
Volume 58
February 2013
ISSN 0167-9473
ELSEVIER
COMPUTATIONAL STATISTICS & DATA ANALYSIS
Incorporating Statistical Software Newsletter
Contains Special Issue:
The Third Special Issue on Statistical Signal Extraction and Filtering
Guest Editors: Stephen G. Pollock, Tommaso Proietti, Esther Ruiz and Howard Weinert
The official journal of
iasc
The International Association for Statistical Computing
A Section of The International Statistical Institute
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Computational Statistics and Data Analysis 58 (2013) 276–282

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A procedure for finding an improved upper bound on the number of optimal design points

Seung Won Hyun [a,*], Min Yang [b], Nancy Flournoy [b]

[a] Department of Statistics, North Dakota State University, Fargo, ND 58102, USA
[b] Department of Statistics, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 28 September 2011
Received in revised form 14 August 2012
Accepted 16 August 2012
Available online 23 August 2012

Keywords:
Carathéodory's theorem
Nonlinear regression
Cardinality of design
Experimental design

ABSTRACT

Knowing an upper bound on the number of optimal design points greatly simplifies the search for an optimal design. Carathéodory's Theorem is commonly used to identify an upper bound. However, the upper bound from Carathéodory's Theorem is relatively loose when there are three or more parameters in the model. In this paper, an alternative approach of finding a sharper upper bound for classical optimality criteria is proposed. Examples are given to demonstrate how to use the new approach.

Published by Elsevier B.V.

1. Introduction

Regression models with at least three parameters are required in many applications. For example, although the two parameter logistic model is widely used to estimate the dose corresponding to the p%th effectiveness (ED_{100p}), four parameter sigmoidal shaped functions, such as the EMAX model (Wagner, 1968), are commonly required when continuous responses are scaled to fall between zero and one and minimum and maximum values are unknown. In addition, many biological science experiments produce dose response functions with a downturn which require nonlinear models with at least three parameters (e.g. Welshons et al. (2003)). In any experiment, the sample space of treatments (e.g. doses) and the distribution of subjects over this treatment sample space must be determined. Let the treatment space be finite, and let $\boldsymbol{\xi} = \{(x_i, w_i)^\top\}_1^K$ denote a design, where x_i is the ith dose and $w_i = n_i/n$ is the proportion of subjects allocated to dose x_i (i.e., the associated design weight).

An optimal design specifies the doses to use and how to distribute subjects over these doses in the manner which minimizes a criterion function, which reflects the goals of the experiment. Because the inverse of the Fisher information matrix is approximately proportional to the variance–covariance matrix of maximum likelihood estimates (under common regularity conditions), this information matrix is typically used to construct criterion functions. Two common examples are the so called D- and A-optimality criteria. D-optimality minimizes the determinant of the inverse of the Fisher information matrix for the parameters of interest, which is asymptotically equivalent to minimizing the confidence ellipsoid for the joint estimation of the parameters; A-optimality minimizes the sum of the asymptotic variances of the maximum likelihood estimates.

As an illustrative model, define $\boldsymbol{\Theta} = (\theta_1, \theta_2, \ldots, \theta_t)$ and let observations be independent given treatment with

$$y_{ij} = \mu(x_i, \boldsymbol{\Theta}) + \epsilon_{ij}; \quad \epsilon_{ij} \sim N(0, \sigma^2), \tag{1}$$

* Corresponding author. Tel.: +1 701 231 8178.
E-mail addresses: Seung.W.Hyun@ndsu.edu (S.W. Hyun), YangMi@Missouri.edu (M. Yang), FlournoyN@Missouri.edu (N. Flournoy).

0167-9473/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.csda.2012.08.012

where $\mu(x_i, \boldsymbol{\Theta})$ is differentiable with respect to $\boldsymbol{\Theta}$ and σ^2 is unknown. Let $\hat{\boldsymbol{\Theta}}$ denote the maximum likelihood estimate of $\boldsymbol{\Theta}$. Then, by the Taylor expansion, an approximate Fisher information matrix for $\boldsymbol{\Theta}$ can be written as

$$\mathbf{M}_{(\boldsymbol{\xi},\boldsymbol{\Theta})} = \left(m_{uv}\right) = \frac{n}{\sigma^2}\sum_{i=1}^{K} w_i \boldsymbol{f}(x_i)^\top \boldsymbol{f}(x_i), \tag{2}$$

where $\boldsymbol{f}(x) = \left(\frac{\partial \mu(x,\boldsymbol{\Theta})^\top}{\partial\theta_1} \quad \frac{\partial \mu(x,\boldsymbol{\Theta})^\top}{\partial\theta_2} \quad \cdots \quad \frac{\partial \mu(x,\boldsymbol{\Theta})^\top}{\partial\theta_t}\right)$, $u, v = 1, 2, \ldots, t$.

Often interest is in $s(1 \le s \le t)$ linear combinations of the parameters, say $\mathbf{A}^\top\boldsymbol{\Theta}$, where $\mathbf{A}$ is a $t \times s$ dimensional matrix. For example, with $s = 1$, $\mathbf{A}^\top = \begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}$ specifies interest in estimating the first parameter with all the rest being nuisance parameters. The information matrix for $\mathbf{A}^\top\boldsymbol{\Theta}$ under model (1) is $\mathbf{C} = \left(\mathbf{A}^\top\mathbf{M}^-_{(\boldsymbol{\xi},\boldsymbol{\Theta})}\mathbf{A}\right)^{-1}$, where $\mathbf{M}^-_{(\boldsymbol{\xi},\boldsymbol{\Theta})}$ is the generalized inverse of $\mathbf{M}_{(\boldsymbol{\xi},\boldsymbol{\Theta})}$. A general, classical, class of optimality criteria is given by

$$\boldsymbol{\phi}_p = \begin{cases} \left(\frac{1}{s} Tr\left[\mathbf{C}^p\right]\right)^{1/p}, & p < 0; \\ (\mathrm{Det}[\mathbf{C}])^{1/s} & p = 0; \\ \lambda_{\min}(\mathbf{C}) & p = -\infty; \end{cases} \tag{3}$$

where $\lambda_{\min}(\mathbf{C})$ denotes the minimum eigenvalue of $\mathbf{C}$. The most well known criteria are D-optimality if $p = 0$ and A-optimality if $p = -1$. Note that the factorization in (2) is not limited to models of the form (1), and the methods described in the paper can be applied to other models as long as the information matrix can be factored as (2).

Analytical methods for obtaining optimal designs are well developed for linear models, See Fedorov (1972), Fedorov and Hackl (1997), Pukelsheim (2006), and Atkinson et al. (2007). A geometric method developed by Elfving (1952) has proven useful for some optimal design criteria for two parameter generalized linear models, See Ford et al. (1992), and Fellman (1999). General analytical methods for two parameter nonlinear models are developed by Yang and Stufken (2009).

In general, nonlinear models with three or more parameters have complicated information matrices making it hard to obtain optimal designs analytically. There are relatively few recent papers on analytically obtaining optimal designs for nonlinear models with three or more parameters, See Li and Majumdar (2008, 2009), Yang (2010), and Dette and Melas (2011). Sitter and Torsney (1995) apply Elving's geometric approach to a three parameter model.

However, the application of optimal design theory for nonlinear models has exploded with the use of numerical algorithms. For example, the exchange algorithm is described by Fedorov and Hackl (1997). Knowing an upper bound on the number of design points limits the number of candidate design points to be considered, making the search more tractable.

Yang (2010) provides a procedure for determining an upper bound on the number of support points for any classical optimal design criterion. The basic idea is to identify a subclass of designs such that for any arbitrary design, one can find a design in that subclass which has an improved information matrix in terms of Lowener ordering. Dette and Melas (2011) generalized Yang (2010), utilizing different methods.

In this paper, we provide yet a different way of determining an upper bound on the number of support points for classical optimal criterion function. Although this approach is not as general as Yang's approach, it is easy to understand and straightforward to implement.

When each subject has a single response and the number of design points is less than the number of parameters, the parameters are unidentifiable. Therefore, one may require the number of design points be at least t, where t is the number of parameters. Finding a upper bound q on the number of design permits the search for optimal design points to be restricted to the range $[t, q]$ and reducing q by even just one can greatly simplify the search for optimal designs. Carathéodory's Theorem (cf. Liski et al. (2002), and Pukelsheim (2006)) is commonly used for obtaining q. The theorem states that the number of points for any optimal design is no more than $t(t+1)/2+1$. When the number of parameters is 2, the upper bound is 4. Thus, the search can be limited to 2, 3 and 4 point designs; since the cardinality of possible optimal design points is very narrow, optimal designs for two parameter models can be obtained relatively easily. However, when the model has more than 3 parameters, Carathéodory's Theorem provides an upper bound that is increasingly far from t and the optimal design problem explodes. In this paper, we present a new approach which reduces q below the bound given by Carathéodory's Theorem. This approach is invariant under any classical optimality criteria, and it can be applied to multiple-stage experiments as well as to single-stage ones.

In Section 2, the new approach for identifying an upper bound on the number of optimal design points is described. In Section 3, the proposed approach is applied to Hyun et al.'s model (2011) and to a probit model with three parameters. Finally, brief conclusions are given.

2. Identifying the upper bound on the number of design points

Let $\boldsymbol{\xi}^*$ denote the optimal design under some ϕ_p-optimality criterion, and let $\mathbf{G}$ denote a generalized inverse of $\mathbf{M}_{(\boldsymbol{\xi}^*,\boldsymbol{\Theta})}$, the information for $\boldsymbol{\xi}^*$. Suppose the research question of interest lies in $s\,(s \le t)$ linear combinations of the parameters $\mathbf{A}^\top\boldsymbol{\Theta}$, where $\mathbf{A}$ is a $t \times s$ matrix of rank s. Then the information matrix for $\mathbf{A}^\top \boldsymbol{\Theta}$ under the model (1) is $(\mathbf{A}^\top \mathbf{G} \mathbf{A})^{-1}$. Kiefer's General Equivalence Theorem (Kiefer (1958); cf. Pukelsheim (2006)) states that $\boldsymbol{\xi}^*$ is the ϕ_p-optimal design for $\mathbf{A}^\top \boldsymbol{\Theta}$ if and only if there is a generalized inverse $\mathbf{G}$ of $\mathbf{M}_{(\boldsymbol{\xi}^*, \boldsymbol{\Theta})}$ such that

$$\boldsymbol{f}(x)\mathbf{C}\boldsymbol{f}(x)^\top \le Tr((\mathbf{A}^\top \mathbf{G}\mathbf{A})^{-p}), \tag{4}$$

where

$$\mathbf{C} = \mathbf{G}\mathbf{A}\left(\mathbf{A}^\top \mathbf{G}_{(\boldsymbol{\xi}^*, \boldsymbol{\Theta})}\mathbf{A}\right)^{-(p+1)} \mathbf{A}^\top \mathbf{G}^\top = \left(c_{ij}\right)_{1,1}^{t,t}. \tag{5}$$

If $\mathbf{G}$ is constructed from any ϕ_p-optimal design $\boldsymbol{\xi}^*$ for $\mathbf{A}^\top \boldsymbol{\Theta}$, equality will be obtained in (4) when x is replaced with any support point x_i from $\boldsymbol{\xi}^*$. So no matter what the support points are, they must be the local maximum points of

$$F(x) = \boldsymbol{f}(x)\mathbf{C}\boldsymbol{f}(x)^\top, \tag{6}$$

and, hence, the number of optimal design points cannot exceed the number of local maxima in $F(x)$. If we can determine an upper bound on the number of local maxima in $F(x)$, then we have an upper bound on the number of optimal design points. Note that c_{ij}, $1 \le j \le i \le t$, is some function of $\boldsymbol{\xi}^*$, $\boldsymbol{\Theta}$, $\mathbf{A}$, and $\mathbf{G}$, but not of x. Therefore, in seeking the maxima in (6), one can treat $\{c_{ij}\}$ as constant. Define $\boldsymbol{\Psi}_{ij} = \boldsymbol{f}(x_i)\boldsymbol{f}(x_j)$, $1 \le j \le i \le t$. Then (6) can be written as

$$F(x) = \sum_{i=1}^{t} c_{ii}\boldsymbol{\Psi}_{ii}(x) + 2\sum_{i=1}^{t}\sum_{j<i} c_{ij}\boldsymbol{\Psi}_{ij}(x). \tag{7}$$

Suppose that $F(x)$ is a smooth function defined on $[A, B]$, where A can be $-\infty$ and B can be ∞. We first give two definitions:

1. A point x is called a down-crossing critical point of $F(x)$ if $F(x) = 0$, $F(x-) > 0$ and $F(x+) < 0$;
2. A point x is called a up-crossing critical point of $F(x)$ if $F(x) = 0$, $F(x-) < 0$ and $F(x+) > 0$.

Lemma 1 gives some basic facts about down-crossing critical points and up-crossing critical points of $F(x)$ which follow straightforwardly from mathematical analysis:

Lemma 1. *Let $F(x)$ be given by (6), and suppose it is a smooth function defined on $[A, B]$ with a finite number of critical points.*

1. *Down-crossing critical points and up-crossing critical points alternate, i.e., any two adjacent down-crossing critical points must have one and only one up-crossing critical point between them and any two adjacent up-crossing critical points must have one and only one down-crossing critical point between them;*
2. *Suppose x_1 is a down-crossing critical point and x_2 is a up-crossing critical point which is adjacent to x_1. If $x_1 < x_2$, then $F(x) < 0$ for $x \in (x_1, x_2)$; if $x_1 > x_2$, then $F(x) > 0$ for $x \in (x_2, x_1)$;*
3. *Suppose x_A is the smallest down-crossing critical point. If there is no up-crossing critical point on $[A, x_A]$, then $F(x) > 0$ for $x \in [A, x_A]$. If there exist a up-crossing critical point on $[A, x_A]$, say x_0, then $F(x) < 0$ for $x \in [A, x_0]$ and $F(x) > 0$ for $x \in [x_0, x_A]$;*
4. *Suppose x_B is the largest down-crossing critical point. If there is no up-crossing critical point on $[x_B, B]$, then $F(x) < 0$ for $x \in [x_B, B]$. If there exist a up-crossing critical point on $[x_B, B]$, say x_0, then $F(x) < 0$ for $x \in [x_B, x_0]$ and $F(x) > 0$ for $x \in [x_0, B]$.*

The following theorem helps to clarify the relationship between the number of down-crossing critical points of $F(x)$ and the number of down-crossing critical points of it's derivative $F'(x)$.

Theorem 1. *For a differentiable function $F(x)$ defined on $[A, B]$, suppose $F'(x)$ has finite number of critical points with at most $M(\ge 1)$ down-crossing critical points on $[A, B]$. Then $F(x)$ has finite number of critical points and the number of down-crossing critical points depends on one of the following four situations:*

1. *if the smallest and largest critical points of $F'(x)$ are both down-crossing, then $F(x)$ has either (i) at most $M - 1$ down-crossing critical points or (ii) M down-crossing critical points and the largest critical point is down-crossing;*
2. *if the largest critical point of $F'(x)$ is down-crossing, then in addition to the two cases in 1, $F(x)$ has either (i) M down-crossing critical points and the smallest critical point is down-crossing or (ii) $M + 1$ down-crossing critical points and both the smallest and largest critical point are down-crossing;*
3. *if the smallest critical point of $F'(x)$ is down-crossing, then $F(x)$ has at most M down-crossing critical points;*
4. *otherwise, $F(x)$ either has (i) at most M down-crossing critical points or (ii) $M + 1$ down-crossing critical points and the smallest critical point is down-crossing.*

Proof. Let $x_{A'}$ and $x_{B'}$ be the smallest and largest down-crossing critical points of $F'(x)$ and let x_A and x_B be the smallest and largest down-crossing critical points of $F(x)$. The conclusions of this theorem are straightforward once we establish the following three facts:

(a) Between two adjacent down-crossing critical points of $F'(x)$, there exists at most one down-crossing critical point of $F(x)$ and at most one up-crossing point of $F(x)$, and there are no other critical points of $F(x)$.

(b) On $[A, x_{A'}]$, if there is no up-crossing critical point of $F'(x)$, then there is no down-crossing critical point of $F(x)$ on $[A, x_{A'}]$, but there may exist an up-crossing critical point on this interval; if there exists an up-crossing critical point of $F'(x)$, then there may exist a down-crossing critical point of $F(x)$ on $[A, x_{A'}]$ and there is no up-crossing critical point before the down-crossing critical point.
(c) On $[x_{B'}, B]$, if there is no up-crossing critical point of $F'(x)$, then there exists at most one down-crossing critical point of $F(x)$ on $[x_{B'}, B]$, and there is no up-crossing critical point on this interval; if there exists an up-crossing critical point of $F'(x)$, then there may exist a down-crossing critical point of $F(x)$ on $[A, x_{A'}]$ and there may also exist an up-crossing critical point after the down-crossing critical point.

The proof of the three facts are similar, so we just prove (a). Let x_1 and x_3 be two adjacent down-crossing critical points of $F'(x)$. Then by 1 of Lemma 1, there exists one and only one up-crossing critical point of $F'(x)$ on (x_1, x_3), say x_2. By 2 of Lemma 1, $F'(x) \leq 0$ for $x \in (x_1, x_2)$, which means $F(x)$ is a strictly decreasing function ($F'(x)$ has finite number of critical points). Thus there exists at most one critical point of $F(x)$ on (x_1, x_2), and if it exists, it must be down-crossing critical point. Similarly we can show that there exists at most one critical point of $F(x)$ on (x_2, x_3), and if it exists, it must be up-crossing critical point. □

Clearly, the number of local maxima of a function $F(x)$ depends on the number of down-crossing critical points of $F'(x)$ and whether there are up-crossing points in $[A, x_{A'}]$ and $[x_{B'}, B]$. By mathematical induction, we have

Lemma 2. *Suppose that $F(x)$ is a function defined on $[A, B]$. Let the number of local maximum points of a function $F(x)$ be R and the number of down-crossing critical points of $F'(x)$ be M.*

1. *if both the smallest and largest critical points of $F'(x)$ are down-crossing, then $R = M$;*
2. *if only the smallest critical point of $F'(x)$ is down-crossing, then $R = M + 1$ and B is one of local maximum points;*
3. *if only the largest critical point of $F'(x)$ is down-crossing, then $R = M + 1$ and A is one of local maximum points;*
4. *if none of the smallest or largest critical point of $F'(x)$ is down-crossing, then $R = M + 2$ and both A and B are local maximum points.*

If we divide $F'(x)$ by some positive function on $[A, B]$, say $P(x)$, then the down-crossing critical points and up-crossing points of $F'(x)$ are the same as those of $F'(x)/P(x)$. Thus we can study $F'(x)/P(x)$ instead of $F'(x)$. If it is still difficult to determine the number of down-crossing critical points of $F'(x)/P(x)$, we can study the down-crossing critical points and up-crossing points of $\left(F'(x)/P(x)\right)'$. This process can be continued until one obtains functions in a simple format, in particular, a quadratic form. The derivatives of $F(x)$ are typically easily obtained by *Mathematica*. Using the following theorem, we can obtain the maximum number of locally maximum points of $F(x)$.

Theorem 2. *Let $F(x)$ be a differentiable function given by (6). Suppose there exist positive functions $P_i(x)$ (possibly equal to 1) such that*

$$\begin{aligned}
G_1(x) &= \frac{dF(x)}{dx} = P_1(x)F_1(x)\\
G_2(x) &= \frac{dF_1(x)}{dx} = P_2(x)F_2(x)\\
&\vdots\\
G_m(x) &= \frac{dF_{m-1}(x)}{dx} = P_m(x)[ax^2 + bx + c],
\end{aligned} \tag{8}$$

where a, b, and c are constants; and m is the number of derivatives required to reach a quadratic form. Let r be a positive integer. Then

1. *If a is negative and $m = 2r$, the upper bound is $r + 1$;*
2. *If a is negative and $m = 2r + 1$, the upper bound is $r + 2$ and $\boldsymbol{\xi}^*$ includes the lowest point in the design space;*
3. *If a is positive and $m = 2r$, the upper bound is $r + 2$ and $\boldsymbol{\xi}^*$ includes both the lowest point and the highest point in the design space;*
4. *If a is positive and $m = 2r + 1$, the upper bound is $r + 2$ and $\boldsymbol{\xi}^*$ includes the highest point in the design space.*

Proof. We only give the proof when a is negative by mathematical induction. The proof when a is positive is similar. Let M' and M be the number of down-crossing critical points of $F'(x)$ and $F(x)$, respectively.

When $m = 1$, consider the behavior of $F'(x)$ when $x \in (-\infty, +\infty)$. Clearly, $F'(x)$ either < 0 for all x or has one down-crossing critical point and the largest critical point is down-crossing. For the first case, $F(x)$ is a strict decreasing function, thus $F(x)$ has one locally maximum point and $M \leq 1$ (if $M = 1$ the largest critical point is down-crossing). For the second case, $F(x)$ has at most two locally maximum points including A by Lemma 2. In addition, as given by situation 2 of Theorem 1, $M \leq 1$ (if $M = 1$ at least one of the smallest and the largest critical point is down-crossing) or $M = 2$ and both the smallest and largest critical point are down-crossing. Thus when $m = 1$, the conclusion follows.

When $m = 2$, by the discussion of $m = 1$, we have $M' = 0$, or $M' = 1$ and at least one of the smallest and the largest critical point is down-crossing, or $M' = 2$ and both the smallest and largest critical point are down-crossing. By Lemma 2, $F(x)$ has at most two locally maximum points. In addition, by applying Theorem 1 and discussing the three cases, we have either (i) $M \le 1$ or (ii) $M = 2$ and the largest critical point is down-crossing.

Assume that when $m = 2r$, $F(x)$ has at most $r + 1$ locally maximum points and two cases for the down-crossing critical points of $F(x)$: (i) $M \le r$ or (ii) $M = r + 1$ and the largest critical point is down-crossing.

When $m = 2r + 1$, by the assumption, down-crossing critical points of $F'(x)$ have two possibilities: (i) $M' \le r$; (ii) $M' = r + 1$ and the largest critical point is down-crossing. By Lemma 2, $F(x)$ has at most $r + 2$ locally maximum points including A. In addition, by applying Theorem 1 and discussing the two possibilities, there are three possibilities for the down-crossing critical points of $F(x)$: (i) $M \le r$; (ii) $M = r + 1$ and at least one of the smallest and the largest critical is down-crossing; (iii) $M = r + 2$ and both the smallest and the largest critical are down-crossing.

When $m = 2r + 2$, by the discussion of $m = 2r + 1$, down-crossing critical points of $F'(x)$ have three possibilities: (i) $M' \le r$; (ii) $M' = r+1$ and at least one of the smallest and the largest critical point is down-crossing; (iii) $M' = r+2$ and both the smallest and the largest critical are down-crossing. By Lemma 2, $F(x)$ has at most $r + 2$ locally maximum points. In addition, by applying Theorem 1 and discussing the three possibilities, there are two possibilities for the down-crossing critical points of $F(x)$: (i) $M \le r + 1$ or (ii) $M = r + 2$ and the largest critical point is down-crossing.

By mathematics induction, our conclusion follows. □

3. Application

In many toxicological assays, response functions have a downturn at high doses. This situation has been described by Margolin et al. (1981), Bretz et al. (2003), and Welshons et al. (2003); and others. Nonlinear models with three or more parameters have been utilized to explain such a response function with downturn. For example, Hyun et al. (2011) adapted Margolin et al.'s three parameter model (1981) to continuous responses. They subsequently found that the probit model with a quadratic term fitted the motivating experimental data very well. In this paper, these two models are used to demonstrate our approach.

Example 1. In Hyun et al.'s model (2011), the mean response function is

$$\mu(x_i, \boldsymbol{\Theta}) = \{1 - \exp[-(\alpha + \beta x_i)]\} \exp[-\gamma x_i], \quad \text{where } \alpha > 0, \beta \ge 0, \gamma \ge 0. \tag{9}$$

The Fisher information matrix for $\hat{\boldsymbol{\Theta}} = \{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\}$ can be written as

$$\mathbf{M}_{(\xi,\boldsymbol{\Theta})} = \frac{n}{\sigma^2} \sum_{i=1}^{K} w_i (\mathbf{B}^\top)^{-1} \mathbf{B}^\top \boldsymbol{f}(x_i)^\top \boldsymbol{f}(x_i) \mathbf{B}\mathbf{B}^{-1},$$

where

$$\boldsymbol{f}(x) = \begin{pmatrix} e^{-(\alpha+\beta x+\gamma x)} & xe^{-(\alpha+\beta x+\gamma x)} & -x\{e^{-\gamma x} - e^{-(\alpha+\beta x+\gamma x)}\} \end{pmatrix};$$

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Now $\mathbf{M}_{(\xi,\boldsymbol{\Theta})}$ can be rewritten as $\mathbf{M}_{(\xi,\boldsymbol{\Theta})} = \frac{n}{\sigma^2} \sum_{i=1}^{K} w_i (\mathbf{B}^\top)^{-1} (\boldsymbol{f}^*(x_i))^\top \boldsymbol{f}^*(x_i) \mathbf{B}^{-1}$, where $\boldsymbol{f}^*(x) = \mathbf{B}\boldsymbol{f}(x) = \begin{pmatrix} e^{-(\alpha+\beta x+\gamma x)} & xe^{-(\alpha+\beta x+\gamma x)} & xe^{-\gamma x} \end{pmatrix}$. The General Equivalence Theorem states

$$\boldsymbol{f}^*(x)\mathbf{C}\boldsymbol{f}^*(x)^\top \le Tr((\mathbf{A}^\top \mathbf{M}^{-}_{(\xi^*,\boldsymbol{\Theta})} \mathbf{A})^{-p}), \tag{10}$$

where $\mathbf{C}$ is given by (5) and equality holds when x is a support point of $\boldsymbol{\xi}^*$. Recall that $\mathbf{C}$ is constant in (10) since $\boldsymbol{\xi}^*$ is a ϕ_p-optimal design.

Eq. (6) becomes

$$F(x) = \boldsymbol{f}^*(x) \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} (\boldsymbol{f}^*(x))^\top$$

$$= e^{-2(\alpha+\beta x+\gamma x)} \{c_{22}x^2 + 2c_{12}x + c_{11} + xe^{\alpha+\beta x}(2c_{13} + 2c_{23}x + c_{33}xe^{\alpha+\beta x})\}. \tag{11}$$

Recall that, in applying Theorem 2, a positive factor in $F'(x)$ does not change the down-crossing critical points and up-crossing critical points of $F'(x)$. Therefore, any positive factors in the derivatives of (11) can be ignored. The seventh derivative in the sequence (8) applied to (11) is a quadratic equation with respect to x:

$$\frac{G_7(x)}{P_7(x)} = -16\beta^6 c_{33}\gamma x^2 + (16\beta^6 c_{33} - 144\beta^5 c_{33}\gamma)x + 72\beta^5 c_{33} - 264\beta^4 c_{33}\gamma = 0. \tag{12}$$

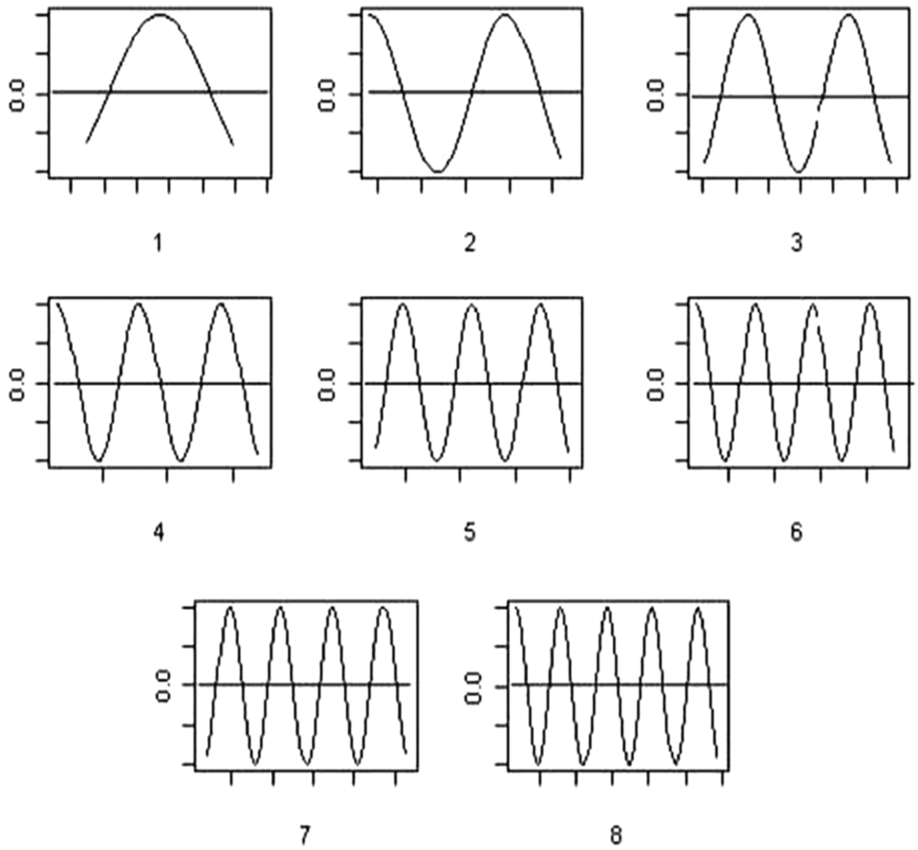

Fig. 1. The change of shapes of the graphs for different differential equations. The graphs show how many extreme values can be observed at most for each of lower order derivatives from the concave down quadratic equation. For example, the graph 1 represents the shape of the concave down quadratic equation and the graph 2 shows that there are 2 extreme values including the lowest point where there is one lower order derivatives from the concave down quadratic equation.

Since c_{33} is a diagonal element of the positive definite matrix (5), c_{33} is positive. Recall $\beta, \gamma \geq 0$ as given in (9). Thus the sign of coefficient for x^2 is negative. By 2 of Theorem 2, there are 5 extreme values and $\boldsymbol{\xi}^*$ includes the lowest point in the design space. Thus the upper bound on the number of design points for a ϕ_p-optimality criterion is 5 of which only four are unknown. The procedure for obtaining this result is sketched in Fig. 1. In Fig. 1, the graph 1 shows the shape of the quadratic equation (12), and the graph 2 represents the shape of one lower order derivative from (12), $\frac{G_6(x)}{P_6(x)}$. In order to come back to $F(x)$, there are seven lower order derivatives from (12) in this case. Thus, the graph 8 in Fig. 1 represents the shape of (11), which shows that there are 5 extreme values including the lowest point.

Example 2. In Hyun and Flournoy (submitted for publication), the probit model with quadratic term fit the data that motivated Hyun et al. (2011) very well when the mean response function is

$$\mu(x_i, \boldsymbol{\Theta}) = \Phi(\alpha + \beta x_i + \gamma x_i^2), \tag{13}$$

where $\alpha < 0, \beta < 0, \gamma < 0$ and Φ is the cumulative distribution function of the standard normal distribution. The Fisher information matrix for $\boldsymbol{\Theta}$ is

$$M_{(\boldsymbol{\xi}, \boldsymbol{\Theta})} = \frac{n}{\sigma^2} \sum_{i=1}^{K} w_i \frac{1}{\sqrt{2\pi}} \boldsymbol{f}(x_i)^\top \boldsymbol{f}(x_i),$$

where

$$\boldsymbol{f}(x) = \exp\left\{-\frac{(\alpha + \beta x_i + \gamma x_i^2)^2}{2}\right\} \begin{pmatrix} 1 & x & x^2 \end{pmatrix}.$$

Eq. (6) becomes

$$\begin{aligned} F(x) &= \boldsymbol{f}(x) \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} (\boldsymbol{f}(x))^\top \\ &= \exp\{-(\alpha + \beta x + \gamma x^2)^2\}\{c_{33}x^4 + 2c_{23}x^3 + (2c_{13} + c_{22})x^2 + 2c_{12}x + c_{11}\}. \end{aligned} \tag{14}$$

The positive term $\exp\{-(\alpha+\beta x+\gamma x^2)^2\}$ can be ignored in $G_1(x)$. Continuing to apply Theorem 2, the sixth derivative in the sequence (8) applied to (14) is a quadratic equation with respect to x:

$$\frac{G_6(x)}{P_6(x)} = -42c_{33}\gamma^2x^2 - 6(4c_{23}\gamma^2 + 3c_{33}\beta\gamma)x - (c_{33}\beta^2 + 6c_{23}\beta\gamma + 4c_{13}\gamma^2 + 2c_{22}\gamma^2 + 2c_{33}\alpha\gamma) = 0. \quad (15)$$

As for Example 1, the sign of coefficients for the both quadratic terms are negative. By 1 of Theorem 2, the upper bound for ϕ_p-optimal designs under the model (13) is 4 design points. This also can be seen in the Fig. 1 again. The procedure is the same as before. In Fig. 1, the graph 1 shows the shape of the quadratic equation (15). In order to come back to $F(x)$, there are six lower order derivatives from (15) in this case. Thus, the graph 7 in Fig. 1 represents the shape of (14), which shows that there exists 4 extreme values.

4. Conclusion

We present a new approach for identifying the maximum number of design points under classical optimality criteria. To illustrate our approach, Hyun et al.'s model (2011) and a probit model with a quadratic term are used to describe the response surface. The approach provides the maximum number of design points under ϕ_p-optimality cannot be greater than 5 and one point is the lowest in the design space under Hyun et al.'s model and cannot be greater than 4 under the probit model. Carathéodory's Theorem provides 7 points as the upper bound when there are 3 parameters in the model. So our approach reduces the upper bound for both examples. In addition, one point is given in some cases and others may have weights equal to zero indicating a smaller optimal set.

As mentioned in the introduction, the proposed approach is not restricted to single-stage experiments. Multi-stage experiments are used when good parameter estimates do not exist. Following the first stage using an initial design, parameter estimates are used to estimate the optimal design, which is then used in the second stage. Continuing in this manner, the design points that are added are increasingly efficient. Let $\boldsymbol{\xi}_1$ denote an initial design and $\boldsymbol{\xi}_2$ denote a design used at the second stage. Then a common measure of information is $\mathbf{M}_{(\xi,\Theta)} = \mathbf{M}_{(\xi_1,\Theta)} + \mathbf{M}_{(\xi_2,\Theta)}$. The corresponding equivalence theorem can be re-written in a format similar to (4). So, identifying the upper bound on the number of optimal design points for $\boldsymbol{\xi}_2$ following $\boldsymbol{\xi}_1$ becomes the same as identifying the upper bound for $\boldsymbol{\xi}_2$ in a single-stage ignoring $\boldsymbol{\xi}_1$.

The approach presented in this paper is illustrated for specific models. If the model changes, the upper bounds may change too, and the bounds proposed in this paper can be better or worse than in the examples given here. Regardless, it will be worthwhile to apply Theorem 2 to determine whether or not it provides a reduced upper bound on the maximum number of design points from Carathéodory's Theorem.

References

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs with SAS. Oxford University Press.

Bretz, F., Hothorn, L.A., Hsu, J.C., 2003. Identifying effective and/or safe doses by stepwise confidence intervals for ratio. Statistics in Medicine 22, 847–858.

Dette, H., Melas, V.B., 2011. A note on the de la Garza phenomenon for locally optimal designs. Annals of Statistics 39, 1266–1281.

Elfving, G., 1952. Optimal allocation in linear regression theory. Annals of Mathematical Statistics 23, 255–262.

Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press.

Fedorov, V.V., Hackl, P., 1997. Model-Oriented Design of Experiments. In: Lecture Notes in Statistics, vol. 125. Springer-Verlag, New York.

Fellman, J., 1999. Gustav Flfving's contribution to the emergence of the optimal experimental design theory. Statistical Science 14, 197–200.

Ford, I., Torsney, B., Wu, C.F.J., 1992. The use of a canonical form in the construction of locally optimal designs for non-linear problems. Journal of the Royal Statistical Society, Series B 54, 569–583.

Hyun, S.W., Flournoy, N., 2012. Optimal designs for a probit model with a quadratic term. Statistics in Biopharmaceutical Research (submitted for publication).

Hyun, S.W., Flournoy, N., Yang, M., 2011. Optimal designs for response functions with a downturn. Journal of Statistical Planning and Inference 141, 559–575.

Kiefer, J., 1958. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. Annals of Mathematical Statistics 29, 675–699.

Li, G., Majumdar, D., 2008. D-optimal designs for logistic models with three and four parameters. Journal of Statistical Planning and Inference 138, 1950–1959.

Li, G., Majumdar, D., 2009. Some results on D-optimal designs for nonlinear models with applications. Biometrika 96, 487–493.

Liski, E.P., Mandal, N.K., Shah, K.R., Siniha, B.K., 2002. Topics in Optimal Design. Springer-Verlag, New York.

Margolin, B.H., Kaplan, N., Zeiger, E., 1981. Statistical analysis of the ames Salmonella/microsome test. Proceedings of the National Academy of Sciences of the United States of America 78, 3779–3783.

Pukelsheim, F., 2006. Optimal design of experiment. In: Society for Industrial and Applied Mathematics. SIAM, Philadelphia, PA.

Sitter, R.R., Torsney, B., 1995. Optimal designs for binary response experiments with two design variables. Statistica Sinica 5, 405–419.

Wagner, J.G., 1968. Kinetics of pharmacologic response, I, proposed relationships between response and drug concentration in the intact animal and man. Journal of Theoretical Biology 20, 173–201.

Welshons, W.V., Thayer, K.A., Judy, B.M., Taylor, J.A., Curran, E.M., Saal, F.S., 2003. Large effects from small exposures, I, mechanisms for endocrine-disrupting chemicals with estrogenic activity. Environmental Health Perspectives 111, 994–1006.

Yang, M., 2010. On the de la Garza phenomenon. Annals of Statistics 38, 2499–2524.

Yang, M., Stufken, J., 2009. Support points of locally optimal designs for nonlinear models with two parameters. Annals of Statistics 37, 518–541.