# Optimal Designs for Some Selected Nonlinear Models

**A.S. Hedayat**[*,†], **Ying Zhou**[†] **Min Yang**[‡]

*Department of Mathematics, Statistics, and Computer Science*
*University of Illinois at Chicago*
*Science and Engineering Offices (M/C 249)*
*851 S. Morgan Street*
*Chicago, IL 60607-7045*

**Abstract:** Some design aspects related to three complex nonlinear models are studied in this paper. For the Klimpel's flotation recovery model, it is proved that regardless of model parameter and optimality criterion, any optimal design can be based on two design points and the right boundary is always a design point. For this model, an analytical solution for a $D$-optimal design is derived. For the 2-parameter chemical kinetics model, it is found that the locally D-optimal design is a saturated design. Under a certain situation, any optimal design under this model can be based on two design points. For the $2n$-parameter compartment model, compared to the upper bound by Carathéodory's theorem, the upper bound of the maximal support size is significantly reduced by the analysis of related Tchebycheff Systems. Some numerically calculated $A$-optimal designs for both Klimpel's flotation recovery model and 2-parameter chemical kinetic model are presented. For each of the three models discussed, the $D$-efficiency when the parameter misspecification happens is investigated. Based on two real examples from the mining industry, it is demonstrated how the estimation precision can be improved if optimal designs would be adopted. A simulation study is conducted to investigate the efficiencies of adaptive designs.

## 1. Introduction

Optimal designs based on nonlinear models have wide and important applications in many areas of science. A good example of the application is the optimal design based on PK/PD models which are widely used in pharmaceutical industries for the examination of absorption, distribution, metabolism, elimination, efficacy and toxicity parameters in drug developments, see Gieschke and Steimer (2000), Meibohm and Derendorf (2002).

We study and explore optimal designs for three complex nonlinear models. These are the 2-parameter chemical kinetic model, the 2n-parameter compartment model and the 2-parameter Klimpel's flotation recovery model. These three models have been shown to have extensive applications in real life situa-

*Corresponding author

†Research is supported by the U.S. National Science Foundation Grants DMS-0904125 and DMS-1306394

‡Research is supported by the U.S. National Science Foundation Grants DMS-0707013 and DMS-1322797

tions (Godfrey (1983), Jacquez (1985), Parekh and Miller (1999)), Saleh (2010) and Yuan et al. (1996).

We concentrate on the nonlinear model $y = \eta(x, \boldsymbol{\theta})+\varepsilon$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T$ is a vector of $k$ unknown parameters and $x$ is the explanatory variable defined on a design space $\chi$ in $\mathbb{R}$. The error $\varepsilon$ is postulated to be distributed as $N(0, \sigma^2)$ and without loss of generality we let $\sigma = 1$. Further, we assume that all observations are independent.

Typically, the optimal nonlinear design studies are under approximate theory, i.e., instead of exact sample sizes for design points, design weights are used. Let $\xi$ be any $n$-point approximate design,

$$\xi = \begin{pmatrix} t_1 & t_2 & \dots t_n \\ \omega_1 & \omega_2 & \dots \omega_n \end{pmatrix}.$$

Here $0 < \omega_i < 1$ represents the proportion of the number of points studied at $t_i$ with $\sum_{i=1}^{n} \omega_i = 1$. The Fisher information matrix for $y = \eta(t, \boldsymbol{\theta}) + \varepsilon$ can be written as

$$I_\xi = \sum_{i=1}^{n} \omega_i (\frac{\partial \eta(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial \eta(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^T. \tag{1.1}$$

How to compare two designs? There are variety of optimality criteria. Two popular optimality criteria are $D$-optimality and $A$-optimality, which are to maximize $|I_\xi|$ and minimize $Tr(I_\xi^{-1})$ over all possible designs, respectively. A $D$-optimal design minimizes the volume of an asymptotic confidence ellipsoid for $\theta$, and an $A$-optimal design minimizes the average of the asymptotic variances for the estimators of the individual parameters.

## 2. Preliminaries

In this paper, we shall use and deal with Extended Tchebycheff systems and Extended Complete Tchebysheff systems. The Tchebycheff systems were first introduced by the Russian mathematicians Chebyshev (1859) and Bernshtein (1937). In Karlin and Studden (1966), the theory of the Tchebycheff systems and its applicability in optimal design of experiments theory is introduced and studied.

Let $\{u_0, u_1, ..., u_n\}$ be $n+1$ continuous real-valued functions on $[a, b]$. $\{u\}_{k=0}^{n} = \{u_0, u_1, ..., u_n\}$ is called a Tchebycheff system(T-system) if the following determinant is strictly positive whenever $a \le t_0 < t_1 < ... < t_n \le b$

$$U \begin{pmatrix} u_0, & u_1, ..., u_n \\ t_0, & t_1, ..., t_n \end{pmatrix} = \begin{vmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_n) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_n) \\ \vdots & \vdots & \vdots & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_n) \end{vmatrix} > 0.$$

$\{u\}_{k=0}^{n}$ is called a Complete Tchebycheff system (CT-system) if $\{u\}_{k=0}^{m}$ is a T-system on $[a,b]$ for each $m = 0, 1, ..., n$. In the sequel, we shall deal with Extended Tchebycheff system and Extended Complete Tchebysheff system, which are defined below:

(i) $\{u\}_{k=0}^{n}$ on $[a,b]$ is called an Extended Tchebycheff system(ET-system) of order $p$, provided $u_i \in C^{p-1}[a,b]$, $i = 0, 1, ..., n$ and

$$U^* \begin{pmatrix} 0, & 1, & \dots, & n \\ t_0, & t_1, & \dots, & t_n \end{pmatrix}$$

$$= \begin{vmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_i) & u_0^{(1)}(t_i) & \cdots & u_0^{(q)}(t_i) & u_0(t_{i+q+1}) & \cdots & u_0(t_n) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_i) & u_1^{(1)}(t_i) & \cdots & u_1^{(q)}(t_i) & u_1(t_{i+q+1}) & \cdots & u_1(t_n) \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_i) & u_n^{(1)}(t_i) & \cdots & u_n^{(q)}(t_i) & u_n(t_{i+q+1}) & \cdots & u_n(t_n) \end{vmatrix} > 0$$

for all choices $a \le t_0 < t_1 < ... < t_{i-1} < t_i = t_{i+1} = ... = t_{i+q} < ... < t_n \le b$, $0 \le q \le p-1$, where equality occurs in groups of at most $p$ consecutive $t_i$ values.

In the above, $u_k^{(j)}(t)$ denotes the jth order derivative of $u_k$.

(ii) $\{u\}_{k=0}^{n}$ is called an Extended Complete Tchebysheff system(ECT-system) if $\{u\}_{k=0}^{m}$ is an ET-system on $[a,b]$ for each $m = 0, 1, ..., n$.

The following results are known and we skip the proofs.

(a) $\{u\}_{k=0}^{n}$ on $[a,b]$ is a Tchebysheff system if and only if every non-trivial linear combination $g(t) = \sum_{i=0}^{n} c_i u_i(t)$ has at most $n$ zeros, where $(c_0, c_1, ..., c_n) \neq (0, 0, ..., 0)$.

(b) $\{u\}_{k=0}^{n}$ on $[a,b]$ is an ECT-system if and only if for $k = 0, 1, ..., n$,

$$W(u_0, u_1, ..., u_k)(t) = \begin{vmatrix} u_0(t) & u_0^{(1)}(t) & \cdots & u_0^{(n)}(t) \\ u_1(t) & u_1^{(1)}(t) & \cdots & u_1^{(n)}(t) \\ \vdots & \vdots & \vdots & \vdots \\ u_n(t) & u_n^{(1)}(t) & \cdots & u_n^{(n)}(t) \end{vmatrix} > 0$$

where $W(u_0, u_1, ..., u_k)(t)$ is the Wronskian determinant, see Hartman(1964).

## 3. Klimpel's Flotation Recovery Model

Flotation model is a gravity separation process that originated from processing of minerals. They are widely used in mining engineering and have found wide application in industrial waste-water treatment. It is also useful in the concentration of a variety of dissolved chemical species often following a sorption process. They are based on an observation that was made in the earliest experimental kinetic studies of flotation, namely, that not all particles will be recovered by flotation no matter how much time they have in the flotation environment. Each particle type has an ultimate recovery that is less than 100 percent. The particles that do float are recovered at a rate that is governed by

a simple first-order kinetic law. Thus two kinetic parameters are required for each type of particle: the ultimate recovery and the kinetic constant. There are three types of flotation recovery models: exponential flotation recovery model, Klimpel's flotation recovery model and Agars flotation recovery model. We will discuss and treat the Klimpel's flotation recovery model.

The 2-parameter Klimpel's flotation recovery model is widely used in environmental science for metal recovery:

$$R(t, R_{max}, k) = R_{max}[1 - \frac{1}{kt}(1 - e^{-kt})] + \varepsilon. \tag{3.1}$$

$R(t, R_{max}, k)$ refers to the recovery of mineral or metal of interest; $t$ refers to time and $R_{max}$ is the ultimate recovery while $k$ is constant first-order rate.

For an easy presentation, we rewrite Model (3.1) in the following form:

$$y = \eta(t, \boldsymbol{\theta}) + \varepsilon = a[1 - \frac{1}{bt}(1 - e^{-bt})] + \varepsilon \tag{3.2}$$

where, $\boldsymbol{\theta} = (a, b)'$.

**Theorem 3.1.** *Under Model (3.2), for any arbitrary design $\xi$, there exists a design $\xi^*$ such that $I_\xi(\boldsymbol{\theta}) \leq I_{\xi^*}(\boldsymbol{\theta})$ (here and elsewhere, matrix inequalities are under the Loewner ordering). Here, $\xi^*$ is based on two design points including the upper bound point $T$ of the design space. When $I_\xi(\theta) \leq I_{\xi^*}(\theta)$, it means design $\xi^*$ is not inferior to $\xi$ under commonly used matrix based optimality criteria. An optimal design under Loewner ordering criterion does not exist in general. We have to consider optimal design under less restrictive criteria.*

*Proof.* For any arbitrary design $\xi = (t_i, \omega_i), i = 1, \ldots, n$, it can be shown that the Fisher information matrix $I_\xi$ can be written in the form:

$$I_\xi = P(\boldsymbol{\theta})C(\xi, \boldsymbol{\theta})P(\boldsymbol{\theta})^T. \tag{3.3}$$

Here

$$P(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 \\ -\frac{a}{b} & \frac{a}{b} \end{pmatrix} \text{ and } C(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \omega_i \begin{pmatrix} (1 + \frac{x_i}{ln(1-x_i)})^2 & x_i(1 + \frac{x_i}{ln(1-x_i)}) \\ x_i(1 + \frac{x_i}{ln(1-x_i)}) & x_i^2 \end{pmatrix},$$

with $x_i = 1 - e^{-bt_i}$. Since $0 < t \leq T$, we have $0 < x_i < 1 - e^{-bT}$.

Let $\Psi_1(x) = x^2$, $\Psi_2(x) = x(1 + \frac{x}{ln(1-x)})$, and $\Psi_3(x) = (1 + \frac{x}{ln(1-x)})^2$. We can verify that for any $0 < x < 1$,

(a) $\Psi_1'(x) > 0$;

(b) $\left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)' > 0$;

(c) $\left(\left(\frac{\Psi_3'(x)}{\Psi_1'(x)}\right)' / \left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)'\right)' > 0$;

(d) $\lim_{x \uparrow (1 - e^{-bT})} \frac{\Psi_2'(x)}{\Psi_1'(x)}(\Psi_1(1 - e^{-bT}) - \Psi_1(x)) = 0$.

The above inequalities are rather difficult to verify by hand since involving manipulation of derivative, therefore we use symbolic computational software MAPLE(Waterloo, Canada) to verify them. By Corollary 3 of Yang and Stufken (2009), there exists a design with two design points including the upper bound $T$, say $\xi^*$, such that one diagonal element and the off-diagonal element of $C(\xi^*, \boldsymbol{\theta})$ are the same as that of $C(\xi, \boldsymbol{\theta})$, and the remaining diagonal element is larger. Thus the conclusion follows. □

Theorem 3.1 shows that no matter what optimal designs we are looking for, we can always restrict ourself to two-points design including the upper bound $T$, irrespective of parameters of interest or the optimality criterion. Consequently, we are able to obtain analytical expression for some specific optimal designs.

**Corollary 3.1.** *Under Model (3.2), $\xi^*$ is the D-optimal design for $\boldsymbol{\theta}$, where, $\xi^* = \{(T, 0.5); (t^*, 0.5)\}$, where $t^* = -ln(1-x^*)/b$ and $x^*$ is the unique solution of the following equation*

$$(1 - e^{-bT}) \left( \frac{ln(1-x) + \frac{x}{1-x}}{ln^2(1-x)} \right) - \left( 1 - \frac{1}{bT}(1 - e^{-bT}) \right) = 0.$$

*Proof.* By Theorem 3.1, a $D$-optimal design, which maximizes determinant of $I_\xi$, for $\boldsymbol{\theta}$ must only have two design points including the upper bound $T$. Consequently, it must have equal weights on each support point, Silvey (1980). Thus a $D$-optimal design for $\boldsymbol{\theta}$ must maximize $g^2(x)$, where

$$g(x) = (1 - e^{-bT})(1 + \frac{x}{ln(1-x)}) - x[1 - \frac{1}{bT}(1 - e^{-bT})].$$

Clearly $g(x) = 0$ when $x = 1 - e^{-bT}$. We can show that $\lim_{x \downarrow 0} g(x) = 0$. Now let us consider the first derivative of $g(x)$,

$$g'(x) = (1 - e^{-bT}) \left( \frac{ln(1-x) + \frac{x}{1-x}}{ln^2(1-x)} \right) - \left( 1 - \frac{1}{bT}(1 - e^{-bT}) \right).$$

With some simple algebra, we can show that (i) $g'(x)$ is a strictly increasing function on $(0, 1)$; (ii) $\lim_{x \downarrow 0} g'(x) < 0$ and (iii) $g'(x) > 0$ when $x = 1 - e^{-bT}$. Thus, there must exist a unique solution $x^*$ such that $g'(x^*) = 0$, $g'(x) < 0$ when $0 < x < x^*$, and $g'(x) > 0$ when $x^* < x \leq 1 - e^{-bT}$. So $g(x)$ is minimized at $x = x^*$. Combining the fact the $g(x) = 0$ when $x = 1 - e^{-bT}$ and $\lim_{x \downarrow 0} g(x) = 0$, $g^2(x)$ must be maximized at $x^*$. □

From Theorem 3.1, we notice that the $D$-optimal design does not depend on parameter $a$. Once the values of upper bound $T$ and the parameter $b$ are given, the value of $x^*$ can be easily computed. For example, when $T = 100$, Table 1 provides the value of $x^*$ for some selected $b$ values.

Although it may not be easy to derive other optimal designs, we may be able to derive any optimal design numerically due to the simple format of $\xi^*$ in Theorem 3.1. For example, let us consider $A$-optimal designs for $\boldsymbol{\theta}$, which minimizes the trace of $I_\xi^{-1}$. With the explicit optimal weight formula provided

TABLE 1
*$x^*$ for the D-optimal design for $\boldsymbol{\theta}$ under Model (3.2)*

| $b$ | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 | 1.6 |
|---|---|---|---|---|---|---|---|---|
| $x^*$ | 8.3817 | 4.3395 | 2.9253 | 2.2060 | 1.7705 | 1.4786 | 1.2693 | 1.1119 |

TABLE 2
*$(x^*, \omega^*)$ for the A-optimal design for $\boldsymbol{\theta}$ under Model (3.2)*

| $b$ | 0.2 | 0.4 | 0.6 | 0.8 |
|---|---|---|---|---|
| $(x^*, \omega^*)$ | (7.1165,0.3903) | (3.6294, 0.5283) | (2.3780, 0.5929) | (1.7594,0.6261) |
| $b$ | 1.0 | 1.2 | 1.4 | 1.6 |
| $(x^*, \omega^*)$ | (1.3951,0.6449) | (1.1558,0.6563) | (0.9869,0.6638) | (0.8612,0.6689) |

by Biedermann, Dette, and Zhu (2006), the numerical search turns out to be one dimension which can be easily carried out. Table 2 provides the optimal designs for $T = 100$, $a = 1$ under some selected $b$ values.

## 4. 2-Parameter Chemical Kinetic Model

Kinetic models related to chemical reactions are widely used in chemical engineering and chemistry. The models are usually in the form of differential equations with two groups of parameters, the rate and the order of reaction; see Boroujerdi (2001).

Consider the consecutive reaction: $A \longrightarrow B$, with reaction order $\lambda$ and reaction rate $\theta$. The kinetic model is given by the differential equation:

$$\frac{d[A]}{dt} = -\theta[A]^{\lambda}, \tag{4.1}$$

where, $t \geq 0$ is the reaction time, $\lambda > 0$ is the reaction order, $\theta > 0$ is the reaction rate.

Given (4.1), with the initial conditions $A = 1, B = 0$ at $t = 0$, the model is determined as follows.

$$[A] = \eta(t, \theta, \lambda) + \varepsilon = [1 - (1 - \lambda)\theta t)]^{\frac{1}{1-\lambda}} + \varepsilon. \tag{4.2}$$

**Theorem 4.1.** *Under model (4.2), a D-optimal design for all parameters is supported on two points.*

*Proof.* By the extended equivalence theorem of Kiefer (1974) established by White (1973), it is enough to verify that there are at most 2 maximal points for $\mathbf{d}(\xi, x, \boldsymbol{\theta})$, directional derivative, which will be introduced shortly.
Let $\frac{1}{(1-\lambda)} = \beta$, $1 - \frac{\theta}{\beta}t = x$, and $\boldsymbol{\theta}' = (\theta, \beta)$. Note that a $D$-optimal design for $\lambda$ and $\theta$ is the same as that for $\beta$ and $\theta$. The Fisher information matrix for $\xi$ is

$$I_\xi = \begin{pmatrix} \frac{1}{\theta^2}\sum_{i=1}^n \omega_i \beta^2 x_i^{2\beta-2}(x_i - 1)^2 & \frac{1}{\theta}\sum_{i=1}^n \omega_i \beta x_i^{2\beta-1}(x_i - 1)(lnx_i + \frac{1}{x_i} - 1) \\ \frac{1}{\theta}\sum_{i=1}^n \omega_i \beta x_i^{2\beta-1}(x_i - 1)(lnx_i + \frac{1}{x_i} - 1) & \sum_{i=1}^n \omega_i (lnx_i + \frac{1}{x_i} - 1)^2 x_i^{2\beta} \end{pmatrix}.$$

We tacitly assume that $n \geq 2$ and that $x_i$'s are all distinct. This ensures the existence of the inverse of the Fisher information matrix,

$$I_\xi^{-1} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

with $m_{12} = m_{21}$.

Then,

$$\mathbf{d}(\xi, x, \boldsymbol{\theta}) = tr \begin{pmatrix} \frac{1}{\theta^2}\beta^2 x^{2\beta-2}(x-1)^2 & \frac{\beta}{\theta}x^{2\beta-1}(x-1)(lnx+\frac{1}{x}-1) \\ \frac{\beta}{\theta}x^{2\beta-1}(x-1)(lnx+\frac{1}{x}-1) & (lnx+\frac{1}{x}-1)^2 x^{2\beta} \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

$$= \frac{m_{11}}{\theta^2}\beta^2 x^{2\beta-2}(x-1)^2 + \frac{2}{\theta}m_{12}\beta x^{2\beta-1}(x-1)(lnx+\frac{1}{x}-1) + m_{22}x^{2\beta}(lnx+\frac{1}{x}-1)^2.$$

We obtain the first order derivative of $\mathbf{d}(\xi, x, \boldsymbol{\theta})$ with respect to $x$, which is denoted by $\mathbf{d}'(\xi, x, \boldsymbol{\theta})$:

$$\mathbf{d}'(\xi, x, \boldsymbol{\theta}) = x^{2\beta-3}\{k_1 x^2 + k_2 x + k_3 + k_4 x^2 lnx + k_5 x^2 (lnx)^2 + k_6 x lnx\} \quad (4.3)$$

where,
$k_1 = 2\frac{1}{\theta^2}m_{11}\beta^3 - 4\frac{1}{\theta}m_{12}\beta^2 + 2\frac{1}{\theta}m_{12}\beta + 2m_{22}\beta - 2m_{22}$
$k_2 = -4\frac{1}{\theta^2}m_{11}\beta^3 + 2\frac{1}{\theta^2}m_{11}\beta^2 + 8\frac{1}{\theta}m_{12}\beta^2 - 6\frac{1}{\theta}m_{12}\beta - 4m_{22}\beta + 4m_{22}$
$k_3 = 2\frac{1}{\theta^2}m_{11}\beta^3 - 2\frac{1}{\theta^2}m_{11}\beta^2 - 4\frac{1}{\theta}m_{12}\beta^2 + 4\frac{1}{\theta}m_{12}\beta + 2m_{22}\beta - 2m_{22}$
$k_4 = 4\frac{1}{\theta}m_{12}\beta^2 - 4m_{22}\beta + 2m_{22}$
$k_5 = 2m_{22}\beta$
$k_6 = -4\frac{1}{\theta}m_{12}\beta^2 + 2\frac{1}{\theta}m_{12}\beta + 4m_{22}\beta - 2m_{22} = 2(1-2\beta)(\frac{\beta}{\theta}m_{12} - m_{22})$.

We notice that $\mathbf{d}'(\xi, x, \boldsymbol{\theta})$ is composed of two Tchebycheff systems (Lemma 1 of Appendix) as indicated below:
$T_1 = \{1, x, x^2, x^2 lnx, x^2 (lnx)^2\}$ and $T_2 = \{1, x, x^2, x^2 lnx, x lnx\}$.

We consider two circumstances for $\beta$ in studying

$$(\frac{\beta}{\theta}m_{12} - m_{22}) = \frac{1}{\theta^2}\frac{1}{|I_\xi|}\sum_{i=1}^{n}\beta^2\omega_i x_i^{2\beta-2}(1-x_i)x_i lnx_i. \quad (4.4)$$

Recall that $\beta(=\frac{1}{1-\lambda})$ is either greater than 1 or smaller than 0. Consequently $(\frac{\beta}{\theta}m_{12} - m_{22})$ is less than 0 for all feasible values of $\beta$. Therefore $k_5$ is positive for $\beta > 1$ and negative for $\beta < 0$ while $k_6$ is positive when $\beta > 1$ and negative when $\beta < 0$.

Since both $T_1$ and $T_2$ are T-systems with positive determinant, any positive linear combination of the two systems is also a T-system.

For $\beta > 1$, $\{1, x, x^2, x^2 lnx, k_5 x^2 (lnx)^2 + k_6 x lnx\}$ is also a T-system with $k_5 > 0, k_6 > 0$.

For $\beta < 0$, $\{1, x, x^2, x^2 lnx, -[k_5 x^2 (lnx)^2 + k_6 x lnx]\}$ is also a T-system with $-k_5 > 0, -k_6 > 0$.

For both cases, there are at most 4 roots for $\mathbf{d}^{'}(\xi, x, \boldsymbol{\theta}) = 0$. Thus there are at most two local maximal points in $(0, T)$. On the other hand,
for $\beta > 1$, when $t \to 0$, $\mathbf{d}(\xi, x, \boldsymbol{\theta}) \to 0$; $t \to T$, $\mathbf{d}(\xi, x, \boldsymbol{\theta}) \to 0$, $T$ is the upper bound. And for $\beta < 0$, when $t \to 0$, $\mathbf{d}(\xi, x, \boldsymbol{\theta}) \to 0$; $t \to T$, $\mathbf{d}(\xi, x, \boldsymbol{\theta}) \to 0$. Consequently, the two boundary points cannot be the support points. Thus, a $D$-optimal design is precisely supported on 2 points. □

When $0 < \lambda < 1$, we are able to extend our result to any arbitrary optimal design. To make Model 4.2 meaningful, $1 - (1-\lambda)\theta t$ needs to be positive, i.e., $0 < t < \frac{1}{(1-\lambda)\theta}$.

For any arbitrary design $\xi = (t_i, \omega_i), i = 1, \ldots, n$, it can be shown that the Fisher information matrix $I_\xi$ can be written as follows:

$$I_\xi = P(\boldsymbol{\theta}) C(\xi, \boldsymbol{\theta}) P(\boldsymbol{\theta})^T. \tag{4.5}$$

$$P(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\beta}{\theta} & 0 \\ -1 & 1 \end{pmatrix} \text{ and } C(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{n} \omega_i \begin{pmatrix} x_i^{2\beta}(1-\frac{1}{x_i})^2 & x_i^{2\beta}(1-\frac{1}{x_i})lnx_i \\ x_i^{2\beta}(1-\frac{1}{x_i})lnx_i & x_i^{2\beta}ln^2 x_i \end{pmatrix},$$

with $x_i = 1 - \frac{\theta}{\beta} t_i$ and $\beta = \frac{1}{1-\lambda}$, where $0 < x_i < 1$.

Let $\Psi_1(x) = \int_0^x 2z^{2\beta-2}(1-1/z)dz$, $\Psi_2(x) = x^{2\beta}(1-\frac{1}{x})^2$, $\Psi_3(x) = x^{2\beta}(1-\frac{1}{x})lnx$, and $\Psi_4(x) = x^{2\beta}ln^2 x$. We can verify that for any $0 < x < 1$,

(a) $\Psi_1'(x) < 0$;
(b) $\left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)' > 0$;
(c) $\left(\left(\frac{\Psi_3'(x)}{\Psi_1'(x)}\right)' / \left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)'\right)' > 0$.

On the other hand, $h(x)$ is not always positive for all $x \in (0, 1)$, where

$$h(x) = \left(\left(\left(\frac{\Psi_4'(x)}{\Psi_1'(x)}\right)' / \left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)'\right)' / \left(\left(\frac{\Psi_3'(x)}{\Psi_1'(x)}\right)' / \left(\frac{\Psi_2'(x)}{\Psi_1'(x)}\right)'\right)'\right)'.$$

However, we can show that $h(x) > 0$ for $x \in [L(\lambda), U(\lambda)]$ with $L(\lambda) > 0$ and $U(\lambda) < 1$. $L(\lambda)$ and $U(\lambda)$ depend on the value of $\lambda$, and consequently of $\beta$. We do not have the explicit expression for $L(\lambda)$ and $U(\lambda)$ due to the rather complicated expression of $h(x)$. However, for a given $\lambda$, it is relatively easy to determine the value of $L(\lambda)$ and $U(\lambda)$ MAPLE(Waterloo, Canada).

Table 3 shows that $L(\lambda)$ is closed to 0 and $U(\lambda)$ is closed to 1. Even when $\lambda$ is closed to either 0 or 1, it seems Table 3 still holds. For example, $[L(\lambda), U(\lambda)] = [0.0110, 0.9959]$ and $[0.0250, 0.9963]$ when $\lambda = 0.001$ and 0.999, respectively. Now using Theorem 2 of Yang (2010), we have the following theorem.

**Theorem 4.2.** *Under Model (4.2), suppose that* $0 < \lambda < 1$ *and the induced design space is* $[L(\lambda), U(\lambda)]$. *For any arbitrary design* $\xi$, *there exists a design* $\xi^*$ *such that* $I_\xi(\boldsymbol{\theta}) \leq I_{\xi^*}(\boldsymbol{\theta})$. *Here,* $\xi^*$ *is based on two design points.*

TABLE 3
*$L(\lambda)$ and $U(\lambda)$ for some selected $\lambda$*

| $\lambda$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| $L(\lambda)$ | 0.0120 | 0.0131 | 0.0142 | 0.0154 | 0.0168 | 0.0182 | 0.0197 | 0.0213 | 0.0231 |
| $U(\lambda)$ | 0.9958 | 0.9959 | 0.9967 | 0.9962 | 0.9961 | 0.9957 | 0.9966 | 0.9959 | 0.9959 |

TABLE 4
*D-optimal designs for $\boldsymbol{\theta}$ for some selected $\lambda$*

| $\lambda$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| $t_1^*$ | 0.6348 | 0.6363 | 0.6372 | 0.6375 | 0.6375 | 0.6371 | 0.6366 | 0.6358 | 0.6350 |
| $t_2^*$ | 1.0927 | 1.1920 | 1.2997 | 1.4167 | 1.5439 | 1.6821 | 1.8324 | 1.9957 | 2.1731 |

Theorem 4.2 allows us to limit our search on two-point designs for any optimality when the induced design space is a subinterval of $[L(\lambda), U(\lambda)]$, regardless of parameters of interest. This makes it relatively easy to obtain a specific optimal design numerically. Although the result is not for entire induced design interval, $[L(\lambda), U(\lambda)]$ is closed to the whole induced space in general. Likely, an optimal design in the restricted space is also optimal in the entire design space. We can derive a specific optimal design in the restricted design space and then verify whether it is optimal in the entire design space via the general equivalence theorem. Notice that $D$-optimal designs have equal weights on the two points. Based on the information matrix, such optimal designs (after transformation) do not depend on the value of $\theta$. For example, let $t_1^*$ and $t_2^*$ be the two support points of the $D$-optimal design for $\boldsymbol{\theta}$ with some $\lambda$ and $\theta_1$. Then for the same $\lambda$ but different $\theta_2$, the two support points of the $D$-optimal design for $\boldsymbol{\theta}$ are $t_1^* \frac{\theta_1}{\theta_2}$ and $t_2^* \frac{\theta_1}{\theta_2}$. For convenience, we choose $\theta = 1$ and use the original scale to present the design (see Table 4). All optimal designs have been verified through general equivalence theorem in the entire design space $(0, \frac{1}{(1-\lambda)\theta})$.

Table 5 gives $A$-optimal designs for $\boldsymbol{\theta}$. Such designs depend on the values of both $\theta$ and $\lambda$. We have taken $\theta = 1$ in preparing Table 5.

## 5. 2n-parameter Compartment Models

Compartment models are important for the evaluation of efficacy and toxicity in drug developments. There is substantial literature investigating the nature

TABLE 5
*A-optimal designs for $\boldsymbol{\theta}$ for some selected $\lambda$*

| $\lambda$ | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|
| $(t_1^*, \omega_1^*)$ | ( 0.5295, 0.6265) | (0.5200, 0.5661) | (0.5108, 0.5162) | (0.5018, 0.4748) |
| $(t_2^*, \omega_2^*)$ | (1.1020, 0.3735) | (1.2171, 0.4339) | (1.3474, 0.4838) | (1.4952, 0.5252) |
| $\lambda$ | 0.5 | 0.6 | 0.7 | 0.8 |
| $(t_1^*, \omega_1^*)$ | (0.4932, 0.4417) | ( 0.4856, 0.4159) | ( 0.4789, 0.3963) | ( 0.4731, 0.3813) |
| $(t_2^*, \omega_2^*)$ | (1.6623, 0.5583) | (1.8509, 0.5841) | (2.0626, 0.6037) | ( 2.2998, 0.6187) |

of locally D-optimal designs for such models, see for example, Li and Majumdar (2008) and Fang and Hedayat (2008) from both theoretical and applied perspective.

$2n$-parameter compartment models is the sum of $n$ open one-compartment model with zero-order input and first-order output. The process is described as follows. Drug is introduced into the compartment by a constant zero-order input, with rate of input $k_0$; then the drug is eliminated from the compartment by the first-order elimination rate, with rate of output $KA$. At the start, the amount of drug is low and gradually the amount increases and the rate of elimination increases accordingly, and eventually the rate of input and the rate of output are equal. Therefore, the amount or concentration of drug remains constant. The following differential equation defines the rate of accumulation of drug in the compartment during a single infusion with constant rate:

$$\frac{dA}{dt} = k_0 - KA. \tag{5.1}$$

Given the initial amount of the drug: $A(t = 0) = 0$, we obtain integrated form:

$$A = \frac{k_0}{K}(1 - e^{-Kt}). \tag{5.2}$$

Compartment models are also extremely useful in modeling HIV dynamics, especially those multiple compartment models with large numbers of parameters. In modeling HIV dynamics within a host, biomathematicians and theoretical biologists have made great advances in the development of mathematical models to study the characteristics of HIV replication and HIV evolution. Most of these models are differential equations or compartmental models. See, for example, Ding and Wu (2000) and Han and Chaloner (2003). Here we start with 4-parameter compartment model (5.3).

Carathéodory's theorem provides an upper bound for any $k$-parameter $D$-optimal design formulation. From the above discussion of chemical kinetic model, we know $\mathbf{d}'(\xi, t, \boldsymbol{\theta})$ is very important in studying the numbers of support points. Here we also follow the same idea and study the upper bounds of the number of support points for compartment models.

$$y = \eta(t, \boldsymbol{\theta}) + \varepsilon = a_1(1 - e^{-\lambda_1 t}) + a_2(1 - e^{-\lambda_2 t}) + \varepsilon, \tag{5.3}$$

where, $0 \leq t$, $\lambda_1, \lambda_2 > 0$, $a_1, a_2 \in \mathbb{R}$, $\boldsymbol{\theta}' = (a_1, a_2, \lambda_1, \lambda_2)$.

**Theorem 5.1.** *Under model (5.3), the minimum upper bound of the number of support points for a locally D-optimal design is 6.*

*Proof.* After simplification, we obtain

$\mathbf{d}'(\xi, t, \boldsymbol{\theta})$
$= e^{-2\lambda_1 t}[-2\lambda_1 m_{33}t^2 + (2m_{33} + 4\lambda_1 m_{13})t - (2m_{13} + 2\lambda_1 m_{11})] + e^{-2\lambda_2 t}[-2\lambda_2 m_{44}t^2 + (2m_{44} + 4\lambda_2 m_{24})t - (2m_{24} + 2\lambda_2 m_{22})] + e^{-(\lambda_1+\lambda_2)t}[-2(\lambda_1 + \lambda_2)m_{34}t^2 + \{4m_{34} + (\lambda_1 + \lambda_2)(2m_{24} + 2m_{23})\}t - \{2m_{14} + 2m_{23} + 2(\lambda_1 + \lambda_2)m_{12}\}] + e^{-\lambda_1 t}[-\lambda_1(2m_{13} +$

$2m_{33})t+\{2m_{13}+2m_{33}+\lambda_1(2m_{11}+2m_{12})\}]+e^{-\lambda_2 t}[-\lambda_2(2m_{14}+2m_{24})t+\{2m_{14}+2m_{24}+\lambda_2(2m_{22}+2m_{12})\}]$.

We follow the same idea of the approach used in Fang and Hedayat (2008) to obtain the upper bound of the number of support points. We need to know the upper bounds for the numbers of roots for $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$. We will simplify $\mathbf{d}'(\xi,t,\boldsymbol{\theta})$ by taking derivative several times and what we obtain after the last derivative step is a linear combination of a T-system. Based on this, then we revisit the problem to determine bounds on the number of roots for $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$.

1) Divide both sides by $e^{-\lambda_1 t}$ and take derivative twice.
2) Divide both sides by $e^{(\lambda_1-\lambda_2)t}$ and take derivative twice.
3) Divide both sides by $e^{-\lambda_1 t}$ and take derivative three times.

Then solving $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$ is equivalent to solving $g_1(t)$, where

$$\begin{aligned} g_1(t) =&\lambda_1^2(\lambda_2-2\lambda_1)^2(\lambda_2-\lambda_1)^3e^{(\lambda_2-\lambda_1)t}[-2\lambda_1 m_{33}t^2+b_1{}'t+c_1{}'] \\ &+(\lambda_2)^2(\lambda_1-2\lambda_2)^2(\lambda_1-\lambda_2)^3e^{(\lambda_1-\lambda_2)t}[-2\lambda_2 m_{44}t^2+b_2{}'t+c_2{}'] \end{aligned} \tag{5.4}$$

Note that $g_1(t)$ is a linear combination of the following T-systems:
$\{e^{(\lambda_2-\lambda_1)t}, te^{(\lambda_2-\lambda_1)t}, t^2e^{(\lambda_2-\lambda_1)t}, e^{(\lambda_1-\lambda_2)t}, te^{(\lambda_1-\lambda_2)t}, t^2e^{(\lambda_1-\lambda_2)t}\}$
and $g_1(t)$ has at most 5 roots (see Appendix).

We now revisit $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$. Since $t=0$ is a local minimum point for $\mathbf{d}(\xi,t,\boldsymbol{\theta})$ and $g_1(t)$ has at most 5 roots, then $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$ has at most (5-1)+2+2+3=11 roots in $(0,\infty)$ excluding $t=0$. Here 2, 2 and 3 represent the numbers of times the derivatives are taken in steps 1), 2) and 3) respectively. Since as $t\to 0$, $\mathbf{d}(\xi,t,\boldsymbol{\theta})\to 0$; $t\to\infty$, $\mathbf{d}(\xi,t,\boldsymbol{\theta})\to c$, where $c=m_{11}+m_{22}+2m_{12}>0$. There are at most 6 support points(5(local maximum)+1(counting in the right boundary as a possible support point)) for the $D$-optimal design. □

This upper bound of 6 which we obtained here is much smaller than the upper bound provided by Carathéodory's theorem, which is $\frac{k(k+1)}{2}+1=\frac{4(4+1)}{2}+1=11$.

Finally we consider the general case of compartment models with $2n$ parameters:

$$y=\eta(t,\boldsymbol{\theta})+\varepsilon=\sum_{i=1}^{n}a_i(1-e^{-\lambda_i t})+\varepsilon \tag{5.5}$$

where, $t\geq 0$, $\lambda_i>0$, $a_i\in\mathbb{R}$, $\boldsymbol{\theta}=(a_1,a_2...a_n,\lambda_1,\lambda_2...\lambda_n)'$.

**Theorem 5.2.** *Under model (5.5), the smallest upper bound of the number of support points for a locally D-optimal design can be as few as* $\frac{3n^2+7n}{4}$ *when* $\frac{3n^2+7n-4}{2}$ *is even and* $\frac{3n^2+7n-2}{4}$ *when* $\frac{3n^2+7n-4}{2}$ *is odd for* $n=1,2,3....$

*Proof.* Borrowing the same idea as in the proof of Theorem 5.1, we need to decide on the number of roots for $\mathbf{d}'(\xi,t,\boldsymbol{\theta})=0$ for the case of general $n$. This

time we have

$$\mathbf{d}^{'}(\xi,t,\boldsymbol{\theta})=\sum_{i=1}^{n}e^{-2\lambda_i t}[at^2+bt+c]+\sum_{1\le i<j\le n}\sum e^{-(\lambda_i+\lambda_j)t}[dt^2+et+f]+\sum_{i=1}^{n}e^{-\lambda_i t}[gt+h], \tag{5.6}$$

where $a$, $b$, $c$, $d$, $e$, $f$, $g$ and $h$ are the coefficients. Numbers of roots for $\mathbf{d}^{'}(\xi,t,\boldsymbol{\theta})=0$ in $(0,\infty]$ is $[5+3(n-2)]-1+3\binom{n}{2}+2n=\frac{3n^2+7n-4}{2}$.

When $\frac{3n^2+7n-4}{2}$ is even, the number of support points is $\frac{3n^2+7n}{4}$.

When $\frac{3n^2+7n-4}{2}$ is odd, the number of support points is $\frac{3n^2+7n-2}{4}$. □

Both the upper bounds $\frac{3n^2+7n}{4}$ and $\frac{3n^2+7n-2}{4}$ are smaller than those from Carathéodory's theorem and the differences are of quadratic order.

## 6. Locally D-optimal Designs and D-efficiency for Parameter Misspecifications

Based on the initial guess of the unknown parameters, we can run the algorithm proposed in Yang, Biedermann, and Tang (2013) to obtain the locally D-optimal design $\xi$ evaluated at this guess value. To speed up the search, we also utilize the results in Sections 3 to 5 to set up the initial design points. The optimality results have been verified by the general equivalence theorem.

The D-efficiency of design $\xi_1$ relative to design $\xi_2$, according to Hedayat, Yan, and Pezzuto (1997), is given by the following index

$$eff_D(\xi_1,\xi_2)=\left(\frac{|\mathbf{I}(\xi_1,\boldsymbol{\theta})|}{|\mathbf{I}(\xi_2,\boldsymbol{\theta})|}\right)^{\frac{1}{k}}. \tag{6.1}$$

The efficiency index (6.1) is a tool to check if the design is robust to minor and major parameter misspecifications.

### *6.1. Klimpel's Flotation Recovery Model*

A *D*-optimal design has 2 support points with equal weights and the right boundary point is always a support point. Table 6 shows the D-efficiency by varying $b$ with $a=89\%$ for the design space of $(0,100]$ (the true value of $b$ is assumed to be 0.1).

### *6.2. 2-parameter Chemical Kinetic Model*

Table 7 shows the D-efficiency by varying $\beta>0$ with $\theta=0.15$ (the true value of $\beta$ is assumed to be 1.5) in the design space of $(0,\frac{\beta}{\theta})$. It is clear the loss in efficiency is substantial if $\beta>1.5$. If the initial guess for $\beta$ is 2.0, then the efficiency has fallen to 40%. However, the loss in efficiency is even worse if $\beta<1.5$. The efficiency for $\beta<0$ in the design space of $(0,400)$ in Table 8 is much better than that of $\beta>0$ since with the same magnitude of changes, we still have about 91% efficiency (assuming $\beta=-0.5$ is the true value).

TABLE 6
*D-efficiency when $b > 0$ deviates from the true value for $a = 89\%$*

| $b$ | Locally $D$-optimal design points | $D$-efficiency |
|---|---|---|
| 0.02 | $t_1^* = 38.31 \quad t_2^* = 100.00$ | 62.55% |
| 0.04 | $t_1^* = 29.31 \quad t_2^* = 100.00$ | 80.57% |
| 0.06 | $t_1^* = 22.91 \quad t_2^* = 100.00$ | 92.67% |
| 0.08 | $t_1^* = 18.61 \quad t_2^* = 100.00$ | 98.44% |
| **0.1** | $t_1^* = 15.51 \quad t_2^* = 100.00$ | **100%** |
| 0.20 | $t_1^* = 8.41 \quad t_2^* = 100.00$ | 86.59% |
| 0.4 | $t_1^* = 4.31 \quad t_2^* = 100.00$ | 58.70% |
| 0.6 | $t_1^* = 2.91 \quad t_2^* = 100.00$ | 43.68% |
| 0.8 | $t_1^* = 2.21 \quad t_2^* = 100.00$ | 34.84% |
| 1.0 | $t_1^* = 1.81 \quad t_2^* = 100.00$ | 29.33% |
| 1.8 | $t_1^* = 1.01 \quad t_2^* = 100.00$ | 17.31% |
| 2.4 | $t_1^* = 0.71 \quad t_2^* = 100.00$ | 12.41% |

TABLE 7
*D-efficiency when $\beta > 0$ deviates from the true value for $\theta = 0.15$*

| $\beta$ | Locally $D$-optimal design points | $D$-efficiency |
|---|---|---|
| 1.3 | $t_1^* = 2.52 \quad t_2^* = 5.00$ | 28.64% |
| **1.5** | $t_1^* = 4.25 \quad t_2^* = 8.92$ | **100%** |
| 1.8 | $t_1 = 4.25 \quad t_2^* = 9.81$ | 57.88% |
| 2.0 | $t_1^* = 4.25 \quad t_2^* = 10.29$ | 43.18% |
| 2.5 | $t_1^* = 4.25 \quad t_2^* = 11.21$ | 24.07% |
| 3.0 | $t_1^* = 4.25 \quad t_2^* = 11.88$ | 15.34% |
| 3.5 | $t_1^* = 4.24 \quad t_2^* = 12.37$ | 10.63% |
| 4.0 | $t_1^* = 4.24 \quad t_2^* = 12.75$ | 7.80% |

TABLE 8
*D-efficiency when $\beta < 0$ deviates from the true value for $\theta = 0.15$*

| $\beta$ | Locally $D$-optimal design points | $D$-efficiency |
|---|---|---|
| -0.15 | $t_1^* = 3.01 \quad t_2^* = 349.91$ | 82.38% |
| -0.2 | $t_1^* = 3.40 \quad t_2^* = 279.91$ | 87.74% |
| -0.25 | $t_1^* = 3.61 \quad t_2^* = 199.91$ | 93.22% |
| -0.3 | $t_1^* = 3.71 \quad t_2^* = 149.91$ | 96.77% |
| **-0.5** | $t_1^* = 4.11 \quad t_2^* = 86.41$ | **100%** |
| $-0.8$ | $t_1^* = 4.11 \quad t_2^* = 45.41$ | 95.11% |
| $-1.0$ | $t_1^* = 4.14 \quad t_2^* = 36.75$ | 91.05% |
| $-1.5$ | $t_1^* = 4.17 \quad t_2^* = 27.73$ | 83.58% |
| $-2.0$ | $t_1^* = 4.21 \quad t_2 = 24.11$ | 78.88% |
| $-3.5$ | $t_1^* = 4.21 \quad t_2^* = 20.11$ | 72.14% |
| $-4.0$ | $t_1^* = 4.21 \quad t_2^* = 19.51$ | 70.93% |

TABLE 9
*D-efficiency when* $\lambda_1$ *deviates from the true value for* $\{a_1 = 0.1, a_2 = 3.5, \lambda_2 = 4.0\}$

| $\lambda_1$ | Locally $D$-optimal design points | | | | $D$-efficiency |
|---|---|---|---|---|---|
| 0.2 | $t_1 = 0.23$ | $t_2^* = 1.04$ | $t_3^* = 4.34$ | $t_4^* = 10$ | 70.19% |
| 0.4 | $t_1 = 0.22$ | $t_2^* = 0.93$ | $t_3^* = 3.33$ | $t_4^* = 10$ | 88.79% |
| 0.6 | $t_1 = 0.21$ | $t_2^* = 0.85$ | $t_3^* = 2.68$ | $t_4^* = 10$ | 97.82% |
| **0.8** | $t_1 = 0.2$ | $t_2^* = 0.78$ | $t_3^* = 2.27$ | $t_4^* = 10$ | **100%** |
| 1.0 | $t_1 = 0.19$ | $t_2^* = 0.73$ | $t_3^* = 2.01$ | $t_4^* = 10$ | 98.88% |
| 1.5 | $t_1^* = 0.17$ | $t_2^* = 0.63$ | $t_3^* = 1.60$ | $t_4^* = 10$ | 90.84% |
| 1.8 | $t_1 = 0.16$ | $t_2^* = 0.55$ | $t_3^* = 1.42$ | $t_4^* = 10$ | 83.28% |
| 2.0 | $t_1^* = 0.16$ | $t_2^* = 0.56$ | $t_3^* = 1.38$ | $t_4^* = 10$ | 82.19% |

### *6.3. 2n-parameter Compartment Models*

For the 4-parameter compartment model, the investigation is done only for nonlinear parameters since the design does not depend on "linear" parameters.

Table 9 shows the performance of the D-efficiency in the design space of $(0, 10)$ by varying the initial guess value of $\lambda_1$ while fixing the other 3 parameter values (assuming the true value of $\lambda_1$ is 0.8). It is obvious that the locally D-optimal design for 4-parameter compartment model is very robust to the parameter misspecification which may happen in both directions. If the initial guess is 1.0, we still have 99% efficiency, which is very high.

For the 6-parameter compartment model, the locally D-optimal design is supported on six points. The right boundary point is always a support point, which is similar to the 2-parameter and 4-parameter compartment models. Table 10 shows the performance of the D-efficiency in the design space of $(0, 10)$ assuming the true value of $\lambda_1$ is 0.5. The locally D-optimal design for 6-parameter compartment model is also robust to the parameter misspecifications in both directions.

For 4-parameter compartment model, we show how the D-efficiency changes when $\lambda_1$ and/or $\lambda_2$ change in Table 11 assuming $\lambda_1 = 0.8$ and $\lambda_2 = 4.0$ is the true value.

## 7. Real Examples in the Mining Industry

### *7.1. Local Designs*

We presents two examples with real-life application in the mineral industry from Saleh (2010) and Yuan et al. (1996). In Saleh (2010), seven models (six flotation models and a 2n-parameter compartment model) were discussed. The iron ore sample was used to evaluate the fitting of six flotation models to experimental data. The optimal flotation model parameters were determined by the criteria of minimization of the absolute sum of squares of the deviation at given time between observed (experiment) and calculated recovery. In Yuan et al. (1996),

TABLE 10
*D-efficiency when $\lambda_1$ deviates from the true value for* $\{a_1 = 0.1, \lambda_2 = 4.0, a_2 = 3.5, a_3 = 0.9, \lambda_3 = 1.7\}$

| $\lambda_1$ | Locally $D$-optimal design points | $D$-efficiency |
|---|---|---|
| 0.1 | $t_1^* = 0.17 \quad t_2^* = 0.61 \quad t_3^* = 1.41 \quad t_4^* = 2.96 \quad t_5^* = 6.34 \quad t_6^* = 10$ | 87.95% |
| 0.2 | $t_1^* = 0.15 \quad t_2^* = 0.55 \quad t_3^* = 1.37 \quad t_4^* = 2.78 \quad t_5^* = 5.93 \quad t_6^* = 10$ | 93.97% |
| 0.3 | $t_1^* = 0.15 \quad t_2^* = 0.53 \quad t_3^* = 1.27 \quad t_4^* = 2.64 \quad t_5^* = 5.56 \quad t_6^* = 10$ | 97.76% |
| **0.5** | $t_1^* = 0.15 \quad t_2^* = 0.53 \quad t_3^* = 1.24 \quad t_4^* = 2.51 \quad t_5^* = 4.97 \quad t_6^* = 10$ | **100%** |
| 0.8 | $t_1^* = 0.50 \quad t_2^* = 0.51 \quad t_3^* = 1.16 \quad t_4^* = 2.28 \quad t_5^* = 4.27 \quad t_6^* = 10$ | 97.56% |
| 1.0 | $t_1^* = 0.50 \quad t_2^* = 0.50 \quad t_3^* = 1.11 \quad t_4^* = 2.14 \quad t_5^* = 3.94 \quad t_6^* = 10$ | 93.91% |
| 1.5 | $t_1^* = 0.13 \quad t_2^* = 0.49 \quad t_3^* = 1.01 \quad t_4^* = 1.92 \quad t_5^* = 3.40 \quad t_6^* = 10$ | 83.95% |
| 2.0 | $t_1^* = 0.12 \quad t_2^* = 0.43 \quad t_3^* = 0.94 \quad t_4^* = 1.75 \quad t_5^* = 3.07 \quad t_6^* = 10$ | 75.10% |
| 2.5 | $t_1^* = 0.11 \quad t_2^* = 0.38 \quad t_3^* = 0.90 \quad t_4^* = 1.62 \quad t_5^* = 2.84 \quad t_6^* = 10$ | 67.16% |
| 3.0 | $t_1^* = 0.11 \quad t_2^* = 0.38 \quad t_3^* = 0.82 \quad t_4^* = 1.53 \quad t_5^* = 2.67 \quad t_6^* = 10$ | 60.91% |

TABLE 11
*D-efficiency when $\lambda_1$ and/or $\lambda_2$ deviates from the true value for* $\{a_1 = 0.1, a_2 = 3.5\}$

| $\lambda_2$ | $\lambda_1$ | | | | |
|---|---|---|---|---|---|
| | 0.2 | 0.6 | **0.8** | 1.2 | 1.6 |
| 2.0 | 42.19% | 72.48% | 80.88% | 88.03% | 92.45% |
| 3.0 | 60.42% | 90.65% | 96.55% | 98.53% | 98.63% |
| **4.0** | 70.19% | 97.79% | **100%** | 98.86% | 94.97% |
| 5.0 | 73.86% | 98.17% | 98.00% | 94.77% | 89.99% |
| 6.0 | 74.42% | 96.10% | 93.70% | 89.13% | 84.75% |

six kinetic flotation models were tested for applicability to batch flotation time-recovery profiles for a complex sulphide ore.

For both examples, assuming the estimated parameters as true value, the corresponding optimal design, A or D-optimal was obtained through the simulation. The optimal design was compared with the design of the conducted experiment by the efficiency index (6.1) to demonstrate if the optimal design is used, how much we can improve the estimation precision. The results are shown in Tables 12 and 13.

TABLE 12
*D- and A- efficiencies for Saleh (2010) paper*

| Parameters | a=0.5221 b=2.0522 | Efficiency |
|---|---|---|
| Design in Saleh (2010) | $t_1^* = 0.5 \quad t_2^* = 1.0 \quad t_3^* = 2.0 \quad t_4^* = 4.0 \quad t_5^* = 6.0 \quad t_6^* = 8.0$ | 81.50%(D-) |
| | $\omega_1^* = \omega_2^* = \omega_3^* = \omega_4^* = \omega_5^* = \omega_6^* = 0.1667$ | 56.22%(A-) |
| D-optimal Design | $t_1^* = 0.8044 \quad t_2^* = 8$ | |
| | $\omega_1^* = 0.5 \quad \omega_2^* = 0.5$ | |
| A-optimal Design | $t_1^* = 0.6173 \quad t_2^* = 8$ | |
| | $\omega_1^* = 0.6814 \quad \omega_2^* = 0.3186$ | |

TABLE 13
*D- and A- efficiencies for Yuan et al. (1996) paper*

| Parameters | a=0.9581 b=5.411 | Efficiency |
|---|---|---|
| Design in Yuan et al. (1996) | $t_1^* = 1 \quad t_2^* = 2.5 \quad t_3^* = 5.5$ | 45.54%(D-) |
| | $\omega_1^* = \omega_2^* = \omega_2^* = 0.333$ | 13.35%(A-) |
| D-optimal Design | $t_1^* = 0.3174 \quad t_2^* = 5.5$ | |
| | $\omega_1^* = 0.5 \quad \omega_2^* = 0.5$ | |
| A-optimal Design | $t_1^* = 0.2427 \quad t_2^* = 5.5$ | |
| | $\omega_1^* = 0.6841 \quad \omega_2^* = 0.3159$ | |

TABLE 14
*Two-Stage D- and A-optimal Designs for Saleh (2010) paper*

| Parameters | a=0.5221 b=2.0522 |
|---|---|
| 1st Stage Design(Existing Design) | $t_1^* = 0.5 \quad t_2^* = 1.0 \quad t_3^* = 2.0 \quad t_4^* = 4.0 \quad t_5^* = 6.0 \quad t_6^* = 8.0$ |
| | $\omega_1^* = \omega_2^* = \omega_3^* = \omega_4^* = \omega_5^* = \omega_6^* = 0.1667$ |
| 2nd Stage D-optimal Design | $t_1^* = 0.7804 \quad t_2^* = 8.0$ |
| | $\omega_1^* = 0.5225 \quad \omega_2^* = 0.4775$ |
| 2nd Stage A-optimal Design | $t_1^* = 0.6173 \quad t_2^* = 8.0$ |
| | $\omega_1^* = 0.7701 \quad \omega_2^* = 0.2299$ |

### *7.2. Adaptive Designs*

Locally optimal design approach is based on the "best guess" of the unknown parameters. How can we obtain most reliable "best guess"? A natural way is to implement adaptive designs. An initial experiment is conducted to obtain the working idea about the unknown parameters. The estimate at one stage can serve as the "best guess" on which the next stage design can be based. At a second stage, the question then becomes how to select design $\xi_2$ such that the total information matrix $I_{\xi_1}(\theta) + rI_{\xi_2}(\theta)$ is optimized under some pre-specified optimality criterion. Here $r$ refers to the ratio of the sample size of the second stage to that of the first stage. We implement two-stage adaptive D- and A-optimal designs for the two real examples by adding 10 points to the given designs, where the first-stage designs are the designs used in the experiment in the papers by Saleh (2010) and Yuan et al. (1996). The results are shown in tables 14 and 15.

TABLE 15
*Two-Stage D- and A-optimal Design for Yuan (1996) paper*

| Parameters | a=0.9581 b=5.411 |
|---|---|
| 1st Stage Design(Existing Design) | $t_1^* = 1 \quad t_2^* = 2.5 \quad t_3^* = 5.5$ |
| | $\omega_1^* = \omega_2^* = \omega_3^* = 0.3333$ |
| 2nd Stage D-optimal Design | $t_1^* = 0.3097 \quad t_2^* = 5.5$ |
| | $\omega_1^* = 0.5884 \quad \omega_2^* = 0.4116$ |
| 2nd Stage A-optimal Design | $t_1^* = 0.2427 \quad t_2^* = 5.5$ |
| | $\omega_1^* = 0.8028 \quad \omega_2^* = 0.1972$ |

### *7.3. Simulation Study*

Clearly the derived adaptive designs depend on the estimated parameter values from the first stage design. This implies that the adaptive designs are not uniquely determined since the estimated parameter values are random variables. A simulation study is conducted to evaluate the performance of the derived adaptive designs. We use the designs in Saleh (2010) and Yuan et al. (1996) as the first stage designs. We also use the estimated parameters in Saleh (2010) and Yuan et al. (1996) as the true parameters values. The simulation study consists of three steps: (i) drawing a random variable from the multivariable normal distribution based on the estimation from the first stage design; (ii) deriving the adaptive optimal design based on the drawn parameter values; and (iii) evaluating the efficiency of the derived adaptive design using a local optimal design as a benchmark. We repeated the process 1000 times and we obtained the distribution of efficiency. The mean and the standard deviation of D- and A-efficiencies are reported respectively for both examples: they are 0.86(0.08)and 0.54(0.02) for Saleh (2010); 0.88(0.07) and 0.38(0.01) for Yuan et al. (1996). The histograms of the efficiencies are presented in Fig.1.

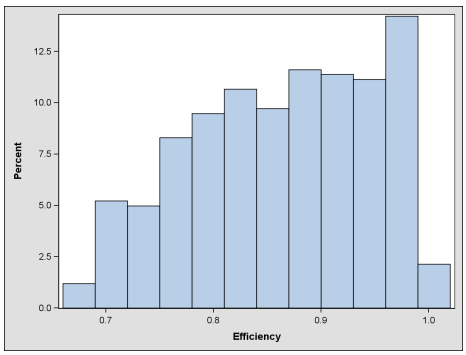
Another interesting question is, what is the relative efficiency if we use the existing design in the paper at both the first and the second stages compared to the situation that if we use the adaptive optimal design at the second stage? We found, after running the similar simulation study, the relative D-efficiencies for Saleh and Yuan's paper are 0.78(0.17) and 0.35(0.10) and the relative A-efficiencies are 0.30(0.07) and 0.14(0.01). Both examples show that using adaptive designs at the second stage is better than using the existing designs at both stages.

## 8. Discussion

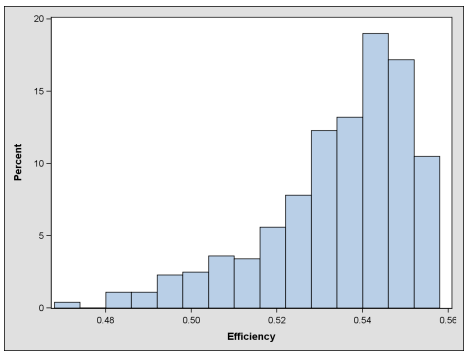
Searching for optimal designs is important and also complicated. Knowing the upper-bound of the number of support points can greatly simplify the search process, numerically or analytically. We have a complete answer for the Klimpel's flotation recovery model. However, we only have partial answer for 2-parameter chemical kinetic model as well as compartment models. Extensive numerical studies suggest that optimal designs can be based on saturated designs. How to prove this conclusion remains an open problem.

Optimal designs for compartment models are less sensitive, while both Klimpel's flotation recovery model and 2-parameter chemical kinetic model suffer significant efficiency loss when parameters are moderately deviated from their true values. Bayesian or minimax optimal designs could be a remedy for this problem.

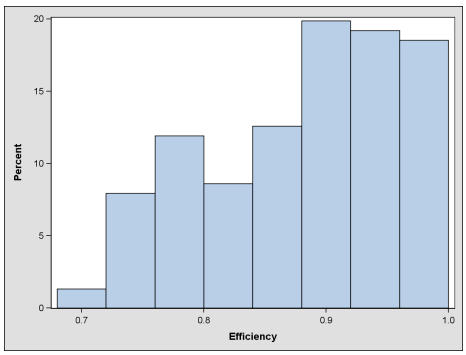
Adaptive designs are promising since with any existing design at the first stage and initial information about parameter values, adaptive optimal designs can be derived for the second stage for certain optimality criterion. Our simulation also shows adaptive designs are comparable to local optimal designs in some cases. Given the first stage existing designs, using adaptive designs at the second stage is also efficient than using the existing design at the second stage.



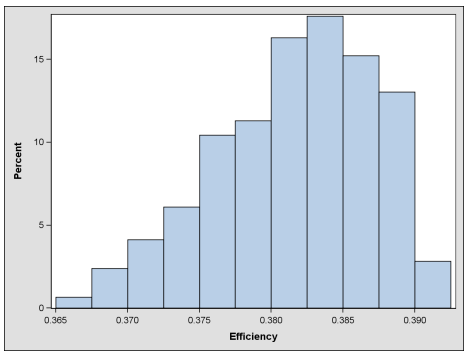

(a) D-efficiency: Saleh (2010)




(b) A-efficiency: Saleh (2010)




(c) D-efficiency: Yuan et al. (1996)




(d) A-efficiency: Yuan et al. (1996)

FIG 1. *Distribution of D- and A- Efficiencies*

## Appendix

**Lemma 1.** *The following two systems are ECT-systems.*

*1)* $T_1 = \{1, x, x^2, x^2 lnx, x^2 (lnx)^2\}$ *is an ECT-system with* $W_1(u_0, u_1, ..., u_k)(x) > 0$.

*2)* $T_2 = \{1, x, x^2, x^2 lnx, xlnx\}$ *is an ECT-system with* $W_2(u_0, u_1, ..., u_k)(x) > 0$.

*Proof.* We give a detailed proof for $T_1$. Similar proof for $T_2$ can be derived similarly. It can be shown that

$W_1(u_0)(x) = 1 > 0$,

$W_1(u_0, u_1)(x) = 1 > 0$,

$$W_1(u_0, u_1, u_2)(x) = \begin{vmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ x^2 & 2x & 2 \end{vmatrix} = 2 > 0,$$

$$W_1(u_0, u_1, u_2, u_3)(x) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ x & 1 & 0 & 0 \\ x^2 & 2x & 2 & 0 \\ x^2 lnx & 2xlnx + x & 2lnx + 3 & \frac{2}{x} \end{vmatrix} = \frac{4}{x} > 0,$$

and $W_1(u_0, u_1, u_2, u_3, u_4)(x)$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ x & 1 & 0 & 0 & 0 \\ x^2 & 2x & 2 & 0 & 0 \\ x^2 lnx & 2xlnx + x & 2lnx + 3 & \frac{2}{x} & -\frac{2}{x^2} \\ x^2 (lnx)^2 & 2x(lnx)^2 + 2xlnx & 2(lnx)^2 + 6lnx + 2 & \frac{4}{x} lnx + \frac{6}{x} & -\frac{4}{x^2} lnx - \frac{2}{x^2} \end{vmatrix} = \frac{16}{x^3} > 0.$$

Therefore, $\{1, x, x^2, x^2 lnx, x^2 (lnx)^2\}$ is a ECT-System, which is also a T-System. □

**Lemma 2.** $g_1(t)$ *defined in (5.4) is a linear combination of T-Systems and has at most 5 roots.*

*Proof.* Let $g_1(t) = a_1 t^2 e^{(\lambda_2 - \lambda_1)t} + b_1 t e^{(\lambda_2 - \lambda_1)t} + c_1 e^{(\lambda_2 - \lambda_1)t} + a_2 t^2 e^{(\lambda_1 - \lambda_2)t} + b_2 t e^{(\lambda_1 - \lambda_2)t} + c_2 e^{(\lambda_1 - \lambda_2)t}$,

which can be re-written as

$g_1(t) = \frac{1}{e^{(\lambda_2 - \lambda_1)t}} [e^{2(\lambda_2 - \lambda_1)t}(a_1 t^2 + b_1 t + c_1) + a_2 t^2 + b_2 t + c_2]$.

To find out the number of roots for $g_1(t) = 0$, consider

$\tilde{g_2}(t) = [e^{2(\lambda_2 - \lambda_1)t}(a_1 t^2 + b_1 t + c_1) + a_2 t^2 + b_2 t + c_2] = 0$.

Since $\frac{d^3 \tilde{g_2}(t)}{dt^3} = e^{2(\lambda_2 - \lambda_1)t}(a_1^{'} t^2 + b_1^{'} t + c_1^{'})$, has at most 2 roots, $g_1(t)$ has at most 5 roots. Therefore, it is a T-system. □

## Acknowledgements

Especially, we want to thank professor Bikas Sinha from Indian Statistical Institute for his generous help on revising the manuscript.

## References


[1] Bernshtein, S. N. (1937). The extremal properties of polynomials and best approximation of continuous functions of a real variable. *Moscow-Leningrad (In Russian).*

[2] Biedermann, S. and Dette, H. and Zhu, W. (2006). Optimal designs for dose-response models with restricted design spaces. *Journal of the American Statistical Association* **101(474)** 747-759.

[3] Boroujerdi, M. (2001). *Pharmacokinetics: Principles and Applications.* McGraw-Hill/Appleton & Lange.

[4] Chebyshev, P. L. (1859). Questions on smallest quantities connected with the approximate representation of functions. *Collected works, Moscow-Leningrad (In Russian)* **2** 151-238.

[5] Ding, A. A. and Wu, H. (2000). A Comparison study of models and fitting procedures for biphasic viral dynamics in HIV-1 infected patients treated with antiviral therapies. *Biometrics* **56(1)** 293-300.

[6] Fang, X. and Hedayat, A. S. (2008).Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics* **36** 428-444.

[7] Gieschke, R. and Steimer, J. L. (2000). Pharmacometrics: modelling and simulation tools to improve decision making in clinical drug development. *European Journal of Drug Metabolism and Pharmacokinetics***25(1)** 49-58.

[8] Godfrey, K. (1983). *Compartmental Models and Their Application.* Academic Press.

[9] Han, C. and Chaloner, K. (2003). D- and C-optimal designs for exponential regression models used in viral dynamics and other applications. *Journal of Statistical Planning and Inference* **115** 585-601.

[10] Hartman, P. (1964). *Ordinary Differential Equations.* Wiley, New York.

[11] Hedayat, A. S. and Yan, B. and Pezzuto, J. M.(1997). Optimum designs for fitting dose-response curves based on raw optical density data. *Journal of the American Statistical Association* **92** 1132-1140.

[12] Jacquez, J. A. (1985). *Compartmental Analysis in Biology and Medicine.* The University of Michigan Press, 2nd edition.

[13] Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Pure and Applied Mathematics XV. Wiley, New York.

[14] Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **5**, 849-879.

[15] Li, G. and Majumdar, D. (2008). D-optimal design for logistic models with three and four parameters. *Journal of Statistical Planning and Inference* **138** 1950-1959.

[16] Meibohm, B. and Derendorf, H. (2002). Pharmacokinetic/pharmacodynamic studies in drug product development. *Journal of Pharmaceutical Sciences***91(1)** 18-31.

[17] Parekh, B. K. and Miller, J. D. (1999). *Advances in Flotation Tech-*

*nology*. Society for Mining Metallurgy & Exploration.

[18] Saleh, A. M. (2010). A study on the performance of second order models and two phase models in iron ore flotation. *Physicochemical Problems of Mineral Processing* **44** 215-230.

[19] Silvey, S. D. (1980). *Optimal Design: An Introduction to The Theory for Parameter Estimation*. Chapman & Hall, London.

[20] White, L. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika* **60** 345-348.

[21] Yang, M. and Biedermann, S. and Tang, E.(2013). On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association* **in press**.

[22] Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Annals of Statistics* **37** 518-541.

[23] Yang, M. (2010). On the de la Garza phenomenon. *Annals of Statistics*, **38** 2499-2524.

[24] Yuan, X. M. and Pålsson, B. I. and Forssberg, K. S. E.(1996). Statistical interpretation of flotation kinetics for a complex sulphide ore. *Minerals Engineering* **9(4)** 429-442.