Discussiones Mathematicae
Probability and Statistics 31 (2011) 59–70
doi:10.7151/dmps.1136

DMPS Page

ON THE OPTIMAL CONTINUOUS EXPERIMENTAL DESIGN PROBLEM

Christos P. Kitsos

Technological Educational Institute of Athens
Department of Mathematics
12210 Egaleo, Athens, Greece

e-mail: xkitsos@teiath.gr

Abstract

The target of this paper is to provide a compact review of the Optimal Experimental Design, the continuous case. Therefore we are referring to the general nonlinear problem in comparison to the linear one.

Keywords: nonlinear experimental design, static, sequential design.

2010 Mathematics Subject Classification: 62K05, 62L05, 62F25.

1. Introduction

The first statistical treatment under a nonlinear function came from the pioneer of Statistics, R.A. Fisher. In Rothamsted Experimental Station, he came across with the dilution series problem, Fisher (1922). Since then the objectives of the experimenter can be:

(i) The response surface problem introduced by Box and Draper (1959).

(ii) The discrimination problem between rival models reviewed early by Hill (1976), see also Atkinson & Donev (1992).

(iii) In some sense, all or a subset of the parameters to be estimated as well as possible. This is the optimal design problem originated by Smith (1918).

The above (ii) and (iii) mentioned objectives are common to linear and nonlinear experimental designs (LED and NLED) i.e when the assumed suitable model is linear or nonlinear, with respect to its parameters, while the RS is tackled for the LED.

Ford *et al.* (1989) review the NLED problem, Kitsos (1989) worked with the sequential procedures, while Fedorov (1972) provided all the useful results on the linear design. We shall not refer to the Bayesian approach to this paper, although certainly needs particular interest, and perhaps another paper. The nonlinear experimental design problem finds applications in many fields: as a regression problem (continuous case) in chemical kinetics (chemistry, biology), or as a binary model (discrete case) in testing explosives, biological assays etc. We are restricted in the continuous case, emphasizing that the discrete case in NLED has not such a theoretical insight as in the LED problems (Number Theory, Gometrical implementations etc). The emphasis will be on the target (3).

Let $U \subseteq \mathbb{R}^k$ be the space in which the covariates $u = (u_1, u_2, \ldots, u_k)$ take their values and is known as experimental region. The parameter space $\Theta \subseteq \mathbb{R}^p$ is the set where the parameters $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_p)$ take their values. Let Ξ be the family of measures ξ such that $\xi(u) \geq 0$ $u \in U$ and $\int \xi(du) = 1$, the design measure, while the pair (U, ξ) will be called the design. The support of the design Supp(ξ), say, is the set of points u for which $\xi(u) > 0$. This is only a theoretical consideration, as it might be a design point optimal, but with zero design measure at this point. Practically speaking the design measure acts as the proportion of observations devoted to the optimal design points.

When the response y is supposed to take any value in Ψ we also suppose that a regression model (in general nonlinear) exists consisting of the deterministic portion $f(u, \vartheta)$ and the stochastic portion, e known as error, linked through the (continuous) regression formulation $y_i = f(u_i, \vartheta) + e_i$. Usually we denote $\eta = E(y)$.

Example 1. The Oxidation of Benzene model is, among the chemical reaction models, one with the most necessary needed information. Includes 4 input variables and 4 parameters:

$$\eta = \frac{\vartheta_1 \exp(-\theta_3 u_3)\, \theta_2 \exp(-\theta_4 u_3)\, u_1 u_2}{\theta_1 \exp(-\theta_3 u_3)\, u_1 + u_4 \vartheta_2 \exp(-\theta_4 u_3)\, u_2},$$

with $u = (u_1, u_2, u_3, u_4) \in U = \Delta_1 \times \Delta_2 \times \Delta_3 \times \{5.75\} \subseteq \mathbb{R}^4$, $\Theta \subseteq \mathbb{R}^4$, $\Delta_1 = [10^{-3}, 16{\cdot}10^{-3}]$, $\Delta_2 = [10^{-3}, 4{\cdot}10^{-3}]$, $\Delta_3 = [623, 673]$ and $\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4$ parameters arising in Arrhenius law, where n: the initial reaction rate, u_1: the concentration of oxygen, u_2: the concentration of benzene, $u_3 = \frac{1}{T} - 0.0015428$, T: the absolute temperature of the reaction and u_4: the observed stoichiometric number.

If we assume that $f(u, \vartheta) = \vartheta^{\mathrm{T}} g(u)$ with g being a (vector) continuous function of u, then the nonlinear problem is reduced to the so called linear problem, when the Least Square Estimates can always obtained under certain conditions, Wu (1981), in both linear and nonlinear case.

In principle, Fishers information measures is strongly related with the involved uncertainty to the physical phenomenon under investigation. In NLEDr problems the variance depends on the unknown parameter we want to estimate, i.e., $\sigma^2 = \sigma^2(u, \vartheta)$, while in LED in the linear case it is assumed independent of the parameter vector ϑ. In practice it may or may not be possible to assume that is "known" provided a guess, or knowledge from a previous experiment. In principle, this is the main difference between the linear and the non-linear case. Let $\nabla\eta$ denote the vector $\nabla\eta = (\frac{\partial\eta}{\partial\vartheta_j})^{\mathrm{T}}$, $j = 1, 2, \ldots, p$. Then for the exponential family of models Fisher's information matrix is defined to be

$$\mathrm{I}(\vartheta, u) = \sigma^{-2}(\nabla\eta)(\nabla\eta)^{\mathrm{T}}. \tag{1}$$

Moreover in many of the nonlinear problems the covariate u and the parameter ϑ appear together linearly in the form $\vartheta^{\mathrm{T}} u$. Thus, the nonlinear model is "*intrinsic linear*" $\eta = \eta(\vartheta^{\mathrm{T}} u)$ as far as the parameter is concerned, then Fishers information matrix can be evaluated proportional to the matrix produced by the input vector u.

The concept of the *average-per-observation information matrix* will play an important role in our scenario for the nonlinear experiment design problem. It is defined for the continuous case

$$\mathrm{M}(\vartheta, \xi) = \int_U \mathrm{I}(\vartheta, u)\xi(du). \tag{2}$$

The idea of Caratheodory's Theorem so essential for the linear experiment design, can be used for the average information matrix in nonlinear problems, Titterington (1980). Now, suppose the matrix $\mathrm{M} = \mathrm{M}(\vartheta, \xi)$ is partitioned in the form $M = (M_{ij})$, with $\mathrm{M}_{11} \in \mathbb{R}^{s\times s}$, $\mathrm{M}_{12} \in \mathbb{R}^{s\times(p-s)}$,

$\mathrm{M}_{22} \in \mathbb{R}^{(p-s)\times(p-s)}$, $1 \leq s < p$. Then we define the matrix: $\mathrm{M}_s = \mathrm{M}_s(\vartheta, \epsilon) = \mathrm{M}_{11} - \mathrm{M}_{12}\mathrm{M}_{22}^{-}\mathrm{M}_{12}^{\mathrm{T}}$, with M_{22}^{-} being a generalized inverse of M_{22}. This partition is helpful when our interest lies in estimating the leading $s < p$ parameters in the vector ϑ. Moreover in NLED, a useful approximation to the covariance matrix is

$$\mathrm{M}(\vartheta, \xi) = \frac{1}{n}(\nabla\eta)^{\mathrm{T}}\nabla\eta = \frac{\sigma^2}{n}\mathrm{C}^{-1}(\hat{\vartheta}, \xi).$$

In both cases, linear and nonlinear, the uncertainty is measured with the entropy which equals to $\log\det \mathrm{C}$.

The idea of the design matrix is essential in LED, while in nonlinear models we can extend the definition approximately through the partial derivatives of ϑ with ϑ taking its "true" value, ϑ_t. We define the $n \times p$ matrix

$$\mathrm{X} = (x_{ij}) \in \mathbb{R}^{n\times p} \text{ with } x_{ij} = \left.\frac{\partial f(u_i, \vartheta)}{\partial \vartheta_j}\right|_{\vartheta=\vartheta_t}. \tag{3}$$

Following the LED pattern an approximation to the covariance matrix can be defined as $\mathrm{C} \cong [\mathrm{X}^{\mathrm{T}}(\vartheta_t)\mathrm{X}(\vartheta_t)]^{-1}\sigma^2$ in NLED problems as well.

2. Locally Optimal Designs

It is the ϑ-dependence which leads to the term "locally optimal": the optimal design depends on the true value of ϑ, therefore it might be optimal "locally".

Suppose we wish to estimate a set of linear combinations of the parameter vector $\vartheta = (\vartheta_1, \ldots, \vartheta_p)$. This might lead to an estimation of the vector ϑ itself, some linear combinations of the p components of ϑ or to $s \leq p$ components. Let $\mathrm{Q} \in \mathbb{R}^{s\times p}$, $1 \leq s \leq p$, be the matrix of the known coefficients defining and the quantities of interest are $\mathrm{Q}\vartheta$. If $\operatorname{rank}\mathrm{Q} = p$, when $s = p$, the matrix Q is nonsingular. If $s < p$ we suppose that $\operatorname{rank}\mathrm{Q} = s$. Then we can define the following operator J_{Q} applied to M, through the above matrix Q:

$$J_{\mathrm{Q}} = \mathrm{Q}\mathrm{M}^{-}(\vartheta, \xi)\mathrm{Q}^{\mathrm{T}}, \tag{4}$$

with M^{-} the generalized inverse of M and $\mathrm{Q}^{\mathrm{T}} \in \mathbb{R}^{p\times s}$.

Given the above notation we need a real valued function, ω say, applied to J_{Q} to be used as an optimality criterion. We choose ω to be a convex decreasing function on the set of nonnegative definite matrices $\mathbb{R}_{+}^{s\times s}$ say.

Definition. The design measure ξ^* is called ω-optimal if and only if (iff):

$$\omega\{J_{\mathrm{Q}}\left(\mathrm{M}(\vartheta,\xi^*)\right)\} = \min\{\omega\left(\mathrm{QM}^{-}(\vartheta,\xi)\mathrm{Q}^{\mathrm{T}}\right), \xi\in\Xi\}. \tag{5}$$

We now examine the special cases of ω and Q which lead us to traditional definitions, adopted from the linear case. The special cases we shall consider for ω are

$$\omega(\cdot) = \begin{cases} \omega_1(\cdot) = \log\det\left(\mathrm{QM}^{-}\mathrm{Q}^{\mathrm{T}}\right), & \mathrm{D}-\text{optimal},\\ \omega_2(\cdot) = \mathrm{tr}\left(\mathrm{QM}^{-}\mathrm{Q}^{\mathrm{T}}\right), & \mathrm{A}-\text{optimal},\\ \omega_3(\cdot) = \max\text{eigenval}\left(\mathrm{QM}^{-}\mathrm{Q}^{\mathrm{T}}\right), & \mathrm{E}-\text{optimal},\\ \omega_4(\cdot) = \sup\mathrm{tr}\left(\mathrm{I}(\vartheta,u)\mathrm{QM}^{-}\mathrm{Q}^{\mathrm{T}}\right), & \mathrm{G}-\text{optimal}, \end{cases} \tag{6}$$

The most often considered cases for Q are

$$\mathrm{Q} = \begin{cases} \mathrm{A}\in\mathbb{R}^{p\times p},\ \mathrm{rank\ A}=p, & \mathrm{D_A}-,\ \mathrm{G_A}-,\ \mathrm{E_A}-\ ,\mathrm{A_A}-\text{optimality},\\ \mathrm{I}\in\mathbb{R}^{p\times p}\setminus\mathbb{I}_p, & \mathrm{D}-,\ \mathrm{G}-,\ \mathrm{E}-,\ \mathrm{A}-\text{optimality},\\ \left[\mathrm{I}_s:\mathbb{O}_{s\times(p-s)}\right], & \mathrm{I}_s\in\mathbb{R}^{p\times p}, \mathrm{D}_s,\ \mathrm{G}_s,\ \mathrm{E}_s,\ \mathrm{A}_s-\text{optimality},\\ c\in\mathbb{R}^{p\times 1}, & \mathrm{c}-\text{optimality}. \end{cases}$$

Actually, under our notation, the traditionally described as φ criterion, Kiefer and Wolfowitz (1960), is $\varphi=\omega\circ J_{\mathrm{I}}$, where $\circ$ denotes the composition of two functions. Corresponding to ω_i, $i=1,2,3,4$ we obtain $\varphi_i=\omega_i\circ J_{\mathrm{I}}$, $i=1,2,3,4$. The optimality criteria $D(\vartheta)$ and $G(\vartheta)$ were introduced by White (1973) who also extended Kiefer and Wolfowitz′s (1960) equivalence theorem between D– and G–optimality criteria. Moreover D–optimality remains invariant to linear transformations in LED and NLED, an essential difference with the other optimality criteria. The equivalence theorem holds for a subset of parameters as well, Karlin and Studden (1966) for LED, White (1973) for NLED, when $\mathrm{A}=[\mathrm{I}_s:\mathbb{O}_{s\times(p-s)}]$ and $J_{\mathrm{A}}(\mathrm{M})=\mathrm{M}_s$. The essential difference that in NLED the stated equivalence theorems hold with the parameter taking its true value!

A(ϑ)–optimality minimizes the sum of approximate variances of the parameter estimates, as in the linear case (Titterington (1980)), while E(ϑ)–optimality seeks to minimize the variance of the worst-estimated linear combination $c^{\mathrm{T}}\vartheta$, with $c^{\mathrm{T}}c = 1$.

Hill (1980) defined when model is partially and Khuri (1984) provided, a sufficient condition for $\mathrm{D}_s(\vartheta)$–optimality. The Michaelis-Menten model is a typical example, Kitsos (1986) of a partially nonlinear model. For the model $f(u, \vartheta) = l(u, \beta_1) + m(u, \beta_2)$ where $\beta_1 = (\vartheta_1, \vartheta_2, \ldots, \vartheta_s)$, $\beta_2 = (\vartheta_{s+1}, \vartheta_{s+2}, \ldots, \vartheta_p)$, i.e., $\vartheta = (\beta_1, \beta_2)$ while $l(u, \beta_1) = \vartheta_0 + \vartheta_1 u + \cdots \vartheta_s u_s$ and $m(u, \beta_2)$ any nonlinear function, the D(ϑ)–optimal and the $\mathrm{D}_s(\vartheta)$–optimal designs depends on β_2. It is in fact convenient, if we have a design criterion, which will remain invariant under certain transformations of the design space. As this is not true for c–optimality the "canonical form" was introduced for it, Kitsos (1986), with application to simple logit model and proved that we can have a group of affine transformations to obtain the canonical form for the p–variable logit, see Kitsos (2011) for details.

For the linear model the geometry was built up not on the design space U, but on its image through g, $U_0 = g(U)$ say, known as induced design space. Furthermore, for this transformation the information M(ξ) is preserved, expressed in terms of the family of design measures $\Xi_0 = \{\xi_0 = \xi g^{-1},\ \ \xi \in \Xi\}$.

Let U_0 be a compact set which spans $\mathbb{R}^p$ (which spans the leading s-dimensional coordinate subspace). Then:

(i) the D–optimal design problem for U_0 is the dual of the minimal ellipsoid problem for U_0 (Sibson, 1972),

(ii) the D_s–optimal design problem is the dual of the thinnest cylinder problem (Silvey & Titterington, 1973).

In both cases the two problems share a common extreme value. When ϑ takes its true value all the geometric aspects covered by Titterington (1980) can be applied to the nonlinear case. The geometry of c–optimality is covered by Elfving (1952). This excellent theorem remain invariant to time, and has been used extensively to the (non-linear problem, eventually of) calibration problem, Kitsos and Kolovos (2010).

Example 2. The geometrical insight of the NLED can be clarified with he cosinor model, $y(t) = n(t, \vartheta) + \epsilon$ with $n(t, \vartheta) = \vartheta_0 + \vartheta_1 \cos(\omega t + \vartheta_2)$, where

$y(t)$: the response at time $t \in [0,1]$, i.e., the biological rhythm we want to study, ϑ_0: the mesor, ϑ_1: the amplitude, ϑ_2: the acrophase, ω: the angular frequency, and ϵ: the error term with the usual normality assumption.

Interest is, mainly, in efficient estimation of the relative stability, i.e., the ratio $g = g(\vartheta_0, \vartheta_1) = \vartheta_1/\vartheta_0 < 1$ is to be estimated as well as possible. For the 4-point D–optimal design the approximate variance, is $V_4 = (\sigma^2/(n\vartheta_0^2))[(\vartheta_1/\vartheta_0)^2 + 2]$, while for the 2-point c–optimal the variance is $V_2 = \vartheta_0^{-2}(\sigma^2/n)$. The corresponding design measure and the optimal design points can be also evaluated. This model, was suitably adopted for the Robotic Total Stations (RTS) when operated in tracking mode, Zarikas *et al.* (2010). Analysis of the results reveals a reduction (of the order of 3% to 10%) in the variance of the effective amplitude if the optimal design method is used.

3. Estimation: Static and Sequential Design

In the linear case, with a typical example the p-term polynomial regression, the D–optimal design has a tendency to use as optimal design points the "end" points of the design space, among others, when $p \geq 2$. Under some considerations, Fedorov (1972, Th. 2.2.3), the design points for D–optimality can be defined as roots of a hypergeometric function (Legendre, Jacobi, Laguerre, Hermite) and the design then allocates measure $1/p$ at these points.

This ϑ–dependence requires the development of alternative strategies for the construction of experimental designs in practice. Two procedures are proposed:

(i) Choose (the optimal) design points. Perform the experiment once at these points. (static design).

(ii) Choose (the possible optimal) initial design points. Perform the experiment at these points and estimate the parameters. Re-assess the (optimal) design points (using the estimates of the parameters). Perform the experiment at these new points and get new estimates. Continue the procedure until a predefined stopping rule is satisfied (sequential design).

An attempt to avoid ϑ–dependence has been made through S–optimality, "averaging" over all possible values of the parameter, was first attempted by Lauter (1974). A discrimination approach was proposed by Stone and Morris (1985), based on two alternative criteria for the static problem. Both criteria accepted a criticism by Ford *et al.* (1989). Another alternative method of avoiding the ϑ–dependence problem is the maximin design approach. That is we solve

$$\max \min\{\Phi(\mathrm{M}(\vartheta,\xi))\square\Phi(\mathrm{M}(\vartheta,\xi^*))\}, \quad \xi\in\Xi, \quad \vartheta\in\Theta, \tag{7}$$

with $\square = /$ or $-$, and $\xi^* = \xi^*(\vartheta)$ the locally optimum design for ϑ. The maxmin efficiency criterion for the calibration problem is equivalent to D–optimality, Kitsos (1989).

Adopting the sequential procedure we choose an initial design, using prior knowledge on ϑ and get an estimate of the parameters. This estimate is useful as an initial guess to redesign, re-estimate and so on. The question is "how we choose the next design point"? The answer is: choose the next design point that, which minimizes the estimators generalized variance. That is a D(ϑ)–optimality criterion is used for choosing the next design point. For the particular case of D(ϑ)–optimality, when the initial design is D–optimal the limiting is also, providing the stochastic approximation approach has been adopted, see Kitsos (1989) for details. One main virtue of the linear theory is the dichotomous convergence theorem of Wu and Wynn (1978) for any criterion function Φ, and for the sequence M(ξ_n).

In NLED problems we try to apply the LED theory to construct the confidence intervals. But the nonlinearity is essential and various attempts have been developed to overpass it. In the nonlinear case confidence regions sometimes appear to have "banana-shapes".

Beale (1960), in his pioneering paper, suggested a dimensionless empirical measure of nonlinearity $\Lambda^* = \Lambda^*(s^2)$ and a theoretical measure of nonlinearity, $\Lambda = \Lambda(\sigma^2)$ say, while the minimum value of the theoretical measure of nonlinearity, Λ_0 say, was named the *intrinsic nonlinearity*, a sort of curvature of the solution locus, of the assumed correct model. Bates and Watts (1980) using ideas from differential geometry proved that Λ_0 is one quarter of the mean square intrinsic curvature. Moreover they proved that by replicating the design r times the curvature at any point in any direction is reduced by a factor $1/\sqrt{r}$. The measure of nonlinearity is used to adjust the confidence region from the usual form, using the following form, introducing

an extra parameter λ,

$$(\vartheta - \hat{\vartheta})^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})(\vartheta - \hat{\vartheta}) \leq \lambda p s^2 F(\alpha; p, n-p),$$

with X as in (3) and $F(\alpha; p, n-p)$ is as usual the $100(1-\alpha)\%$ of the F distribution with p and $n-p$ degrees of freedom. The new parameter λ is given by

$$(8)\ \lambda = \begin{cases} 1, & \text{linearization without Beale}'\text{s assumption,} \\ 1 + \frac{n}{n-1}\Lambda_0, & \text{if } p = 1, \\ 1 + \frac{n(p+2)}{(n-p)p}\Lambda_0, & \text{if } p \geq 2, \\ B = 1 + \frac{n}{n-2}F^{-1/2}, & \text{for every } p, \end{cases}$$

where B is the supremum value of Beales measure of nonlinearity, see Kitsos (2001) for details, and can be applied in any case.

4. Discussion

Experimental design in the linear case started as an optimum allocation of the observations at the treatment points. In his classical book, Fedorov (1972) summarized and extended all the linear work. The main target through this theoretical framework is to obtain methods –possibly based on algorithms– to get the optimum design measure for estimating. He provided the first algorithm, but it was only in Wu & Wynn (1978) that a general dichotomous convergence theorem was obtained, concerning the convergence of the sequence of design measures. The theoretical framework in the linear case is completed by the duality theory which first came to light in Lagrangian theory (Silvey (1972), Sibson (1972), Silvey & Titterington (1973), Pukelsheim & Titterington (1983)) while Fornius (2008) worked on optimality concerning quadratic logistic.

Geometry can be really very helpful to understand the optimality criteria: D–optimality minimizes the volume of the confidence ellipsoid, and remains invariant to linear transformation. This is not the case for the G–, A– optimality, as the volume of the confidence ellipsoid might remain constant, but not the axes of the ellipsoid which are related to these criteria. The c–optimality is based on the geometrical oriented theorem of Elfvings. The sequential nature of the design, Ford *et al.* (1985) among others, in

the linear case, is rather based on the augmentation of the data than an approach to estimate the parameters and not a method for point estimation. The design points are very well defined at the linear case, while in the nonlinear case are heavily depending on the parameters (which is itself the problem in nonlinear cases), and therefore we are referring to local optimal design. In general, the NLED was constructed as a generalization of LED with ϑ being true. The LED it is not a special case of NLED, i.e. it cannot be provided from a reduced NLED theory. The main problem, we believe, is the study of the sequence $\mathrm{M}(\vartheta_n, \xi_n)$ in NLED problems.

References

[1] K.M. Abdelbasit and R.L. Plackett, *Experimental design for binary data*, JASA **78** (1983) 90–98.

[2] A.C. Atkinson and A.N. Donev, Optimum Experimental Designs (Clarendon Press, Oxford, 1992).

[3] D.M. Bates and D.G. Watts, *Relative curvature measures of nonlinearity*, JRSSB **42** (1980) 1–25.

[4] E.M.L. Beale, *Confidence regions in non-linear estimation*, JRSSB **22** (1960) 41–88.

[5] G. Elfving, *Optimum allocation in linear regression theory*, Ann. Math. Stat. **23** (1952) 255–262.

[6] V.V. Fedorov, Theory of Optimal Experiments (Academic Press, 1972).

[7] R.A. Fisher, *On the mathematical foundation of theoretical statistics*, Phil. Trans. Roy. Soc. London (Ser. A) **22** (1922) 309–368.

[8] I. Ford, C.P. Kitsos and D.M. Titterington, *Recent advances in nonlinear experimental design*, Technometrics **31** (1989) 49–60.

[9] I. Ford, D.M. Titterington and C.F.J. Wu, *Inference and sequential design*, Biometrika **72** (1985) 545–551.

[10] F.E. Fornius, Optimal Design of Experiments for the Quadratic Logistic Model (PhD Thesis, Stockholm University, Sweden, 2008).

[11] D.C Hamilton, D.G. Watts and D.M. Bates, *Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions*, Ann. Stat. **10** (1982) 386–393.

[12] P.D.H. Hill, *D-optimal designs for partially nonlinear regression models*, Technometrics **22** (1980) 275–276.

[13] P.D.H. Hill, Optimal Experimental Design for Model Discrimination (PhD Thesis, University of Glasgow, UK, 1976).

[14] S. Karlin and W.J. Studden, *Optimal experimental designs*, Ann. Math. Stat. **37** (1966) 783–815.

[15] A.I. Khuri, *A note on D-optimal designs for partially nonlinear regression models*, Technometrics **26** (1984) 59–61.

[16] J. Kiefer, *General equivalence theory for optimum designs (approximate theory)*, Ann. Stat. **2** (1974) 849–879.

[17] J. Kiefer and J. Wolfowitz, *The equivalence of two extreme problems*, Canadian J. Math. **12** (1960) 363–366.

[18] C.P. Kitsos, *Invariant canonical form for the multiple logistic regression*, Math. in Eng. Science and Aerospace (MESA) **2(3)** (2011) 267–275.

[19] C.P. Kitsos, *Design aspects for the Michaelis-Menten model*, Biometrical Letters **38** (2001) 53–66.

[20] C.P. Kitsos, *Fully sequential procedures in nonlinear design problems*, Comp. Stat. and Data Analysis **8** (1989) 13–19.

[21] C.P. Kitsos, Design and Inference in Nonlinear Problems (PhD Thesis, University of Glasgow, UK, 1986).

[22] C.P. Kitsos and K.G. Kolovos, *An optimal calibration design for pH meters*, Instrumentation Science and Technology **38(6)** (2010) 436–447.

[23] E. Lauter, *A method of designing experiments for nonlinear models*, Math. Operationsforsch U. Stat. **5** (1974) 697–708.

[24] F. Pukeisheim and D.M. Titterington, *General differential and Lagrangian theory for optimal experiment design*, Ann. Stat. **11** (1983) 1060–1068.

[25] R. Sibson, *Discussion of paper by H.P. Wynn*, JRSSB **34** (1972) 181–183.

[26] S.D. Silvey, *Discussion of paper by H.P. Wynn*, JRSSB **34** (1972) 174–175.

[27] S.D. Silvey, Optimal Design (Chapman and Hall, 1980).

[28] S.D. Silvey and D.M. Tittertington, *A geometric approach to optimal design theory*, Biometrika **60** (1973) 21–32.

[29] K. Smith, *On the standard deviation of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations*, Biometrika **12** (1918) 1–85.

[30] M. Stone and J.A. Morris (1985), *Explanatory studies in non-sequential nonlocal non-linear design*, J. of Statistical Planning and Inference **12** (1985) 1–9.

[31] D.M. Titterington, *Aspects of optimal design in dynamic systems*, Technometrics **22** (1980) 287–299.

[32] D.M. Titterington, *Optimal design: Some geometrical aspects of D–optimality*, Biometrika **62** (1975) 313–320.

[33] L.V. White, *An extension to the general equivalence theorem for nonlinear models*, Biometrika **60** (1973) 345–348.

[34] C.F.J. Wu, *Asymptotic theory of nonlinear least squares estimates*, Ann. Stat. **9** (1981) 501–513.

[35] C.F.J. Wu and H.P. Wynn, *The convergence of general step-length algorithms for regular optimum design criteria*, Ann. Stat. **6** (1978) 1273–1285.

[36] V. Zarikas, V. Gikas and C.P. Kitsos, *Evaluation of the optimal design cosinor model for enhancing the potential of robotic theodolite kinematic observation*, Measurement **43(10)** (2010) 1416–1424.

Received 18 May 2011