma^2\Lambda| - \frac{1}{2n}(R + A_8 - A_1 - A_2).$$

Hence, using the equivalence $e^x \sim 1 + x$, as $x \to 0$, we obtain the second order entropy criterion which can be motivated by the results in Pronzato and Pazman (1994), i.e.

$$Q_{PP}(\xi_n) = |\sigma^2\Lambda| \cdot \left(1 + E_{2n} n^{-1}\right), \tag{4.7}$$

where

$$E_{2n} = A_1 + A_2 - R - A_8. \tag{4.8}$$

Table 2. Coefficients α_k of the linear combination $L_{2n} = \sum_{k=1}^{8} \alpha_k A_k$ in the second order optimality criterion (4.5)

Criterion		α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
		Factors $L_{2n} \in \Psi_1$							
1. Bias	B_{2n}				$\frac{1}{4}$				
2. Variation	V_{2n}	1	2		1	$-\frac{1}{2}$	-1	1	-1
3. Deviation	Z_{2n}	1	2		$\frac{5}{4}$	$-\frac{1}{2}$	-1	1	-1
4. Entropy	E_{2n}	1	1		1	-1	-1	1	-1
		Factors $L_{2n} \in \Psi_2$							
5. Kullback-Leibler	$K_{2n}^{(1)}$	1	1	1	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{2}$	-1
6. Wald	$K_{2n}^{(2)}$	4	4	4	1	2	-1	-2	-3
7. Rao	$K_{2n}^{(3)}$	1	1	1					-1
8. Pazman	$K_{2n}^{(4)}$				$\frac{1}{2}$	1	$-\frac{1}{2}$	-1	

Table 2 gives the coefficients α_k of the linear combination

$$L_{2n} = \sum_{k=1}^{8} \alpha_k A_k \tag{4.9}$$

for various second order optimality criteria. Note that all criteria of the considered class Ψ have nonnegative coefficients α_1, α_2, α_3, α_4 and nonpositive coefficients α_6, α_8 in the representation (4.9) of L_{2n}. The criteria of the class Ψ_1 obtained by the asymptotic second order approximation of the bias, covariance and entropy satisfy additionally $\alpha_3 = 0$, $\alpha_5 \le 0$ and $\alpha_7 \ge 0$. On the other hand the criteria in the class Ψ_2 (developed by Theorem 1) have nonnegative coefficient α_5 and nonpositive coefficient α_7. Moreover, the factors $K_{2n}^{(1)}$ and $K_{2n}^{(2)}$ corresponding to the Kullback-Leibler and Wald criterion are related by

$$K_{2n}^{(2)} = 4K_{2n}^{(1)} + A_8$$

and are proportional if the third order derivatives are ignored. Similary, if $m = n$ Beale's measure of intrinsic nonlinearity vanishes and the factors $K_{2n}^{(3)}$ and $K_{2n}^{(1)}$ in Rao's and the Kullback-Leibler criterion coincide, i.e.

$$K_{2n}^{(1)} = K_{2n}^{(3)} = A_1 + A_2 + A_3 - A_8 \qquad (m = n).$$

If, additionally, third order derivatives are ignored, we obtain the criterion of Hamilton and Watts (1985).

A different approach to second order optimality criteria for the design of experiment was suggested by O'Brien and Rawlings (1996). Analogous to the bias measure of Box (1971) these authors consider bias-variance ratio, given by

$$m \times (BVR) = \frac{\left|\Delta_n^{ij}\right| - \left|D_n^{ij}\right|}{\left|D_n^{ij}\right|} = D_{kl} b_n^k b_n^l n^{-1}, \tag{4.10}$$

where D_n^{ij} and Δ_n^{ij} are defined in (2.8), (2.12). Since

$$D_{kl} = (D^{ij})^{-1} = \sigma^{-2} \Pi_{(k)(l)} (1 + O(n^{-1})), \qquad n \to \infty,$$

Table 3. γ-coefficients for the general criterion in (5.2) in the one-dimensional case.

Criterion	κ^2	γ-coefficients			$\Sigma\gamma_i$
		γ_1	γ_2	γ_3	
Bias	3	0	0	$\frac{1}{4}$	$\frac{1}{4}$
Variation	3	0	-1	$\frac{7}{2}$	$\frac{5}{2}$
Deviation	3	0	-1	$\frac{15}{4}$	$\frac{11}{4}$
Entropy	3	0	-1	2	1
Kullback-Leibler	$\chi^2_{1-\alpha}(1)$	$-\frac{1}{4}$	$-\frac{1}{3}$	$\frac{5}{4}$	$\frac{2}{3}$
modified Wald	$\chi^2_{1-\alpha}(1)$	-1	-1	5	3
modified Rao	$\chi^2_{1-\alpha}(1)$	0	$-\frac{1}{3}$	1	$\frac{2}{3}$
Pazman	$\chi^2_{1-\alpha}(1)$	$-\frac{1}{2}$	0	$\frac{1}{2}$	0

we draw a conclusion that from asymptotic point of view

$$m \times (BVR) \sim B_{2n}$$

as $n \to \infty$. Therefore, bias-variance-ratio criterion should be considered in the framework of n^{-2}-order asymptotic theory.

5. Examples: A one- and two-parameter model

Example 1. Our first example discusses the criterion (4.6) in detail for the one-dimensional case ($m = 1$). To this end we introduce here the additional notation

$$\tilde{\Pi}_{(k)(\ell)} = \langle F_{\underbrace{1\ldots 1}_{k}}, F_{\underbrace{1\ldots 1}_{\ell}} \rangle.$$

From the definition of $A_1 - A_8$ we obtain

$$\begin{aligned} A_1 &= \cdots = A_5 = \sigma^2 \Lambda^3 \tilde{\Pi}^2_{(1)(2)}, \\ A_6 &= A_7 = \sigma^2 \Lambda^2 \tilde{\Pi}_{(2)(2)}, \\ A_8 &= \sigma^2 \Lambda^2 \tilde{\Pi}_{(1)(3)}, \end{aligned}$$

which yields by (2.6), (2.11), (4.8) and (3.27) to the expressions

$$\begin{aligned} B_{2n} &= \frac{\sigma^2}{4} \Lambda^3 \tilde{\Pi}^2_{(1)(2)}, \\ V_{2n} &= \sigma^2 \left(\frac{7}{2} \Lambda^3 \tilde{\Pi}^2_{(1)(2)} - \Lambda^2 \tilde{\Pi}_{(1)(3)} \right), \\ E_{2n} &= \sigma^2 (2\Lambda^3 \tilde{\Pi}^2_{(1)(2)} - \Lambda^2 \tilde{\Pi}_{(1)(3)}), \\ K_{2n} &= 3\sigma^2 \{ c_1^2 \Lambda^3 \tilde{\Pi}^2_{(1)(2)} - c_2 (\Lambda^2 \tilde{\Pi}_{(2)(2)} - \Lambda^3 \tilde{\Pi}^2_{(1)(2)}) - c_3 \Lambda^2 \tilde{\Pi}_{(1)(3)} \}. \end{aligned}$$

Consequently, the quantity L_{2n} in (4.6) can be written as

$$L_{2n} = 3\sigma^2 \{ \gamma_1 \Lambda^2 \tilde{\Pi}_{(2)(2)} + \gamma_2 \Lambda^2 \tilde{\Pi}_{(1)(3)} + \gamma_3 \Lambda^3 \tilde{\Pi}^2_{(1)(2)} \}. \tag{5.1}$$

From (4.6) and (5.1) it now follows that for $m = 1$ our criteria are given by

$$Q(\xi_n) = \sigma^2 \Lambda \left\{ \beta + (1 - \beta) \frac{\sigma^2 \kappa^2}{n} (\gamma_1 \Lambda^2 \tilde{\Pi}_{(2)(2)} + \gamma_2 \Lambda^2 \tilde{\Pi}_{(1)(3)} + \gamma_3 \Lambda^3 \tilde{\Pi}^2_{(1)(2)}) \right\}, \tag{5.2}$$

where $\kappa^2 \in \{3, \chi^2_{1-\alpha}(1)\}$.

For all criteria, which have been discussed in this paper, the coefficients γ_k in the representation (5.2) are given in Table 3. Note that for the statistics $T_n^{(r)}$ defined in (3.12)–(3.15) we have

$$\gamma_1 = -c_2, \qquad \gamma_2 = -c_3, \qquad \gamma_3 = c_1^2 + c_2$$

and that the (refined) criterion of Hamilton and Watts (1985) coincides with Kullback-Leibler criterion by $b_{i_1,\dots,i_k} = 0$.

In order to illustrate the impact of different second order optimality criteria we consider the one-dimensional exponential regression model

$$\eta(x,\theta) = e^{-\theta x}, \qquad \theta \in (0,\infty), \qquad x \in [0,\infty).$$

Let $\xi_n = \{x_n\}$ denote a one-point design. In this case we obtain for the quantities

$$\tilde{\Pi}_{(k)(\ell)} = (-x_n)^{k+\ell} e^{-2x_n\theta} \tag{5.3}$$

and (5.2) can be rewritten as

$$Q(\xi_n) = \frac{\sigma^2\theta^2}{y_n^2}\{\beta e^{2y_n} + (1-\beta)v_n e^{4y_n}\}, \tag{5.4}$$

where

$$y_n = \theta x_n, v_n = \frac{\kappa^2\sigma^2}{n}(\gamma_1+\gamma_2+\gamma_3). \tag{5.5}$$

It follows from (5.4) that the second order D-optimal design of experiments is concentrated at the point $x_n^* = y_n^*\theta^{-1}$, where y_n^* is a zero of the equation

$$\beta(y-1) + (1-\beta)v_n(2y-1)e^{2y} = 0. \tag{5.6}$$

From (5.6) we obtain the following five different cases:

- $v_n > 0, \beta \in (0,1)$. The second order D-optimal one-point design is concentrated at the point x_n^* in the interval $\left(\frac{1}{2\theta}, \frac{1}{\theta}\right)$;
- $v_n < 0, \beta \in (0,1)$. The second order D-optimal one-point design is concentrated at the point x_n^* in the interval $\left(0, \frac{1}{2\theta}\right)$;
- $\beta = 0, v_n > 0$. The bias optimal one-point design is concentrated at the point $x_n^* = \frac{1}{2\theta}$;
- $\beta = 1$. The first order D-optimal one-point design is concentrated at the point $x_n^* = \frac{1}{\theta}$;
- $v_n = 0$ (Pazman's criterion, $\gamma_1+\gamma_2+\gamma_3 = 0$). In this case the first and second order one-point designs coincide and are concentrated at the point $x_n^* = \frac{1}{\theta}$.

In all cases the point $y_n^* = \theta x_n^*$ is the unique root of (5.6).

Example 2. In general optimal designs with respect to second order optimality criteria have to be calculated numerically. As an illustrative example we consider the two-dimensional intermediate product model investigated in Box and Lucas (1959)

$$\eta(x,\theta) = \frac{\theta_1}{\theta_1-\theta_2}\left(e^{-\theta_2 x} - e^{-\theta_1 x}\right), \qquad \theta_1,\theta_2 \in (0,\infty), \qquad x \in [0,\infty),$$

Table 4. Optimal designs with respect to various second order optimality criteria in the intermediate product model, $n = 2$

Criterion	σ	Optimal design	$\|\sigma^2\Lambda\|$	$Q(\xi_n^*)$
1. Bias	0.025	(1.22917, 6.858761)	0.0000023791	0.0000023938
	0.050	(1.228108, 6.861766)	0.0000380649	0.0000390030
	0.075	(1.226584, 6.866164)	0.0001927050	0.0002033888
	0.100	(1.224727, 6.871301)	0.0006090521	0.0006690692
2. Variation	0.025	(1.221077, 6.784996)	0.0000023797	0.0000024808
	0.050	(1.201620, 6.621477)	0.0000381732	0.0000444754
	0.075	(1.180213, 6.448154)	0.0001944000	0.0002645194
	0.100	(1.161554, 6.301695)	0.0006191777	0.0010068212
3. Deviation	0.025	(1.220843, 6.787037)	0.0000023797,	0.0000024954
	0.050	(1.201602, 6.635525)	0.0000381628	0.0000454210
	0.075	(1.181662, 6.486092)	0.0001941315	0.0002754444
	0.100	(1.165379, 6.369063)	0.0006170278	0.0010692640
4. Entropy	0.025	(1.225660, 6.835870)	0.0000023791	0.0000024296
	0.050	(1.215638, 6.835870)	0.0000380795	0.000024296
	0.075	(1.202435, 6.702388)	0.0001948948	0.0002292756
	0.100	(1.188739, 6.184514)	0.0006111208	0.0008134570
5. Kullback-Leibler	0.025	(1.223305, 6.771860)	0.0000023798	0.0000024970
	0.050	(1.208397, 6.577811)	0.0000382015	0.0000454916
	0.075	(1.191085, 6.368178)	0.0001948948	0.0002757634
	0.100	(1.175087, 6.184514)	0.0006225665	0.0010682144
6. Wald	0.025	(1.204230, 6.534296)	0.0000023906	0.0000028684
	0.050	(1.166023, 6.090162)	0.0000391851	0.0000680358
	0.075	(1.137103, 5.777111)	0.0002046462	0.0005239032
	0.100	(1.115480, 5.554974)	0.0006666019	0.0024339430
7. Rao	0.025	(1.223305, 6.771860)	0.0000023798	0.0000024970
	0.050	(1.208397, 6.577811)	0.0000382015	0.0000454916
	0.075	(1.191085, 6.368178)	0.0001948948	0.0002757634
	0.100	(1.175087, 6.184514)	0.0006225665	0.0010682144
8. Pazman	0.025	(1.229471, 6.857689)	0.0000023791	0.0000023790
	0.050	(1.229471, 6.857689)	0.0000380648	0.0000380648
	0.075	(1.229471, 6.857689)	0.0001927033	0.0001927032
	0.100	(1.229471, 6.857689)	0.0006090376	0.0006090376
9. D-optimal	—	(1.229471, 6.857689)	0.0000023791	—

where θ_1, θ_2 are constants measuring the specific rates of first and second decompositions of a substance in a consecutive chemical reaction. Following Box and Lucas (1959) we chose $\theta = (0.7, 0.2)$ as initial parameter values. In Tables 4 and 5 we show the optimal design with respect to the second order optimality criterion (4.5) for various choices of κ^2 and L_{2n}. Following O'Brien (1992) we used the values $\sigma = 0.0, 0.025, 0.05, 0.075$ and 0.1 for the standard deviation and $n = 2, 3$. The parameter κ^2 was chosen as the 95% quantile of the χ^2-distribution with two degrees of freedom, i.e. $\kappa^2 = 5.991$. For example, for $n = 2$, $L_{2n} = K_{2n}^{(1)}$ and $\sigma = 0.025$ the design $\xi_n^* = (1.2233, 6.7718)$ minimizes the second order criterion (4.5) with $|\sigma^2\Lambda| = 0.2677 \cdot 10^{-5}$ and $Q(\xi_n^*) = 0.2897 \cdot 10^{-5}$.

Table 5. Optimal designs with respect to various second order optimality criteria in the intermediate product model, $n = 3$

Criterion	σ	Optimal design	$\lvert\sigma^2\Lambda\rvert$	$Q(\xi_n^*)$
1. Bias	0.025	(1.2285, 6.8594, 6.8594)	0.0000026764	0.0000026846
	0.050	(1.2256, 6.8645, 6.8645)	0.0000428235	0.0000433439
	0.075	(1.2210, 6.8725, 6.8725)	0.0002168044	0.0002227178
	0.100	(1.2150, 6.8826, 6.8826)	0.0006852874	0.0007184118
2. Variation	0.025	(1.2179, 6.8200, 6.8200)	0.0000026768	0.0000027501
	0.050	(1.1904, 6.7266, 6.7266)	0.0000428983	0.0000474740
	0.075	(1.1593, 6.6144, 6.6144)	0.0002180750	0.0002688969
	0.100	(1.1313, 6.5088, 6.5088)	0.0006933992	0.0009734413
3. Deviation	0.025	(1.2170, 6.8219, 6.8219)	0.0000026769,	0.0000027583
	0.050	(1.1881, 6.7356, 6.7356)	0.0000428999	0.0000479912
	0.075	(1.1564, 6.6360, 6.6360)	0.0002180508	0.0002747867
	0.100	(1.1287, 6.5466, 6.5466)	0.0006929454	0.0010067003
4. Entropy	0.025	(1.2243, 6.8548, 6.8548)	0.0000026765	0.0000027131
	0.050	(1.2106, 6.8466, 6.8466)	0.0000428339	0.0000451618
	0.075	(1.1922, 6.8349, 6.8349)	0.0002170092	0.0002432946
	0.100	(1.1727, 6.8215, 6.8215)	0.0006867847	0.0008332602
5. Kullback-Leibler	0.025	(1.2175, 6.8085, 6.8085)	0.0000026770	0.0000027691
	0.050	(1.1905, 6.6876, 6.6876)	0.0000429186	0.0000486661
	0.075	(1.1616, 6.5433, 6.5433)	0.0002183862	0.0002822518
	0.100	(1.1370, 6.4072, 6.4072)	0.0006952589	0.0010474303
6. Wald	0.025	(1.1863, 6.6508, 6.6508)	0.0000026846	0.0000030615
	0.050	(1.1285, 6.3174, 6.3174)	0.0000436751	0.0000664786
	0.075	(1.0919, 6.0667, 6.0667)	0.0002259096	0.0004784440
	0.100	(1.0704, 5.9002, 5.9002)	0.0007272996	0.0021282692
7. Rao	0.025	(1.2175, 6.8085, 6.8085)	0.0000026770	0.0000027691
	0.050	(1.1905, 6.6876, 6.6876)	0.0000429186	0.0000486661
	0.075	(1.1616, 6.5432, 6.5432)	0.0002183863	0.0002822518
	0.100	(1.1370, 6.4072, 6.4072)	0.0006952589	0.0010474303
8. Pazman	0.025	(1.2295, 6.8577, 6.8577)	0.0000026764	0.0000026764
	0.050	(1.2295, 6.8577, 6.8577)	0.0000428230	0.0000428230
	0.075	(1.2295, 6.8577, 6.8577)	0.0002167912	0.0002167912
	0.100	(1.2295, 6.8577, 6.8577)	0.0006851673	0.0006851673
9. D-optimal	—	(1.2295, 6.8577, 6.8577)	0.0000026764	—

The analysis of the Tables 4 and 5 shows rather small differences between the optimal designs for all considered second order optimality criteria. This can be partially explained by the small variances considered in our study. As expected an increasing variance produces larger deviations compared to the first order D-optimal design. For example, if $n = 3$, $\sigma = 0.175$ the optimal design with respect to second order optimality criterion of Rao is supported at the points 1.0918, 6.1207, and 6.1207, while the criterion of Wald yields a second order D-optimal design supported at the points 1.0429, 5.6591 and 5.6591. Except for the Bias-criterion all optimal designs show a similar behaviour as a function of the variance. The support points are decreasing with an increasing variance. For the Bias-criterion we observe an decreasing left and an increasing right support point. We also note, that for $n = 3$ all optimal designs with respect to the considered second order optimality criteria in Table 5 are two-point designs. Moreover, optimal designs with respect to Pazman's criterion are independent of the noise. This reflects theoretical findings, because a cumbersome calculation shows that in the case $n = m = 2$ Beale's measure of intrinsic nonlinearity vanishes (see Ivanov (1997), pp. 256–265). A similar result was obtained by our numerical calculations in the case $n = 3$.

6. Discussion

In this paper we studied a unified approach to second order optimality criteria in the design of experiments for nonlinear models using a new technique for calculating second order bias, variance and volumina. Most of the commonly used criteria are covered by our approach and new criteria are proposed. Our work generalizes and improves the second order criterion introduced by Hamilton and Watts (1985), connects this approach with a criterion focussing on second-order MSE and unifies some fragmented concepts from tensor analysis such as the work of Ricci, Efron and Beale. It is demonstrated in the intermediate product model that for a sufficiently small variance the total influence of the second order term is rather small. This fact was already noted by Box (1971) in the context of bias and by O'Brien (1992) for the not u_n-representable criterion of Hamilton and Watts (1985).

There are several open and interesting questions which require future research. For example, it follows from Box's inequality (see Box (1971)) that

$$B^T \geq \frac{4}{\sigma^2}\frac{m+2}{m}B_{2n}$$

where

$$B^T = \sum_{i=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m}(A_{i,rr}A_{i,ss} + 2A_{i,rs}A_{i,rs}).$$

Since the contribution of Beale's intrinsic nonlinearity to the total nonlinearity $B + B^T$ is usually less than 10% (see Grigoriev (1994)) it would be interesting to investigate if there exist similar inequalities for other factors of L_{2n}. In this case attention should be given to the minimisation of B^T. A different interesting extension of the proposed methods is the consideration of a nonsymmetric error distribution. In this case the second order approximation of the variance matrix D_n^{ij} requires additional terms which depend on the skewness of the error distribution (see Grigoriev (1994) or Ivanov (1982, 1997)).

Finally, one more direction of possible future research is the application of the Bartlett's adjustment to the radius $\sigma^2\chi^2_{1-\alpha}(m)$ of a confidence region. This correction is used to improve the accuracy of the asymptotic equality (1.6) and it is written as $\rho_n^2 = 1 + \Delta_n n^{-1}$. If we denote $R_n = \sigma^2\chi^2_{1-\alpha}(m)\rho_n^2$ and $C_{1-\alpha}(\Delta_n) = \{\theta \in \Theta \mid T_n(\hat{\theta}_n) \leq R_n\}$, then, by an appropriate choice of Δ_n the accuracy of the confidence region can be improved by the factor n^{-1}, i.e.

$$P\{\theta \in C_{1-\alpha}(\Delta_n)\} = 1 - \alpha + o(n^{-1}), \qquad n \to \infty.$$

The Bartlett's adjustments are calculated for all criteria considered in this paper (see Grigoriev (1994), Ivanov (1997)) and can be used for the construction of alternative second order optimality criteria for the design of experiment, as it is made in the given work.

Acknowledgements

The authors would like to thank the DAAD (Deutscher Akademischer Austauschdienst) for the travel support which made this visit possible and I. Gottschlich who typed this manuscript with considerable technical expertise. We are deeply indepted to V. Tschedkoldin for his computational assistance and to two referees for a number of valuable suggestions which produced a substantial improvement of an earlier version of this paper.

Appendix. The proof of Theorems 1 and 2

Theorem 1 will be proved by calculating the integral

$$V_n = \int_{U_n(x)} du_n, \tag{A.1}$$

where

$$U_n(x) = \left\{ u_n \in \mathbb{R}^m : \frac{1}{\sigma^2} T_n'(u_n) < x \right\}.$$

To this end we use a method introduced by Bardadym and Ivanov (1985). By the substitution of variables $u_n^i \to \sigma \Lambda^{1/2} y^i = \sigma D_r^i y^r$ and the introduction of polar coordinates $y \to (r, \varphi) = (r, \varphi_0, \varphi_1, \ldots, \varphi_{m-2})$ we obtain

$$y^i = r \left(\prod_{\alpha=i-1}^{m-2} \cos\varphi_\alpha \right) \sin\varphi_{i-2}, \quad i = 1, \ldots, m, \tag{A.2}$$

$$\varphi_{-1} = \frac{\pi}{2}, \qquad \varphi_0 \in [0, 2\pi), \qquad \varphi_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right), \qquad i = 1, \ldots, m-2.$$

Throughout this proof we use the notation

$$Q^i(\varphi) = \left(\prod_{\alpha=i-1}^{m-2} \cos\varphi_\alpha \right) \sin\varphi_{i-2}, \qquad Q^{i_1,\ldots,i_k} = \prod_{\alpha=1}^{k} Q^{i_\alpha},$$

$$\bar U(m) = [0, 2\pi) \times \left[-\frac{\pi}{2}, \frac{\pi}{2} \right)^{m-2}$$

$$\bar U_n(x) = \{ r \geq 0, \varphi \in \bar U(m) \mid T_n'(r, \varphi) < x \}.$$

According to (A.3) the function T_n' (used in the set $U_n(x)$) is transformed into the form

$$T_n'(r, \varphi) = r^2 + \sum_{\nu=1}^{2} n^{-\nu/2} r^{\nu+2} \sigma^\nu \pi_{rn}(\varphi),$$

where

$$\pi_{1n}(\varphi) = c_1 \Pi_{(i)(jk)} D_r^i D_s^j D_t^k Q^{rst}(\varphi), \tag{A.3}$$

$$\pi_{2n}(\varphi) = (c_2 \Pi_{(ij)(k\ell)} + c_3 \Pi_{(i)(jk\ell)} + c_4 \Lambda^{\alpha\beta} \Pi_{(\alpha)(ij)} \Pi_{(\beta)(k\ell)}) D_r^i D_s^j D_t^k D_u^\ell Q^{rstu}(\varphi) \tag{A.4}$$

are trigonometric polynomials in the variables $\varphi_0, \ldots, \varphi_{m-2}$. The set $U_n(x)$ is transformed into the set $\bar U_n(x)$. The integrand in the integral (A.1) now takes the form $|\sigma^2 \Lambda|^{1/2} J(r, \varphi)$, where

$$J(r, \varphi) = r^{m-1} \prod_{i=1}^{m-2} (\cos\varphi_i)^i \tag{A.5}$$

is the Jacobian of the polar coordinate transformation. Consequently,

$$V_n = |\sigma^2\Lambda|^{1/2} \int_{\bar{U}(m)} J(1,\varphi)d\varphi \int_{\bar{U}_n(x)} r^{m-1}dr. \tag{A.6}$$

For fixed φ let

$$\Psi_n(\rho,\varphi) : \mathbb{R}^1 \to \mathbb{R}^1$$

denote the inverse of the function $T_n'(r,\varphi)$, i.e. if $\rho = H_n(r) = T_n'(r,\varphi)$, then

$$r = \Psi_n(\rho,\varphi).$$

The function Ψ_n can be found formally in the form of a series in half-integer powers of ρ, where the coefficients are trigonometric polynomials in $\varphi_0, \ldots, \varphi_{m-2}$ and uniformly bounded in $\theta \in \Theta$ and n, i.e.

$$r = \Psi_n(\rho,\varphi) = \rho^{1/2} + \sum_{\nu=1}^{\infty} n^{-\nu/2}\rho^{(\nu+1)/2}\sigma^\nu \Delta_{\nu n}(\varphi). \tag{A.7}$$

The coefficients of $\Delta_{\nu n}$ in this expansion are calculated by the substitution of $T_n'(r,\varphi)$ in the series (A.7) (i.e. instead $\rho = H_n(r)$ we substitute into (A.7) $T_n'(r,\varphi)$ and ignore all coefficients corresponding to the powers of order $r > 1$).

In order to find the quantities Δ_{1n} and Δ_{2n} in (A.7) we denote $\rho = T_n'(r,\varphi)$, $\tau = r\sigma$ and obtain

$$\rho^{1/2} = r(1 + \tau\pi_{1n}n^{-1/2} + \tau^2\pi_{2n}n^{-1})^{1/2}.$$

An application of the binomial expansion gives

$$\rho^{1/2} = r + \frac{1}{2}\sigma\pi_{1n}r^2n^{-1/2} + \left(\frac{1}{2}\pi_{2n} - \frac{1}{8}\pi_{1n}^2\right)\sigma^2r^3n^{-1} + o(n^{-1}). \tag{A.8}$$

A substitution of (A.8) in (A.7) now yields recurrence relations for the coefficients of $\Delta_{\nu n}$ i.e.

$$\frac{1}{2}\pi_{1n} + \Delta_{1n} = 0,$$
$$\frac{1}{2}\pi_{2n} - \frac{1}{8}\pi_{1n}^2 + \Delta_{1n}\pi_{1n} + \Delta_{2n} = 0$$

etc. In particular we find

$$\Delta_{1n} = -\frac{1}{2}\pi_{1n},$$
$$\Delta_{2n} = \frac{5}{8}\pi_{1n}^2 - \frac{1}{2}\pi_{2n}.$$

From arguments linked to the solution of the problem of the inversion of a power series it follows (Fikhtengolts (1966)), that for small values of $t = \sigma\rho^{1/2}n^{-1/2}$ the series in (A.7) is convergent.

A substitution of $r = \Psi_n(\rho,\varphi)$ in (A.6) gives

$$V_n = |\sigma^2\Lambda|^{1/2} \int_{\bar{U}(m)} J(1,\varphi)d\varphi \int_0^x \Psi_n^{m-1}(\rho,\varphi)\frac{\partial\Psi_n(\rho,\varphi)}{\partial\rho}d\rho. \tag{A.9}$$

For the calculation of $\Psi_n^{m-1}(\rho,\varphi)$ and $\partial\Psi_n/\partial\rho$ we use (A.7):

$$\Psi_n^{m-1} = \rho^{(m-1)/2}\Big\{1 + (m-1)b_{1n}\rho^{1/2}n^{-1/2} + \left[(m-1)b_{2n} + \frac{(m-1)(m-2)}{2}b_{1n}^2\right]\rho n^{-1}\Big\} + \cdots,$$

where $b_{\nu n} = \sigma^\nu \Delta_{\nu n}$, $\nu = 1, 2$, and

$$\frac{\partial\Psi_n}{\partial\rho} = \frac{1}{2}\rho^{-1/2} + b_{1n}n^{-1/2} + \frac{3}{2}b_{2n}\rho^{1/2}n^{-1}.$$

We can multiply Ψ_n^{m-1} and $\frac{\partial\Psi_n}{\partial\rho}$ now and obtain

$$\text{(A.10)}\qquad \Psi_n^{m-1}\cdot\frac{\partial\Psi_n}{\partial\rho} = \frac{1}{2}\rho^{(m-2)/2}[1 + (m+1)b_{1n}]\rho^{1/2}n^{-1/2} + \left[(m+2)b_{2n} + \frac{1}{2}(m-1)(m+2)b_{1n}^2\right]\rho n^{-1}.$$

Substituting (A.10) in (A.9) yields

$$\text{(A.11)}\qquad V_n = \frac{|\sigma^2\Lambda|^{1/2}}{2}\int_0^x \rho^{(m-2)/2}d\rho\int_{\bar U(m)} J(1,\varphi) \cdot\left[(m+2)b_{2n} + \frac{(m-1)(m+2)}{2}b_{1n}^2\right]\frac{\rho}{n}d\varphi$$
$$= \frac{|\sigma^2\Lambda|^{1/2}}{2}\int_0^x \rho^{(m-2)/2} \cdot\left\{I_0 + (m+1)I_1\frac{\rho^{1/2}}{n^{1/2}} + \left[(m+2)I_2 + \frac{(m-1)(m+2)}{2}I_3\right]\frac{\rho}{n}\right\}d\rho,$$

where

$$I_0 = \int_{\bar U(m)} J(1,\varphi)d\varphi = (\text{volume of the unit ball}) = \frac{2\pi^{m/2}}{\Gamma\left(\frac{m}{2}\right)},$$
$$I_1 = \int_{\bar U(m)} J(1,\varphi)b_{1n}(\varphi)d\varphi = -\frac{1}{2}\sigma\int_{\bar U(m)} J(1,\varphi)\pi_{1n}(\varphi)d\varphi = 0,$$
$$I_2 = \int_{\bar U(m)} J(1,\varphi)b_{2n}(\varphi)d\varphi = \int_{\bar U(m)} J(1,\varphi)\sigma^2\left(\frac{5}{8}\pi_{1n}^2 - \frac{1}{2}\pi_{2n}\right)d\varphi = \frac{5}{8}\sigma^2\bar I_2 - \frac{1}{2}\sigma^2\bar I_3,$$
$$I_3 = \int_{\bar U(m)} J(1,\varphi)b_{1n}^2(\varphi)d\varphi = \int_{\bar U(m)} J(1,\varphi)\frac{\sigma^2}{4}\pi_{1n}^2(\varphi)d\varphi = \frac{\sigma^2}{4}\bar I_2.$$

and the last two equalities define $\bar I_2$ and $\bar I_3$. Substituting I_2 and I_3 in (A.11) we obtain

$$\text{(A.12)}\quad V_n = \frac{|\sigma^2\Lambda|^{1/2}}{2}\int_0^x \rho^{(m-2)/m}\left\{I_0 + \sigma^2\left[\frac{(m+2)(m+4)}{8}\bar I_2 - \frac{m+2}{2}\bar I_3\right]\rho n^{-1}\right\}d\rho$$
$$= \frac{|\sigma^2\Lambda|^{1/2}}{2}\left\{\frac{2}{m}I_0\cdot x^{m/2} + \sigma^2\left[\frac{(m+4)}{4}\bar I_2 - \bar I_3\right]x^{(m+2)/2}n^{-1}\right\}.$$

Finally, the integrals

$$\bar{I}_2 = \int_{\bar{U}(m)} J(1,\varphi)\pi_{1n}^2(\varphi)d\varphi, \tag{A.13}$$

$$\bar{I}_3 = \int_{\bar{U}(m)} J(1,\varphi)\pi_{2n}(\varphi)d\varphi \tag{A.14}$$

have to be calculated. Substituting (A.3) in (A.14) and (A.4) in (A.15) we obtain

$$\begin{aligned} \bar{I}_2 = {} & c_1^2 D_{j_1}^{i_1} D_{j_2}^{i_2} D_{j_3}^{i_3} D_{j_4}^{i_4} D_{j_5}^{i_5} D_{j_6}^{i_6} \Pi_{(i_1)(i_2 i_3)} \Pi_{(i_4)(i_5 i_6)} \\ & \cdot \int_{\bar{U}(m)} J(1,\varphi) \prod_{\alpha=1}^{6} Q^{j_\alpha}(\varphi)d\varphi, \end{aligned} \tag{A.15}$$

$$\begin{aligned} \bar{I}_3 = {} & \left\{ c_2 \Pi_{(i_1 i_2)(i_3 i_4)} + c_3 \Pi_{(i_1)(i_2 i_3 i_4)} + c_4 \Lambda^{\alpha\beta} \Pi_{(\alpha)(i_1 i_2)} \Pi_{(\beta)(i_3 i_4)} \right\} \\ & \times D_{j_1}^{i_1} D_{j_2}^{i_2} D_{j_3}^{i_3} D_{j_4}^{i_4} \int_{\bar{U}(m)} I(1,\varphi) \prod_{\alpha=1}^{4} Q^{j_\alpha}(\varphi)d\varphi. \end{aligned} \tag{A.16}$$

The integrals in (A.16) and (A.17) are given by

$$\begin{aligned} I^{i_1 \ldots i_{2\nu}} &= \int_{\bar{U}(m)} J(1,\varphi) \prod_{\alpha=1}^{2\nu} Q^{i_\alpha}(\varphi)d\varphi \\ &= \begin{cases} 0, & \text{if at least one index appear with odd multiplicity,} \\ \dfrac{\pi^{m/2}}{2^{\nu-1}\Gamma\left(\dfrac{m+2\nu}{2}\right)}, & \text{else.} \end{cases} \end{aligned}$$

Consequently,

$$\bar{I}_2 = \frac{2\pi^{m/2}}{\sigma^2 m(m+2)(m+4)\Gamma\left(\frac{m}{2}\right)} \cdot K_1, \tag{A.17}$$

where

$$K_1 = c_1^2(4A_1 + 4A_2 + 4A_3 + A_4 + 2A_5)$$

and

$$\bar{I}_3 = \frac{2\pi^{m/2}}{\sigma^2 m(m+2)\Gamma\left(\frac{m}{2}\right)} \cdot K_2, \tag{A.18}$$

where

$$K_2 = c_2(A_6 + 2A_7) + c_3(3A_8) + c_4(A_4 + 2A_5).$$

Substituting (A.17) and (A.18) in (A.12) and observing that $x = a^2 = \chi_{1-\alpha}^2(m)$, and $I_0 = 2\pi^{m/2}/\Gamma(\frac{m}{2})$, we obtain

$$V_n = \frac{(a\sqrt{\pi})^m}{\Gamma\left(\dfrac{m+2}{2}\right)} \cdot |\sigma^2\Lambda|^{1/2} \cdot \left\{1 + \frac{a^2}{2(m+2)} K_{2n}\right\},$$

where

$$K_{2n} = \frac{1}{4}K_1 - K_2$$
$$= c_1^2(A_1 + A_2 + A_3) - c_2(A_6 + 2A_7) + \left(\frac{c_1^2}{4} - c_4\right)(A_4 + 2A_5) - \frac{3c_3}{2}A_8,$$

which proves the first part of Theorem 1. Finally, the c-property (3.10) yields the result (3.27). For a proof of Theorem 2 we must use (3.20) instead of (3.10). In this case we have instead K_1 and K_2

$$K_1' = K_1 + 6c_0c_1(2B_1 + 2B_2 + B_3) + 3c_0^2(B_4 + 2B_5),$$
$$K_2' = K_2 + 3c_5B_6$$

and obtain

$$\frac{1}{4}K_1' - K_2' = K_{2n} + K_{2n}',$$

where K_{2n}' is defined in (3.30).

References

Bardadym, T. A. and Ivanov, A. V. (1985). An asymptotic expansion related to an empirical regression function, *Theory of Mathematical Statistics*, **30**, 7–13.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of non-linearity, *J. Roy. Statist. Soc. B*, **42**, 1–25.

Beale, E. M. L. (1960). Confidence regions in non-linearity estimation (with discussion), *J. Roy. Statist. Soc. B*, **22**, 41–88.

Borovkov, A. A. (1998). *Mathematical Statistics*, Gordon and Breach, London.

Box, G. E. P. and Lucas, H. (1959). Design of experiments in nonlinear situations, *Biometrika*, **41**, 77–90.

Box, M. J. (1971). Bias in nonlinear estimation (with discussion), *J. Roy. Statist. Soc. B*, **32**, 171–201.

Chernoff, H. (1953). Locally optimal designs for estimating parameters, *Ann. Math. Statist*, **24**, 586–602.

Clarke, G. P. Y. (1980). Moments of the least squares estimators in a nonlinear regression model, *J. Roy. Statist. Soc. B*, **42**, 227–237.

Efron, B. (1975). Defining the curvature of a statistical problem with applications to second order efficiency (with discussion), *Ann. Statist.*, **3**, 1199–1242.

Fikhtengolts, G. M. (1966). *A Course of Differential and Integral Calculus*, Vol. 2, Nauka, Moscow (in Russian).

Ford, I., Torsney, B. and Wu, C. F. Y. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems, *J. Roy. Statist. Soc. B*, **54**, 569–583.

Grigoriev, Yu. D. (1993). The asymptotic second order criteria of optimality for designs of experiments, *Electronic Technic, Ser. 7*, 1 (176), 22–25 (in Russian).

Grigoriev, Yu. D. (1994). Development and investigation of algorithms in the analysis of nonlinear regression models, Doctor of Science Thesis, State Technical University, Novosibirsk.

Grigoriev, Yu. D. and Ivanov, A. V. (1987*a*). Asymptotic expansions in nonlinear regression analysis, *Zavodckaja Laboratoria*, **53**, 48–51. (in Russian)

Grigoriev, Yu. D. and Ivanov, A. V. (1987*b*). On measures of non-linearity of regression models, *Zavodckaja Laboratoria*, **53**, 57–61 (in Russian).

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

Ivanov, A. V. (1982). An asymptotic expansion of moments of least squares estimator for a vector parameter of nonlinear regression, *Ukrainian Math. J.*, **34**, 134–139.

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

O'Brien, T. E. (1992). A note on quadratic designs for nonlinear regression models, *Biometrika*, **79**, 847–849.

O'Brien, T. E. (1996). Robust design strategies for nonlinear regression models, *Versuchsplanung in der Industrie*, (eds. H. Toutenberg, and R. Gössl), 41–52, Prentice Hall, München.

O'Brien, T. E. and Rawlings J. O. (1996). A nonsequential design procedure for parameter estimation and model discrimination in nonlinear regression models, *J. Stat. Plann. Inference*, **55**, 77–93.

Pazman, A. (1989). On information matrices in nonlinear experimental design, *J. Stat. Plann. Inference*, **21**, 253–263.

Pazman, A. (1990). Small-sample distribution of properties of nonlinear regression estimators (a geometric approach), *Statistics*, **21**, 323–367.

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

Pazman, A. (1993). Nonlinear Statistical Models, Kluwer, Dordrecht.

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

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Rashevsky, P. C. (1967). *Riemannian Geometry and Tensor Analysis*, Nauka, Moscow (in Russian).

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.