Comparison of Optimal Design Methods in Inverse Problems

H. T. Banks[1,2], Kathleen Holm[1,2] and Franz Kappel[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
and
[3]Institute for Mathematics and Scientific Computation
University of Graz
Graz, Austria A8010

July 6, 2010

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 from the optimal sampling distribution. We compare here three different optimal design methods. These are are two standard designs, D-optimal, and E-optimal, and a recently [5] proposed one, the SE-optimal design. 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 [7], the standard harmonic oscillator model [7] and a popular glucose regulation model [10, 12, 19].

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

Key Words: 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 or physical systems. If the parameters in the model are known, the model can be used for prediction. However, typically we do not know the parameter values. Instead, we must estimate the parameters using experimental data. The predictive capabilities of the model depend on the accuracy of the parameter estimates. Traditional optimal design methods (D-optimal, E-optimal, c-optimal) use information from the model to find the sampling distribution for the data that minimizes a design criterion, quite often a function of the Fisher Information Matrix (FIM). Experimental data taken on this optimal mesh is supposed to result in accurate parameter estimates.

We compare three different optimal design methods for the Verhulst-Pearl logistic population model, a harmonic oscillator model and a simple glucose regulation model. The optimal design methods in these comparisons are SE-optimal, D-optimal, and E-optimal design. SE-optimal design (standard error-optimal design) was introduced in [5]. 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 parameters. D-optimal and E-optimal design methods minimize the variance in the parameter estimates [1, 8, 13]. 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.

For each optimal design method, standard errors are computed by various method using the optimal mesh. The optimal design methods are compared based on the standard errors. We expect that SE-optimal design will result in smaller standard errors compared with the other optimal design method since SE-optimal design optimizes directly on the standard errors themselves while the D-optimal and E-optimal methods minimize functions of the standard errors.

2 Optimal Design Formulations

Following [5], 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

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

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 $\mathcal{C}$ 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 [4], one also requires a *statistical model*. 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 [4, 5, 18], θ_0 is a hypothesized "true" value of the unknown parameters. We make the following standard assumptions on the random variable $\mathcal{E}(t)$:

$$\begin{aligned} E(\mathcal{E}(t)) &= 0, \quad t \in [0,T], \\ \mathrm{Var}\mathcal{E}(t) &= \sigma_0^2, \quad t \in [0,T], \\ \mathrm{Cov}(\mathcal{E}(t)\mathcal{E}(s)) &= \sigma_0^2 \delta(t-s), \quad t,s \in [0,T], \end{aligned}$$

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

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

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

We introduce a generalized weighted least squares criterion

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

where P is a general measure on $[0,T]$. We seek the parameter estimate $\hat{\theta}$ 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 $\{a\}$, we obtain

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

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* (FIM)

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

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

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

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

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

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 [5]. We do not consider this design here, but rather focus on the SE-optimal design also proposed in [5] 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 Fisher information matrix $F(P,\theta)$ 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 FIM is given. The *optimal design problem* associated with $\mathcal{J}$ is one of finding a probability measure $\hat{P} \in \mathcal{P}(0,T)$ such that

$$\mathcal{J}\big(F(\hat{P},\theta_0)\big) = \min_{P\in\mathcal{P}(0,T)} \mathcal{J}\big(F(P,\theta_0)\big). \tag{9}$$

A general theoretical framework for existence and approximation in the context of $\mathcal{P}(0,T)$ taken with the Prohorov metric [2, 11, 14, 17] is given for these problems in Section 4 of [5]. In particular, this theory permits development of computational methods using weighted discrete measures (i.e., weighted versions of (4)). This formulation incorporates all strategies for optimal design which try to optimize 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, the optimal design methods we consider are SE-optimal design, D-optimal design, and E-optimal design. The design cost functional for the SE-optimal design method is given by (see [5])

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

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 esitmated. Note 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 [8, 13])

$$\mathcal{J}(F) = \det(F^{-1}).$$

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 [1, 8])

$$\mathcal{J}(F) = \max \frac{1}{\lambda_i},$$

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

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 $0 \leq t_1 \leq \cdots \leq t_i \leq t_{i+1} \leq \cdots \leq t_N \leq T$.

These optimal design methods were implemented using constrained optimization algorithms, either MATLAB's *fmincon* or SolvOpt, developed by A. Kuntsevich and F. Kappel [15], with four variations on the constraint implementation. We denote these different *constraint implementations* (which result in different parameter and SE outcomes even in cases where the $\{t_i\}$ are initially required to satisfy similar constraints) by $(C1) - (C4)$. Complete details of the differences in the algorithms are given in an appendix.

$(C1)$ 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 N 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 $t_i = t_{i-1} + \nu_i$, $i = 2, \ldots, N-1$, $t_1 = 0$ and $t_N = T$, with $\nu_i \geq 0$, $i = 2, \ldots, N-1$, and $\nu_2 + \ldots + \nu_{N-1} \leq T$. 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, $t_i = t_{i-1} + \nu_i$, $i = 2, \ldots, N$, and $t_1 = 0$ with $\nu_i \geq 0$, $i = 2, \ldots, N$, and $\nu_2 + \ldots + \nu_N = T$. This constraint is implemented by defining $\nu_N = T - \sum_{i=2}^{N-1} \nu_i$. 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 [4]. Since 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

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

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

$$\hat{\theta}_{\mathrm{OLS}} = \hat{\theta}_{\mathrm{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [y_j - f(t_j, \theta)]^2.$$

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

$$\chi_{j,k} = \frac{\partial(\mathcal{C}x(t_j))}{\partial \theta_k} = \frac{\partial f(t_j, \theta)}{\partial \theta_k}, \text{ for } j = 1, \ldots N, k = 1, \ldots p,$$

where p is the number of parameters being estimated. Note that $\chi = \chi^N$ is an $N \times p$ matrix. The true constant variance

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

can be estimated by

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

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

$$\Sigma_0^N \approx \sigma_0^2 [\chi^T(\theta_0)\chi(\theta_0)]^{-1}.$$

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

$$F = F(\tau) = F(\tau, \theta_0) = (\Sigma_0^N)^{-1}, \tag{10}$$

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

$$\Sigma_0^N \approx \hat{\Sigma}^N(\hat{\theta}_{\mathrm{OLS}}) = \hat{\sigma}_{\mathrm{OLS}}^2 [\chi^T(\hat{\theta}_{\mathrm{OLS}})\chi(\hat{\theta}_{\mathrm{OLS}})]^{-1}. \tag{11}$$

The corresponding FIM can be estimated by

$$\hat{F}(\tau) = \hat{F}(\tau, \hat{\theta}_{\mathrm{OLS}}) = (\hat{\Sigma}^N(\hat{\theta}_{\mathrm{OLS}}))^{-1}.$$

The asymptotic standard errors are given by

$$SE_k(\theta_0) = \sqrt{(\Sigma_0^N)_{kk}}, \; k = 1, \ldots, p. \tag{12}$$

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

$$SE_k(\hat{\theta}_{\mathrm{OLS}}) = \sqrt{(\hat{\Sigma}^N(\hat{\theta}_{\mathrm{OLS}}))_{kk}}, \; k = 1, \ldots, p. \tag{13}$$

It can be shown, under certain conditions, for $N \to \infty$, that the estimator Θ_{OLS}^N is asymptotically normal [18]; i.e., for N large

$$\Theta_{\mathrm{OLS}}^N \sim \mathcal{N}_p(\theta_0, \Sigma_0^N).$$

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 $\{t_j\}_{j=1}^N$,

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

where $\theta^0 = 1.4\theta_0$, where θ_0 are the true parameter values and the ϵ_j are realizations of $\mathcal{E}_j \sim \mathcal{N}(0, \sigma^2)$ for $j = 1, \ldots, N$. Parameters are estimated using OLS and standard errors are estimated using asymptotic theory (13). 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 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 [6]. 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

$$Y_j = f(t_j, \theta_0) + \mathcal{E}_j, \tag{14}$$

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

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

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 below.

The following algorithm [6] can be used to compute the bootstrapping estimate $\hat{\theta}_{boot}$ of θ_0 and its empirical distribution.

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

2. Using this estimate define the standardized residuals:

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

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

$$E(\bar{r}_j | F_N) = N^{-1} \sum_{j=1}^N \bar{r}_j = 0, \qquad \mathrm{var}(\bar{r}_j | F_N) = N^{-1} \sum_{j=1}^N \bar{r}_j^2 = \hat{\sigma}^2.$$

   Set $m = 0$.

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

4. Create bootstrap sample points
$$y_j^m = f(t_j, \hat{\theta}^0) + r_j^m,$$
where $j = 1, \ldots, N$.

5. Obtain a new estimate $\hat{\theta}^{m+1}$ from the bootstrap sample $\{y_j^m\}$ using OLS. Add $\hat{\theta}^{m+1}$ into the vector Θ, where Θ is a vector of length Mp 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 Mp.

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

$$\begin{aligned} \hat{\theta}_{boot} &= \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \hat{\theta}^m, \\ Cov(\hat{\theta}_{boot}) &= \tfrac{1}{M-1} \textstyle\sum_{m=1}^{M} (\hat{\theta}^m - \hat{\theta}_{boot})^T (\hat{\theta}^m - \hat{\theta}_{boot}), \\ SE_k(\hat{\theta}_{boot}) &= \sqrt{Cov(\hat{\theta}_{boot})_{kk}}. \end{aligned} \tag{15}$$

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:

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

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 $\mathcal{E} \sim \mathcal{N}(0, \sigma_0^2)$ to generate simulated data (with for use in the Monte Carlo calculations. A realization of the observation process is given by

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

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 implementation ($C2$) on the time points: $t_1 = 0$, $t_N = T$ and $t_i \leq t_{i+1}$, such that the optimal mesh contains 0 and T. For this example, we took $T = 25$ and $N = 10$ or $N = 15$. 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 $\theta_0 = (K, r, x_0) = (17.5, 0.8, 0.1)$, and variance $\sigma_0^2 = 0.16$. The average estimates and standard errors are based on $M = 1000$ Monte Carlo trials. Average Monte Carlo asymptotic estimates and standard errors were also computed on the uniform mesh.

We report the average estimates and standard errors in Tables 1 and 2 ($N = 10$, and $N = 15$). Figures 1 and 5 contain plots of the resulting optimal distribution of time points for the different optimal design methods, along with the uniform mesh, plotted on the logistic curve, for $N = 10$ and $N = 15$, respectively.

Histograms for the Monte Carlo standard errors of K are given in Figs. 2 and 6, for r in Figs. 3 and 7, and for x_0 in Figs. 4 and 8. Each of the histogram figures contain the curves for the uniform points and the three optimal design methods. However, the standard errors for the four different methods have different magnitude in many cases. Within Figs. 2 - 4 and 6 - 8 subfigures display the same results but for different scales.

Table 1: Average estimates and standard errors using SolvOpt, $N = 10$, $M = 1000$, and $\theta_0 = (17.5, 0.8, 0.1)$. Optimization with constraint implementation ($C2$).

Method	Parameter	Average Estimate	Average Standard Error
Uniform	K	17.5001	1.789×10^{-1}
	r	0.7044	5.019×10^{-2}
	x_0	0.1040	3.736×10^{-2}
SE-optimal	K	17.4972	2.000×10^{-1}
	r	0.7010	3.468×10^{-2}
	x_0	0.1023	2.457×10^{-2}
D-optimal	K	8.3550×10^{11}	1.600×10^{25}
	r	139.7750	5.348×10^{4}
	x_0	0.1081	3.238×10^{-1}
E-optimal	K	8.244×10^{10}	6.649×10^{23}
	r	133.9957	2.291×10^{4}
	x_0	0.1087	2.701×10^{-1}

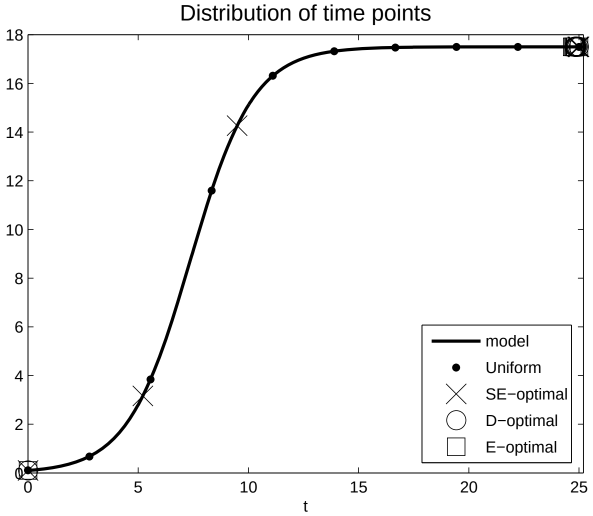
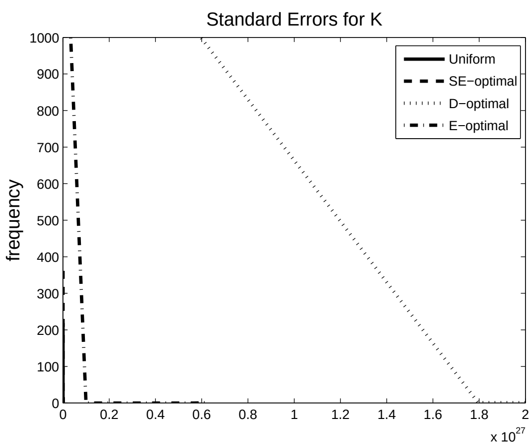
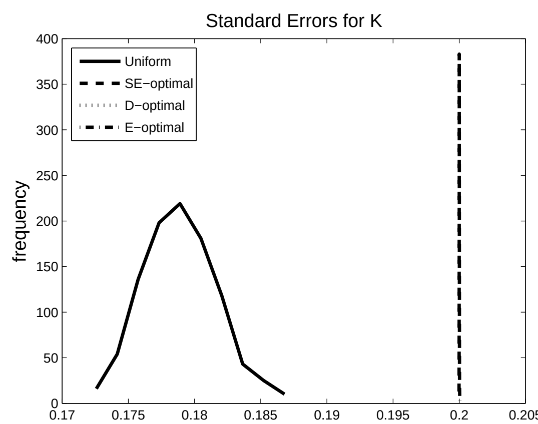


Figure 1: The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points obtained using SolvOpt, with $N = 10$, and the optimal design methods SE-optimality, D-optimality, and E-optimality. Optimization with constraint implementation ($C2$).



(a)



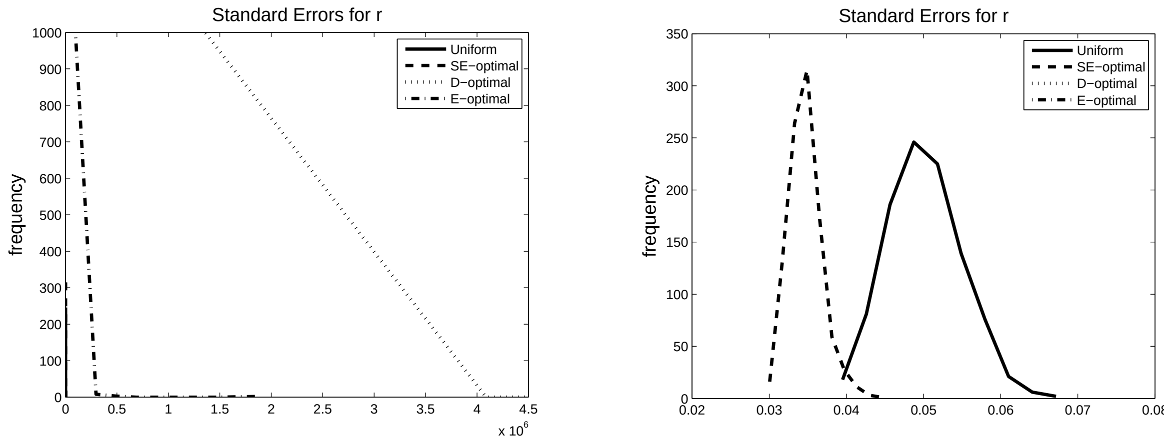
(b)

Figure 2: Using SolvOpt, with $N = 10$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_K, shown in panels (a) and (b) for various scales. Optimization with constraint implementation ($C2$).

4.3 Discussion of Logistic Results

The average asymptotic estimates from the uniform distribution and SE-optimality are both very close to the true values, θ_0. The average asymptotic estimates for x_0 from D-optimality and E-optimality are close to the true value. When $N = 10$ (Table 1), the average estimates from D-optimal and E-optimal are much worse for K and r . When $N = 15$ (Table 2), D-optimality's average estimate for K is closer to the true value (but not as close as SE-optimal), but it is worse for r compared to the other optimal design methods.

Comparing the standard errors (average standard errors and histogram plots), we find that SE-



(a) (b)

Figure 3: Using SolvOpt, with $N = 10$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_r, shown in panels (a) and (b) for various scales. Optimization with constraint implementation ($C2$).

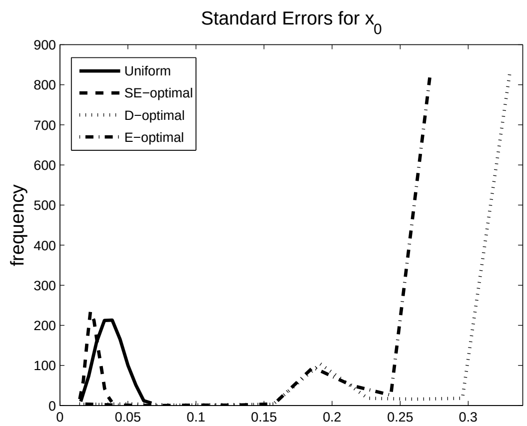


Figure 4: Using SolvOpt, with $N = 10$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_{x_0}. Optimization with constraint implementation ($C2$).

optimal has consistently smaller average standard errors than both D-optimal and E-optimal design. The uniform grid has smaller standard errors compared to SE-optimal for K (when $N = 10$ and $N = 15$), but SE-optimal has smaller standard errors than the uniform grid for r and x_0. E-optimal design has smaller standard errors than D-optimal except for K when $N = 15$.

SE-optimal design is often comparable to the uniform mesh in terms of average parameter estimates and standard errors. Though SE-optimal has smaller standard errors for r and x_0 compared to the uniform mesh. SE-optimal design has average parameter estimates closer to the true value and smaller average standard errors compared to both D and E-optimal design for this logistic example.

Table 2: Average estimates and standard errors using SolvOpt, $N = 15$, $M = 1000$, and $\theta_0 = (17.5, 0.8, 0.1)$. Optimization with constraint implementation ($C2$).

Method	Parameter	Average Estimate	Average Standard Error
Uniform	K	17.5037	1.467×10^{-1}
	r	0.7020	4.119×10^{-2}
	x_0	0.1033	3.049×10^{-2}
SE-optimal	K	17.5044	1.633×10^{-1}
	r	0.7004	2.735×10^{-2}
	x_0	0.1018	2.002×10^{-2}
D-optimal	K	17.8309	1.129×10^{0}
	r	53.8564	3.045×10^{3}
	x_0	0.1070	2.385×10^{-1}
E-optimal	K	1.186×10^{12}	9.094×10^{24}
	r	22.5871	8.811×10^{2}
	x_0	0.0924	1.307×10^{-1}

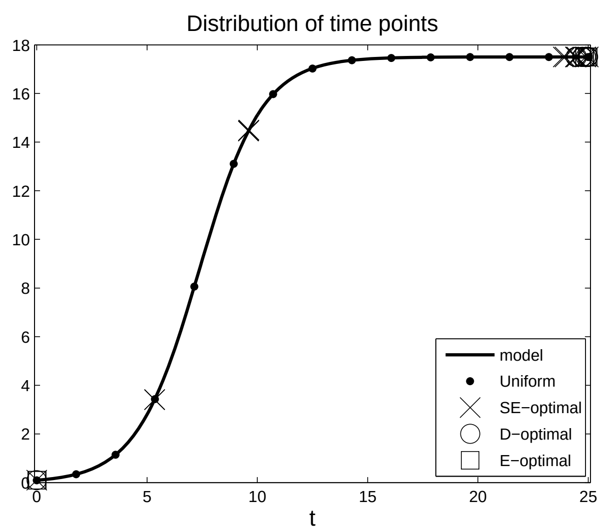
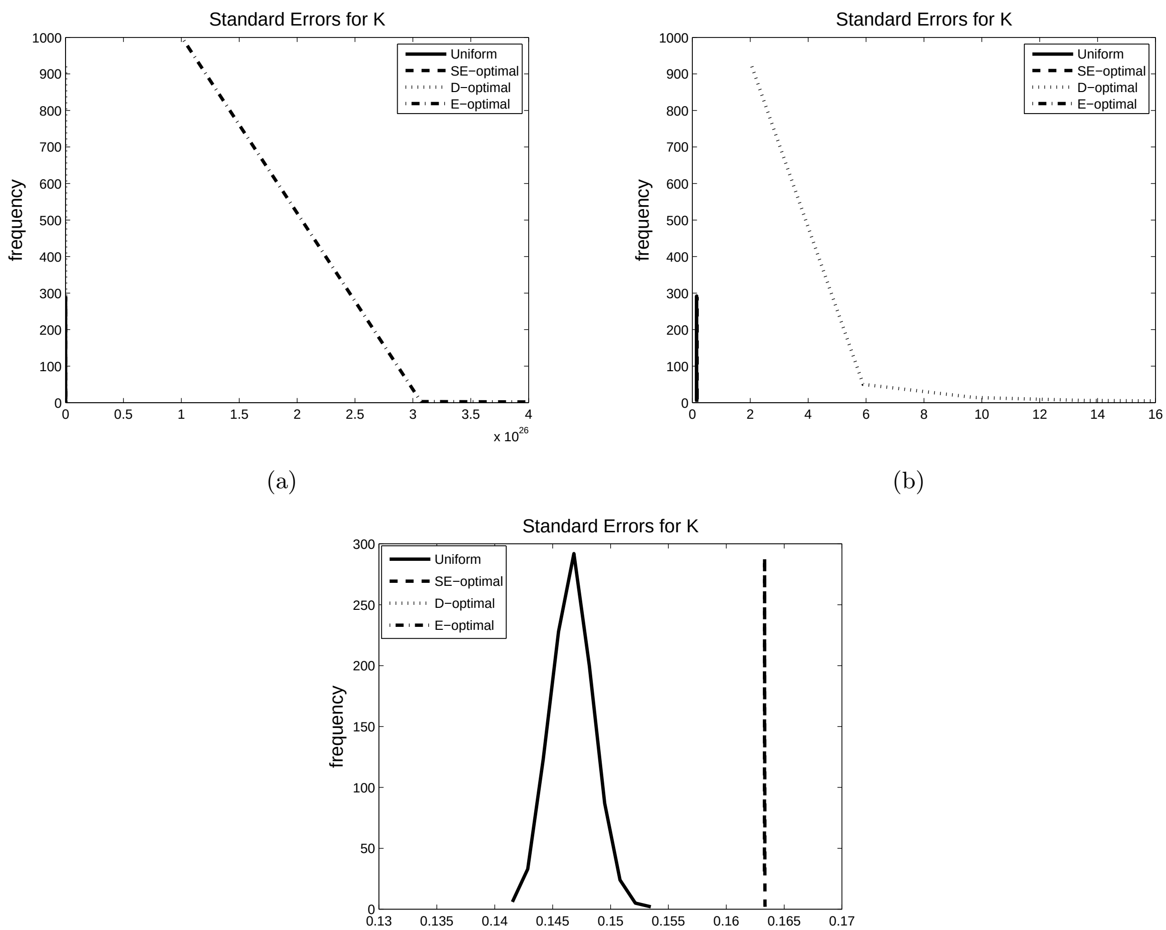
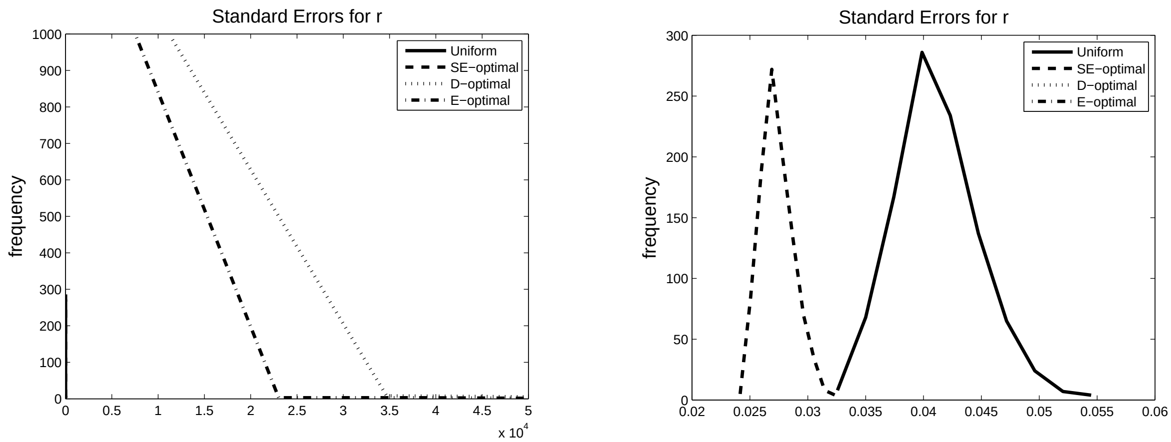


Figure 5: The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points obtained using SolvOpt, with $N = 15$, and the optimal design methods SE-optimality, D-optimality, and E-optimality. Optimization with constraint implementation ($C2$).



(a) (b) (c)

Figure 6: Using SolvOpt, with $N = 15$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_K, shown in panels (a) - (c) for various scales. Optimization with constraint implementation ($C2$).



(a) (b)

Figure 7: Using SolvOpt, with $N = 15$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_r, shown in panels (a) and (b) for various scales. Optimization with constraint implementation ($C2$).

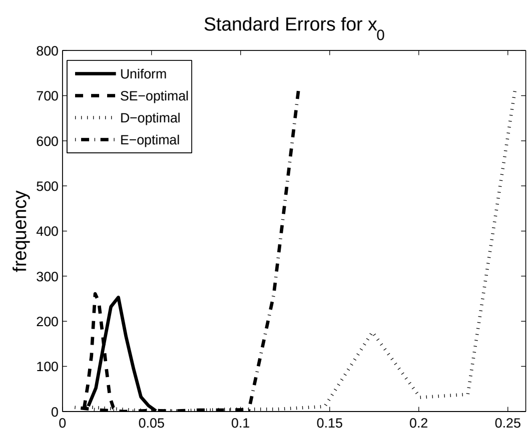


Figure 8: Using SolvOpt, with $N = 15$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_{x_0}. Optimization with constraint implementation ($C2$).

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 mass-balance (see [7]) and is given by

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

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.

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

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,

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

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 (12). Recall that uncertainty is quantified by constructing confidence intervals using parameter estimate with the asymptotic standard error. Since our main focus here is the width of the confidence intervals, we can forgo the obtaining of the parameter estimates themselves which 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. 9, the second constraint implementation ($C2$) in Fig. 10, the third constraint implementation ($C3$) in Fig. 11, and the last constraint implementation ($C4$) in Fig. 12. The standard errors (12) from the asymptotic theory corresponding to these optimal meshes are given in Table 3-6, respectively for the four different constraints.

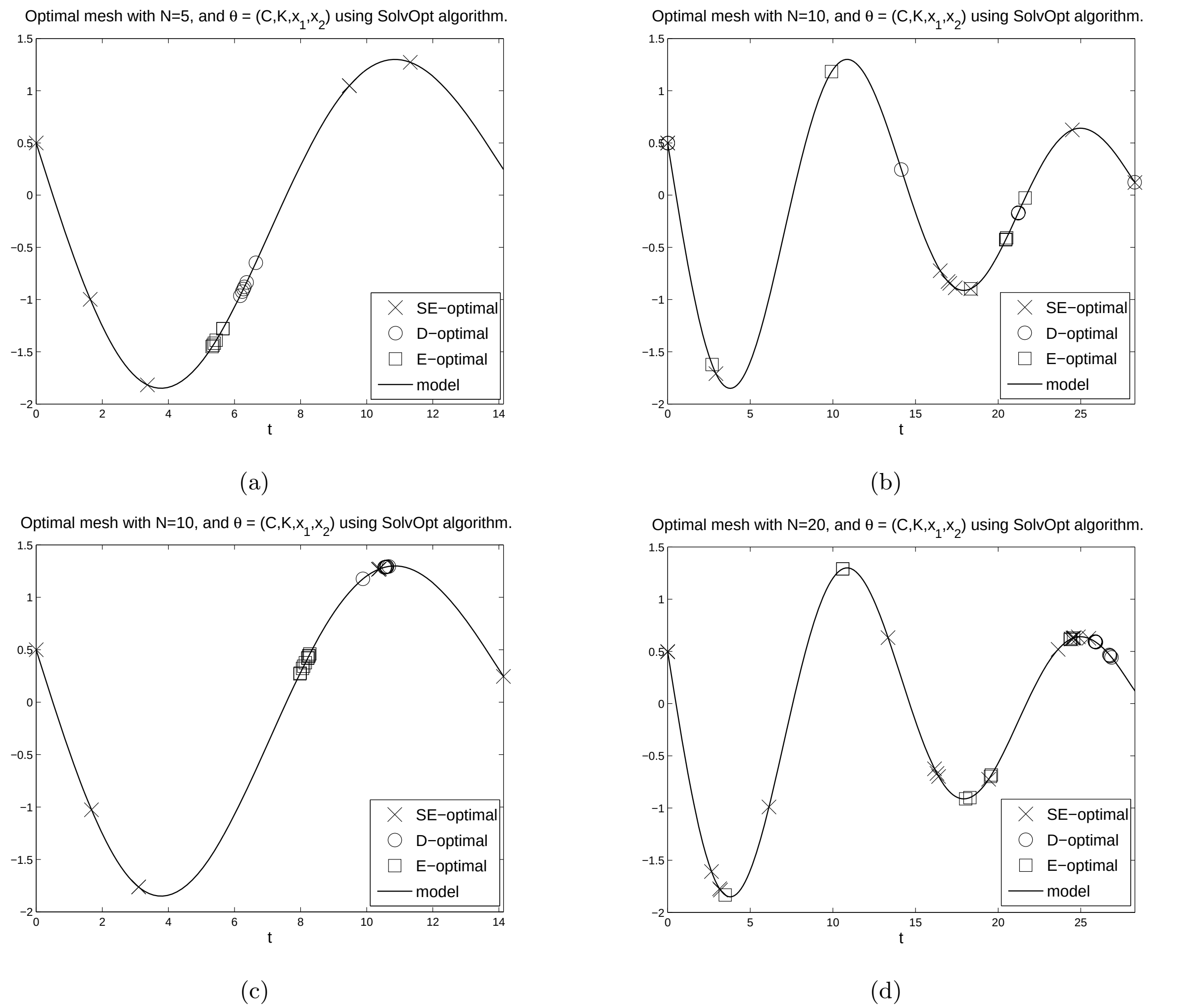


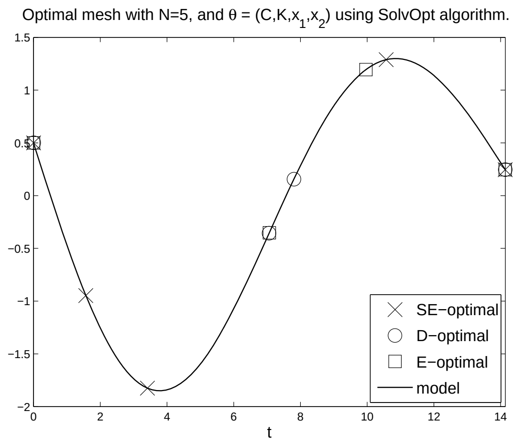
Figure 9: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (a)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (b)). Optimization with constraint implementation $(C1)$.

Table 3: Approximate asymptotic standard errors from the asymptotic theory (12) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C1$).

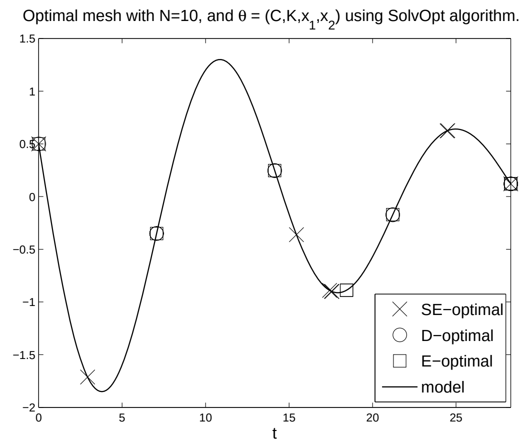
T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14 (1-period)	5	SE-optimal	7.603×10^{-2}	4.318×10^{-2}	2.869×10^{-1}	3.714×10^{-1}
	5	D-optimal	1.940×10^{2}	9.372×10^{2}	2.917×10^{3}	1.657×10^{4}
	5	E-optimal	4.218×10^{3}	2.597×10^{3}	1.710×10^{4}	5.139×10^{4}
14.14 (1-period)	10	SE-optimal	5.526×10^{-2}	2.519×10^{-2}	2.113×10^{-1}	2.717×10^{-1}
	10	D-optimal	2.516×10^{3}	2.379×10^{2}	1.192×10^{4}	2.907×10^{3}
	10	E-optimal	2.612×10^{3}	1.634×10^{3}	5.468×10^{3}	2.586×10^{4}
28.28 (2-periods)	10	SE-optimal	4.040×10^{-2}	2.055×10^{-2}	2.605×10^{-1}	2.308×10^{-1}
	10	D-optimal	8.718×10^{1}	4.458×10^{0}	9.164×10^{2}	1.513×10^{0}
	10	E-optimal	3.632×10^{1}	5.021×10^{1}	7.621×10^{2}	2.717×10^{3}
28.28 (2-periods)	20	SE-optimal	2.509×10^{-2}	1.440×10^{-2}	1.497×10^{-1}	1.750×10^{-1}
	20	D-optimal	3.195×10^{2}	1.465×10^{2}	5.558×10^{3}	1.156×10^{4}
	20	E-optimal	6.012×10^{-1}	1.453×10^{0}	7.891×10^{0}	7.002×10^{1}

Table 4: Approximate asymptotic standard errors from the asymptotic theory (12) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C2$).

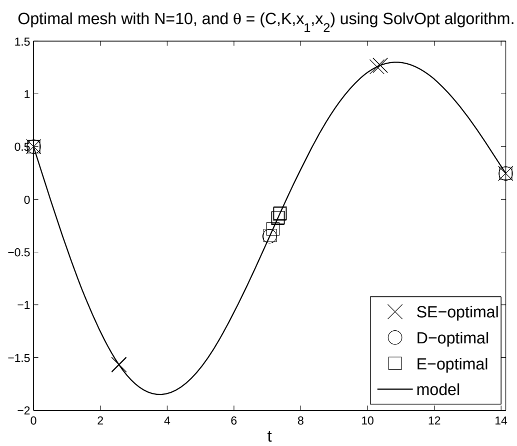
T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14 (1-period)	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	1.647×10^{4}	8.422×10^{2}	5.935×10^{4}	3.970×10^{-1}
	5	E-optimal	1.639×10^{3}	8.387×10^{1}	7.547×10^{3}	2.819×10^{-1}
14.14 (1-period)	10	SE-optimal	5.667×10^{-2}	2.483×10^{-2}	1.964×10^{-1}	2.309×10^{-1}
	10	D-optimal	1.841×10^{2}	9.418×10^{0}	7.515×10^{2}	3.216×10^{-1}
	10	E-optimal	4.636×10^{1}	2.373×10^{0}	1.579×10^{2}	2.680×10^{-1}
28.28 (2-periods)	10	SE-optimal	3.888×10^{-2}	1.928×10^{-2}	2.559×10^{-1}	2.304×10^{-1}
	10	D-optimal	6.242×10^{1}	3.199×10^{0}	5.901×10^{2}	3.259×10^{-1}
	10	E-optimal	9.495×10^{1}	4.859×10^{0}	8.095×10^{2}	2.297×10^{-1}
28.28 (2-periods)	20	SE-optimal	2.632×10^{-2}	1.788×10^{-2}	1.359×10^{-1}	1.413×10^{-1}
	20	D-optimal	2.794×10^{3}	1.433×10^{2}	3.645×10^{4}	4.000×10^{-1}
	20	E-optimal	1.385×10^{1}	7.126×10^{-1}	1.039×10^{2}	2.028×10^{-1}



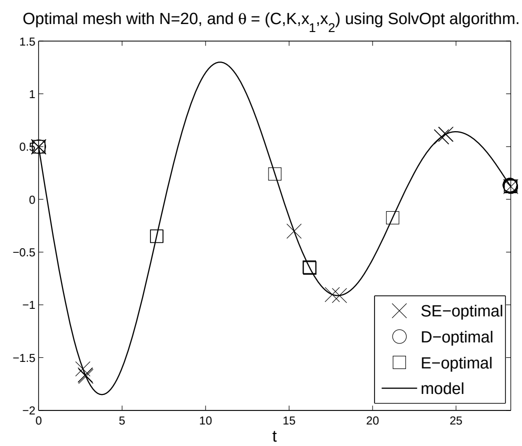
(a)



(b)



(c)



(d)

Figure 10: 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 $(C2)$.

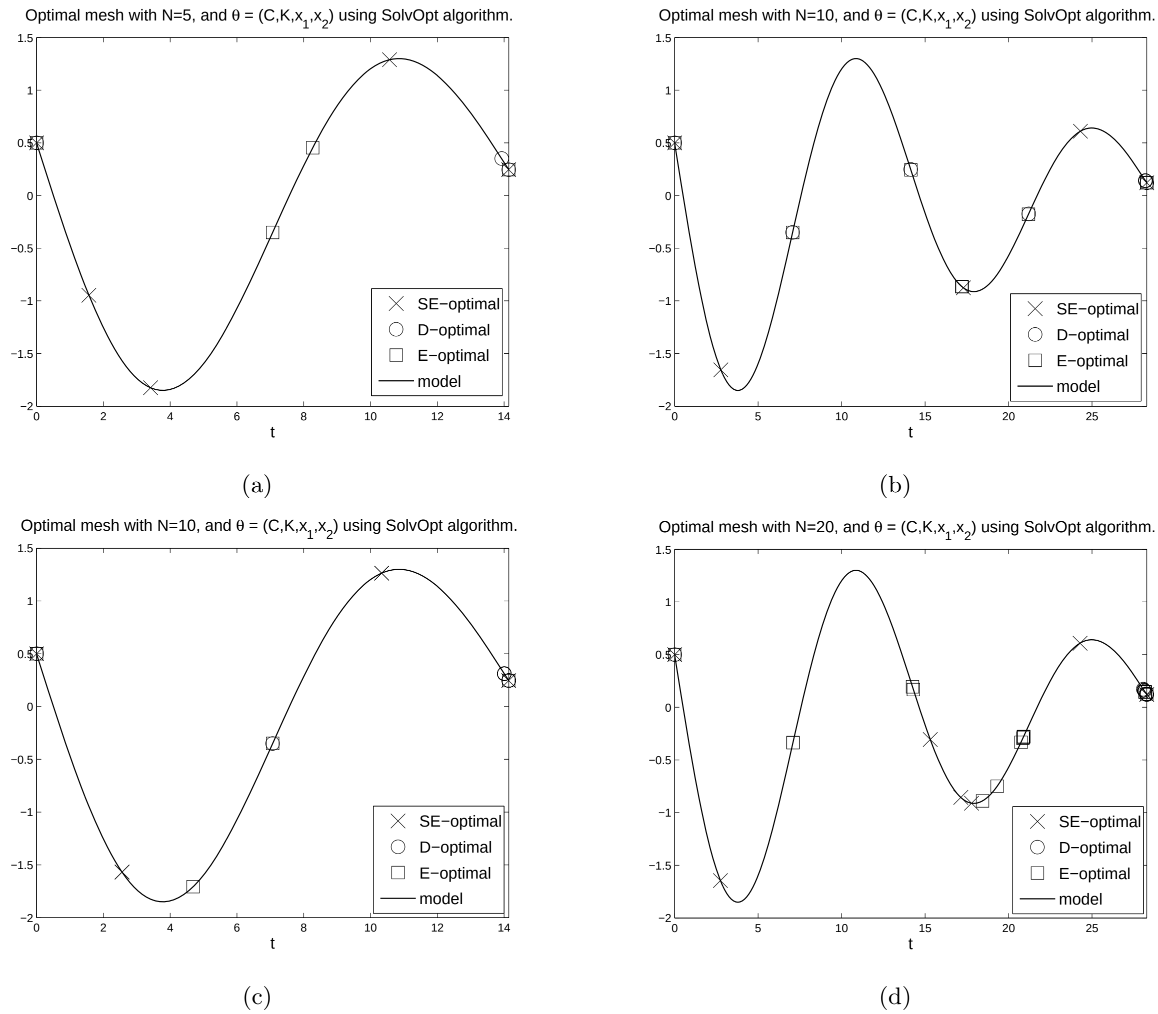


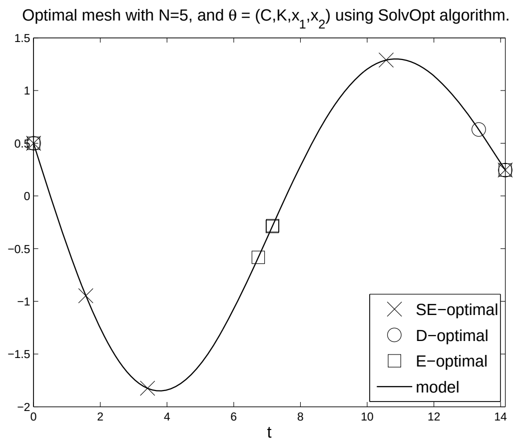
Figure 11: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (d)). Optimization with constraint implementation $(C3)$.

Table 5: Approximate asymptotic standard errors from the asymptotic theory (12) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C3$).

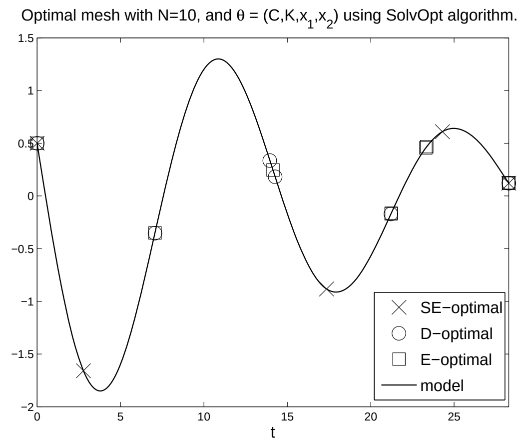
T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14 (1-period)	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	5.923×10^{4}	3.037×10^{3}	3.813×10^{5}	4.000×10^{-1}
	5	E-optimal	2.771×10^{2}	1.419×10^{1}	1.059×10^{3}	2.797×10^{-1}
14.14 (1-period)	10	SE-optimal	5.666×10^{-2}	2.484×10^{-2}	1.963×10^{-1}	2.309×10^{-1}
	10	D-optimal	1.146×10^{4}	5.879×10^{2}	7.421×10^{4}	3.798×10^{-1}
	10	E-optimal	6.908×10^{2}	3.527×10^{1}	1.498×10^{3}	3.669×10^{-1}
28.28 (2-periods)	10	SE-optimal	3.673×10^{-2}	2.399×10^{-2}	1.925×10^{-1}	2.000×10^{-1}
	10	D-optimal	3.569×10^{3}	1.830×10^{2}	4.650×10^{4}	3.419×10^{-1}
	10	E-optimal	3.236×10^{1}	1.664×10^{0}	2.573×10^{2}	2.701×10^{-1}
28.28 (2-periods)	20	SE-optimal	2.671×10^{-2}	1.812×10^{-2}	1.368×10^{-1}	1.413×10^{-1}
	20	D-optimal	4.071×10^{2}	2.088×10^{1}	5.284×10^{3}	4.000×10^{-1}
	20	E-optimal	1.892×10^{0}	1.094×10^{-1}	1.619×10^{1}	3.117×10^{-1}

Table 6: Approximate asymptotic standard errors from the asymptotic theory (12) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C4$).

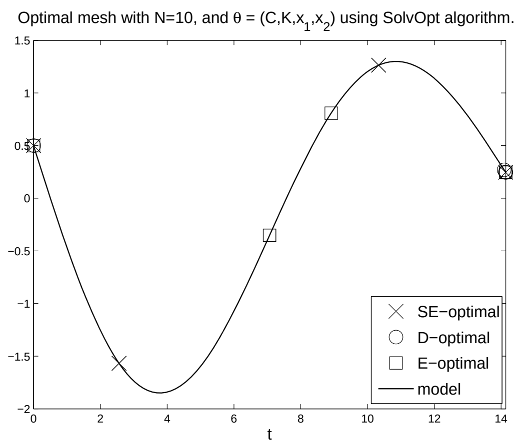
T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14 (1-period)	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	4.672×10^{5}	2.395×10^{4}	2.881×10^{6}	4.000×10^{-1}
	5	E-optimal	4.780×10^{1}	2.428×10^{0}	1.507×10^{2}	2.827×10^{-1}
14.14 (1-period)	10	SE-optimal	5.666×10^{-2}	2.484×10^{-2}	1.963×10^{-1}	2.309×10^{-1}
	10	D-optimal	1.314×10^{4}	6.738×10^{2}	8.556×10^{4}	4.000×10^{-1}
	10	E-optimal	2.104×10^{2}	1.078×10^{1}	8.668×10^{2}	3.634×10^{-1}
28.28 (2-periods)	10	SE-optimal	3.554×10^{-2}	2.395×10^{-2}	1.906×10^{-1}	2.000×10^{-1}
	10	D-optimal	2.553×10^{4}	1.301×10^{3}	1.663×10^{5}	3.645×10^{-1}
	10	E-optimal	2.655×10^{1}	1.367×10^{0}	2.862×10^{2}	3.569×10^{-1}
28.28 (2-periods)	20	SE-optimal	2.474×10^{-2}	1.683×10^{-2}	1.347×10^{-1}	1.510×10^{-1}
	20	D-optimal	8.138×10^{2}	4.174×10^{1}	1.056×10^{4}	4.000×10^{-1}
	20	E-optimal	4.629×10^{0}	2.514×10^{-1}	3.763×10^{1}	3.295×10^{-1}



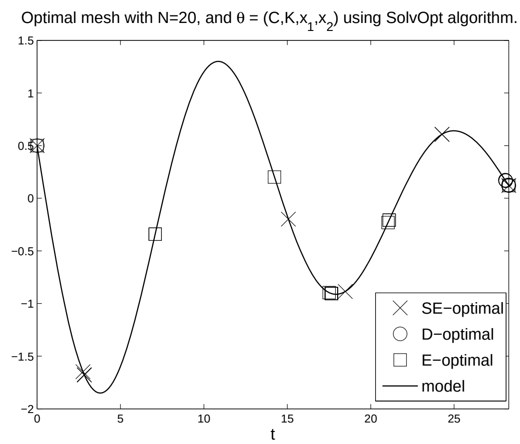
(a)



(b)



(c)



(d)

Figure 12: 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 $(C4)$.

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.

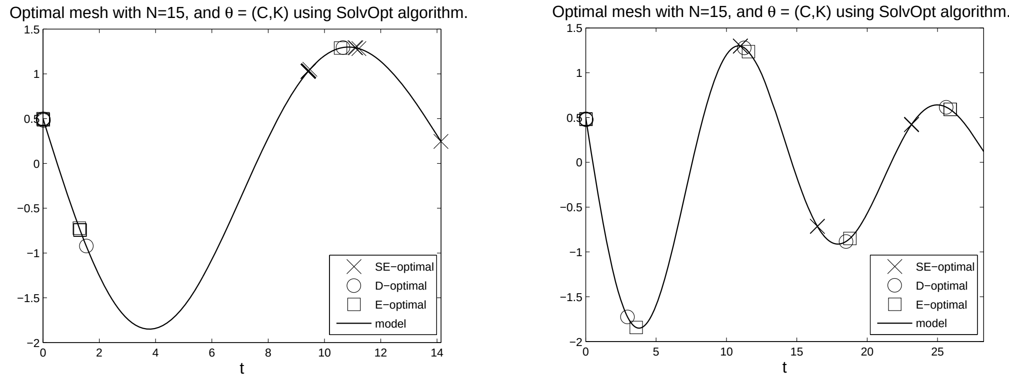
In most cases, optimal meshes with a larger number of points were nested in the optimal meshes with a reduced the number of points. However, in Fig. 9 (a) and (c), when the numbers of points were reduced to $N = 5$, and $N = 10$, respectively, the new optimal design points were often clustered at a different time points. There was not a consistent benefit in a choice of N based on the standard errors.

From the tables of standard errors (Tables 3-6), we see not surprisingly that SE-optimal design produces smaller standard errors than D and E-optimal design, except in some cases with $SE(x_2)$ when E-optimal design is smaller. SE-optimal design optimizes directly on the standard errors, whereas D and E-optimal design minimize a function of the standard errors. 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 use the inverse problem with the OLS formulation to obtain parameter estimates and standard errors from both asymptotic theory (13) and the bootstrapping method (15). 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 $\mathcal{E}_j \sim \mathcal{N}(0, \sigma_0^2)$. 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 $\mathrm{Cov}(C, K)$ according to asymptotic theory (11) and bootstrapping (15). 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 $T = 28.28$ for $N = 15$ under the first constraint implementation ($C1$) in Fig. 13, the second constraint implementation ($C2$) in Fig. 14, the third constraint implementation ($C3$) in Fig. 15, and the last constraint implementation ($C4$) in Fig. 16. The estimates, standard errors, and covariance between parameters are estimated from the asymptotic theory (13) corresponding to these optimal meshes are given in Table 7, 9, 11, and 13, respectively for the four different constraint implementations. The estimates, standard errors, and covariance between parameters are estimated from the bootstrapping method (15) corresponding to these optimal meshes are given in Table 8, 10, 12, and 14, respectively for the four different constraints. In each of the tables are also results on the uniform grid of time points for the same T and N. Since this is unaffected by constraints, the results for the uniform grid are repeated in the tables.



(a) (b)

Figure 13: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, for $N = 15$ with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation ($C1$).

Table 7: Estimates and standard errors from the asymptotic theory (13) 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 ($C1$).

T	Method	$\hat{C}_{asy}$	$\hat{SE}(\hat{C}_{asy})$	$\hat{K}_{asy}$	$\hat{SE}(\hat{K}_{asy})$	$\hat{\text{Cov}}(\hat{C}_{asy}, \hat{K}_{asy})$
14.14	SE-optimal	0.0816	1.614×10^{-2}	0.1948	1.327×10^{-2}	-4.268×10^{-5}
14.14	D-optimal	-0.0705	1.745×10^{-1}	0.3323	4.859×10^{-2}	-8.387×10^{-3}
14.14	E-optimal	-0.1141	1.588×10^{-1}	0.3392	4.784×10^{-2}	-7.561×10^{-3}
14.14	Uniform	0.0963	3.574×10^{-2}	0.2066	1.317×10^{-2}	-2.940×10^{-4}
28.28	SE-optimal	0.0612	1.402×10^{-2}	0.1890	8.164×10^{-3}	1.510×10^{-5}
28.28	D-optimal	0.1129	1.929×10^{2}	0.2007	1.201×10^{2}	-2.316×10^{4}
28.28	E-optimal	0.1250	7.304×10^{1}	0.2014	3.780×10^{1}	-2.761×10^{3}
28.28	Uniform	0.0938	2.620×10^{-2}	0.2024	1.044×10^{-2}	-1.342×10^{-4}

5.4 Discussion of Oscillator Results with the Inverse Problem

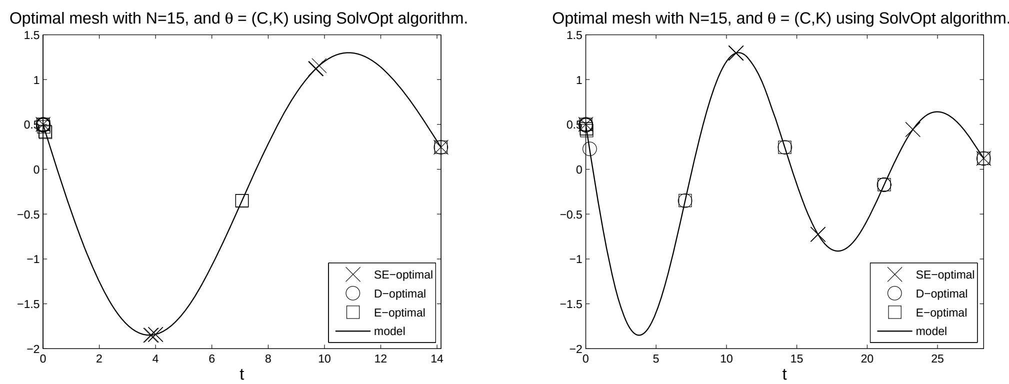
The simulated data was created using the "true" parameter values $\theta_0 = (C, K) = (0.1, 0.2)$. So we can compare the optimal design methods based on how close the parameter estimates are as well as how large the estimates of the standard errors and covariances are.

For asymptotic estimates:

In general, the results from the uniform grid and SE-optimal were similar, with SE-optimal often with smaller estimates of estimated standard error and covariance ($\text{Cov}(\hat{C}, \hat{K})$) compared to the uniform grid. SE-optimal had closer estimates relative to the true values, except for in the following cases: (1) for the constraint implementation ($C1$) (Table 7, $T = 28.28$) D and E-optimal both had closer parameter estimates than SE-optimal, (2) for the constraint implementation ($C2$) (Table 9, $T = 28.28$) E-optimal had a closer estimate of C than either D or SE-optimal and D-optimal

Table 8: Estimates and standard errors from the bootstrap method (15) 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 ($C1$).

T	Method	$\hat{C}_{boot}$	$\hat{SE}(\hat{C}_{boot})$	$\hat{K}_{boot}$	$\hat{SE}(\hat{K}_{boot})$	$\hat{\mathrm{Cov}}(\hat{C}_{boot}, \hat{K}_{boot})$
14.14	SE-optimal	0.0776	1.803×10^{-2}	0.1963	1.280×10^{-2}	1.337×10^{-4}
14.14	D-optimal	-0.1397	1.241×10^{-1}	0.3492	3.736×10^{-2}	-4.342×10^{-3}
14.14	E-optimal	-0.0897	1.145×10^{-1}	0.3175	5.363×10^{-2}	-5.726×10^{-3}
14.14	Uniform	0.1053	2.758×10^{-2}	0.2042	1.374×10^{-2}	-1.461×10^{-4}
28.28	SE-optimal	0.0755	1.736×10^{-2}	0.1935	1.193×10^{-2}	5.841×10^{-5}
28.28	D-optimal	0.0984	2.634×10^{-2}	0.2013	2.569×10^{-2}	-1.541×10^{-4}
28.28	E-optimal	0.1109	2.642×10^{-2}	0.1970	2.811×10^{-2}	-1.955×10^{-4}
28.28	Uniform	0.0942	2.200×10^{-2}	0.2024	9.930×10^{-3}	-2.210×10^{-6}



(a) (b)

Figure 14: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, for $N = 15$ with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation ($C2$).

had a closer estimate of K than the other two optimal design methods, (3) for the third constraint implementation ($C3$) (Table 11, $T = 28.28$) D-optimal had a closer estimate of C than either E or SE-optimal and E-optimal had a closer estimate of K, (4) for the last constraint implementation ($C4$) (Table 13, $T = 14.14$) D-optimal had a closer estimate of C than either E or SE-optimal and E-optimal had a closer estimate of K. SE-optimal design resulted in smaller estimated standard error and covariance as compared to both D-optimal and E-optimal design.

For bootstrap estimates:

SE-optimal design resulted in standard error and covariance estimates similar to (and often smaller than) the uniform mesh. Comparing SE-optimal to D-optimal and C-optimal, we find that SE-optimal had the smallest estimated standard errors, except for the 2nd constraint implementation ($C2$) (Table 10, $T = 28.28$) for K where D-optimal had smaller standard errors compared to SE-optimal and E-optimal. Comparing the parameter estimates with the true parameter values, SE-

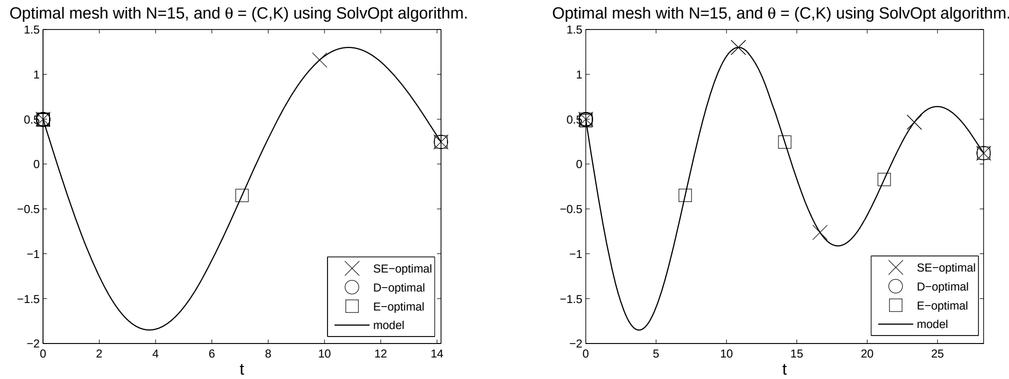
Table 9: Estimates and standard errors from the asymptotic theory (13) resulting from different optimal design methods for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation ($C2$).

T	Method	$\hat{C}_{asy}$	$\hat{SE}(\hat{C}_{asy})$	$\hat{K}_{asy}$	$\hat{SE}(\hat{K}_{asy})$	$\hat{\text{Cov}}(\hat{C}_{asy}, \hat{K}_{asy})$
14.14	SE-optimal	0.0813	1.9611×10^{-2}	0.1992	1.117×10^{-2}	-8.609×10^{-5}
14.14	D-optimal	0.3002	1.174×10^{4}	0.1789	2.001×10^{3}	-2.350×10^{7}
14.14	E-optimal	0.2304	3.246×10^{-1}	0.1804	3.366×10^{-2}	-9.933×10^{-3}
14.14	Uniform	0.0963	3.574×10^{-2}	0.2066	1.317×10^{-2}	-2.940×10^{-4}
28.28	SE-optimal	0.1109	1.808×10^{-2}	0.1919	1.242×10^{-2}	-1.834×10^{-5}
28.28	D-optimal	0.0037	1.036×10^{0}	0.1984	9.012×10^{-2}	-9.332×10^{-2}
28.28	E-optimal	0.0972	3.494×10^{-1}	0.1931	3.906×10^{-2}	-1.344×10^{-2}
28.28	Uniform	0.0938	2.620×10^{-2}	0.2024	1.044×10^{-2}	-1.342×10^{-4}

Table 10: Estimates and standard errors from the bootstrap method (15) 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	$\hat{C}_{asy}$	$\hat{SE}(\hat{C}_{boot})$	$\hat{K}_{boot}$	$\hat{SE}(\hat{K}_{boot})$	$\hat{\text{Cov}}(\hat{C}_{boot}, \hat{K}_{boot})$
14.14	SE-optimal	0.0777	1.824×10^{-2}	0.2004	1.028×10^{-2}	4.404×10^{-5}
14.14	D-optimal	0.2169	2.358×10^{-1}	0.1508	6.569×10^{-2}	-5.697×10^{-3}
14.14	E-optimal	0.2545	1.798×10^{-1}	0.1693	3.039×10^{-2}	-4.141×10^{-3}
14.14	Uniform	0.1053	2.758×10^{-2}	0.2042	1.374×10^{-2}	-1.461×10^{-4}
28.28	SE-optimal	0.1102	2.112×10^{-2}	0.1958	1.826×10^{-2}	1.938×10^{-5}
28.28	D-optimal	0.0050	4.591×10^{-2}	0.1972	7.157×10^{-3}	-2.032×10^{-4}
28.28	E-optimal	0.1122	1.347×10^{-1}	0.1858	3.919×10^{-2}	-1.622×10^{-3}
28.28	Uniform	0.0942	2.200×10^{-2}	0.2024	9.930×10^{-3}	-2.210×10^{-6}

optimal design resulted in closer parameter estimates, except for in the following cases: (1) for the constraint implementation ($C1$) (Table 8, $T = 28.28$) D and E-optimal both had closer parameter estimates than SE-optimal, (2) for the 2nd constraint implementation ($C2$) (Table 10, $T = 28.28$) D-optimal had a closer estimate of K than the other two optimal design methods, (3) for the 3rd constraint implementation ($C3$) (Table 12, $T = 28.28$) D-optimal had closer parameter estimates than SE-optimal and E-optimal. SE-optimal design resulted in smaller estimated covariance as compared to both D-optimal and E-optimal design.



(a) (b)

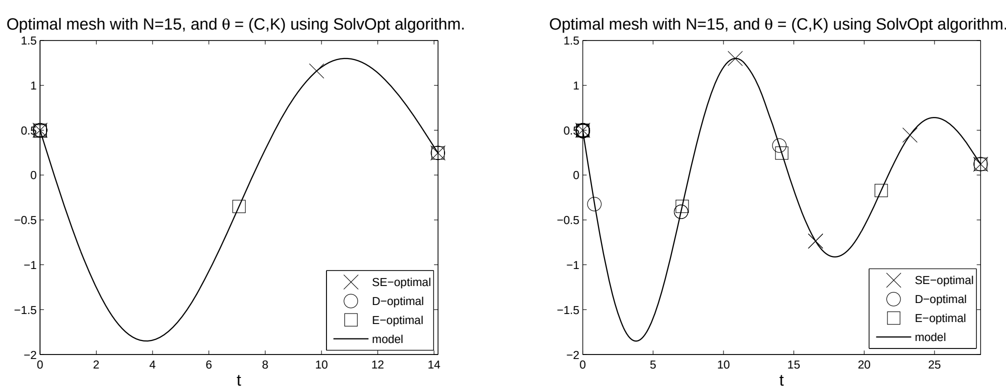
Figure 15: Plot of model with optimal time points resulting from different optimal design methods for $\theta = (C, K)$, for $N = 15$ with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C3)$.

Table 11: Estimates and standard errors from the asymptotic theory (13) resulting from different optimal design methods for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C3)$.

T	Method	$\hat{C}_{asy}$	$\hat{SE}(\hat{C}_{asy})$	$\hat{K}_{asy}$	$\hat{SE}(\hat{K}_{asy})$	$\hat{\text{Cov}}(\hat{C}_{asy}, \hat{K}_{asy})$
14.14	SE-optimal	0.0913	1.757×10^{-2}	0.2268	1.572×10^{-2}	-1.842×10^{-4}
14.14	D-optimal	0.4323	1.823×10^{4}	0.1603	1.033×10^{4}	-1.883×10^{8}
14.14	E-optimal	-0.2744	1.348×10^{-1}	0.2295	1.847×10^{-2}	-2.487×10^{-3}
14.14	Uniform	0.0963	3.574×10^{-2}	0.2066	1.317×10^{-2}	-2.940×10^{-4}
28.28	SE-optimal	0.0732	1.646×10^{-2}	0.1881	9.153×10^{-3}	8.010×10^{-6}
28.28	D-optimal	0.0918	4.210×10^{4}	0.1879	8.192×10^{3}	-3.449×10^{8}
28.28	E-optimal	0.0428	1.951×10^{2}	0.1980	1.516×10^{1}	-2.957×10^{3}
28.28	Uniform	0.0938	2.620×10^{-2}	0.2024	1.044×10^{-2}	-1.342×10^{-4}

Table 12: Estimates and standard errors from the bootstrap method (15) 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 $(C3)$.

T	Method	$\hat{C}_{boot}$	$\hat{SE}(\hat{C}_{boot})$	$\hat{K}_{boot}$	$\hat{SE}(\hat{K}_{boot})$	$\hat{\mathrm{Cov}}(\hat{C}_{boot}, \hat{K}_{boot})$
14.14	SE-optimal	0.0878	1.275×10^{-2}	0.2275	1.520×10^{-2}	3.120×10^{-5}
14.14	D-optimal	0.3863	5.523×10^{-1}	-0.0164	4.185×10^{-1}	-1.926×10^{-1}
14.14	E-optimal	-0.2173	1.649×10^{-1}	0.2279	1.708×10^{-2}	-2.598×10^{-3}
14.14	Uniform	0.1053	2.758×10^{-2}	0.2042	1.374×10^{-2}	-1.461×10^{-4}
28.28	SE-optimal	0.0643	1.718×10^{-2}	0.1879	7.581×10^{-3}	6.410×10^{-5}
28.28	D-optimal	0.0663	1.211×10^{-1}	0.1928	3.366×10^{-2}	1.599×10^{-4}
28.28	E-optimal	0.1659	2.274×10^{0}	0.2282	7.772×10^{-1}	1.480×10^{0}
28.28	Uniform	0.0942	2.200×10^{-2}	0.2024	9.930×10^{-3}	-2.210×10^{-6}



(a) (b)

Figure 16: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, for $N = 15$ with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C4)$.

Table 13: Estimates and standard errors from the asymptotic theory (13) resulting from different optimal design methods for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C4)$.

T	Method	$\hat{C}_{asy}$	$\hat{SE}(\hat{C}_{asy})$	$\hat{K}_{asy}$	$\hat{SE}(\hat{K}_{asy})$	$\hat{\text{Cov}}(\hat{C}_{asy}, \hat{K}_{asy})$
14.14	SE-optimal	0.0952	1.751×10^{-2}	0.1891	1.336×10^{-2}	-1.626×10^{-5}
14.14	D-optimal	0.0976	3.560×10^{5}	0.2282	6.690×10^{4}	2.382×10^{10}
14.14	E-optimal	0.0242	4.244×10^{-1}	0.2095	8.172×10^{-3}	-2.031×10^{-3}
14.14	Uniform	0.0963	3.574×10^{-2}	0.2066	1.317×10^{-2}	-2.940×10^{-4}
28.28	SE-optimal	0.0905	1.873×10^{-2}	0.1916	1.242×10^{-2}	-7.062×10^{-6}
28.28	D-optimal	0.2631	1.822×10^{-1}	0.4349	1.546×10^{-1}	-2.083×10^{-2}
28.28	E-optimal	0.2980	1.255×10^{-1}	0.1518	2.542×10^{-2}	-2.101×10^{-3}
28.28	Uniform	0.0938	2.620×10^{-2}	0.2024	1.044×10^{-2}	-1.342×10^{-4}

Table 14: Estimates and standard errors from the bootstrap method (15) 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 $(C4)$.

T	Method	$\hat{C}_{boot}$	$\hat{SE}(\hat{C}_{boot})$	$\hat{K}_{boot}$	$\hat{SE}(\hat{K}_{boot})$	$\hat{\text{Cov}}(\hat{C}_{boot}, \hat{K}_{boot})$
14.14	SE-optimal	0.0968	2.009×10^{-2}	0.1910	1.363×10^{-2}	1.584×10^{-4}
14.14	D-optimal	0.0849	1.688×10^{-1}	0.2407	4.146×10^{-2}	3.286×10^{-3}
14.14	E-optimal	0.1950	3.197×10^{0}	0.3154	7.302×10^{-1}	2.269×10^{0}
14.14	Uniform	0.1053	2.758×10^{-2}	0.2042	1.374×10^{-2}	-1.461×10^{-4}
28.28	SE-optimal	0.0798	1.863×10^{-2}	0.1887	1.0006×10^{-2}	8.754×10^{-5}
28.28	D-optimal	0.3434	2.475×10^{0}	0.7302	5.633×10^{0}	1.382×10^{1}
28.28	E-optimal	0.3080	1.239×10^{-1}	0.1424	3.064×10^{-2}	-2.448×10^{-3}
28.28	Uniform	0.0942	2.200×10^{-2}	0.2024	9.930×10^{-3}	-2.210×10^{-6}

6 A Simple Glucose Regulation Model

Next we will consider a well-known model for the intervenous glucose tolerance test (IVGTT). This model is referred to as the *minimal model* in the literature [10, 12, 19]. 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 of glucose into the blood, p_0, 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 is 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 cell of the pancreas to secrete insulin. Insulin helps the fat and muscle cells to uptake glucose from the blood, either for fuel or for fuel 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 [3]. 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 [10, 12, 19] for details):

$$\dot{G}(t) = -p_1\big(G(t) - G_b\big) - X(t)G(t), \quad G(0) = p_0, \tag{16}$$

$$\dot{X}(t) = -p_2 X(t) + p_3\big(I(t) - I_b\big), \quad X(0) = 0, \tag{17}$$

$$\dot{I}(t) = p_4 t \max\big(0, G(t) - p_5\big) - p_6\big(I(t) - I_b\big), \quad I(0) = p_7 + I_b, \tag{18}$$

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

We use the following approximate max function in equation (18) since it is continuously differentiable:

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

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

An interpretation of the parameters is given in Table 15.

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

<u>Equation (16) (Glucose concentration in plasma)</u>

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.

<u>Equation (17) (Insulin-dependent glucose uptake activity)</u>

Table 15: 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

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 (18) (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 [10, 19] 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 which is given by

$$\begin{aligned} S_I &= \frac{p_3}{p_2}, \\ S_G &= p_1, \\ \phi_1 &= \frac{I_{\max} - I_b}{p_6\big(p_0 - G_b\big)}, \end{aligned} \tag{19}$$

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

Bergman et al. [9] 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 [16]. 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 $\vec{\mathcal{E}}(t)$:

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

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

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

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 $V_0 = var(\vec{\mathcal{E}}_i) = \mathrm{diag}(\sigma_{0,G}^2, \sigma_{0,I}^2)$.

A subset of the parameters is estimated by inverse problem methodology [4]. 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

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

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

$$\hat{\theta}_{\mathrm{OLS}} = \hat{\theta}_{\mathrm{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [\vec{y}_j - \vec{f}(t_j, \theta)] V_0^{-1} [\vec{y}_j - \vec{f}(t_j, \theta)].$$

The definition of variance gives

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

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

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

which is solved simultaneously with the estimate $\hat{\theta} = \hat{\theta}_{\mathrm{OLS}}$, 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

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

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

$$F = F(\tau) = F(\tau, \theta_0) = (\Sigma_0^N)^{-1}, \tag{20}$$

and is explicitly dependent on the optimal sampling times τ.

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

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

The corresponding FIM can be estimated by

$$\hat{F}(\tau) = \hat{F}(\tau, \hat{\theta}_{\text{OLS}}) = (\hat{\Sigma}^N(\hat{\theta}_{\text{OLS}}))^{-1}.$$

The asymptotic standard errors are given by

$$SE_k(\theta_0) = \sqrt{(\Sigma_0^N)_{kk}}, \; k = 1, \ldots, p. \tag{21}$$

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

$$SE_k(\hat{\theta}_{\text{OLS}}) = \sqrt{(\hat{\Sigma}^N(\hat{\theta}_{\text{OLS}}))_{kk}}, \; k = 1, \ldots, p. \tag{22}$$

It can be shown, under certain conditions, for $N \to \infty$, that the estimator Θ_{OLS}^N is asymptotically normal (vector observation problems can be treated by arguments similar to those in [18]); i.e., for N large

$$\Theta_{\text{OLS}}^N \sim \mathcal{N}_p(\theta_0, \Sigma_0^N).$$

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 sampled with replacement separately. The bootstrap algorithm modified for a system with vector observations is outlined for completeness.

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

2. Using this estimate define the standardized residuals:

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

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

for $j = 1, \ldots, N$. Then $\{\bar{r}_{G,1}, \ldots, \bar{r}_{G,N}\}$,$\{\bar{r}_{I,1}, \ldots, \bar{r}_{I,N}\}$ are realizations of *iid* random variables from the empirical distribution $\vec{F}_N$, 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) $\{\bar{r}_{G,1}, \ldots, \bar{r}_{G,N}\}$ and $\{\bar{r}_{I,1}, \ldots, \bar{r}_{I,N}\}$ to form the bootstrap samples $\{r^m_{G,1}, \ldots, r^m_{G,N}\}$ and $\{r^m_{I,1}, \ldots, r^m_{I,N}\}$.

4. Create bootstrap sample points

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

$$y^m_{2,j} = f_2(t_j, \hat{\theta}^0) + r^m_{I,j},$$

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

5. Obtain a new estimate $\hat{\theta}^{m+1}$ from the bootstrap samples $\{\vec{y}^m_j\}$ using OLS. Add $\hat{\theta}^{m+1}$ into the vector Θ, where Θ is a vector of length Mp 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 Mp.

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

$$\begin{aligned} \hat{\theta}_{boot} &= \tfrac{1}{M}\textstyle\sum_{m=1}^{M} \hat{\theta}^m, \\ Cov(\hat{\theta}_{boot}) &= \tfrac{1}{M-1}\textstyle\sum_{m=1}^{M} (\hat{\theta}^m - \hat{\theta}_{boot})^T(\hat{\theta}^m - \hat{\theta}_{boot}), \\ SE_k(\hat{\theta}_{boot}) &= \sqrt{Cov(\hat{\theta}_{boot})_{kk}}. \end{aligned} \tag{23}$$

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 (22) 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 one of the constrained minimization algorithms SolvOpt or MATLAB's *fmincon*, depending on which scheme returned optimal time points within the constraints. 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 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 ($C1$) (using optimization algorithm SolvOpt) in Fig. 17, the second constraint implementation ($C2$) (with SolvOpt) in Fig. 18, the third constraint implementation ($C3$) (with MATLAB's *fmincon*) in Fig. 19, and the last constraint implementation ($C4$) (with MATLAB's *fmincon*) in Fig. 20. The standard errors (22) from the asymptotic theory corresponding to these optimal meshes are given in Table 16-19, respectively for the four different constraint implementations.

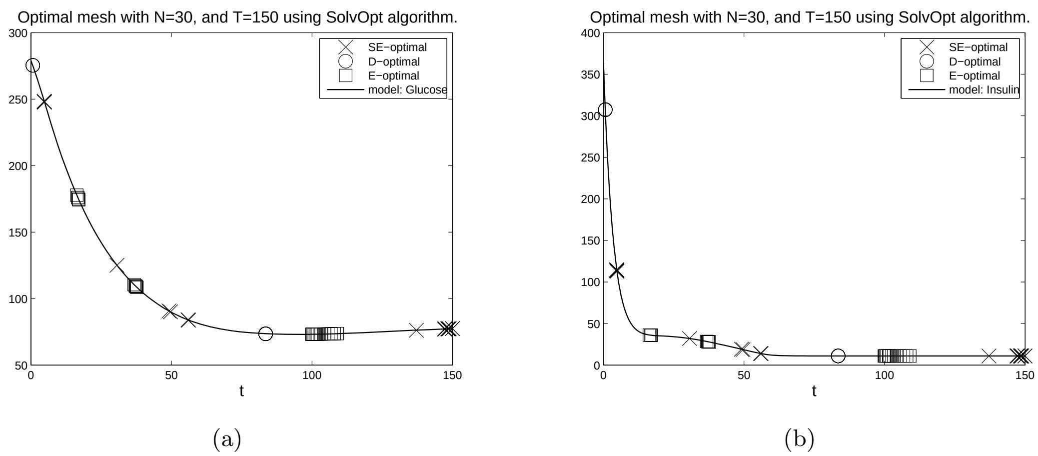
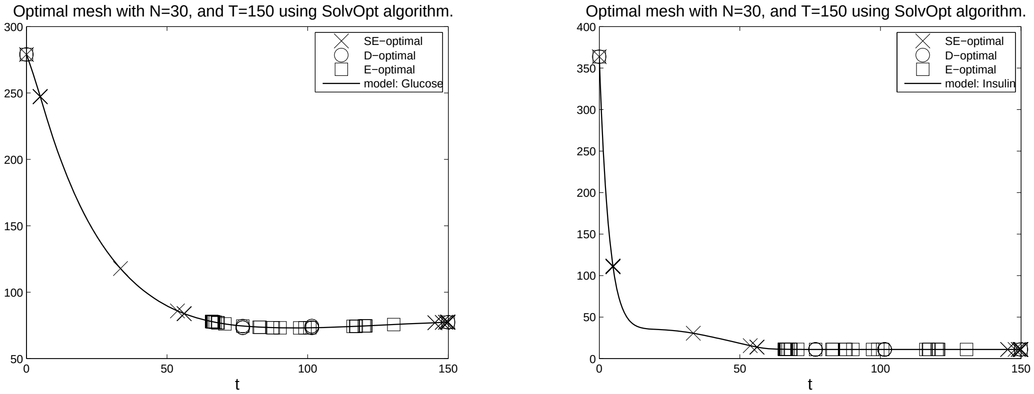


Figure 17: Plot of model with optimal time points resulting from different optimal design methods for $\theta = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation ($C1$).

Table 16: Approximate asymptotic standard errors from the asymptotic theory (21) 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	4.361×10^{-3}	6.609×10^{-3}	3.190×10^{-6}	3.150×10^{-4}
D-optimal	7.685×10^{0}	1.272×10^{1}	$0.007.533 \times 10^{-3}$	2.242×10^{-2}
E-optimal	1.148×10^{-1}	1.763×10^{-1}	1.095×10^{-4}	2.560×10^{-4}

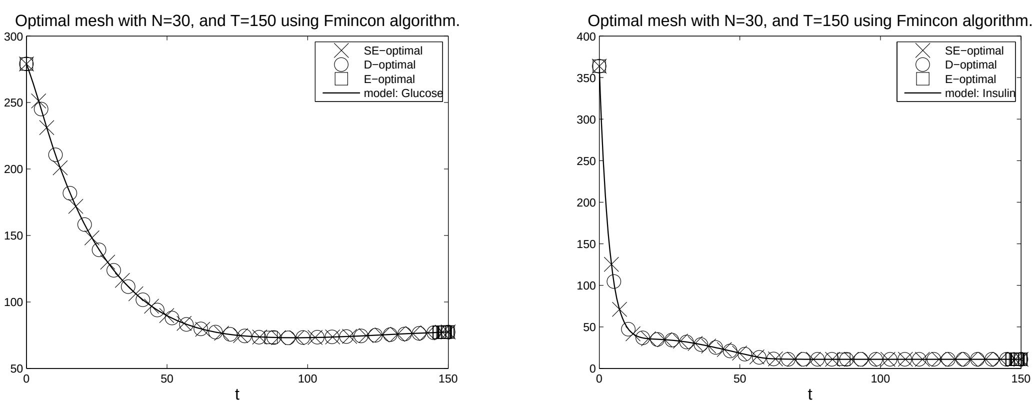


(a) (b)

Figure 18: 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 ($C2$).

Table 17: Approximate asymptotic standard errors from the asymptotic theory (21) 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	4.587×10^{-3}	6.784×10^{-3}	3.341×10^{-6}	3.406×10^{-4}
D-optimal	5.902×10^{-2}	1.191×10^{-1}	7.033×10^{-5}	1.324×10^{-3}
E-optimal	1.181×10^{-1}	1.306×10^{-1}	8.203×10^{-5}	1.423×10^{-2}



(a) (b)

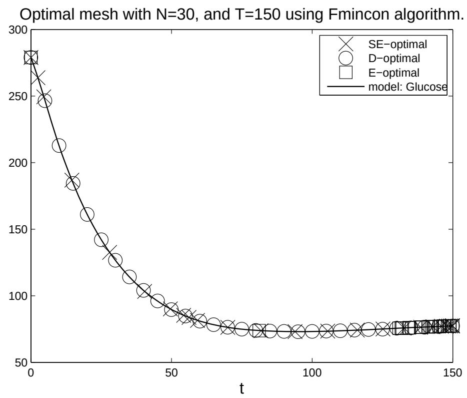
Figure 19: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using fmincon, with constraint implementation ($C3$).

6.4 Discussion for the Glucose Regulation Model

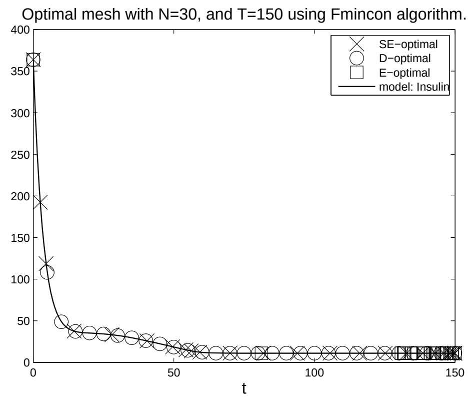
Comparing the optimal design methods using approximate asymptotic standard errors, we find that in most cases SE-optimal has the smallest standard errors, with a few exceptions. In Table 16

Table 18: Approximate asymptotic standard errors from the asymptotic theory (21) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using fmincon, with constraint implementation $(C3)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	5.654×10^{-3}	8.776×10^{-3}	4.222×10^{-6}	1.866×10^{-4}
D-optimal	8.218×10^{-3}	1.191×10^{-2}	5.939×10^{-6}	1.919×10^{-4}
E-optimal	2.167×10^{1}	1.849×10^{1}	2.174×10^{-2}	8.082×10^{0}



(a)



(b)

Figure 20: 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 fmincon, with constraint implementation $(C4)$.

Table 19: Approximate asymptotic standard errors from the asymptotic theory (21) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using fmincon, with constraint implementation $(C4)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.952×10^{-3}	7.558×10^{-3}	3.639×10^{-6}	2.533×10^{-4}
D-optimal	8.172×10^{-3}	1.194×10^{-2}	5.926×10^{-6}	1.887×10^{-4}
E-optimal	4.916×10^{0}	4.270×10^{0}	4.723×10^{-3}	1.692×10^{0}

(constraint implementation $(C1)$), E-optimal has a smaller standard error for p_4 than SE-optimal. In Table 19 (constraint implementation $(C4)$), D-optimal has a smaller standard error for p_4 than SE-optimal. Using this method of comparison, SE-optimal is the favorable method for this glucose regulation model.

Since the asymptotic standard errors are part of the cost function for SE-optimal, it is not surprising that comparing the methods on their asymptotic standard errors we find that SE-optimal has the smallest standard errors most of time. 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 (22) and the bootstrapping method (23). 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 15 and *iid* noise with $\vec{\mathcal{E}}_j \sim \mathcal{N}(0, \vec{\sigma}_0^2)$. We assume the true variances: $\sigma_{0,G}^2 = 25$ and $\sigma_{0,I}^2 = 4$. In this section we 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 (22) and bootstrapping (23). 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. 17-20 for the four different constraints. The parameter estimates, and standard errors are estimated from the asymptotic theory (22) corresponding to these optimal meshes are given in Tables 20, 22, 24, and 26, respectively for the four different constraints. The parameter estimates, standard errors, and covariance between parameters are estimated from the bootstrapping method (23) corresponding to these optimal meshes are given in Tables 21, 23, 25, and 27, respectively for the four different constraints. In each of the tables are also results on the uniform grid of time points.

Table 20: Estimates, standard errors, and covariances between parameters from the asymptotic theory (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using SolvOpt, with constraint implementation ($C1$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	1.542×10^{-2}	-3.673×10^{-4}	1.408×10^{-2}	1.947×10^{-2}
$\hat{SE}(\hat{p}_1)$	4.982×10^{-3}	2.918×10^{-1}	1.488×10^{-1}	9.445×10^{-3}
$\hat{p}_2$	4.658×10^{-2}	3.570×10^{-2}	4.669×10^{-2}	3.346×10^{-2}
$\hat{SE}(\hat{p}_2)$	7.480×10^{-3}	3.376×10^{0}	2.286×10^{-1}	1.373×10^{-2}
$\hat{p}_3$	2.238×10^{-5}	6.561×10^{-5}	2.698×10^{-5}	1.795×10^{-5}
$\hat{SE}(\hat{p}_3)$	3.873×10^{-6}	1.906×10^{-3}	1.418×10^{-4}	6.964×10^{-6}
$\hat{p}_4$	4.564×10^{-3}	-7.837×10^{-3}	4.007×10^{-3}	4.328×10^{-3}
$\hat{SE}(\hat{p}_4)$	7.806×10^{-4}	7.115×10^{-2}	5.477×10^{-4}	4.308×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-3.326×10^{-5}	-9.806×10^{-1}	-3.401×10^{-2}	-1.263×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-1.846×10^{-8}	-5.545×10^{-4}	-2.110×10^{-5}	-6.526×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	5.159×10^{-7}	-6.735×10^{-5}	9.711×10^{-6}	6.889×10^{-7}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	2.711×10^{-8}	6.422×10^{-3}	3.241×10^{-5}	9.446×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	3.110×10^{-7}	1.574×10^{-2}	-1.399×10^{-5}	-4.271×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	-4.657×10^{-10}	4.866×10^{-7}	-8.936×10^{-9}	-4.483×10^{-10}

Table 21: Estimates, standard errors, and covariances between parameters from the bootstrap method (23) 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 SolvOpt, with constraint implementation ($C1$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	1.454×10^{-2}	-1.637×10^{-4}	1.588×10^{-2}	1.934×10^{-2}
$\hat{SE}(\hat{p}_1)$	5.425×10^{-3}	1.203×10^{-2}	7.675×10^{-3}	8.408×10^{-3}
$\hat{p}_2$	4.965×10^{-2}	5.599×10^{-2}	4.425×10^{-2}	3.334×10^{-2}
$\hat{SE}(\hat{p}_2)$	1.117×10^{-2}	8.294×10^{-2}	1.163×10^{-2}	1.236×10^{-2}
$\hat{p}_3$	2.416×10^{-5}	9.243×10^{-5}	2.554×10^{-5}	1.899×10^{-5}
$\hat{SE}(\hat{p}_3)$	6.939×10^{-6}	8.058×10^{-5}	9.513×10^{-6}	7.600×10^{-6}
$\hat{p}_4$	4.513×10^{-3}	-1.204×10^{-2}	4.001×10^{-3}	4.300×10^{-3}
$\hat{SE}(\hat{p}_4)$	3.859×10^{-4}	1.557×10^{-2}	2.526×10^{-4}	1.854×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-5.569×10^{-5}	-1.619×10^{-4}	-8.179×10^{-5}	-1.014×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-3.629×10^{-8}	-1.276×10^{-7}	-7.150×10^{-8}	-6.273×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	-1.935×10^{-8}	-2.056×10^{-5}	3.121×10^{-7}	5.499×10^{-9}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	7.506×10^{-8}	5.734×10^{-6}	1.056×10^{-7}	9.261×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	6.836×10^{-7}	1.768×10^{-4}	-2.726×10^{-8}	2.413×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	1.356×10^{-10}	-4.111×10^{-7}	-1.523×10^{-10}	1.217×10^{-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 and different constraints, 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.

Asymptotic theory: parameter estimates.

For the constraint implementation ($C1$) using asymptotic theory (Table 20), the estimates for p_1, p_2, and p_3 on the uniform mesh are closest to the true values compared to the optimal design methods. For p_4, E-optimal's estimate is closest to the true value. Comparing among the optimal design methods, SE-optimal has the closest estimates for p_1 and p_3, followed by E-optimal. D-optimal has the closest estimate for p_2, followed by SE-optimal. For p_4, E-optimal has the closest estimate, followed by SE-optimal.

For the second constraint implementation ($C2$) (Table 22), E-optimal has the closest estimates compared to the uniform mesh and the other optimal design meshes. The uniform mesh has closer estimates than SE-optimal and D-optimal, and SE-optimal was closer to the true estimates than D-optimal.

Table 22: Estimates, standard errors, and covariances between parameters from the asymptotic theory (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using SolvOpt, with constraint implementation $(C2)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	3.082×10^{-2}	4.879×10^{-2}	2.727×10^{-2}	1.947×10^{-2}
$\hat{SE}(\hat{p}_1)$	5.230×10^{-3}	3.571×10^{0}	2.217×10^{-1}	9.445×10^{-3}
$\hat{p}_2$	1.414×10^{-2}	4.530×10^{-3}	2.730×10^{-2}	3.346×10^{-2}
$\hat{SE}(\hat{p}_2)$	7.655×10^{-3}	3.231×10^{0}	2.291×10^{-1}	1.373×10^{-2}
$\hat{p}_3$	8.369×10^{-6}	5.864×10^{-6}	1.290×10^{-5}	1.795×10^{-5}
$\hat{SE}(\hat{p}_3)$	4.100×10^{-6}	3.173×10^{-3}	1.689×10^{-4}	6.964×10^{-6}
$\hat{p}_4$	3.322×10^{-3}	2.498×10^{-3}	4.009×10^{-3}	4.328×10^{-3}
$\hat{SE}(\hat{p}_4)$	8.531×10^{-4}	1.034×10^{0}	4.117×10^{-2}	4.308×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-3.606×10^{-5}	-1.154×10^{1}	-5.056×10^{-2}	-1.263×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-2.049×10^{-8}	-1.133×10^{-2}	-3.722×10^{-5}	-6.526×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	7.047×10^{-7}	3.688×10^{0}	8.172×10^{-3}	6.889×10^{-7}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	2.932×10^{-8}	1.025×10^{-2}	3.796×10^{-5}	9.446×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	2.053×10^{-7}	-3.333×10^{0}	-8.032×10^{-3}	-4.271×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	-6.656×10^{-10}	-3.279×10^{-3}	-6.516×10^{-6}	-4.483×10^{-10}

For the third constraint implementation $(C3)$ (Table 24), the estimates for p_1 and p_2 are closest using the mesh from E-optimal, followed by SE-optimal and then the uniform mesh. The estimate for p_3 is closest from the uniform mesh, followed by D-optimal and then SE-optimal. The estimate for p_4 is closest to the true value using SE-optimal's mesh, followed by the uniform mesh, and then D-optimal.

For the fourth constraint implementation $(C4)$ (Table 26), the estimates for p_1, p_2, and p_3 are the closest using the mesh from E-optimal, followed by SE-optimal, and then the uniform mesh. For p_4 the uniform mesh gives the closest estimate, followed by D-optimal and then SE-optimal.

Asymptotic theory: standard errors.

Here we compare the optimal design methods based on which has the smallest standard error estimates.

For the first and second constraint implementations $(C1)$ and $(C2)$ (Tables 20 and 22) comparing the standard error estimates, we find that SE-optimal has the smallest standard error estimates for p_1, p_2, p_3, followed by the uniform mesh and then E-optimal. The standard error estimate for p_4 is smallest for the uniform mesh, followed by E-optimal and then SE-optimal.

For the third constraint implementation $(C3)$ (Table 24), the standard error estimates for p_1, p_2, p_3 are the smallest using the mesh from SE-optimal, followed by D-optimal and then the uniform mesh. The standard error estimate for p_4 is smallest for D-optimal, followed by the uniform mesh and then SE-optimal.

For the fourth constraint implementation $(C4)$ (Table 26), the standard error estimates for p_1 and p_2 are smallest from SE-optimal's mesh, followed by the uniform mesh and then D-optimal.

Table 23: Estimates, standard errors, and covariances between parameters from the bootstrap method (23) 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 SolvOpt, with constraint implementation ($C2$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	3.035×10^{-2}	6.206×10^{-2}	3.491×10^{-2}	1.934×10^{-2}
$\hat{SE}(\hat{p}_1)$	4.164×10^{-3}	5.696×10^{-2}	3.022×10^{-2}	8.408×10^{-3}
$\hat{p}_2$	1.507×10^{-2}	1.133×10^{-2}	2.629×10^{-2}	3.334×10^{-2}
$\hat{SE}(\hat{p}_2)$	5.654×10^{-3}	1.834×10^{-2}	1.979×10^{-2}	1.236×10^{-2}
$\hat{p}_3$	8.768×10^{-6}	2.064×10^{-5}	1.582×10^{-5}	1.899×10^{-5}
$\hat{SE}(\hat{p}_3)$	3.141×10^{-6}	2.259×10^{-4}	1.718×10^{-5}	7.600×10^{-6}
$\hat{p}_4$	3.408×10^{-3}	1.725×10^{-3}	3.667×10^{-3}	4.300×10^{-3}
$\hat{SE}(\hat{p}_4)$	3.756×10^{-4}	6.804×10^{-2}	1.006×10^{-2}	1.854×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-2.130×10^{-5}	-4.667×10^{-4}	-3.486×10^{-4}	-1.014×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-1.202×10^{-8}	1.094×10^{-7}	-1.263×10^{-7}	-6.273×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	6.269×10^{-8}	-1.618×10^{-4}	-8.121×10^{-5}	5.499×10^{-9}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	1.696×10^{-8}	1.421×10^{-6}	9.081×10^{-8}	9.261×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	2.499×10^{-7}	-3.325×10^{-4}	1.084×10^{-4}	2.413×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	9.168×10^{-11}	-1.517×10^{-5}	-8.986×10^{-8}	1.217×10^{-10}

The standard error estimates for p_3 are smallest using the mesh from SE-optimal, followed by the uniform mesh and then E-optimal. The standard error estimate for p_4 is smallest for the uniform mesh, followed by D-optimal and then SE-optimal.

For all of the constraints, the standard error estimates are the smallest for SE-optimal for p_1, p_2 and p_3, but not for p_4.

Asymptotic theory: covariance estimates.

We also compare the optimal design methods based on which has the smallest covariance estimates in absolute value.

For the first and second constraint implementations ($C1$) and ($C2$) (Table 20 and 22), the absolute values of the covariance estimates are very close for the uniform mesh and SE-optimal, both are smaller than the covariance estimates from E-optimal (which are smaller than those from D-optimal).

For the third constraint implementation ($C3$) (Table 24), the absolute value of the covariance estimates are smallest for SE-optimal. The absolute value of the covariance estimates are similar between the uniform mesh and D-optimal, which are both smaller than those of E-optimal.

For the fourth constraint implementation ($C4$) (Table 26), the smallest (in absolute value) covariance estimates are from SE-optimal, followed by the uniform mesh and then D-optimal.

Comparing the absolute values of the covariance estimates among the optimal design methods, we find SE-optimal's are consistently smaller.

Table 24: Estimates, standard errors, and covariances between parameters from the asymptotic theory (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation $(C3)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	3.121×10^{-2}	8.646×10^{-3}	2.031×10^{-2}	1.947×10^{-2}
$\hat{SE}(\hat{p}_1)$	6.511×10^{-3}	9.265×10^{-3}	5.937×10^{1}	9.445×10^{-3}
$\hat{p}_2$	1.818×10^{-2}	4.967×10^{-2}	3.183×10^{-2}	3.346×10^{-2}
$\hat{SE}(\hat{p}_2)$	1.004×10^{-2}	1.336×10^{-2}	5.050×10^{1}	1.373×10^{-2}
$\hat{p}_3$	9.235×10^{-6}	2.858×10^{-5}	8.739×10^{-6}	1.795×10^{-5}
$\hat{SE}(\hat{p}_3)$	4.962×10^{-6}	6.809×10^{-6}	5.984×10^{-2}	6.964×10^{-6}
$\hat{p}_4$	4.271×10^{-3}	4.357×10^{-3}	3.587×10^{-3}	4.328×10^{-3}
$\hat{SE}(\hat{p}_4)$	4.439×10^{-4}	4.273×10^{-4}	2.231×10^{1}	4.308×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-6.169×10^{-5}	-1.205×10^{-4}	-2.998×10^{3}	-1.263×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-3.176×10^{-8}	-6.259×10^{-8}	-3.552×10^{0}	-6.526×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	2.695×10^{-7}	6.983×10^{-7}	1.323×10^{3}	6.889×10^{-7}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	4.849×10^{-8}	8.983×10^{-8}	3.020×10^{0}	9.446×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	1.795×10^{-7}	-4.555×10^{-7}	-1.125×10^{3}	-4.271×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	-1.442×10^{-10}	-4.574×10^{-10}	-1.335×10^{0}	-4.483×10^{-10}

Bootstrapping: parameter estimates.

For the first constraint implementation $(C1)$ using bootstrapping (Table 21), the estimates for p_1 and p_2 are closest to the true values for the uniform mesh, followed by E-optimal and then SE-optimal. For p_3 the uniform mesh results in the closest estimate, followed by SE-optimal and then E-optimal. The estimate for p_4 is closest to the true value using the mesh from E-optimal, followed by the uniform mesh and then the mesh from SE-optimal.

For the second constraint implementation $(C2)$ (Table 23), the uniform mesh results in the closest estimate for p_1, followed by SE-optimal and then E-optimal. For p_2 the mesh from E-optimal results in the closest estimate, followed by the uniform mesh and then SE-optimal. The closest estimate for p_3 resulted from E-optimal's mesh, followed by the uniform mesh and then D-optimal. The closest estimate for p_4 came from the uniform mesh, followed by E-optimal and then SE-optimal.

For the third constraint implementation $(C3)$ (Table 25), the estimates for p_1 and p_2 are closest to the true values using the mesh from SE-optimal, followed by E-optimal and then the uniform mesh. The estimate for p_3 is closest to the true value for E-optimal, followed by SE-optimal and then the uniform mesh. The closest estimate of p_4 from the mesh from SE-optimal, followed by the uniform mesh then E-optimal.

For the last constraint implementation $(C4)$ (Table 27), the estimates for p_1 and p_3 are closest to the true values for SE-optimal, followed by the uniform mesh and then D-optimal. For p_2 the mesh from E-optimal results in the closest estimates, followed by SE-optimal and then the uniform mesh. The estimate for p_4 is closest using the uniform mesh, followed by D-optimal and then SE-optimal.

None of the optimal design methods are consistent with parameter estimates that are the closest to the true values for all cases.

Table 25: Estimates, standard errors, and covariances between parameters from the bootstrap method (23) 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 ($C3$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	3.047×10^{-2}	9.841×10^{-3}	1.975×10^{-2}	1.934×10^{-2}
$\hat{SE}(\hat{p}_1)$	5.624×10^{-3}	7.345×10^{-3}	1.542×10^{-2}	8.408×10^{-3}
$\hat{p}_2$	1.917×10^{-2}	4.874×10^{-2}	3.198×10^{-2}	3.334×10^{-2}
$\hat{SE}(\hat{p}_2)$	8.988×10^{-3}	1.188×10^{-2}	4.622×10^{-2}	1.236×10^{-2}
$\hat{p}_3$	1.014×10^{-5}	2.805×10^{-5}	1.363×10^{-5}	1.899×10^{-5}
$\hat{SE}(\hat{p}_3)$	4.476×10^{-6}	7.776×10^{-6}	3.177×10^{-5}	7.600×10^{-6}
$\hat{p}_4$	4.282×10^{-3}	4.494×10^{-3}	4.421×10^{-3}	4.300×10^{-3}
$\hat{SE}(\hat{p}_4)$	2.095×10^{-4}	2.227×10^{-4}	1.774×10^{-2}	1.854×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-4.789×10^{-5}	-8.462×10^{-5}	-3.547×10^{-4}	-1.014×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-2.426×10^{-8}	-5.603×10^{-8}	-2.733×10^{-7}	-6.273×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	5.858×10^{-8}	-2.638×10^{-7}	6.963×10^{-6}	5.499×10^{-9}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	3.900×10^{-8}	9.125×10^{-8}	1.034×10^{-6}	9.261×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	1.405×10^{-7}	7.430×10^{-7}	2.927×10^{-5}	2.413×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	5.720×10^{-11}	4.049×10^{-10}	-1.591×10^{-8}	1.217×10^{-10}

Bootstrapping: standard errors.

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

Comparing the standard error estimates from the first constraint implementation ($C1$) (Table 21) for p_1 and p_2, we find the optimal mesh from SE-optimal has the smallest standard error estimates, followed by E-optimal and then the uniform mesh. For p_3 the smallest standard error estimate is from the mesh from SE-optimal, followed by the uniform mesh and then E-optimal. The smallest standard error estimate for p_4 is from the uniform mesh, followed by E-optimal and then SE-optimal.

For the second constraint implementation ($C2$) (Table 23), the smallest standard error estimates for p_1 and p_2 are from SE-optimal's mesh, followed by the uniform mesh and then D-optimal. The smallest standard error estimates for p_3 are from SE-optimal, followed by the uniform mesh and then E-optimal. The smallest standard error estimate for p_4 is from the uniform mesh, followed by D-optimal and then SE-optimal.

For the third constraint implementation ($C3$) (Table 25), the standard error estimates for p_1 and p_2 are smallest from SE-optimal's mesh, followed by D-optimal and then the uniform mesh. The standard error estimate for p_3 are smallest using the mesh from SE-optimal, followed by the uniform mesh and then D-optimal. The standard error estimate for p_4 is smallest using the uniform mesh, followed by SE-optimal and then D-optimal.

For the last constraint implementation ($C4$) (Table 27), the standard error estimates for p_1, p_2,

Table 26: Estimates, standard errors, and covariances between parameters from the asymptotic theory (21) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation ($C4$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	2.249×10^{-2}	1.946×10^{-2}	2.830×10^{-2}	1.947×10^{-2}
$\hat{SE}(\hat{p}_1)$	4.656×10^{-3}	9.760×10^{-3}	1.121×10^{1}	9.445×10^{-3}
$\hat{p}_2$	3.081×10^{-2}	3.435×10^{-2}	2.704×10^{-2}	3.346×10^{-2}
$\hat{SE}(\hat{p}_2)$	7.075×10^{-3}	1.419×10^{-2}	9.729×10^{0}	1.373×10^{-2}
$\hat{p}_3$	1.588×10^{-5}	1.912×10^{-5}	1.102×10^{-5}	1.795×10^{-5}
$\hat{SE}(\hat{p}_3)$	3.553×10^{-6}	7.196×10^{-2}	1.078×10^{-2}	6.964×10^{-6}
$\hat{p}_4$	4.560×10^{-3}	4.413×10^{-3}	4.738×10^{-3}	4.328×10^{-3}
$\hat{SE}(\hat{p}_4)$	5.129×10^{-4}	4.451×10^{-4}	3.863×10^{0}	4.308×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$	-3.016×10^{-5}	-1.349×10^{-4}	-1.091×10^{2}	-1.263×10^{-4}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$	-1.602×10^{-8}	-6.968×10^{-8}	-1.208×10^{-1}	-6.526×10^{-8}
$\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$	3.096×10^{-7}	7.356×10^{-7}	4.329×10^{1}	6.889×10^{-7}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$	2.405×10^{-8}	1.009×10^{-7}	1.048×10^{-1}	9.446×10^{-8}
$\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$	1.663×10^{-7}	-4.560×10^{-7}	-3.757×10^{1}	-4.271×10^{-7}
$\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$	-2.209×10^{-10}	4.786×10^{-10}	-4.164×10^{-2}	-4.483×10^{-10}

p_3 are the smallest using the mesh from SE-optimal, followed by the uniform mesh and then D-optimal. The standard error estimate for p_4 is smallest for the uniform mesh, followed by D-optimal and then SE-optimal.

The standard error estimates are the smallest for SE-optimal, for all the constraints, for p_1, p_2 and p_3 but not always for p_4.

Bootstrapping: covariance estimates.

For the first constraint implementation ($C1$) (Table 21), the smallest absolute value covariance estimates are from SE-optimal for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$, $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$, and $\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$, the uniform mesh for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$ and $\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$, and E-optimal for $\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$. The absolute value of the covariance estimates from SE-optimal, E-optimal, and the uniform mesh are all smaller than those of D-optimal.

For the second constraint implementation ($C2$) (Table 23), the smallest absolute value covariance estimates are from SE-optimal or the uniform mesh (for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$ and $\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$).

For the third constraint implementation ($C3$) (Table 25), the smallest absolute value covariance estimates are from SE-optimal, except for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$ which is smallest for the uniform mesh.

For the last constraint implementation ($C4$) (Table 27), the smallest absolute value covariance estimates are from SE-optimal for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_2)$, $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_3)$, and $\hat{\text{Cov}}(\hat{p}_2, \hat{p}_3)$, the uniform mesh for $\hat{\text{Cov}}(\hat{p}_1, \hat{p}_4)$, and D-optimal for $\hat{\text{Cov}}(\hat{p}_2, \hat{p}_4)$ and $\hat{\text{Cov}}(\hat{p}_3, \hat{p}_4)$. The absolute value of the covariance estimates from SE-optimal, E-optimal, and the uniform mesh are all smaller than those of E-optimal.

Comparing the optimal design methods based on the bootstrapping covariance estimates, we find there is not one method that is always favorable.

Table 27: Estimates, standard errors, and covariances between parameters from the bootstrap method (23) 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 ($C4$).

	SE-optimal	D-optimal	E-optimal	Uniform
$\hat{p}_1$	2.223×10^{-2}	1.885×10^{-2}	3.858×10^{-2}	1.934×10^{-2}
$\hat{SE}(\hat{p}_1)$	5.723×10^{-3}	9.131×10^{-3}	3.427×10^{-2}	8.408×10^{-3}
$\hat{p}_2$	3.196×10^{-2}	3.633×10^{-2}	2.729×10^{-2}	3.334×10^{-2}
$\hat{SE}(\hat{p}_2)$	1.013×10^{-2}	1.357×10^{-2}	2.087×10^{-2}	1.236×10^{-2}
$\hat{p}_3$	1.663×10^{-5}	2.057×10^{-5}	2.439×10^{-5}	1.899×10^{-5}
$\hat{SE}(\hat{p}_3)$	5.722×10^{-6}	8.355×10^{-6}	5.318×10^{-5}	7.600×10^{-6}
$\hat{p}_4$	4.795×10^{-3}	4.678×10^{-3}	3.806×10^{-3}	4.300×10^{-3}
$\hat{SE}(\hat{p}_4)$	2.898×10^{-4}	2.195×10^{-4}	1.606×10^{-2}	1.854×10^{-4}
$\hat{\mathrm{Cov}}(\hat{p}_1, \hat{p}_2)$	-5.585×10^{-5}	-1.213×10^{-4}	-2.284×10^{-4}	-1.014×10^{-4}
$\hat{\mathrm{Cov}}(\hat{p}_1, \hat{p}_3)$	-3.204×10^{-8}	-7.499×10^{-8}	9.371×10^{-7}	-6.273×10^{-8}
$\hat{\mathrm{Cov}}(\hat{p}_1, \hat{p}_4)$	-1.487×10^{-7}	9.594×10^{-8}	-2.270×10^{-4}	5.499×10^{-9}
$\hat{\mathrm{Cov}}(\hat{p}_2, \hat{p}_3)$	5.710×10^{-8}	1.120×10^{-7}	2.842×10^{-7}	9.261×10^{-8}
$\hat{\mathrm{Cov}}(\hat{p}_2, \hat{p}_4)$	5.814×10^{-7}	1.278×10^{-7}	5.352×10^{-5}	2.413×10^{-7}
$\hat{\mathrm{Cov}}(\hat{p}_3, \hat{p}_4)$	2.330×10^{-10}	5.354×10^{-11}	-4.754×10^{-7}	1.217×10^{-10}

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 parameter estimates to their true values, we found that SE-optimal's average parameter estimates were always close, but D-optimal and E-optimal had poor estimates for the parameters in some cases. The estimated standard errors from the SE-optimal mesh were much smaller than either D-optimal or E-optimal. SE-optimal's estimated standard error were often similar to those from the uniform mesh. Between D-optimal and E-optimal, there was not a consistently more favorable method for this example.

For the harmonic example, comparing the approximate asymptotic standard errors, we found that SE-optimal had the smallest standard errors except in a couple of cases. We compared methods using the inverse problem with simulated data and asymptotic theory and bootstrapping. No optimal design method had parameter estimates that were consistently the closest to the true values. Use of asymptotic theory resulted in SE-optimal having the smallest estimated standard errors and covariance. Using bootstrapping, we found SE-optimal had the smallest estimated standard errors, except in one case, and it had the smallest estimated covariances.

For the glucose regulation model, comparing the approximate asymptotic standard errors, we again found SE-optimal had the smallest standard error except for in a few cases. 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. For the asymptotic and bootstrap estimated standard errors, SE-optimal had smaller standard errors for p_1, p_2 and p_3 but not for p_4. The asymptotic estimated covariances for SE-optimal were smaller than either D-optimal or E-optimal. For the bootstrapping estimated covariances, none of the optimal design methods were consistently favorable.

The best choice of optimal design method depends on the complexity of the model, the type of constraint one is using, 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 many 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.

Appendix-Constraints and Implementation of Constrained Optimization

We used several constrained optimization algorithms to solve the grid selection minimization problem of the form

$$\vec{\nu}^* = \min_{\vec{\nu}} J(\vec{\nu}),$$

subject to the constraint(s)

$$A\vec{\nu} \leq b, \text{ and/or } A_{eq}\vec{\nu} = b_{eq},$$

where $\vec{\nu}$ is a N-vector, A is a $(N+1) \times N$ matrix, b is a $(N+1)$-vector, A_{eq} is a $N \times N$ matrix, and b_{eq} is a scalar.

For our problem, we have the constraint

$$0 \leq \nu_1 \leq \nu_2 \leq \ldots \leq \nu_N \leq 1,$$

where $\vec{t} = (t_1, ..., t_N) = \vec{\nu}T = (\nu_1 T, ..., \nu_n T)$, then

$$0 \leq t_1 \leq t_2 \leq \ldots \leq t_N \leq T.$$

To express this constraint in the form

$$A\vec{\nu} \leq b, \text{ and/or } A_{eq}\vec{\nu} = b_{eq},$$

we have several options in algebraic formulations. Our four different constraint implementations are detailed below and the differences in the implementations of the constrained optimization algorithm account for the differences in the optimal meshes generated. As is explained, a primary difference in carrying out the optimizations is the number of points over which we optimize (i.e., the number of degrees of freedom in the problem).

Constraint implementation $(C1)$: For this constraint implementation, it differs from the other three in that it is not required that the end points are included in the optimal mesh. For this constraint we define the $(N+1) \times N$ matrix,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

We define the $(N+1)$-vector

$$b = [0, \cdots, 0, 1]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$0 \leq \nu_1 \leq \nu_2 \leq \ldots \leq \nu_N \leq 1.$$

Setting $\vec{t} = \vec{\nu}T$, we obtain

$$0 \leq t_1 \leq t_2 \leq \ldots \leq t_N \leq T.$$

In this case we optimize over N points.

Constraint implementation $(C2)$:

For this constraint implementation, we require that the end points are included in the optimal mesh. We optimize over the remaining mesh points $(t_2, \dots t_{N-1})$. For this constraint we define the $(N-1) \times (N-2)$ matrix,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

We define the $(N-1)$-vector

$$b = [0, \cdots, 0, 1]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$0 = \nu_1 \leq \nu_2 \leq \nu_2 \leq \ldots \leq \nu_{N-1} \leq \nu_N = 1.$$

Upon setting $\vec{t} = \vec{\nu}T$, we obtain

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

In this case we optimize over $N-2$ points.

Constraint implementation (C3):

For the third constraint implementation, we include the end points in the optimal mesh. We optimize over the remaining mesh points $(t_2, \dots t_{N-1})$. For this constraint we define the $(N-1) \times (N-2)$ matrix,

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & -1 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

We define the $(N-1)$-vector,

$$b = [0, \cdots, 0, T]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$\nu_i \geq 0, \text{ for } i = 2, \ldots, N-1$$

and

$$\nu_2 + \nu_3 \ldots + \nu_{N-1} \leq T.$$

To form $\vec{t}$ from $\vec{\nu}$, we first must define the $(N-2) \times (N-2)$ matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

Setting $t_1 = 0$, $t_N = T$ and

$$[t_2, \ldots, t_{N-1}]^{\mathrm{T}} = B[\nu_2, \ldots, \nu_{N-1}]^{\mathrm{T}},$$

which implies that

$$t_k = \sum_{j=2}^{k} \nu_j, \text{ for all } k = 2, \ldots N-1.$$

Then

$$0 = t_1 \leq \nu_2 \leq \nu_2 + \nu_3 \leq \ldots \leq (\nu_2 + \nu_3 + \ldots + \nu_{N-1}) \leq t_N = T,$$

or equivalently,

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

We again optimize over $N-2$ points.

Constraint implementation (C4):

For the fourth constraint, we include the end points in the optimal mesh. For this constraint we define the $(N-1) \times (N-1)$ matrix,

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & -1 \end{pmatrix}.$$

We define the $(N-1)$-vector,

$$b = [0, \cdots, 0]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$\nu_i \geq 0, \text{ for } i = 2, \ldots, N.$$

In addition, we define the $(N-1)$-row vector

$$A_{eq} = [1, 1, \ldots, 1],$$

and the scalar $b_{eq} = T$. The additional constraint, $A_{eq}\nu = b_{eq}$, implies

$$\sum_{j=2}^{N} \nu_j = T$$

To form $\vec{t}$ from $\vec{\nu}$, we first must define the $(N-1) \times (N-1)$ matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

Setting $t_1 = 0$ and

$$[t_2, \ldots, t_N]^{\mathrm{T}} = B[\nu_2, \ldots, \nu_N]^{\mathrm{T}},$$

which implies that

$$t_k = \sum_{j=2}^{k} \nu_j, \text{ for all } k = 2, \ldots N.$$

Then

$$0 = t_1 \leq \nu_2 \leq \nu_2 + \nu_3 \leq \ldots \leq (\nu_2 + \nu_3 + \ldots + \nu_N) = t_N = T,$$

or equivalently,

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

In this algorithm we again effectively optimize over $N - 2$ points.

References

[1] A.C. Atikinson and A.N. Donev, *Optimum Experimental Designs,* Oxford University Press, New York, 1992.

[2] H. T. Banks and K. L. Bihari, Modeling and estimating uncertainty in parameter estimation, *Inverse Problems*, **17** (2001), 95–111.

[3] H.T. Banks and C.A. Carter, Mathematical modeling of the glucose homeostatic system in humans, CRSC Technical Report, CRSC-TR10-09, NCSU, Raleigh, May, 2010.

[4] H.T. Banks, M. Davidian, J.R. Samuels Jr., and K.L. Sutton, An inverse problem statistical methodology summary, CRSC Technical Report, CRSC-TR08-01, NCSU, January, 2008; Chapter 11 in *Statistical Estimation Approaches in Epidemiology*, (edited by Gerardo Chowell, Mac Hyman, Nick Hengartner, Luis M.A. Bettencourt and Carlos Castillo-Chavez), Springer, Berlin Heidelberg New York, 2009, 249–302.

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

[6] H.T. Banks, K. Holm and D. Robbins, Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping, CRSC-TR09-13, June, 2009; Revised, August 2009; *Mathematical and Computer Modeling*, to appear.

[7] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes,* Chapman and Hall/ CRC, Boca Raton, FL, 2008.

[8] M.P.F. Berger and W.K. Wong (Editors), *Applied Optimal Designs,* John Wiley & Sons, Chichester, UK, 2005.

[9] R.N. Bergman, L.S. Phillips and C. Cobelli, 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,* **68(6)** (1981), 1456–1467.

[10] R.N. Bergman, Y.Z. Ider, C.R. Bowden and C. Cobelli, Quantitative estimation of insulin sensitivity, *Am. J. Physiol.,* **236** (1979), E667–E677.

[11] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, NY, 1968.

[12] A. De Gaetano and O. Arino, Mathematical modeling of the intravenous glucose tolerance test, *J. Math. Biology,* **40** (2000), 136–168.

[13] V.V. Fedorov and P. Hackel, *Model-Oriented Design of Experiments,* Springer-Verlag, New York, NY, 1997.

[14] P. J. Huber, *Robust Statistics*, John Wiley & Sons, Inc., New York, NY, 1981.

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

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

[17] Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.,* **1** (1956), 157–214.

[18] G. A. F. Seber and C. J. Wild, *Nonlinear Regression,* John Wiley & Sons, New York, NY, 1989.

[19] G. Toffolo, R.N. Bergman, D.T. Finegood, C.R. Bowden and C. Cobelli, Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal of insulin kinetics in the dog, *Diabetes,* **29** (1980), 979–990.