Published in final edited form as:
Inverse Probl. 2011 July 1; 27(7): . doi:10.1088/0266-5611/27/7/075002.

Comparison of Optimal Design Methods in Inverse Problems

H. T. Banks[1,2], **Kathleen Holm**[1,2], and **Franz Kappel**[1,2,3]

[1]Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8213

[2]Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8212

[3]Institute for Mathematics and Scientific Computation, University of Graz, Graz, Austria A8010

Abstract

Typical optimal design methods for inverse or parameter estimation problems are designed to choose optimal sampling distributions through minimization of a specific cost function related to the resulting error in parameter estimates. It is hoped that the inverse problem will produce parameter estimates with increased accuracy using data collected according to the optimal sampling distribution. Here we formulate the classical optimal design problem in the context of general optimization problems over distributions of sampling times. We present a new Prohorov metric based theoretical framework that permits one to treat succinctly and rigorously any optimal design criteria based on the Fisher Information Matrix (FIM). A fundamental approximation theory is also included in this framework. A new optimal design, *SE*-optimal design (*standard error* optimal design), is then introduced in the context of this framework. We compare this new design criteria with the more traditional *D*-optimal and *E*-optimal designs. The optimal sampling distributions from each design are used to compute and compare standard errors; the standard errors for parameters are computed using asymptotic theory or bootstrapping and the optimal mesh. We use three examples to illustrate ideas: the Verhulst-Pearl logistic population model [13], the standard harmonic oscillator model [13] and a popular glucose regulation model [16, 19, 29].

Keywords

Optimal design methods; Least squares inverse problems; Fisher Information matrix; D-optimal; E-optimal; SE-optimal

1 Introduction

Mathematical models are used to describe dynamics arising from biological, physical and engineering systems. If the parameters in the model are known, the model can be used for simulation, prediction, control design, etc. However, typically one does not have accurate values for the parameters. Instead, one must estimate the parameters using experimental data. The simulation and predictive capabilities of the model depend on the accuracy of the parameter estimates. A major question that experimentalists and inverse problem investigators alike often face is how to best collect the data to enable one to efficiently and accurately estimate model parameters. This is the well-known and widely studied *optimal design* problem.

AMS subject classifications: 93B51, 62B10, 62B15, 62G08, 62H12, 90C31.

Traditional optimal design methods (D-optimal, E-optimal, c-optimal) [1, 2, 14, 20, 21] use information from the model to find the sampling distribution or mesh for the observation times (and/or locations in spatially distributed problems) that minimizes a design criterion, quite often a function of the Fisher Information Matrix (FIM). Experimental data taken on this optimal mesh is then expected to result in accurate parameter estimates.

Here we formulate the classical optimal design problem in the context of general optimization problems over distributions of sampling times. We present a new Prohorov metric based theoretical framework that allows one to treat succinctly and rigorously any optimal design criteria based on the Fisher Information Matrix (FIM). A fundamental approximation theory is also included in this framework. A new optimal design, *SE*-optimal design (*standard error* optimal design), is then introduced in the context of this framework. We compare this new design criteria with the more traditional *D*-optimal and *E*-optimal designs. We consider the performance of these three different optimal design methods for three different dynamical systems: the Verhulst-Pearl logistic population model, a harmonic oscillator model and a simple glucose regulation model. *SE*-optimal design was first introduced in [9]. The goal of *SE*-optimal design is to find the observation times $\tau = \{t_i\}$ that minimize the sum of squared normalized standard errors of the estimated parameters as defined by asymptotic distribution results from statistical theories [7, 11, 18, 28]. *D*-optimal and *E*-optimal design methods minimize functions of the covariance in the parameter estimates [2, 14, 21]. *D*-optimal design finds the mesh that minimizes the volume of the confidence interval ellipsoid of the asymptotic covariance matrix. *E*-optimal design minimizes the largest principle axis of the confidence interval ellipsoid of the asymptotic covariance matrix.

The three examples we use were chosen for several reasons. First, they offer qualitatively different solutions: a monotone increasing trajectory, an oscillatory one and a vector observation or model output. Moreover, the first two can be studied with analytical solutions while the third requires a computational solution for the output. While analytical solutions are often available in drug design studies employing simple compartmental models, the latter is more common in many applications of interest. It is well known that computational errors of the model output can play a significant role in inverse and optimization problems. However, this problem has been frequently studied in the mathematical literature on inverse problems (see [12] for some early theoretical and computational studies along with reassuring convergence concepts on such questions). By adopting high resolution computational methods (as we have done in our third example here), one can in many applications essentially alleviate difficulties that may arise from using computational representations of output in both design and implementation of inverse problems. Indeed, the error in measurement and observations is usually much more of a concern. Our analysis of the third example below did not reveal any concerns related with use of a computational model output as compared to analytical model outputs in carrying out optimal design comparisons.

In an effort to provide a reasonably fair comparison, for each optimal design method, standard errors are computed by several methods using the optimal mesh. The optimal design methods are compared based on these standard errors. Not surprisingly, we find that *SE*-optimal design often results in smaller standard errors compared with the other optimal design method; this is likely because *SE*-optimal design optimizes directly on the standard errors themselves while the *D*-optimal and *E*-optimal methods minimize other functions related to the standard errors through the FIM.

2 Optimal Design Formulations

Following [9], we introduce a formulation of *ideal* inverse problems in which continuous in time observations are available; while not practical, the associated considerations provide valuable insight. A major question in this context is how to choose sampling distributions in an intelligent manner. Indeed, this is the fundamental question treated in the optimal design literature and methodology.

Underlying our considerations is a *mathematical model*

$$\begin{aligned} \dot{x}(t) &= g(t, x(t), q), \\ x(0) &= x_0, \\ f(t,\theta) &= C(x(t,\theta)), \qquad t \in [0,T], \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the vector of state variables of the system, $f(t,\theta) \in \mathbb{R}^m$ is the vector of observable or measurable outputs, $q \in \mathbb{R}^r$ are the system parameters, $\theta = (q, x_0) \in \mathbb{R}^p$, $p = r + n$ is the vector of system parameters plus initial conditions x_0, while g and C are mappings $\mathbb{R}^{1+n+r} \to \mathbb{R}^n$ and $\mathbb{R}^n \to \mathbb{R}^m$, respectively. To consider measures of uncertainty in estimated parameters, one also requires a *statistical model* [7]. Our statistical model is given by the stochastic process

$$Y(t) = f(t,\theta_0) + \mathcal{E}(t). \tag{2}$$

Here $\mathcal{E}$ is a noisy random process representing measurement errors and, as usual in statistical formulations [7, 9, 18, 28], θ_0 is a hypothesized "true" value of the unknown parameters. We make the following standard assumptions on the random variable $\mathcal{E}(t)$:

$$\begin{aligned} E(\mathcal{E}(t)) &= 0, \qquad t \in [0,T], \\ \mathrm{Var}\mathcal{E}(t) &= \sigma(t)^2 I, \qquad t \in [0,T], \\ \mathrm{Cov}(\mathcal{E}(t)\mathcal{E}(s)) &= \sigma(t)^2 I \delta(t-s), \qquad t, s \in [0,T], \end{aligned}$$

where $\delta(s) = 1$ for $s = 0$ and $\delta(s) = 0$ for $s \neq 0$. A realization of the observation process is given by

$$y(t) = f(t,\theta_0) + \varepsilon(t), \quad t \in [0,T],$$

where the measurement error $\varepsilon(t)$ is a realization of $\mathcal{E}(t)$.

The statistical model we assume is a classical first model widely found in regression analysis [18, 28] and is often employed in development of ideas (asymptotic distribution theory, statistical comparison tests, etc.) as well as in certain applications. Two (the second and third) of the assumptions above are sometimes violated in repeated measurement data problems, especially in economic data and in biological data with high frequency sampling using model response dependent assays. The independence and zero correlation statistical model assumptions can often be checked/supported via residual plots [7, 13, 18, 28] (after carrying out the inverse or regression problem fits with ordinary least squares). More sophisticated error models (e.g., nonconstant variance, nontrivial correlation) might also require different inverse problem formulations [7] (e.g., generalized least squares or maximum likelihood if sufficient information is known about the distributions of errors). Our choice of this simple statistical model was motivated primarily for two reasons: (i) it is simple, widely known/used and easy to use in comparison of designs, and (ii) our own

efforts (e.g., see [4, 6, 8, 12, 13] and the references therein) in numerous inverse problems with longitudinal data for HIV patients, data for insect and marine populations, and vibrational engineering data often have dealt with data in which independence of sampling appeared to be reasonable. Even when correlation might be present, it is often assumed fast decaying with respect to distance/time between observations (this is a standard correlation model-see [18, 28]) and independence is a reasonable assumption for low frequency sampling times. However, there are numerous situations in practice where independence is readily violated (e.g., economics and biology) and techniques in the presence of nontrival data correlations are an active area of optimal design research-see [24, 26, 30] and the references therein. Comparison of the design methods investigated here in the context of nontrivial correlation models such as those discussed in [24, 30] and [28, Chap 6], are the focus of our current research efforts and methodology results will be reported in a forthcoming paper. We note that all of the ideas presented here are readily applicable to problems with correlated noise in the data although the definition of the generalized Fisher matrix given below must be modified to include a correlation matrix factor corresponding to the assumed noise model (see [24, 30]). We remark that [30] also contains a nice survey of the vast statistical literature on design of experiments of which computational design methods for correlated observation problems comprise only a very small part.

We introduce a generalized weighted least squares criterion

$$J(y,\theta)=\int_0^T \frac{1}{\sigma(t)^2}|y(t)-f(t,\theta)|^2 dP(t), \tag{3}$$

where P is a general measure on $[0, T]$. We seek the parameter estimate $\widehat{\theta}$ by minimizing $J(y, \theta)$ for θ. Since P represents a weighting of the difference between data and model output, we can, without loss of generality, assume that P is a bounded measure on $[0, T]$.

If, for points $\tau = \{t_i\}$, $t_1 < \cdots < t_N$ in $[0, T]$, we take

$$P_\tau=\sum_{i=1}^{N}\Delta_{t_i}, \tag{4}$$

where Δ_a denotes the Dirac delta distribution with atom at $\{a\}$, we obtain

$$J_d(y,\theta)=\sum_{i=1}^{N}\frac{1}{\sigma(t_i)^2}|y(t_i)-f(t_i,\theta)|^2, \tag{5}$$

which is the weighted least squares cost functional for the case where we take a finite number of measurements in $[0, T]$. Of course, the introduction of the measure P allows us to change the weights in (5) or the weighting function in (3). For instance, if P is absolutely continuous with density $m(\cdot)$ the error functional (3) is just the weighted L^2-norm of $y(\cdot) - f(\cdot, \theta)$ with weight $m(\cdot)/\sigma(\cdot)^2$.

To facilitate our discussions we introduce the *Generalized Fisher Information Matrix* (GFIM)

$$F(P,\theta_0)=\int_0^T \frac{1}{\sigma^2(s)}\nabla_\theta^{\mathrm{T}} f(s,\theta_0)\,\nabla_\theta f(s,\theta_0)\,dP(s), \tag{6}$$

where ∇_θ is a row vector given by $(\partial_{\theta_1}, \ldots, \partial_{\theta_p})$ and hence $\nabla_\theta f$ is an $m \times p$ matrix. It follows that the usual discrete FIM corresponding to P_τ as in (4) is given by

$$F(\tau) = F(P_\tau, \theta_0) = \sum_{j=1}^{N} \frac{1}{\sigma^2(t_j)} \nabla_\theta f(t_j, \theta_0)^\top \nabla_\theta f(t_j, \theta_0). \tag{7}$$

Subsequently we simplify notation and use $\tau = \{t_i\}$ to represent the dependence of $P = P_\tau$ on τ when it has form (4). When one chooses P as simple Lebesgue measure then the GFIM reduces to the continuous FIM

$$F_C = \int_0^T \frac{1}{\sigma^2(s)} \nabla_\theta f(s, \theta_0)^\top \nabla_\theta f(s, \theta_0)\, ds. \tag{8}$$

The major question in optimal design of experiments is how to best choose P in some family $\mathcal{P}(0, T)$ of observation distributions. We observe that one optimal design formulation we might employ is a criterion that chooses the times $\tau = \{t_i\}$ for P_τ in (6) so that (7) best approximates (8)–i.e., one minimizes $|F_C - F(\tau)|$ over τ where $|\cdot|$ is the norm in $\mathbb{R}^{p\times p}$–see [9]. We do not consider this design here, but rather focus on the SE-optimal design also proposed in [9] and its comparison to more traditional designs.

The introduction of the measure P above allows for a unified framework for optimal design criteria which incorporates all the popular design criteria mentioned in the introduction. As already noted, the GFIM $F(P, \theta)$ introduced in (6) depends critically on the measure P. We also remark that we can, without loss of generality, further restrict ourselves to probability measures on $[0, T]$. Thus, let $\mathcal{P}(0, T)$ denote the set of all probability measures on $[0, T]$ and assume that a functional $\mathcal{J}: \mathbb{R}^{p\times p} \to \mathbb{R}^+$ of the GFIM is given. The *optimal design problem* associated with $\mathcal{J}$ is one of finding a probability measure $\widehat{P} \in \mathcal{P}(0, T)$ such that

$$\mathcal{J}\left(F\left(\widehat{P}, \theta_o\right)\right) = \min_{P\in\mathcal{P}(0,T)} \mathcal{J}(F(P, \theta_0)). \tag{9}$$

A general theoretical framework for existence and approximation in the context of $\mathcal{P}(0, T)$ taken with the Prohorov metric [3, 17, 22, 27] is given for these problems in Section 4 of [9]. In particular, this theory permits development of computational methods using weighted discrete measures (i.e., weighted versions of (4)).

2.1 Theoretical Summary

To summarize and further develop the theoretical considerations that are the basis of our efforts here, we first recall that the Prohorov metric ρ on the space $\mathcal{P}(0, T)$ of probability measures on the Borel subsets of $[0, T]$ can be defined [3, 17, 22, 27] in terms of probabilities on closed subsets of $[0, T]$ and their neighborhoods. However for our uses here it is far more useful to work with an equivalent characterization in terms of convergences when viewing the probability measures $\mathcal{P}(0, T)$ as a subset of the topological dual $C_b[0, T]^*$ of the bounded continuous functions on $[0, T]$ taken with the supremum norm. More precisely, ρ-convergence is equivalent to weak*-convergence on $\mathcal{P}(0, T)$ when considering $\mathcal{P}(0, T)$ as a subset of $C_b[0, T]^*$. It is then known that $(\mathcal{P}(0, T), \rho)$ is a complete, compact and separable metric space. (We will hereafter just denote this space by $\mathcal{P}(0, T)$ since the ρ will be understood.)

Our first observation is that the GFIM as defined in (6) is ρ continuous on $\mathcal{P}(0,T)$ for problems in which the observation functions $f(\cdot, \theta)$ are continuously differentiable on [0, T]. Thus, whenever $\mathcal{J}:\mathbb{R}^{p\times p} \to \mathbb{R}^+$ is continuous we find that $P \to \mathcal{J}(F(P,\theta))$ is continuous from $\mathcal{P}(0,T)$ to $\mathbb{R}^+$. Since $\mathcal{P}(0,T)$ is ρ compact, we obtain immediately the existence of solutions for the optimization problems

$$\widehat{P}_{\mathcal{J}} \equiv \operatorname{argmin}_{P\in\mathcal{P}(0,T)} \mathcal{J}(F(P,\theta_0)). \tag{10}$$

Our second observation is related to the separability of $\mathcal{P}(0,T)$ and in particular to the density of finite convex combinations over rational coefficients of Dirac measures Δ_a with atoms at a. Specifically, one can prove [3] that the set

$$\mathcal{P}_0(0,T) := \left\{ P \in \mathcal{P}(0,T) \,|\, P = \sum_{j=1}^{k} p_j \Delta_t, k \in \mathbb{N}^+, t_j \in \mathcal{T}_0, p_j \geq 0, p_j \quad \text{rational}, \quad \sum_{j=1}^{k} p_j = 1 \right\}$$

is dense in $\mathcal{P}(0,T)$ in the Prohorov metric ρ. Here $\mathcal{T}_0 = \{t_j\}_{j=1}^{\infty}$ is a countable, dense subset of [0, T]. In short, the set of $P \in \mathcal{P}(0,T)$ with finite support in $\mathcal{T}_0$ and rational masses is dense in $\mathcal{P}(0,T)$. This leads, for a given choice $\mathcal{J}$, to approximation schemes for $\widehat{P}_{\mathcal{J}}$ as defined in (10). To implement these for a given choice of $\mathcal{J}$ (examples are discussed below) would require approximation by $P^N_{\{p_j,t_j\}} = \Sigma_{j=1}^N \quad p_j \Delta_{t_j}$ in the GFIM (6) and then optimization over appropriate sets of $\{p_j, t_j\}$ in (10) with P replaced by $P^N_{\{p_j,t_j\}}$. For a fixed N, existence of minima in these problems follow from the theory outlined above. In standard optimal designs these problems are approximated even further by fixing the weights or masses p_j as $p_j = \frac{T}{N}$ (which then becomes simply a scale factor in the sum) and searching over the $\{t_j\}$. This, of course, is equivalent to replacing the $P^N_{\{p_j,t_j\}}$ by P_τ of (4) in (6) and searching over the $\tau = \{t_j\}$ for a fixed number N of grid points. This embodies assumption of equal value of the observations the at each of the times $\{t_j\}$. We observe that weighting of information at each of the observation times is carried out in the inverse problems via the weights $\sigma(t_j)$ for observation variances in (5). We further observe that the weights $\{p_j\}$ in $P^N_{\{p_j,t_j\}}$ are related to the value of the observations as a function of the model sensitivities $\nabla_\theta f(t_j, \theta_0)$ in the FIM while the weights $\frac{1}{\sigma(t_j)^2}$ are related to the reliability in the data measurement processes. We note that all of our remarks on theory related to existence above in the general probability measure case also hold for this discrete minimization case.

The formulation (10) incorporates all strategies for optimal design which entail optimization of a functional depending continuously on the elements of the Fisher information matrix. In case of the traditional design criteria mentioned in the introduction, $\mathcal{J}$ is the determinant (D-optimal), the smallest eigenvalue (E-optimal), or a quadratic form (c-optimal), respectively, of the inverse of the Fisher Information Matrix. Specifically, this includes the optimal design methods we consider here: *SE*-optimal design, *D*-optimal design, and *E*-optimal design. The design cost functional for the *SE*-optimal design method is given by (see [9])

$$\mathcal{J}_{SE}(F) = \sum_{i=1}^{p} \frac{1}{\theta_{0,i}^2} \left(F^{-1}\right)_{ii}, \tag{11}$$

where $F = F(\tau)$ is the FIM, defined above in (7), θ_0 is the true parameter vector, and p is the number of parameters to be estimated. Note that both inversion and taking the trace of a matrix are continuous operations. We observe that $F_{ii}^{-1} = SE_i(\theta_0)^2$. Therefore, *SE*-optimal design minimizes the sum of squared normalized standard errors.

D-optimal design minimizes the volume of the confidence interval ellipsoid for the covariance matrix ($\Sigma_0^N = F^{-1}$). The design cost functional for *D*-optimal design is given by (see [14, 21])

$$\mathcal{J}_D(F) = \det\left(F^{-1}\right). \tag{12}$$

Again we note that taking the determinant is a continuous operation on matrices so that $\mathcal{J}_D$ is continuous in F as required by the theory.

E-optimal design minimizes the principle axis of the confidence interval ellipsoid of the covariance matrix (defined in the asymptotic theory summarized in the next section). The design cost functional for *E*-optimal design is given by (see [2, 14])

$$\mathcal{J}_E(F) = \max_i \frac{1}{\lambda_i}, \tag{13}$$

where λ_i, $i = 1 \ldots p$, are the eigenvalues of F (which are continuous functions of F). Therefore $\frac{1}{\lambda_i}$, $i = 1 \ldots p$, corresponds to the eigenvalues of the asymptotic covariance matrix $\Sigma_0^N = F^{-1}$.

2.2 Constrained Optimization and Implementations

Each optimal design computational method we employ is based on constrained optimization to find the mesh of time points $\tau^* = \{t_i^*\}$, $i = 1, \ldots, N$, that satisfy

$$\mathcal{J}(F(\tau^*, \theta_0,)) = \min_{\tau \in \mathcal{T}} \mathcal{J}(F(\tau, \theta_0)),$$

where $\mathcal{T}$ is the set of all time meshes such that $0 \le t_1 \le \cdots \le t_i \le t_{i+1} \le \cdots \le t_N \le T$.

These optimal design methods were implemented using constrained optimization algorithms, either MATLAB's *fmincon* or SolvOpt, developed by A. Kuntsevich and F. Kappel [23], with four variations on the constraint implementation. We denote these different *constraint implementations* (which most often do result in different parameter and standard error outcomes even in cases where the $\{t_i\}$ are initially required to satisfy similar constraints) by (*C*1) – (*C*4). Complete details of the precise implementation differences in the algorithms are given in an appendix of [10].

(C1) The first constraint implementation on the time points is given by $t_1 \geq 0$, $t_N \leq T$ and $t_i \leq t_{i+1}$, such that the optimal mesh may or may not contain 0 and T. In this case we optimize over variables.

(C2) The second constraint implementation is carried out in the same manner as the first, except that the optimal mesh contains 0 and T. Hence we effectively optimize over $N - 2$ variables.

(C3) The third constraint implementation on the time points is given by $t_i = t_{i-1}+v_i$, $i = 2, \ldots, N - 1$, $t_1 = 0$ and $t_N = T$, with $v_i \geq 0$, $i = 2, \ldots, N - 1$, and $v_2 + \ldots + v_{N-1} \leq T$. Note that the optimal mesh always contains 0 and T as we optimize over $N - 2$ variables using slightly different inequality constraints.

(C4) The last constraint implementation on the time points is given by $t_i = t_{i-1} + v_i$, $i = 2, \ldots, N$, and $t_1 = 0$ with $v_i \geq 0$, $i = 2, \ldots, N$, and $v_2 + \ldots + v_N = T$. This constraint is implemented by defining $v_N = T - \Sigma_{i=2}^{N-1} v_i$. The optimal mesh again contains 0 and T, and we also optimize over $N - 2$ variables but an equality constraint is added to the constraint system.

3 Standard Error Methodology

We begin by finding the optimal discrete sampling distribution of time points $\tau = \{t_i\}_{i=1}^N$, for a fixed number N of points in a fixed interval $[0, T]$, using one of three optimal design methods described above. These three optimal design methods are then compared based on the standard errors computed for parameters using these sampling times. Since there are different ways to compute standard errors, we will compare the optimal design method using different techniques for computing the standard errors. In the following sections we will describe the methods for computing standard errors. First we consider the scalar observation case ($m = 1$).

3.1 Asymptotic Theory for Computing Standard Errors

Once we have an optimal distribution of time points we will obtain data or simulated data, $\{y_i\}_{i=1}^N$, a realization of the random process $\{y_i\}_{i=1}^N$, corresponding to the optimal time points, $\tau = \{t_i\}_{i=1}^N$. Parameters are then estimated using inverse problem formulations as described in [7]. The variance $\mathrm{Var}(\mathcal{E}(t)) = \sigma_0^2$ is assumed to be constant, the inverse problem is formulated using ordinary least squares (OLS). The OLS estimator is defined by

$$\Theta_{\mathrm{OLS}} = \Theta_{\mathrm{OLS}}^N = \mathrm{argmin}_\theta \sum_{j=1}^N \left[Y_j - f\left(t_j, \theta\right) \right]^2.$$

The estimate $\widehat{\theta}_{\mathrm{OLS}}$ is defined as

$$\widehat{\theta}_{\mathrm{OLS}} = \widehat{\theta}_{\mathrm{OLS}}^N = \mathrm{argmin}_\theta \sum_{j=1}^N \left[y_j - f\left(t_j, \theta\right) \right]^2.$$

To compute the standard errors of the estimated parameters, we first must compute the sensitivity matrix

$$\chi_{j,k}=\frac{\partial Cx\left(t_j\right)}{\partial\theta_k}=\frac{\partial f\left(t_j,\theta\right)}{\partial\theta_k}\quad\text{for}\quad j=1,\ldots,N,k=1,\ldots p.$$

Note that $\chi=\chi^N$ is an $N\times p$ matrix. The true constant variance

$$\sigma_0^2=\frac{1}{N}E\left[\sum_{j=1}^{N}\left[Y_j-f\left(t_j,\theta_0\right)\right]^2\right],$$

can be estimated by

$$\widehat{\sigma}_{\text{OLS}}^2=\frac{1}{N-p}\sum_{j=1}^{N}\left[y_j-f\left(t_j,\widehat{\theta}_{\text{OLS}}\right)\right]^2.$$

The true covariance matrix is approximately (asymptotically as $N\to\infty$) given by,

$$\Sigma_0^N\approx\Sigma_0^2\left[\chi^T\left(\theta_0\right)\chi\left(\theta_0\right)\right]^{-1}.$$

Note that the approximate Fisher Information Matrix (FIM) is defined by

$$F=F\left(\tau\right)=F\left(\tau,\theta_0\right)=\left(\Sigma_0^N\right)^{-1},\tag{14}$$

and is explicitly dependent on the sampling times τ.

When the true values, θ_0 and σ_0^2, are unknown, the covariance matrix is estimated by

$$\Sigma_0^N\approx\widehat{\Sigma}^N\left(\widehat{\theta}_{\text{OLS}}\right)=\widehat{\Sigma}_{\text{OLS}}^2\left[\chi^T\left(\widehat{\theta}_{\text{OLS}}\right)\chi\left(\widehat{\theta}_{\text{OLS}}\right)\right]^{-1}.\tag{15}$$

The corresponding FIM can be estimated by

$$\widehat{F}\left(\tau\right)=\widehat{F}\left(\tau,\widehat{\theta}_{\text{OLS}}\right)=\left(\widehat{\Sigma}^N\left(\widehat{\theta}_{\text{OLS}}\right)\right)^{-1}.\tag{16}$$

The asymptotic standard errors are given by

$$SE_k\left(\theta_0\right)=\sqrt{\left(\Sigma_0^N\right)_{kk}},\quad k=1,\ldots,p.\tag{17}$$

These standard errors are estimated in practice (when θ_0 and σ_0 are not known) by

$$SE_k\left(\widehat{\theta}_{\text{OLS}}\right)=\sqrt{\left(\widehat{\Sigma}^N\left(\widehat{\theta}_{\text{OLS}}\right)\right)_{kk}},\quad k=1,\ldots,p.\tag{18}$$

It can be shown, under certain conditions, for $N \to \infty$, that the estimator Θ_{OLS}^N is asymptotically normal [28]; i.e., for N large

$$\Theta_{\text{OLS}}^N \tilde{\mathcal{N}}_p\left(\theta_0, \Sigma_0^N\right). \tag{19}$$

3.2 Monte Carlo Method for Asymptotic Standard Errors

To account for the variability in the asymptotic standard errors due to the variability in the residual errors in the simulated data, we use Monte Carlo trials to examine the average behavior. For a single Monte Carlo trial, we generate simulated data on the optimal mesh $\{t_j\}_{j=1}^N$,

$$y_j = f\left(t_j, \theta^0\right) + \epsilon_j, \quad j = 1, \ldots N,$$

where $\theta^0 = 1.4\theta_0$, where θ_0 are the true parameter values and the ϵ_j are realizations of $\mathcal{E}_j \tilde{\mathcal{N}}\left(0, \sigma^2\right)$ for $j = 1, \ldots, N$. Parameters are estimated using OLS and standard errors are estimated using asymptotic theory (18). The parameter estimates and their estimated standard errors are stored, and the process is repeated with new simulated data corresponding to the optimal mesh for $M = 1000$ Monte Carlo trials. The average (and median) of the $M = 1000$ parameter estimates and standard errors are used to compare the optimal design methods in one of our examples.

3.3 The Bootstrapping Method

An alternative way of computing parameter estimates and standard errors uses the bootstrapping method [11]. Again we outline this for the case of scalar ($m = 1$) observations.

As in the previous section, assume we are given experimental data $(y_1, t_1), \ldots, (y_N, t_N)$ from the following underlying observation process

$$Y_j = f\left(t_j, \theta_0\right) + \mathcal{E}_j, \tag{20}$$

where $j = 1, \ldots, N$ and the $\mathcal{E}_j$ are independent identically distributed *(iid)* from a distribution $\mathcal{F}$ with mean zero $\left(E\left(\mathcal{E}_j\right) = 0\right)$ and constant variance σ_0^2, and θ_0 is the "true" parameter value. Associated corresponding realizations of Y_j are given by

$$y_j = f\left(t_j, \theta_0\right) + \epsilon_j.$$

The bootstrapping algorithm is presented for sample points corresponding to the t_j, $j = 1 \ldots N$. To compare the optimal design methods based on their bootstrapping standard errors, we will take our sample points corresponding to the optimal time distribution ($\tau = \{t_i\}_{i=1}^N$). The different optimal design methods are described above.

The following algorithm [11] can be used to compute the bootstrapping estimate $\widehat{\theta}_{boot}$ of θ_0 and its empirical distribution.

1. First estimate $\widehat{\theta}^0$ from the entire sample, using OLS.

2. Using this estimate define the standardized residuals:

$$\bar{r}_j = \sqrt{\frac{N}{(N-p)}}\left(y_j - f\left(t_j, \widehat{\theta}^0\right)\right)$$

for $j = 1, \ldots, N$. Then $\left\{\bar{r}_1, \ldots, \bar{r}_N\right\}$ are realizations of *iid* random variables $\bar{R}_j$ from the empirical distribution $\mathcal{F}_N$, and p for the number of parameters. Observe that

$$E\left(\bar{r}_j|\mathcal{F}_N\right) = N^{-1}\sum_{j=1}^{N}\bar{r}_j = 0, \qquad \mathrm{Var}\left(\bar{r}_j|\mathcal{F}_N\right) = N^{-1}\sum_{j=1}^{N}\bar{r}_j^2 = \widehat{\sigma}^2.$$

Set $m = 0$.

3. Create a bootstrap sample of size N using random sampling with replacement from the data (realizations) $\left\{\bar{r}_1, \ldots, \bar{r}_N\right\}$ to form a bootstrap sample $\left\{r_1^m, \ldots, r_N^m\right\}$.

4. Create bootstrap sample points

$$y_j^m = f\left(t_j, \widehat{\theta}^0\right) + r_j^m,$$

where $j = 1, \ldots, N$.

5. Obtain a new estimate $\widehat{\theta}^{m+1}$ from the bootstrap sample $\left\{y_j^m\right\}$ using OLS. Add $\widehat{\theta}^{m+1}$ into the vector Θ, where Θ is a vector of length M_p (M is the number of bootstrap samples) which stores the bootstrap estimates.

6. Set $m = m + 1$ and repeat steps 3–5.

7. Carry out the above iterative process M times where M is large (e.g., M=1000), resulting in a vector Θ of length M_p.

8. We then calculate the mean, standard error, and confidence intervals from the vector Θ using the formulae

$$\begin{aligned} \widehat{\theta}_{boot} &= \tfrac{1}{M}\Sigma_{m=1}^{M}\widehat{\theta}^m, \\ Cov\left(\widehat{\theta}_{boot}\right) &= \tfrac{1}{M-1}\Sigma_{m=1}^{M}\left(\widehat{\theta}^m - \widehat{\theta}_{boot}\right)^T\left(\widehat{\theta}^m - \widehat{\theta}_{boot}\right), \\ SE_k\left(\widehat{\theta}_{boot}\right) &= \sqrt{Cov\left(\widehat{\theta}_{boot}\right)_{kk}}. \end{aligned} \tag{21}$$

We will compare the optimal design methods using the standard errors resulting from the optimal time points each method proposes. Since there are different ways to compute the standard errors we will present results for several of these computational methods.

4 The Logistic Growth Example

We first compare the optimal design methods for the logistic example using the Monte Carlo method for asymptotic estimates and standard errors.

4.1 Logistic Model

The Verhulst-Pearl logistic population model describes a population that grows at an intrinsic growth rate until it reaches its carrying capacity. It is given by the differential equation:

$$\dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{K}\right), \quad x(0) = x_0,$$

where K is the carrying capacity of the population, r is the intrinsic growth rate, and x_0 is the initial population size. The analytical solution to the differential equation above is given by,

$$x(t) = f(t, \theta_0) = \frac{K}{1 + (K/x_0 - 1)\, e^{-rt}},$$

where $\theta_0 = (K, r, x_0)$ is the true parameter vector. Our statistical model is given by

$$Y(t) = f(t, \theta_0) + \mathcal{E}(t),$$

where we choose $\mathcal{EN}\left(0, \sigma_0^2\right)$ to generate simulated data (for use in the Monte Carlo calculations). A realization of the observation process is given by

$$y(t) = f(t, \theta_0) + \varepsilon(t), \quad t \in [0, T].$$

4.2 Logistic Results

For the logistic model, we use SolvOpt to solve for the optimal mesh for each of the optimal design methods (*D*-optimal, *E*-optimal and *SE*-optimal), using the second constraint (*C*2) on the time points: $t_0 \geq 0$, $t_n \leq T$ and $t_i \leq t_{i+1}$, such that the optimal mesh contains 0 and T. For this example, we took $T = 25$ and $N = n + 1 = 10$. Figure 1 contains the plot of the resulting optimal distribution of time points for the different optimal design methods, along with the uniform mesh, plotted on the logistic curve.

These optimal design methods are compared based on their average Monte Carlo asymptotic estimates and standard errors. The simulated data was generated assuming the true parameter values $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$, and variance $\sigma_0^2 = 0.16$. The average estimates and standard errors are based on $M = 1000$ Monte Carlo trials. Since we obtain histograms of estimates and standard errors from this Monte Carlo analysis, we can also gain information for comparison from the median of these histograms or sampling distributions. Monte Carlo asymptotic estimates and standard errors were also computed on the uniform mesh. We report the average and median estimates and standard errors in Table 1 (sample standard error histograms are depicted graphically in [10]).

4.3 Discussion of Logistic Results

The average asymptotic estimates from the uniform distribution and each of the optimal design methods are very close to the true values, θ_0. For $N = 10$ (Table 1), *SE*-optimal has the closest average and median estimates, followed by *D*-optimal (for r and x_0) and *E*-optimal (for K). Comparing the average and median estimates, we see that for all cases the averages and medians are very close, indicating that the parameter distributions are symmetric. However, the averages were slightly larger than the medians for r and x_0 for all

methods, implying that those parameter distributions are slightly skewed to the right (see Table 1).

Comparing the standard errors (Table 1): For K, we find that E-optimal has the smallestaverage standard errors, then the uniform grid, then SE-optimal. For r and x_0, SE-optimal has the smallest average standard errors, followed by D-optimal, then the uniform grid. The average and median standard errors are very close. However the distribution of standard errors for r and E-optimal seem to be slightly right-skewed.

In conclusion, all of the optimal design methods produce parameter estimates that are close to the true value. In addition, the standard error estimates are similar comparing the different optimal design methods. Based on the standard errors, E-optimal is more favorable for the accuracy of K, and SE-optimal is more favorable for the accuracy of r and x_0 (followed closely by D-optimal).

5 The Harmonic Oscillator Model

In our next example, we consider the harmonic oscillator, also known as the spring-mass-dashpot model. The model for the harmonic oscillator can be derived using Hooke's Law and Newton's Second Law of Motion (see [13]) and is given by

$$m\ddot{x}+c\,\dot{x}+kx=0, \qquad \dot{x}(0)=x_1, \qquad x(0)=x_2.$$

Here, m is mass, c is damping, and k is the spring constant. Dividing through by m, and defining $C = c/m$ and $K = k/m$, we can reduce the number of parameters.

$$\ddot{x}+C\,\dot{x}+Kx=0, \qquad \dot{x}(0)=x_1, \qquad x(0)=x_2.$$

The analytical solution for the position at time t can be obtained and is given by

$$x(t)=e^{-at}\left(C_1\cos bt+C_2\sin bt\right),$$

where $C_1 = x_2$, $C_2 = (x_1 + ax_2)/b$, $a=\frac{1}{2}C$, and $b=\sqrt{K-\frac{1}{4}C^2}$. Substituting in C_1 and C_2, we obtain,

$$x(t)=x(t,\theta_0)=f(t,\theta_0)=e^{-at}\left(x_2\cos bt+\frac{x_1+ax_2}{b}\sin bt\right), \qquad \text{for} \quad 0\le t\le T,$$

where for our considerations the true parameter vector is given by $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ in our examples here.

5.1 Results for the Oscillator Model

The first way we will compare these optimal design methods, given that we know $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ and $\sigma_0^2=0.16$, is to simply use their corresponding standard errors from the asymptotic theory, i.e., the values of $SE(\theta_0)$ given in (17). Recall that uncertainty is quantified by constructing confidence intervals using parameter estimates with the asymptotic standard error. Since our main focus here is the width of the confidence intervals, we first forgo the obtaining of the parameter estimates themselves which, for now,

we tacitly assume may be similar for the three data sampling distributions we investigate here.

The optimal time points for each of the three optimal design methods are plotted with the model for different T and N under the first constraint implementation ($C1$) in Fig. 2, the second constraint implementation ($C2$) in Fig. 3, the third constraint implementation ($C3$) in Fig. 4, and the last constraint implementation ($C4$) in Fig. 5. The standard errors (17) from the asymptotic theory corresponding to the ($C2$) optimal meshes are given in Table 2. Results for the other constraints can be found in the report [10].

5.2 Discussion for the Oscillator Model

The constrained optimization algorithm, SolvOpt, was chosen over MATLAB's *fmincon* for comparisons using the harmonic oscillator example because it overall resulted in more well-behaved standard errors (real and finite values), and *fmincon* often did not.

In most cases, optimal meshes with a larger number of points were nested in the optimal meshes with a reduced number of points. In some cases for $T = 28.28$ (Figs 2 and 3) doubling the number of points result in extra points being dispersed to otherwise empty regions, while other points were nested in the optimal mesh with fewer points. Often the larger number of points in the optimal mesh resulted in smaller standard errors.

From the table of standard errors (Table 2), we find that different optimal sampling distributions produced the smallest standard errors for different parameters, with no optimal design method having consistently smaller standard errors. For *C, SE*-optimal had the smallest standard error, then *D*-optimal. For *K, D*-optimal had the smallest standard error, followed by *SE*-optimal. For x_1, no optimal sampling distribution had consistently smaller standard errors. For x_2, most of the time *E*-optimal had the smallest standard errors, then *SE*-optimal.

Results are similar for the other constraints, as can be seen in [10]. The standard errors from the different optimal design methods were usually on the same order of magnitude. No method was always the best while comparing asymptotic standard errors, though for specific parameters some optimal sampling distributions were favorable.

Since the asymptotic standard errors appear explicitly in the cost function we are minimizing for *SE*-optimal design, it may not be fair to compare these methods based on their asymptotic standard errors. To account for any possible bias in our comparison, we will compare these optimal design methods in the next section using simulated data and the inverse problem to estimate parameters using asymptotic theory and bootstrapping. In these computations, we will compare the optimal design methods based on how close their parameter estimates are to the true parameters, and the values of their estimated standard errors and covariances.

5.3 Results for the Oscillator Model - with the Inverse Problem

We solve the inverse problem with the OLS formulation to obtain parameter estimates and standard errors from both asymptotic theory (18) and the bootstrapping method (21). We create simulated noisy data (in agreement with our statistical model (2)) corresponding to the optimal time meshes using true values $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ and *iid* noise with $\mathcal{E}_j \tilde{\mathcal{N}}\left(0, \sigma_0^2\right)$. In this section we only estimate a subset of the parameters $\theta = (C, K)$. In addition to the estimates and standard errors, we also report the estimated Cov(C, K) according to asymptotic theory (15) and bootstrapping (21). For comparison purposes we also present these results for a uniform grid using the same T and N.

The optimal time points for each of the three optimal design methods are plotted with the model for $T = 14.14$ and $T = 28.28$ and for $N = 15$ under the second constraint implementation ($C2$) in Fig. 6. The estimates, standard errors, and covariance between parameters as estimated from the asymptotic theory (18) corresponding to the ($C2$) optimal meshes are given in Table 3. The estimates, standard errors, and covariance between parameters when estimated from the bootstrapping method (21) corresponding to the ($C2$) optimal meshes are given in Table 4. In each of the tables are also results on the uniform grid of time points for the same T and N. Results for the three other constraints can be found in [10].

5.4 Discussion of Oscillator Results with the Inverse Problem

The simulated data was created using the "true" parameter values $\theta_0 = (C, K) = (0.1, 0.2)$. So we can compare the optimal design methods based on how close the parameter estimates are as well as how large the estimates of the standard errors and covariances are.

For asymptotic estimates—Comparing optimal design methods based on which has parameter estimates closest to the true values, there is no method that is always the best. From constraint implementation ($C2$) (Table 3), either D-optimal or E-optimal had the closest parameter estimates to the true values. Comparing the optimal design methods based on the estimated standard errors and covariance between parameters, we again find that no method is always the best. For constraint implementation ($C2$) (Table 3), the smallest standard errors and covariances came from E-optimal when $T = 14.14$ and SE-optimal when $T = 28.28$, followed by D-optimal in both cases.

For bootstrap estimates—Comparing optimal design methods based on which has bootstrapping parameter estimates closest to the true values, again no method is always the best. For constraint implementation ($C2$) (Table 4), either D-optimal or E-optimal had parameter estimates closest to the true values. For $T = 14.14$, the parameter estimate for K was in fact closest from the uniform mesh, followed by D-optimal. Comparing optimal design methods based on which method produces the smallest bootstrapping estimated standard errors and parameter estimates, no method is consistently favorable. For constraint implementation ($C2$) (Table 4), when $T = 14.14$ the smallest standard errors and covariances come from E-optimal, when $T = 28.28$ either SE-optimal or the uniform grim had the smallest standard errors and covariances, followed by D-optimal.

In conclusion, all of the optimal design methods are favorable under specific conditions. In many of the cases the parameter estimates, standard errors, and covariances are on the same order of magnitude resulting from different optimal design criteria.

6 A Simple Glucose Regulation Model

Next we will consider a well-known model for the intravenous glucose tolerance test (IVGTT). This model is referred to as the *minimal model* in the literature [16, 19, 29]. Prior to the IVGTT the patient is asked to fast. When the patient comes in for the IVGTT, measurements of their baseline glucose and insulin concentrations, G_b and I_b, respectively, are first taken. The IVGTT procedure consists of injecting a bolus resulting in an initial concentration p_0 of glucose into the blood, and measuring the glucose and insulin concentrations in the blood at various time points after the injection.

The body carefully regulates the glucose concentration in the blood within a narrow range. Extremely high blood glucose concentration is referred to as hyperglycemia, whereas hypoglycemia results when the blood glucose concentration is too low. The IVGTT initially brings the blood glucose concentration to hyperglycemic levels. In normal healthy patients,

the high level of glucose in the blood signals the beta cells of the pancreas to secrete insulin. Insulin helps the fat and muscle cells to uptake glucose from the blood, either for fuel or for storage as glycogen. When the blood glucose concentration is too low, the pancreas secretes glucagon which releases glucose stored in the liver into the blood. Glucagon is another dynamic variable [5] during the IVGTT. Though glucagon is not included in this model, it is acknowledged that the liver can regulate glucose independently from insulin through glucagon.

6.1 Model

The minimal model is given by the following system of ordinary differential equations (see [16, 19, 29] for details):

$$\dot{G}(t) = -p_1(G(t) - G_b) - X(t)G(t), \qquad G(0) = p_0, \tag{22}$$

$$\dot{X}(t) = -p_2 X(t) + p_3(I(t) - I_b), \qquad X(0) = 0, \tag{23}$$

$$\dot{I}(t) = p_4 t\max(0, G(t) - p_5) - p_6(I(t) - I_b), \qquad I(0) = p_7 + I_b, \tag{24}$$

where $G(t)$ is the glucose concentration (in mg/dl) in plasma at time t, $I(t)$ is the insulin concentration (in μU/ml) in plasma at time t and $X(t)$ represents insulin-dependent glucose uptake activity (proportional to a remote insulin compartment) in units 1/min.

We use the following approximate max function in equation (24) since it is continuously differentiable:

$$\text{maxfunc}_1(v) = \begin{cases} v & \text{for } v > \epsilon_0, \\ 0 & \text{for } v < -\epsilon_0, \\ \frac{1}{4\epsilon_0}(v+\epsilon_0)^2 & \text{for } v \in [-\epsilon_0, \epsilon_0], \end{cases}$$

where $\epsilon_0 > 0$ is chosen sufficiently small (for example, $\epsilon_0 = 10^{-5}$).

An interpretation of the parameters is given in Table 5.

In the following we will describe the model and its underlying assumptions.

Equation (22) (Glucose concentration in plasma)—At $t = 0$ a bolus of glucose is injected such that the initial glucose concentration in the blood is p_0. The first term represents hepatic glucose balance, which occurs independent of insulin level. The second term is the loss of glucose due to insulin-dependent uptake by peripheral tissues.

Equation (23) (Insulin-dependent glucose uptake activity)—At $t = 0$ there is no glucose uptake activity. Spontaneously tissue loses the ability to uptake glucose, even in the presence of insulin. Glucose uptake activity increases proportionally to the amount by which insulin concentration is greater than baseline insulin concentration.

Equation (24) (Insulin concentration in the plasma)—At $t = 0$ the initial insulin concentration is at some level over baseline, given by $p_7 + I_b$. The increase in insulin concentration is proportional to the amount by which glucose concentration exceeds some

threshold, p_5, and the amount of time that has elapsed since the glucose injection. There is a loss of insulin to degradation in the plasma. The pancreas secretes low levels of insulin, even in hypoglycemic conditions, to maintain insulin concentration at or above baseline I_b.

The analysis of this model found in [16, 29] gives a metabolic portrait for the first phase sensitivity to glucose (Φ_1) (corresponding to initial secretion of insulin), the second phase glucose sensitivity (S_G) (corresponding to a secondary phase of insulin secretion), and the insulin sensitivity index (S_I). The metabolic portrait is given by

$$S_I = \frac{p_3}{p_2}, \qquad S_G = p_1, \qquad \phi_1 = \frac{I_{\max} - I_b}{p_6 (p_0 - G_b)}, \tag{25}$$

where $I_{\max}$ is the maximal value of insulin concentration in plasma.

Bergman et al. [15] suggest the use of this model in the clinical IVGTT setting. Parameters from the model are estimated using patient-specific data. The parameter estimates are then used in the metabolic portrait for diabetes diagnosis purpose for that patient. This process was made readily available to clinicians in the computer software MINMOD [25]. Since the estimation of these parameters plays such a crucial role in the diagnosis, it appears that optimal design methods would be of great assistance. Data sampled at the optimal time points would result in the most accurate metabolic portrait produced by this mathematical model.

Next we will describe the corresponding statistical model for this system involving vector observations. We obtain numerical solutions using MATLAB's *ode45* since there does not exist an analytical solution to this system of differential equations. Let $\vec{z}(t, \theta_0) = (G(t, \theta_0), X(t, \theta_0), I(t, \theta_0))^T$, represent our model solution. Since we can observe realizations of $G(t, \theta_0)$ and $I(t, \theta_0)$, but not $X(t, \theta_0)$, our observation process is given by

$$\vec{y}(t) = \vec{f}(t, \theta_0) = (G(t, \theta_0), I(t, \theta_0))^T.$$

Our statistical model is given by the stochastic process

$$\vec{Y}(t) = \vec{f}(t, \theta_0) + \vec{\mathcal{E}}(t),$$

where $\vec{\mathcal{E}}(t)$ is a noisy vector random process. We assume the following about the vector random variable $\vec{\mathcal{E}}(t)$:

$$\begin{aligned}
E\left(\vec{\mathcal{E}}(t)\right) &= 0, \quad t \in [0, T], \\
\mathrm{Var}\vec{\mathcal{E}}(t) &= \mathrm{diag}\left(\sigma^2_{0,G}, \sigma^2_{0,I}\right), \quad t \in [0, T], \\
\mathrm{Cov}(\mathcal{E}_1(t)\mathcal{E}_1(s)) &= \sigma^2_{0,G}\delta(t-s), \quad t, s \in [0, T], \\
\mathrm{Cov}(\mathcal{E}_2(t)\mathcal{E}_2(s)) &= \sigma^2_{0,I}\delta(t-s), \quad t, s \in [0, T], \\
\mathrm{Cov}(\mathcal{E}_1(t)\mathcal{E}_2(s)) &= 0, \quad t, s \in [0, T].
\end{aligned}$$

We assume constant variance, $\sigma^2_{0,G} = 25$ and $\sigma^2_{0,I} = 4$ for the results given here. A realization of the observation process is given by

$$\vec{y}(t)=\vec{f}(t,\theta_0)+\vec{\varepsilon}(t),\quad t\in[0,T],$$

where the measurement error $\vec{\varepsilon}(t)$ is a realization of $\vec{\mathcal{E}}(t)$.

6.2 Methods

Though the vector methodology is similar to that in the scaler case, for completeness we outline it here for a system of differential equations such as the simple glucose regulation model.

We begin by finding the optimal discrete sampling distribution of time points $\tau=\{t_i\}_{i=1}^N$, for a fixed number of points, N, and a fixed final time, T, using either *SE*-optimal, *D*-optimal, or *E*-optimal. These three optimal design methods are then compared based on the asymptotic standard errors formulae for parameters using these sampling times.

More specifically, once we have an optimal distribution of time points we will obtain data or simulated data, $\{\vec{y}_i\}_{i=1}^N$, a realization of the random process $\{\vec{Y}_i\}_{i=1}^N=\{(G_i,I_i)^T\}_{i=1}^N$ given by

$$\vec{Y}_i=\vec{f}(t_i,\theta_0)+\vec{\mathcal{E}}_i,$$

corresponding to the optimal time points, $\tau=\{t_i\}_{i=1}^N$, where $\vec{\mathcal{E}}_i=\vec{\mathcal{E}}(t_i)$. Define $V_0=\mathrm{Var}(\vec{\mathcal{E}}_i)=\mathrm{diag}(\sigma_{0,G}^2,\sigma_{0,I}^2)$.

A subset of the parameters is estimated using the inverse problem methodology [7]. Since the variance is assumed to be constant, the inverse problem is formulated using ordinary least squares (OLS). The OLS *estimator* for a vector system is defined by

$$\Theta_{\mathrm{OLS}}=\Theta_{\mathrm{OLS}}^N=\mathrm{argmin}_\theta\sum_{j=1}^N\left[\vec{Y}_j-\vec{f}(t_j,\theta)\right]^T V_0^{-1}\left[\vec{Y}_j-\vec{f}(t_j,\theta)\right].$$

For a given realization $\{y_j\}$, the OLS *estimate* $\widehat{\theta}_{\mathrm{OLS}}$ is defined as

$$\widehat{\theta}_{\mathrm{OLS}}=\widehat{\theta}_{\mathrm{OLS}}^N=\mathrm{argmin}_\theta\sum_{j=1}^N\left[\vec{y}_j-\vec{f}(t_j,\theta)\right]V_0^{-1}\left[\vec{y}_j-\vec{f}(t_j,\theta)\right].$$

The definition of variance gives

$$V_0=\mathrm{diag}E\left(\frac{1}{N}\sum_{j=1}^N\left[\vec{Y}_j-\vec{f}(t_j,\theta_0)\right]\left[\vec{Y}_j-\vec{f}(t_j,\theta_0)\right]^T.\right).$$

In the case that V_0 is unknown, an unbiased estimate can be obtained from the realizations $\{\vec{y}_i\}_{i=1}^N$ and $\widehat{\theta}$ by

$$V_0 \approx \widehat{V}=\text{diag}\left(\frac{1}{N-p}\sum_{j=1}^{N}\left[\vec{y}_j-\vec{f}\left(t_j,\widehat{\theta}\right)\right]\left[\vec{y}_j-\vec{f}\left(t_j,\widehat{\theta}\right)\right]^T\right),$$

which is solved simultaneously (in an iterative procedure - see [7]) with normal equations for the estimate $\widehat{\theta}=\widehat{\theta}_{OLS}$, where p is the number of parameters being estimated.

To compute the standard errors of the estimated parameters, we first must compute the $2 \times p$ sensitivity matrices $D_j(\theta)=D_j^N(\theta)$ which are given by

$$D_j=\begin{pmatrix} \frac{\partial f_1(t_j,\theta)}{\partial\theta_1} & \frac{\partial f_1(t_j,\theta)}{\partial\theta_2} & \cdots & \frac{\partial f_1(t_j,\theta)}{\partial\theta_p} \\ \frac{\partial f_2(t_j,\theta)}{\partial\theta_1} & \frac{\partial f_2(t_j,\theta)}{\partial\theta_2} & \cdots & \frac{\partial f_2(t_j,\theta)}{\partial\theta_p} \end{pmatrix},$$

for $j = 1, \ldots, N$. For this system we can rewrite D_j in terms of $(G(t_j, \theta), I(t_j, \theta))^T$ (since $(f_1(t_j, \theta), f_2(t_j, \theta))^T = (G(t_j, \theta), I(t_j, \theta))^T$). We have

$$D_j=\begin{pmatrix} \frac{\partial G(t_j,\theta)}{\partial\theta_1} & \frac{\partial G(t_j,\theta)}{\partial\theta_2} & \cdots & \frac{\partial G(t_j,\theta)}{\partial\theta_p} \\ \frac{\partial I(t_j,\theta)}{\partial\theta_1} & \frac{\partial I(t_j,\theta)}{\partial\theta_2} & \cdots & \frac{\partial I(t_j,\theta)}{\partial\theta_p} \end{pmatrix}.$$

The true covariance matrix is approximately (asymptotically as $N \to \infty$) given by

$$\Sigma_0^N \approx \left(\sum_{j=1}^{N} D_j^T(\theta_0)\, V_0^{-1} D_j(\theta_0)\right)^{-1}.$$

When the true values, θ_0 and V_0, are unknown, the covariance matrix is estimated by

$$\Sigma_0^N \approx \widehat{\Sigma}^N=\left(\sum_{j=1}^{N} D_j^T\left(\widehat{\theta}_{OLS}\right)\widehat{V}^{-1} D_j\left(\widehat{\theta}_{OLS}\right)\right)^{-1}.$$

The corresponding FIM, asymptotic standard errors and asymptotic distribution are again given by (16), (17), (18), and (19), respectively.

6.2.1 The Bootstrap Method for a system—The bootstrap method for a system of differential equations is the same as described in the previous section, except that each state variable has its own residuals that must be separately sampled with replacement. The first four steps of the bootstrap algorithm of Section 3.3 modified for a system with vector observations is outlined here for completeness.

1. First estimate $\widehat{\theta}^0$ from the entire sample, using OLS.
2. Using this estimate define the standardized residuals:

$$\bar{r}_{G,j}=\sqrt{\frac{N}{(N-p)}}\left(y_{1,j}-f_1\left(t_j,\widehat{\theta}^0\right)\right),$$

$$\bar{r}_{1,j} = \sqrt{\frac{N}{(N-p)}}\left(y_{2,j} - f_2\left(t_j, \widehat{\theta}^0\right)\right),$$

for $j = 1, \ldots, N$. Then $\left\{\bar{r}_{G,1}, \ldots, \bar{r}_{G,N}\right\}, \left\{\bar{r}_{1,1}, \ldots, \bar{r}_{1,N}\right\}$ are realizations of *iid* random variables from the empirical distribution $\overrightarrow{\mathcal{F}}_N$, and p for the number of parameters. Set $m = 0$.

3. Create a two different bootstrap sample of size N using random sampling with replacement from the data (realizations) $\left\{\bar{r}_{G,1}, \ldots, \bar{r}_{G,N}\right\}$ and $\left\{\bar{r}_{1,1}, \ldots, \bar{r}_{1,N}\right\}$ to form the bootstrap samples $\left\{r^m_{G,1}, \ldots, r^m_{G,N}\right\}$ and $\left\{r^m_{1,1}, \ldots, r^m_{1,N}\right\}$.

4. Create bootstrap sample points

$$y^m_{1,j} = f_1\left(t_j, \widehat{\theta}^0\right) + r^m_{G,j},$$

$$y^m_{2,j} = f_2\left(t_j, \widehat{\theta}^0\right) + r^m_{1,j},$$

where $j = 1, \ldots, N$.

5. Steps 5-8 are the same as those of the algorithm for scalar observations in Section 3.3.

We compute the optimal time mesh using *SE*-optimality, *D*-optimality, and *E*-optimality, as defined in the previous section, for a subset of the parameters $\theta = (p_1, p_2, p_3, p_4)$, and a fixed number of time points ($N = 30$) and a final time of $T = 150$ minutes. We remark that a subset of parameters was chosen to avoid an ill-conditioned FIM. The subset of parameters was chosen based on the traditional sensitivity functions. The glucose and insulin model solutions were most sensitive to $\theta = (p_1, p_2, p_3, p_4)$. The approximate asymptotic standard errors (17) or (18) for $\theta = (p_1, p_2, p_3, p_4)$ were computed on the optimal mesh corresponding to an optimal design method.

The optimal design methods were implemented using the constrained minimization algorithm SolvOpt. The variations on the constraint employed were the same as in the previous section. We compare *SE*-optimal, *D*-optimal and *E*-optimal design methods based on these approximate asymptotic standard errors.

6.3 Results for the Glucose Regulation Model

The optimal time points (found using the SolvOpt algorithm) for each of the three optimal design methods are plotted with the model for $T = 150$ minutes and $N = 30$ under the first constraint implementation ($C1$) in Fig. 7, the second constraint implementation ($C2$) in Fig. 8, the third constraint implementation ($C3$) in Fig. 9, and the last constraint implementation ($C4$) in Fig. 10. The standard errors (17) from the asymptotic theory corresponding to these optimal meshes are given in Table 6-9, respectively for the four different constraint implementations.

Note that for constraint implementations ($C2$) and ($C4$) initializing SolvOpt with the uniform mesh resulted in a terminal error for D-optimal, stating that the gradient at the starting point was zero. In these cases other initial mesh points were chosen such that D-optimal's initial gradient was non-zero, and optimization could be achieved. To be consistent, all three methods were initialized by the same non-uniform mesh. For ($C2$) the initial mesh was $\tau^0 = \{0, \dots, 0, 10, 37, 150, \dots, 150\}$, and for ($C4$) it was $\tau^0 = \{5, 15, 19, 21, 24, 26, 42, 59, 63, 73, 82, 95, 98, 98, 102, 111, 114, 119, 120, 122, 127, 136, 137, 137, 140, 144, 144, 144, 145, 146\}$. Optimal design methods are guaranteed to converge in a local sense.

6.4 Discussion for the Glucose Regulation Model

Comparing the optimal design methods using approximate asymptotic standard errors, we find that the optimal design methods that are best for (p_1, p_2, p_3) are different than the ones best for the standard error of p_4. For constraint implementation ($C1$) (Table 6), SE-optimal followed by E-optimal had the smallest standard errors for (p_1, p_2, p_3), and D-optimal followed by SE-optimal had the smallest standard errors for p_4. For constraint implementation ($C2$) (Table 7), the smallest standard errors were from E-optimal followed by SE-optimal for (p_1, p_2, p_3), and for p_4 it was D-optimal followed by SE-optimal. For constraint implementations ($C3$) and ($C4$) (Tables 8 and 9), SE-optimal followed by E-optimal had the smallest standard errors for (p_1, p_2, p_3), and D-optimal followed by E-optimal had the smallest standard errors for p_4.

In conclusion, D-optimal tended to have the smallest standard errors for p_4, whereas SE-optimal or E-optimal had the smallest standard errors for (p_1, p_2, p_3). In the next section we compute the estimated standard errors from simulated data using asymptotic theory and bootstrapping as a different method of comparing the optimal design methods.

6.5 Result for the Glucose Regulation Model with the Inverse Problem

As in the harmonic oscillator example, we use the inverse problem with the OLS formulation to obtain parameter estimates and standard errors from both asymptotic theory (18) and the bootstrapping method (21). We create simulated noisy data corresponding to the optimal time meshes (presented in the previous section) in agreement with our statistical model (absolute error, with independent error processes for G and I) assuming true values θ_0 to be the parameter values found in Table 5 and *iid* noise with $\vec{\mathcal{E}}_j \tilde{\mathcal{N}}\left(0, \vec{\sigma}_0^2\right)$. We assume the true variances: $\sigma_{0,G}^2 = 25$ and $\sigma_{0,I}^2 = 4$. In this section we again only estimate a subset of the parameters $\theta = (p_1, p_2, p_3, p_4)$. In addition to the estimates and standard errors, we also report the estimated covariance between estimated parameters according to asymptotic theory (18) and bootstrapping (21). For comparison purposes we also present these results for a uniform grid using the same $T = 150$ and $N = 30$.

The optimal time points for each of the three optimal design methods are the same as computed in the previous results section, and are plotted with the model in Figs. 7-10 for the four different constraints. The parameter estimates, standard errors and covariances are estimated from the asymptotic theory (18) corresponding to the ($C2$) optimal meshes are given in Table 10. The parameter estimates, standard errors, and covariance between parameters estimated from the bootstrapping method (21) corresponding to the ($C2$) optimal meshes are given in Table 11. In each of the tables are also results on the uniform grid of time points. Similar results for the other constraints can be found in [10].

6.6 Discussion for the Glucose Regulation Model with the Inverse Problem

Comparing the resulting parameter estimates from simulated data on the different optimal meshes to the true parameter values, $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, we find there is no optimal design method that is always favorable. Using either asymptotic theory or bootstrapping to compute parameter estimates for different optimal design methods, we examine how close the parameter estimates are to the true values. Often (but not always) these parameter estimates from the different optimal meshes are the same order of magnitude as the true values.

The results for the uniform mesh are given for comparison. In most cases, the optimal design methods produce closer parameter estimates with smaller standard errors and covariances (as estimated by asymptotic theory and bootstrapping) than the uniform mesh.

Asymptotic theory: parameter estimates—We compare the optimal design methods based on how close their parameters are to the true values. For constraint implementation ($C2$) (Table 10), parameters estimates of (p_1, p_2, p_3) were closest to the true values for the optimal sampling distributions from *D*-optimal and then *SE*-optimal. For p_4 the closest parameter estimates were from the uniform mesh, followed by *D*-optimal.

Asymptotic theory: standard errors—Here we compare the optimal design methods based on which has the smallest standard error estimates as estimated by asymptotic theory.

For the constraint implementation ($C2$) (Table 10), the smallest standard error for parameters (p_1, p_2, p_3) come from *SE*-optimal followed by *E*-optimal. For p_4, the smallest standard error estimates are from the uniform grid followed by *D*-optimal.

Asymptotic theory: covariance estimates—We also compared the optimal design methods based on which has the smallest covariance estimates in absolute value.

For constraint implementation (C_2) (Table 10), *SE*-optimal or *D*-optimal have the smallest in absolute value covariance estimates.

Bootstrapping: parameter estimates—Here we compare the optimal design methods based on which had bootstrapping parameter estimates closest to the true values. Often these results are different for the different parameters.

For constraint implementation ($C2$) (Table 11), parameter estimates for (p_1, p_2) the closest parameter estimates came from *E*-optimal followed by *D*-optimal (for p_1), and the uniform grid followed by *SE*-optimal (for p_2). For p_3, the uniform grid then *SE*-optimal had the closest parameter estimates to the true value. For p_4 the closest estimate came from *D*-optimal followed by *E*-optimal.

Bootstrapping: standard errors—We compare the optimal design methods based on how small their standard errors are as estimated by the bootstrap method.

For the second constraint implementation ($C2$) (Table 11), the smallest standard errors for parameters (p_1, p_3) are from *E*-optimal followed by *SE*-optimal. For p_2, the uniform grid has the smallest standard errors, followed by *SE*-optimal. For p_4, the uniform grid has the smallest standard errors followed by *D*-optimal.

Bootstrapping: covariance estimates—For constraint implementation ($C2$) (Table 11), *D*-optimal or *E*-optimal have the smallest in absolute value covariance estimates, except

for $\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$ where *SE*-optimal is the smallest and $\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$ where the uniform grid is the smallest.

Results for the other constraints are similar to those for (*C*2) in that no one optimal design method is superior in every case; see [10] for details of use of the constraint implementation (*C*1), (*C*3) and (*C*4) in each of the design situations considered in this paper.

7 Conclusions

We compared *D*-optimal, *E*-optimal and *SE*-optimal design methods for a simple differential equation model: the logistic population model, a second order differential equation: the harmonic oscillator model, and a vector system for glucose regulation. *D*-optimal and *E*-optimal design methods are more established in the literature. Our comparisons test the performance of *SE*-optimal design, which is a relatively newer method.

For the logistic example, the optimal design methods were compared using the Monte Carlo method for asymptotic standard errors. Comparing the average and median parameter estimates to their true values, we find that *SE*-optimal has closest parameter estimates for $N = 10$ time points. For $N = 15$, no method had estimates that were always closest to the true values. In all cases each optimal design methods produced estimates close to the true values. The average and median standard errors for K were smallest from the optimal mesh from *E*-optimal. For parameters r and x_0, *SE*-optimal had the smallest average and median standard errors. Overall, no optimal design method is consistently favorable for this logistic example.

For the harmonic example, comparing the approximate asymptotic standard errors, we found that different optimal design methods were favorable for different parameters. *D*-optimal often had the smallest standard errors for K and x_1. *SE*-optimal often had the smallest standard errors for C. For x_2, either *SE*-optimal or *E*-optimal had the smallest standard errors. We also compared methods using the inverse problem with simulated data and asymptotic theory and bootstrapping. Comparing methods based on who's parameter estimates were closest to the true values, and who had the smallest standard errors or covariances, there was no method that was preferred over the others. In each comparison, the best optimal design method often depended on the constraint implementation, the choice of $T = 14.14$ or $T = 28.28$, and the parameter.

For the glucose regulation model, comparing the approximate asymptotic standard errors, we found that for parameters (p_1, p_2, p_3) either *SE*-optimal or *E*-optimal had the smallest standard errors. *D*-optimal tended to have the smallest standard errors for p_4. We also compared the optimal design methods for the inverse problem using asymptotic theory and bootstrapping. Comparing the parameter estimates to their true values, none of the optimal design methods were consistently closer. Comparing the optimal design methods based on which had the smallest standard errors and covariances we found that no method was preferable over the others. However, the optimal design methods often had smaller standard errors and covariances than the uniform mesh. The constraint implementation, parameter, and choice of asymptotic theory or bootstrapping influenced which optimal design method would be favorable for this example.

The best choice of optimal design method depends on the complexity of the model, the type of constraint one is using, the subset of parameters you are estimating, and even the choice of N and T. The examples in this comparison provide evidence that *SE*-optimal design is competitive with *D*-optimal and *E*-optimal design, and in some cases *SE*-optimal design is a more favorable method.

Acknowledgments

This research was supported in part by the U.S. Air Force Office of Scientific Research under grant AFOSR-FA9550-09-1-0226 and in part by the National Institute of Allergy and Infectious Disease under grant NIAID 9R01AI071915. The authors are grateful to reviewers for their constructive comments.

References

[1]. Atkinson AC, Bailey RA. One hundred years of the design of experiments on and off the pages of Biometrika. Biometrika. 2001; 88:53–97.

[2]. Atkinson, AC.; Donev, AN. Optimum Experimental Designs. Oxford University Press; New York: 1992.

[3]. Banks HT, Bihari KL. Modeling and estimating uncertainty in parameter estimation. Inverse Problems. 2001; 17:95–111.

[4]. Banks, HT.; Botsford, LW.; Kappel, F.; Wang, C. *Proc. 2nd Course on Math. Ecology* (Trieste, December, 1986). World Scientific Press; Singapore: 1988. Modeling and estimation in size structured population models, LCDS/CCS Rep. 87-13, March, 1987, Brown Univ.; p. 521-541.

[5]. Banks, HT.; Carter, CA. CRSC Technical Report, CRSC-TR10-09. NCSU; Raleigh: May. 2010 Mathematical modeling of the glucose homeostatic system in humans.

[6]. Banks HT, Davidian M, Shuhua Hu, Kepler Grace M. Rosenberg ES. Modeling HIV immune response and validation with clinical data, CRSC-TR07-09, March, 2007. J. Biological Dynamics. 2008; 2:357–385.

[7]. Banks, HT.; Davidian, M.; Samuels, JR., Jr.; Sutton, KL. An inverse problem statistical methodology summary, CRSC Technical Report, CRSC-TR08-01, NCSU, January, 2008. In: Chowell, Gerardo, et al., editors. Chapter 11 in *Statistical Estimation Approaches in Epidemiology*. Springer; Berlin Heidelberg New York: 2009. p. 249-302.

[8]. Banks HT, Davis JL, Ernstberger SL, Hu S, Artimovich E, Dhar AK. Experimental design and estimation of growth rate distributions in size-structured shrimp populations, CRSC-TR08-20, November, 2008. Inverse Problems. Sept.2009 25:095003. 28pp.

[9]. Banks HT, Dediu Sava, Ernstberger SL, Kappel F. A new approach to optimal design problems, CRSC-TR08-12, September, 2008. J. Inverse and Ill-posed Problems. 2010; 18:25–83. (Revised), November, 2009.

[10]. Banks, HT.; Holm, K.; Kappel, F. CRSC Technical Report, CRSC-TR10-11. NCSU; Raleigh: July. 2010 Comparison of optimal design methods in inverse problems.

[11]. Banks HT, Holm K, Robbins D. Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping, CRSC-TR09-13, June, 2009. Mathematical and Computer Modeling. 2010; 52:1610–1625. Revised, August 2009.

[12]. Banks, HT.; Kunisch, K. Estimation Techniques for Distributed Parameter Systems. Birkhaüser Boston: 1989.

[13]. Banks, HT.; Tran, HT. Mathematical and Experimental Modeling of Physical and Biological Processes. Chapman and Hall/ CRC; Boca Raton, FL: 2009.

[14]. Berger, MPF.; Wong, WK., editors. Applied Optimal Designs. John Wiley & Sons; Chichester, UK: 2005.

[15]. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. J Clin Invest. 1981; 68(6):1456–1467. [PubMed: 7033284]

[16]. Bergman RN, Ider YZ, Bowden CR, Cobelli C. Quantitative estimation of insulin sensitivity. Am. J. Physiol. 1979; 236:E667–E677. [PubMed: 443421]

[17]. Billingsley, P. Convergence of Probability Measures. John Wiley & Sons; New York, NY: 1968.

[18]. Davidian, M.; Giltinan, D. Nonlinear Models for Repeated Measurement Data. Chapman & Hall; London: 1998.

[19]. De Gaetano A, Arino O. Mathematical modeling of the intravenous glucose tolerance test. J. Math. Biology. 2000; 40:136–168.

[20]. Fedorov, VV. Theory of Optimal Experiments. Academic Press; New York and London: 1972.

[21]. Fedorov, VV.; Hackel, P. Model-Oriented Design of Experiments. Springer-Verlag; New York, NY: 1997.

[22]. Huber, PJ. Robust Statistics. John Wiley & Sons, Inc.; New York, NY: 1981.

[23]. Kuntsevich, A.; Kappel, F. SolvOpt. Retrieved December 2009, from http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/

[24]. Müller W, Stehlik M. Issues in the optimal design of computer simulation experiments. Appl. Stochastic Models in Business and Industry. 2009; 25:163–177.

[25]. Pacini G, Bergman RN. MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. Comput Methods Programs Biomed. 1986; 23(2):113–122. [PubMed: 3640682]

[26]. Patan M, Bogacka B. Optimum experimental designs for dynamic systems in the presence of correlated errors. Computational Statistics and Data Analysis. 2007; 51:5644–5661.

[27]. Prohorov, Yu. V. Convergence of random processes and limit theorems in probability theory. Theor. Prob. Appl. 1956; 1:157–214.

[28]. Seber, GAF.; Wild, CJ. Nonlinear Regression. John Wiley & Sons; New York, NY: 1989.

[29]. Toffolo G, Bergman RN, Finegood DT, Bowden CR, Cobelli C. Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal of insulin kinetics in the dog. Diabetes. 1980; 29:979–990. [PubMed: 7002673]

[30]. Ucinski D, Atkinson AC. Experimental design for time-dependent models with correlated observations. Studies in Nonlinear Dynamics and Econometrics. 2004; 8(2) Article 13: The Berkeley Electronic Press.

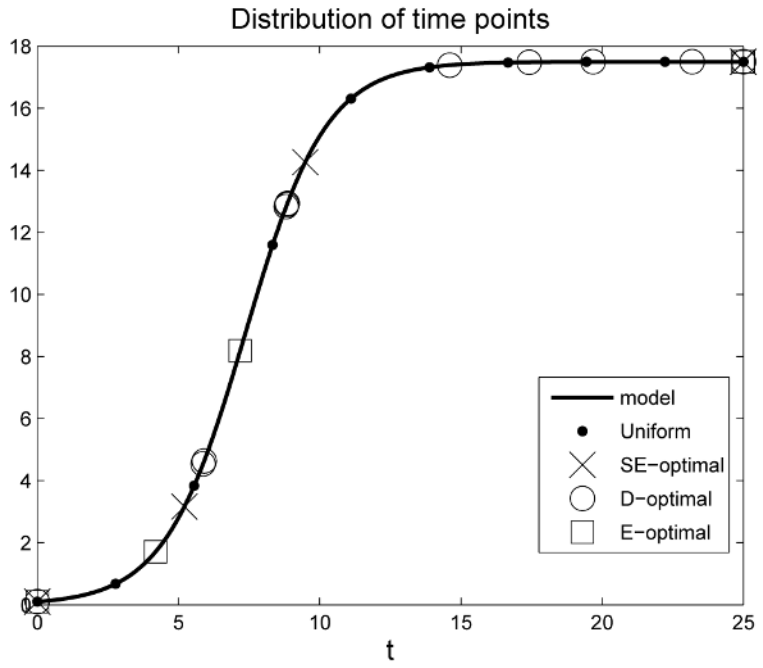

Figure 1.
The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points obtained using SolvOpt, with $N = 10$, and the optimal design methods *SE*-optimality, *D*-optimality, and *E*-optimality. Optimization with constraint implementation ($C2$).

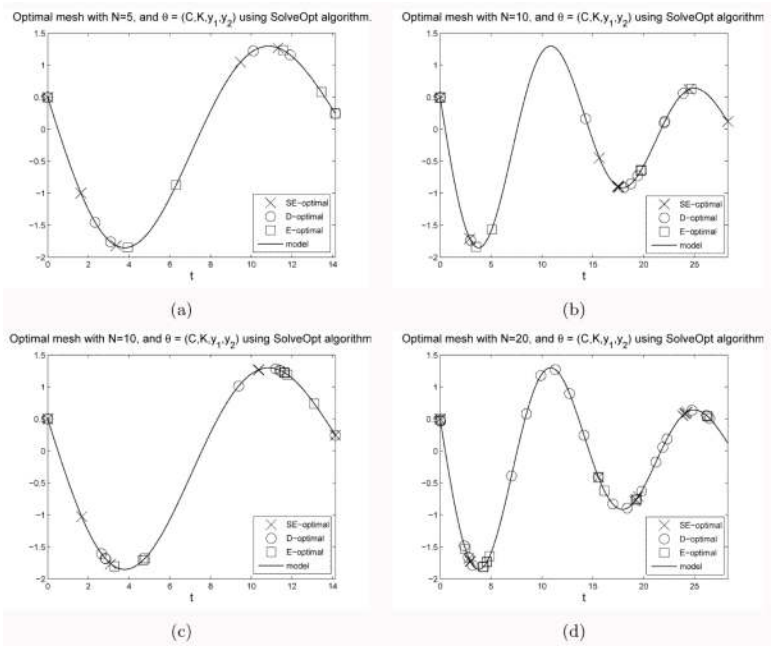

Figure 2.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$(panel (d)). Optimization with constraint implementation ($C1$).

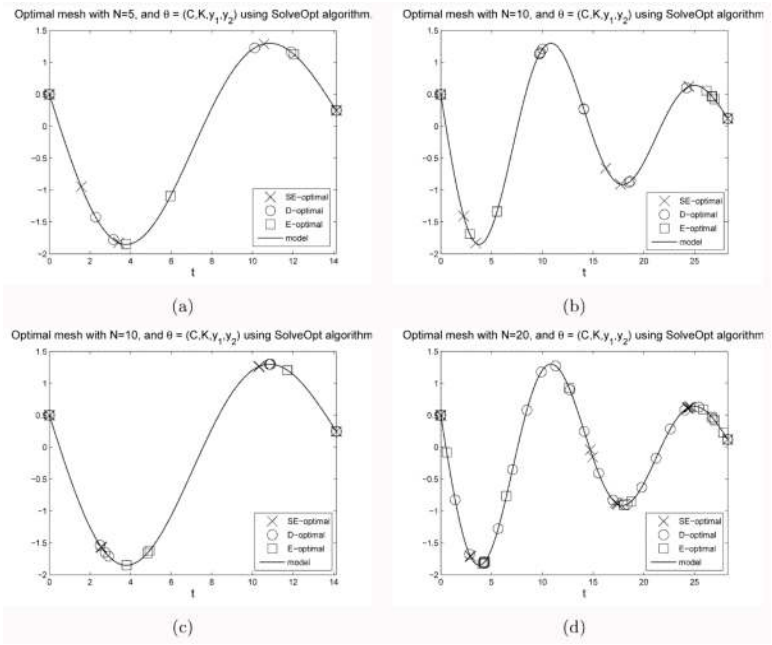

Figure 3.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$(panel (d)). Optimization with constraint implementation ($C2$).

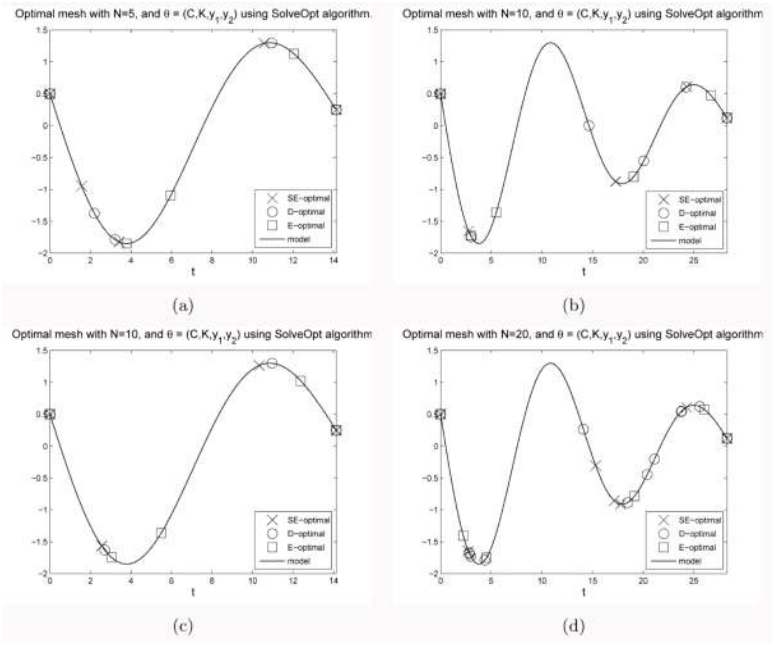

Figure 4.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$(panel (d)). Optimization with constraint implementation ($C3$).

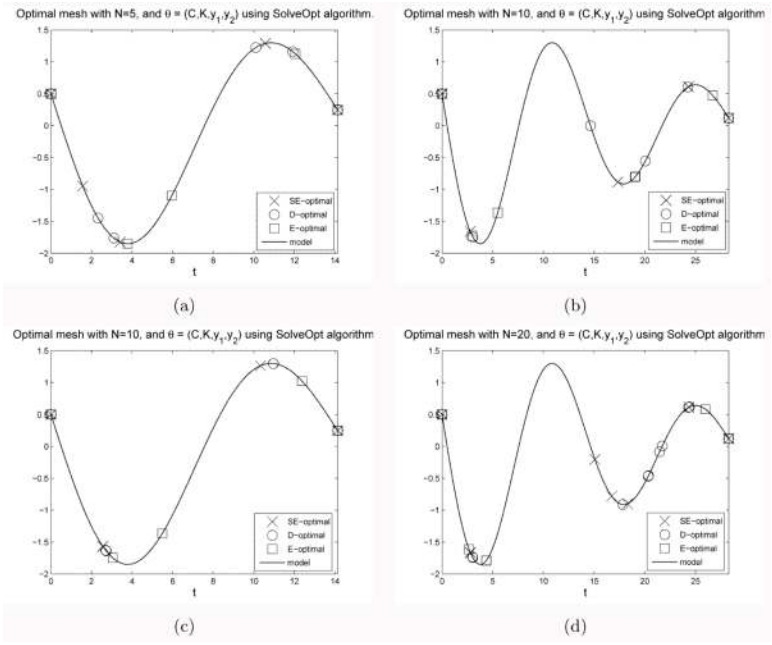

Figure 5.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$(panel (d)). Optimization with constraint implementation ($C4$).

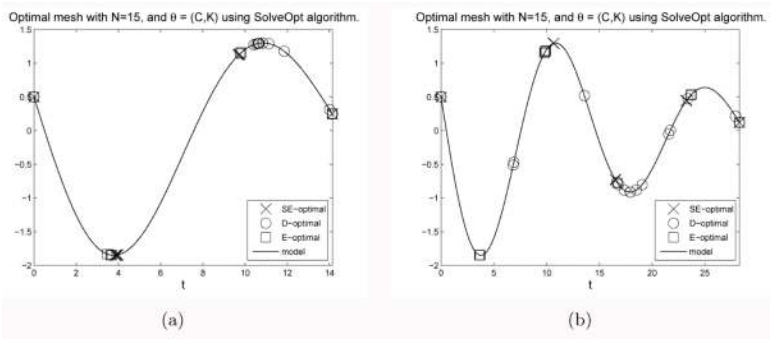

Figure 6.
Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, $N = 15$, with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation ($C2$).

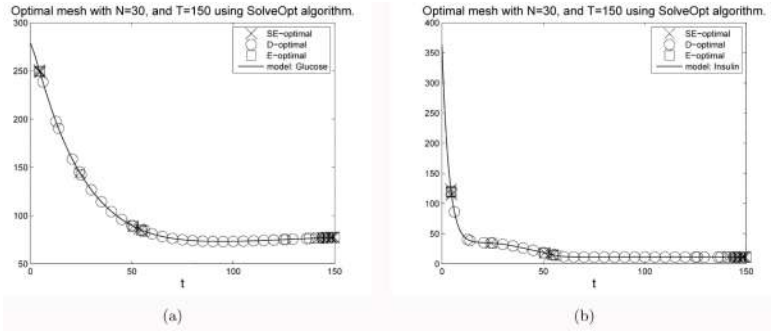

Figure 7.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation ($C1$).

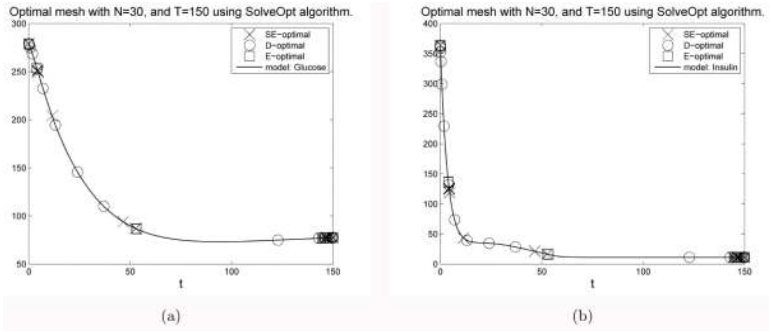

Figure 8.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation ($C2$).

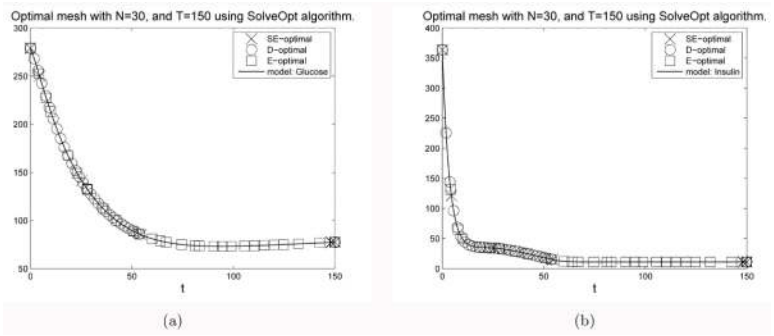

Figure 9.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation ($C3$).

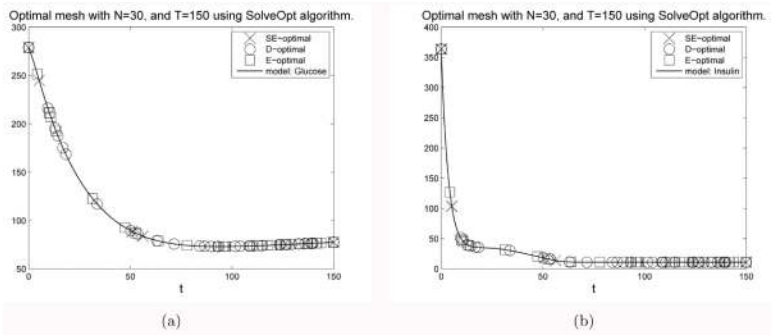

Figure 10.

Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation ($C4$).

Table 1

Average and Median estimates and standard errors using SolvOpt, $N = 10$, $M = 1000$, and $\theta_0 = (17.5, 0.7, 0.1)$. Optimization with constraint implementation ($C2$).

Parameter	Method	Average Estimate	Median Estimate	Average SE	Median SE
K	Unif	17.4978	17.4954	1.789×10^{-1}	1.789×10^{-1}
	SE-opt	17.4985	17.4995	2.000×10^{-1}	2.000×10^{-1}
	D-opt	17.5066	17.5024	2.039×10^{-1}	2.038×10^{-1}
	E-opt	17.4959	17.4957	1.512×10^{-1}	1.512×10^{-1}
r	Unif	0.7042	0.6996	5.020×10^{-2}	4.983×10^{-2}
	SE-opt	0.7019	0.7000	3.473×10^{-2}	3.444×10^{-2}
	D-opt	0.7020	0.7029	3.821×10^{-2}	3.816×10^{-2}
	E-opt	0.7139	0.7033	9.696×10^{-2}	9.090×10^{-2}
x_0	Unif	0.1037	0.0999	3.730×10^{-2}	3.696×10^{-2}
	SE-opt	0.1018	0.1002	2.448×10^{-2}	2.432×10^{-2}
	D-opt	0.1025	0.0982	2.947×10^{-2}	2.859×10^{-2}
	E-opt	0.1103	0.0977	6.417×10^{-2}	6.174×10^{-2}

Table 2

Approximate asymptotic standard errors from the asymptotic theory (17) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C2$).

T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14 (1-period)	5	*SE*-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	*D*-optimal	8.251×10^{-2}	2.541×10^{-2}	2.561×10^{-1}	3.921×10^{-1}
	5	*E*-optimal	1.371×10^{-1}	2.900×10^{-2}	3.583×10^{-1}	3.736×10^{-1}
14.14 (1-period)	10	*SE*-optimal	5.667×10^{-2}	2.484×10^{-2}	1.964×10^{-1}	2.310×10^{-1}
	10	*D*-optimal	6.055×10^{-2}	1.648×10^{-2}	1.986×10^{-1}	2.822×10^{-1}
	10	*E*-optimal	8.507×10^{-2}	2.657×10^{-2}	2.211×10^{-1}	2.283×10^{-1}
28.28 (2-periods)	10	*SE*-optimal	3.430×10^{-2}	2.149×10^{-2}	1.970×10^{-1}	2.274×10^{-1}
	10	*D*-optimal	7.445×10^{-2}	1.711×10^{-2}	4.314×10^{-1}	3.919×10^{-1}
	10	*E*-optimal	8.826×10^{-2}	2.532×10^{-2}	2.132×10^{-1}	2.169×10^{-1}
28.28 (2-periods)	20	*SE*-optimal	2.457×10^{-2}	1.500×10^{-2}	1.516×10^{-1}	1.784×10^{-1}
	20	*D*-optimal	3.254×10^{-2}	1.166×10^{-2}	1.722×10^{-1}	2.867×10^{-1}
	20	*E*-optimal	5.135×10^{-2}	1.628×10^{-2}	1.451×10^{-1}	1.492×10^{-1}

Table 3

Estimates and standard errors from the asymptotic theory (18) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation ($C2$).

T	Method	$\hat{C}_{asy}$	$\widehat{SE}(\hat{C}_{asy})$	$\hat{K}_{asy}$	$\widehat{SE}(\hat{K}_{asy})$	$\widehat{Cov}(\hat{C}_{asy}, \hat{K}_{asy})$
14.14	*SE*-optimal	0.0841	2.852×10^{-2}	0.2314	1.996×10^{-2}	-3.540×10^{-4}
14.14	*D*-optimal	0.0934	2.635×10^{-2}	0.2054	9.968×10^{-3}	-1.414×10^{-4}
14.14	*E*-optimal	0.1076	2.220×10^{-2}	0.1952	1.060×10^{-2}	-9.008×10^{-5}
14.14	Uniform	0.1300	3.529×10^{-2}	0.1938	1.278×10^{-2}	-2.803×10^{-4}
28.28	*SE*-optimal	0.0649	1.440×10^{-2}	0.1842	7.006×10^{-3}	2.883 × 10−6
28.28	*D*-optimal	0.1088	1.888×10^{-2}	0.2086	8.425×10^{-3}	-6.880×10^{-5}
28.28	*E*-optimal	0.1115	2.397×10^{-2}	0.2046	2.073×10^{-2}	-1.256×10^{-4}
28.28	Uniform	0.0854	1.792×10^{-2}	0.2122	7.326×10^{-3}	-6.219×10^{-5}

Table 4

Estimates and standard errors from the bootstrap method (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$, $M = 1000$ bootstraps and $N = 15$, optimization with constraint implementation ($C2$).

T	Method	$\hat{C}_{boot}$	$\widehat{SE}(\hat{C}_{boot})$	$\hat{K}_{boot}$	$\widehat{SE}(\hat{K}_{boot})$	$\widehat{Cov}(\hat{C}_{boot}, \hat{K}_{boot})$
14.14	*SE*-optimal	0.0783	2.076×10^{-2}	0.2320	1.628×10^{-2}	4.751×10^{-5}
14.14	*D*-optimal	0.0976	2.243×10^{-2}	0.2070	9.921×10^{-3}	-4.040×10^{-5}
14.14	*E*-optimal	0.1031	1.930×10^{-2}	0.1956	9.636×10^{-3}	3.043×10^{-5}
14.14	Uniform	0.1170	2.469×10^{-2}	0.1989	1.009×10^{-2}	-4.978×10^{-5}
28.28	*SE*-optimal	0.0576	1.479×10^{-2}	0.1842	6.057×10^{-3}	3.937×10^{-5}
28.28	*D*-optimal	0.1194	1.694×10^{-2}	0.2105	8.317×10^{-3}	4.750×10^{-6}
28.28	*E*-optimal	0.0947	2.161×10^{-2}	0.2045	1.927×10^{-2}	1.499×10^{-4}
28.28	Uniform	0.0837	1.475×10^{-2}	0.2122	6.350×10^{-3}	-4.436×10^{-6}

Table 5

Description of model parameters and typical values.

θ	**Description**	**value**
G_b	basal pre-injection level of glucose	83.7 mg/dl
I_b	basal pre-injection level of insulin	11 μU/ml
p_0	the theoretical glucose concentration in plasma at time $t = 0$	279 mg/dl
p_1	the rate constant of insulin-independent glucose uptake in muscles, and adipose tissue	2.6×10^{-2} min^{-1}
p_2	the rate constant for decrease in tissue glucose uptake ability	0.025 min^{-1}
p_3	the rate constant for the insulin-dependent increase in glucose uptake ability in tissue per unit of insulin concentration above I_b	1.25×10^{-5} min^{-2}(μU/ml)$^{-1}$
p_4	the rate constant for insulin secretion by the pancreatic β-cells after the glucose injection and with glucose concentration above p_5	4.1×10^{-3} (μU/ml) min^{-2}(mg/dl)$^{-1}$
p_5	the threshold value of glucose in plasma above which the pancreatic β-cells secrete insulin	83.7 mg/dl
p_6	the first order decay rate for insulin in plasma	0.27 min^{-1}
p_7	$p_7 + I_b$ is the theoretical insulin concentration in plasma at time $t = 0$	352.7 μU/ml

Table 6

Approximate asymptotic standard errors from the asymptotic theory (17) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation ($C1$).

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.173×10^{-3}	6.501×10^{-3}	3.100×10^{-6}	2.959×10^{-4}
D-optimal	8.411×10^{-3}	1.236×10^{-2}	6.133×10^{-6}	1.714×10^{-4}
E-optimal	4.381×10^{-3}	6.520×10^{-3}	3.182×10^{-6}	4.941×10^{-4}

Table 7

Approximate asymptotic standard errors from the asymptotic theory (17) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation ($C2$).

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.019×10^{-3}	6.451×10^{-3}	3.088×10^{-6}	3.452×10^{-4}
D-optimal	8.322×10^{-3}	1.103×10^{-2}	6.230×10^{-6}	2.748×10^{-4}
E-optimal	3.882×10^{-3}	6.284×10^{-3}	3.063×10^{-6}	5.390×10^{-4}

Table 8

Approximate asymptotic standard errors from the asymptotic theory (17) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation ($C3$).

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.205×10^{-3}	6.535×10^{-3}	3.151×10^{-6}	3.041×10^{-4}
D-optimal	7.434×10^{-3}	1.517×10^{-2}	6.171×10^{-6}	1.181×10^{-4}
E-optimal	7.528×10^{-3}	1.123×10^{-2}	5.509×10^{-6}	1.833×10^{-4}

Table 9

Approximate asymptotic standard errors from the asymptotic theory (17) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation ($C4$).

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.921×10^{-3}	6.995×10^{-3}	3.633×10^{-6}	4.796×10^{-4}
D-optimal	8.767×10^{-3}	1.249×10^{-2}	6.405×10^{-6}	1.965×10^{-4}
E-optimal	7.154×10^{-3}	1.020×10^{-2}	5.253×10^{-6}	2.302×10^{-4}

Table 10

Estimates, standard errors, and covariances between parameters from the asymptotic theory (18) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation ($C2$).

	***SE*-optimal**	***D*-optimal**	***E*-optimal**	**Uniform**
$\hat{p}_1$	2.118×10^{-2}	2.232×10^{-2}	2.116×10^{-2}	2.045×10^{-2}
$\widehat{SE}(\hat{p}_1)$	5.063×10^{-3}	8.596×10^{-3}	5.298×10^{-3}	1.056×10^{-2}
$\hat{p}_2$	3.509×10^{-2}	3.337×10^{-2}	4.356×10^{-2}	3.607×10^{-2}
$\widehat{SE}(\hat{p}_2)$	8.020×10^{-3}	1.139×10^{-2}	8.465×10^{-3}	1.536×10^{-2}
$\hat{p}_3$	1.772×10^{-5}	1.628×10^{-5}	1.958×10^{-5}	1.766×10^{-5}
$\widehat{SE}(\hat{p}_3)$	4.247×10^{-6}	6.573×10^{-6}	4.874×10^{-6}	7.787×10^{-6}
$\hat{p}_4$	4.486×10^{-3}	3.993×10^{-3}	4.249×10^{-3}	4.027×10^{-3}
$\widehat{SE}(\hat{p}_4)$	9.537×10^{-4}	5.919×10^{-4}	1.607×10^{-3}	4.817×10^{-4}
$\widehat{Cov}(\hat{p}_1, \hat{p}_2)$	-3.569×10^{-5}	-9.416×10^{-5}	-3.811×10^{-5}	-1.579×10^{-4}
$\widehat{Cov}(\hat{p}_1, \hat{p}_3)$	-2.036×10^{-8}	-5.566×10^{-8}	-2.376×10^{-8}	-8.160×10^{-8}
$\widehat{Cov}(\hat{p}_1, \hat{p}_4)$	6.620×10^{-7}	1.227×10^{-7}	1.774×10^{-6}	8.615×10^{-7}
$\widehat{Cov}(\hat{p}_2, \hat{p}_3)$	3.131×10^{-8}	7.280×10^{-8}	3.585×10^{-8}	1.181×10^{-7}
$\widehat{Cov}(\hat{p}_2, \hat{p}_4)$	4.626×10^{-7}	4.238×10^{-7}	9.670×10^{-7}	-5.341×10^{-7}
$\widehat{Cov}(\hat{p}_3, \hat{p}_4)$	-6.824×10^{-10}	9.532×10^{-13}	-2.353×10^{-9}	-5.605×10^{-10}

Table 11

Estimates, standard errors, and covariances between parameters from the bootstrap method (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, $M = 1000$ bootstraps and $N = 30$, optimization, using fmincon, with constraint implementation ($C2$).

	***SE*-optimal**	***D*-optimal**	***E*-optimal**	**Uniform**
$\hat{p}_1$	1.874×10^{-2}	1.883×10^{-2}	2.104×10^{-2}	1.973×10^{-2}
$\widehat{SE}(\hat{p}_1)$	6.619×10^{-3}	8.291×10^{-3}	6.397×10^{-3}	8.563×10^{-3}
$\hat{p}_2$	4.034×10^{-2}	4.249×10^{-2}	4.337×10^{-2}	3.730×10^{-2}
$\widehat{SE}(\hat{p}_2)$	1.305×10^{-2}	1.458×10^{-2}	1.409×10^{-2}	1.279×10^{-2}
$\hat{p}_3$	2.124×10^{-5}	2.069×10^{-5}	2.075×10^{-5}	1.916×10^{-5}
$\widehat{SE}(\hat{p}_3)$	8.241×10^{-6}	8.799×10^{-6}	7.733×10^{-6}	8.017×10^{-6}
$\hat{p}_4$	4.341×10^{-3}	3.920×10^{-3}	3.988×10^{-3}	3.984×10^{-3}
$\widehat{SE}(\hat{p}_4)$	4.228×10^{-4}	3.107×10^{-4}	6.192×10^{-4}	2.098×10^{-4}
$\widehat{Cov}(\hat{p}_1, \hat{p}_2)$	-8.157×10^{-5}	-1.149×10^{-4}	-7.952×10^{-5}	-1.053×10^{-4}
$\widehat{Cov}(\hat{p}_1, \hat{p}_3)$	-5.272×10^{-8}	-7.128×10^{-8}	-4.687×10^{-8}	-6.722×10^{-8}
$\widehat{Cov}(\hat{p}_1, \hat{p}_4)$	1.240×10^{-8}	1.275×10^{-7}	-5.657×10^{-7}	5.716×10^{-8}
$\widehat{Cov}(\hat{p}_2, \hat{p}_3)$	1.048×10^{-7}	1.249×10^{-7}	1.042×10^{-7}	9.990×10^{-8}
$\widehat{Cov}(\hat{p}_2, \hat{p}_4)$	4.220×10^{-7}	8.311×10^{-8}	2.226×10^{-6}	2.310×10^{-7}
$\widehat{Cov}(\hat{p}_3, \hat{p}_4)$	6.133×10^{-11}	-2.764×10^{-11}	9.390×10^{-10}	8.308×10^{-11}