

Parameter Estimation in Distributed Systems: Optimal Design

H.T. Banks and K.L. Rehm

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

May 17, 2014

Abstract

We formulate a very general optimal design problem for the selection of best states to observe and optimal sampling times and locations (i.e., selections of what, when, and where to observe) for parameter estimation or inverse problems involving complex nonlinear partial differential systems. A theoretical framework along with a detailed iterative algorithm for implementation of the resulting methodology are proposed and the algorithm's successful use on several examples of wide interest are noted.

1 Introduction

Usual discussions of inverse problems in the presence of uncertainty have been in the context of *a given set or sets* of data carried out under various assumptions on how (e.g., independent sampling, absolute measurement error, relative measurement error) the data were collected. For many years now [4, 7, 20, 21, 22, 25, 26, 28] scientists (and especially engineers) have been actively involved in designing experimental protocols to best study engineering systems that include parameters describing mechanisms. Recently with increased involvement of scientists working in collaborative efforts with biologists and quantitative life scientists, renewed interest in design of the “best” experiments to elucidate mechanisms has been seen [9, 11, 12, 13, 15, 16]. Thus, a major question that experimentalists and inverse problem investigators alike often face is how to best collect the data to enable one to efficiently and accurately estimate model parameters. This is the well-known and widely studied *optimal design* problem for parameter estimation and is a most important step leading up to *control design* (sensor and actuator design and placement).

Traditional optimal design methods (D-optimal, E-optimal, c-optimal) [7, 20, 21, 22] use information from the model to find the sampling distribution or mesh for the observation times and locations in spatially distributed problems that minimizes a design criterion, quite often a function of the Fisher Information Matrix (FIM). Experimental data taken on this optimal mesh are then expected to result in accurate parameter estimates. In many scientific fields where mathematical modeling is utilized, mathematical models grow increasingly complex over time, possibly containing more state variables and parameters, as the underlying governing processes of a system are better understood and refinements in mechanisms are considered. Additionally, as technology invents and improves devices to measure physical and biological phenomena, new data become available to inform mathematical modeling efforts. The world is approaching an era in which the vast amounts of information available to researchers may be overwhelming or even counterproductive to efforts. We outline a framework based on the FIM for a system of partial differential equations (PDEs) to determine *when and where an experimenter should take samples* and *what variables to measure* when collecting information on a physical or biological process modeled by such a dynamical system.

Inverse problem methodologies are usually discussed in the context of a dynamical system or mathematical model when a sufficient number of observations of one or more states (variables) are available. The choice of method depends on assumptions the modeler makes on the form of the error between the model and the observations (the statistical model). The most prevalent source of readily studied error is observation

error, which is made when collecting data. (One can also consider model error, which originates from the differences between the model and the underlying process that the model describes. But this is often quite difficult to quantify.) This measurement error is most often discussed in the context of statistical models. The three techniques commonly addressed are maximum likelihood estimation (MLE), used when the probability distribution form of the error is known; ordinary least squares (OLS), for unknown error distributions with constant variance across observations; and generalized least squares (GLS), used when the variance of the data can be expressed as a nonconstant function. Uncertainty quantification is also described for optimization problems of this type, namely in the form of observation error covariances, standard errors, residual plots, and sensitivity matrices. Techniques to approximate the variance of the error are also included in these discussions.

In [12, 13], the authors develop an experimental design theory using the FIM to identify optimal sampling times for experiments on physical processes modeled by an ordinary differential equation (ODE) system in which scalar or vector data is taken. The experimental design technique developed there is applied in numerical simulations to the logistic curve, a simple ODE model describing glucose regulation, and a harmonic oscillator example.

In addition to *when* and *where* to take samples, the question of *what variables* to measure is also very important in designing effective experiments, especially when the number of state variables is large. Use of such a methodology to optimize what to measure would further reduce testing costs by eliminating extra experiments to measure variables neglected in previous trials. In [6], the best set of variables for an ODE system modeling the Calvin cycle [29] is identified using two methods. The first, an ad-hoc statistical method, determines which variables directly influence an output of interest at any one particular time. Such a method does not utilize the information on the underlying time-varying processes given by the dynamical system model. The second method is based on optimal design ideas. Extension of this method is developed in [15, 16]. Specifically, in [15] the authors compare the SE-optimal design introduced in [12] and [13] with the well-known methods of D-optimal and E-optimal design on a six-compartment HIV model [1, 2, 3] and a thirty-eight dimensional model of the Calvin Cycle [29]. Such models where there may be a wide range of variables to possibly observe are not only ideal on which to test the proposed methodology, but also are widely encountered in applications. For example these methods have been recently used in [17, 18, 19] to design optimal data collection in terms of the location of sensors and the number needed for optimal design in electroencephalography (EEG) in the recording of electrical activity along the scalp. The underlying models in these applications are nonhomogeneous second order elliptic partial differential equations. We turn to an outline of our design methodology for best times and locations and best variables in estimation for general nonlinear PDE models.

2 Mathematical and Statistical Models

The methodology we present can be readily applied to problems involving ordinary, partial and delay differential equations dynamics. Here we explore our experimental design questions using a PDE **mathematical model** (we illustrate with a partial differential equation that is first order in time and second order in space)

$$\frac{\partial \mathbf{u}}{\partial t} = \mathcal{F}(t, x, \mathbf{u}, \frac{\partial \mathbf{u}}{\partial x}, \frac{\partial^2 \mathbf{u}}{\partial x^2}, \boldsymbol{\theta}), \quad t \in [t_0, t_f], x \in [x_0, x_f] \quad (1)$$

with appropriate boundary and initial conditions, where $\mathbf{u}(t, x; \boldsymbol{\theta})$ is the m -vector of state variables of the system generated using a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{\kappa_\theta}$. We define a corresponding **observation process**

$$\mathbf{f}(t, x; \boldsymbol{\theta}) = C\mathbf{u}(t, x; \boldsymbol{\theta}), \quad (2)$$

where C is an observation operator that maps $\mathbb{R}^m \rightarrow \mathbb{R}^M$, where $M \leq m$ is the number of variables observed at a single sampling time and location. Of course the full state observation involves $M = m$; however, this is most often not feasible (due to the impossibility of or the expense in measuring all state variables such as in the plant metabolite example with $m = 38$ states studied in [15]). In other cases (such as the HIV examples studied in [9, 15]) we may be able to directly observe only combinations (e.g., total CD4+ cell counts including both uninfected and infected cells) of the states.

In order to discuss uncertainty in parameter estimates, we formulate a **statistical model** of the form (this corresponds to an ordinary or weighted least squares optimal fit to data formulation)

$$\mathbf{Y}(t, x) = \mathbf{f}(t, x; \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{E}}(t, x), \quad t \in [t_0, t_f], x \in [x_0, x_f], \quad (3)$$

where $\boldsymbol{\theta}_0$ is the hypothesized true values of the unknown parameters and $\boldsymbol{\mathcal{E}}$ is a vector random process that represents observation error for the measured variables. We make the standard assumptions that the errors are uncorrelated and independent with diagonal covariances $\text{Var}(\boldsymbol{\mathcal{E}}(t, x)) = V_0(t, x) = \text{diag}(\sigma_{0,1}(t, x)^2, \sigma_{0,2}(t, x)^2, \dots, \sigma_{0,M}(t, x)^2)$, $t \in [t_0, t_f]$, $x \in [x_0, x_f]$. Realizations of the statistical model (3) are written

$$\mathbf{y}(t, x) = \mathbf{f}(t, x; \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}(t, x), \quad t \in [t_0, t_f], x \in [x_0, x_f].$$

When collecting experimental data, it is often difficult to take continuous measurements of the observed variables; i.e., continuous measurements in t and x are often not possible or prove to be prohibitively expensive. Instead, we assume that we have $I \times J = IJ$ observations at sampling points (t_i, x_j) , $i = 1, \dots, I$, $j = 1, \dots, J$, in periods $[t_0, t_f] \times [x_0, x_f]$. We then write the observation process (2) as

$$\mathbf{f}(t_i, x_j; \boldsymbol{\theta}) = C\mathbf{u}(t_i, x_j; \boldsymbol{\theta}), \quad i = 1, \dots, I, j = 1, \dots, J, \quad (4)$$

the discrete statistical model as

$$\mathbf{Y}_{ij} = \mathbf{f}(t_i, x_j; \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{E}}(t_i, x_j), \quad i = 1, \dots, I, j = 1, \dots, J, \quad (5)$$

and a realization of the discrete statistical model as

$$\mathbf{y}_{ij} = \mathbf{f}(t_i, x_j; \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}(t_i, x_j), \quad i = 1, \dots, I, j = 1, \dots, J.$$

Given a set of data \mathbf{y}_{ij} , we could attempt to estimate $\boldsymbol{\theta}_0$ in a process known as the **inverse problem**. We will use this mathematical and statistical framework to outline a methodology to identify sampling variables that provide the most information pertinent to estimating a given set of parameters as well as the most informative times and locations at which the samples should be taken. To simplify notation below we shall introduce the notation $z_{ij} = (t_i, x_j)$ for discrete data taken from $z \in \Omega_Z = [t_0, t_f] \times [x_0, x_f]$. The corresponding data (continuous and discrete) will take the form

$$\mathbf{Y}(z) = \mathbf{f}(z; \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{E}}(z), \quad (6)$$

and

$$\mathbf{Y}_{ij} = \mathbf{f}(z_{ij}; \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{E}}(z_{ij}), \quad (7)$$

respectively. Below we will assume the data sets have been renumbered so that we may write $\{z_{ij}\}_{i,j=1}^{IJ} = \{z_l\}_{l=1}^L \in \Omega_Z$. Since the errors may in general be nonconstant, the above formulation leads naturally to a weighted least squares formulation.

2.1 Formulation of the Optimal Design Problem

Several methods exist to solve the inverse problem. A major factor in determining which method to use is additional assumptions made about $\boldsymbol{\mathcal{E}}(z)$. It is common practice to make the assumption that realizations of $\boldsymbol{\mathcal{E}}(z)$ at particular time/spatial points are independent and identically distributed (i.i.d.). If, additionally, the distributions describing the behavior of the components of $\boldsymbol{\mathcal{E}}(z)$ are known, then a maximum likelihood estimation method may be used to find an estimate of $\boldsymbol{\theta}_0$. On the other hand, if the distributions for $\boldsymbol{\mathcal{E}}(z)$ are not known but the covariance matrix $V_0(z)$ (also unknown) is assumed to vary over time and space, weighted least squares (WLS) methods are often used. We propose an optimal design problem formulation using a general weighted least squares criterion.

Let $\mathbb{P}_1(\Omega_Z)$ denote the set of all bounded distributions on the set Ω_Z . We consider the generalized weighted least squares cost functional for systems with vector output

$$J_{\text{WLS}}(\boldsymbol{\theta}; \mathbf{y}) = \int_{\Omega_Z} [\mathbf{y}(z) - \mathbf{f}(z; \boldsymbol{\theta})]^T V_0^{-1}(z) [\mathbf{y}(z) - \mathbf{f}(z; \boldsymbol{\theta})] dP_1(z), \quad (8)$$

where $P_1 \in \mathbb{P}_1(\Omega_Z)$ is a general measure on the set Ω_Z . For a given continuous data set $\mathbf{y}(z)$, we search for a parameter $\hat{\boldsymbol{\theta}}$ that minimizes $J_{\text{WLS}}(\boldsymbol{\theta}; \mathbf{y})$.

We next consider the more usual case of observations collected at discrete sampling points (often it is quite unreasonable and/or expensive to take a large number of samples in either time and/or space. If we choose a set of L time/spatial points $\mathcal{Z} = \{z_{ij}\}_{i,j=1}^{IJ} = \{z_l\}_{l=1}^L$, where $z_l \in \Omega_Z$ and take

$$P_1 = P_{\mathcal{Z}}^L = \sum_{l=1}^L \Delta_{z_l}, \quad (9)$$

where Δ_a represents the Dirac measure with atom at a , then the weighted least squares criterion (8) for a finite number of observations becomes

$$J_{\text{WLS}}^L(\boldsymbol{\theta}; \mathbf{y}) = \sum_{l=1}^L [\mathbf{y}(z_l) - \mathbf{f}(z_l; \boldsymbol{\theta})]^T V_0^{-1}(z_l) [\mathbf{y}(z_l) - \mathbf{f}(z_l; \boldsymbol{\theta})].$$

Note here we do not normalize the time/space “distributions” such as (9) by a factor of $\frac{1}{L}$ so that they are not the usual cumulative distribution functions but would be if we normalized each distribution by the integral of its corresponding density to obtain a true probability measure. A similar remark holds for the “variables” observation operator distributions introduced below where without loss of generality we could normalize by a factor of $\frac{1}{K}$ when using K 1-dimensional sampling maps.

To select a useful distribution of sampling points and set of observation variables, we introduce the M by $\kappa_{\boldsymbol{\theta}}$ sensitivity matrices $\frac{\partial \mathbf{f}(z; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and the m by $\kappa_{\boldsymbol{\theta}}$ sensitivity matrices $\frac{\partial \mathbf{u}(z; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ that are determined using the differential operator in row vector form $(\partial_{\theta_1}, \partial_{\theta_2}, \dots, \partial_{\theta_{\kappa_{\boldsymbol{\theta}}}})$ represented by $\nabla_{\boldsymbol{\theta}}$ and the observation operator defined in (2),

$$\nabla_{\boldsymbol{\theta}} \mathbf{f}(z; \boldsymbol{\theta}) = \frac{\partial \mathbf{f}(z; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = C \frac{\partial \mathbf{u}(z; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = C \nabla_{\boldsymbol{\theta}} \mathbf{u}(z; \boldsymbol{\theta}) = C \nabla \mathbf{u}(z; \boldsymbol{\theta}). \quad (10)$$

Using the sensitivity matrix $\nabla_{\boldsymbol{\theta}} \mathbf{f}(z; \boldsymbol{\theta}_0)$, we may formulate the Generalized Fisher Information Matrix (GFIM). Consider the set (assumed compact) $\Omega_C \subset \mathbb{R}^{1 \times m}$ of admissible 1-dimensional observation maps and let $\mathbb{P}_2(\Omega_C)$ represent the set of all bounded distributions P_2 on Ω_C . Then the GFIM may be written

$$\begin{aligned} F(P_{\mathcal{Z}}^L, P_2, \boldsymbol{\theta}_0) &\equiv \int_{\Omega_Z} \int_{\Omega_C} \frac{1}{\sigma^2(z, c)} \nabla_{\boldsymbol{\theta}}^T \mathbf{f}(z; \boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} \mathbf{f}(z; \boldsymbol{\theta}_0) dP_2(c) dP_{\mathcal{Z}}^L(z) \\ &= \int_{\Omega_Z} \int_{\Omega_C} \frac{1}{\sigma^2(z, c)} \nabla_{\boldsymbol{\theta}}^T (C \mathbf{u}(z; \boldsymbol{\theta}_0)) \nabla_{\boldsymbol{\theta}} (C \mathbf{u}(z; \boldsymbol{\theta}_0)) dP_2(c) dP_{\mathcal{Z}}^L(z). \end{aligned} \quad (11)$$

In fact we shall be interested in collections of K 1-dimensional “variable” observation operators and a choice of which K variables provide best information to estimate the desired unknown parameters in a given model. Thus taking K different sampling maps in Ω_C represented by the $1 \times m$ -dimensional matrices \mathcal{C}_k ,

$k = 1, 2, \dots, K$, we construct the discrete distribution on $\Omega_C^K = \bigotimes_{i=1}^K \Omega_C$ (the k -fold cross products of Ω_C)

$$P_S^K = \sum_{k=1}^K \Delta_{\mathcal{C}_k}, \quad (12)$$

where Δ_a represents the Dirac measure with atom at a . Using P_S^K in (11), we obtain the GFIM for multiple

discrete observation methods taken continuously over Ω_Z given by

$$\begin{aligned}
F(P_Z^L, P_S^K, \boldsymbol{\theta}_0) &= \int_{\Omega_Z} \sum_{k=1}^K \frac{1}{\sigma^2(z, \mathcal{C}_k)} \nabla_{\boldsymbol{\theta}}^T (\mathcal{C}_k \mathbf{u}(z; \boldsymbol{\theta}_0)) \nabla_{\boldsymbol{\theta}} (\mathcal{C}_k \mathbf{u}(z; \boldsymbol{\theta}_0)) dP_Z^L(z) \\
&= \int_{\Omega_Z} \sum_{k=1}^K \frac{1}{\sigma^2(z, \mathcal{C}_k)} \nabla_{\boldsymbol{\theta}}^T \mathbf{u}(z; \boldsymbol{\theta}_0) \mathcal{C}_k^T \mathcal{C}_k \nabla_{\boldsymbol{\theta}} \mathbf{u}(z; \boldsymbol{\theta}_0) dP_Z^L(z) \\
&= \int_{\Omega_Z} \sum_{k=1}^K \nabla_{\boldsymbol{\theta}}^T \mathbf{u}(z; \boldsymbol{\theta}_0) \mathcal{C}_k^T \frac{1}{\sigma^2(z, \mathcal{C}_k)} \mathcal{C}_k \nabla_{\boldsymbol{\theta}} \mathbf{u}(z; \boldsymbol{\theta}_0) dP_Z^L(z) \\
&= \int_{\Omega_Z} \nabla_{\boldsymbol{\theta}}^T \mathbf{u}(z; \boldsymbol{\theta}_0) \sum_{k=1}^K \left(\mathcal{C}_k^T \frac{1}{\sigma^2(z, \mathcal{C}_k)} \mathcal{C}_k \right) \nabla_{\boldsymbol{\theta}} \mathbf{u}(z; \boldsymbol{\theta}_0) dP_Z^L(z) \\
&= \int_{\Omega_Z} \nabla_{\boldsymbol{\theta}}^T \mathbf{u}(z; \boldsymbol{\theta}_0) (\mathcal{S}^T V_K^{-1}(z) \mathcal{S}) \nabla_{\boldsymbol{\theta}} \mathbf{u}(z; \boldsymbol{\theta}_0) dP_Z^L(z), \tag{13}
\end{aligned}$$

where $\mathcal{S} = \text{column}(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K) \in \mathbb{R}^{K \times m}$ is the set of observation operators defined above and $V_K(z) = \text{diag}(\sigma^2(z, \mathcal{C}_1), \dots, \sigma^2(z, \mathcal{C}_K))$ is the corresponding covariance matrix for K 1-dimensional observation operators. Applying the distribution P_Z^L as described in (9) to the GFIM (13) for discrete observation operators measured continuously yields the discrete $\kappa_\theta \times \kappa_\theta$ Fisher Information Matrix (FIM) for discrete observation operators measured at discrete times and locations given by

$$\begin{aligned}
F(\mathcal{Z}, \mathcal{S}, \boldsymbol{\theta}_0) &\equiv F(P_Z^L, P_S^K, \boldsymbol{\theta}_0) = \\
&\sum_{l=1}^L \nabla_{\boldsymbol{\theta}}^T \mathbf{u}(z_l; \boldsymbol{\theta}_0) \mathcal{S}^T V_K^{-1}(z_l) \mathcal{S} \nabla_{\boldsymbol{\theta}} \mathbf{u}(z_l; \boldsymbol{\theta}_0). \tag{14}
\end{aligned}$$

This describes the amount of information about the κ_θ parameters of interest that is captured by the observed quantities described by the sampling maps \mathcal{C}_k , $k = 1, 2, \dots, K$, defining \mathcal{S} , when they are measured at the L time/spatial points in Ω_Z .

The questions of determining the best (in some sense) \mathcal{S} and L sampling points \mathcal{Z} are the important questions in the optimal design of an experiment. Recall that the set of sampling points \mathcal{Z} has an associated distribution $P_Z^L \in \tilde{\mathbb{P}}_1(\Omega_Z)$, where $\tilde{\mathbb{P}}_1(\Omega_Z)$ is the set of all bounded discrete distributions on Ω_Z . Similarly, the set of sampling maps \mathcal{S} has an associated bounded discrete distribution $P_S^K \in \tilde{\mathbb{P}}_2(\Omega_C^K)$. Define the space of bounded discrete distributions $\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K) = \tilde{\mathbb{P}}_1(\Omega_Z) \times \tilde{\mathbb{P}}_2(\Omega_C^K)$ with elements $P = (P_Z^L, P_S^K) \in \tilde{\mathbb{P}}$. We may, without loss of generality, assume that $\Omega_C^K \subset \mathbb{R}^{K \times m}$ is closed and bounded, and assume that there exists a functional $\mathcal{J} : \mathbb{R}^{\kappa_\theta \times \kappa_\theta} \rightarrow \mathbb{R}^+$ of the GFIM (13). Then the **optimal design problem** associated with \mathcal{J} is selecting a discrete distribution $\hat{P} \in \tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K)$ such that

$$\mathcal{J}(F(\hat{P}, \boldsymbol{\theta}_0)) = \min_{P \in \tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K)} \mathcal{J}(F(P, \boldsymbol{\theta}_0)), \tag{15}$$

where \mathcal{J} is assumed to depend continuously on the elements of $F(P, \boldsymbol{\theta}_0)$.

The Prohorov Metric Framework [8] developed over the past decade and completed in [14] provides a general theoretical framework for the existence of $\hat{P} = \hat{P}^{L,K}$ in $\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K)$ (a general theoretical framework with proofs is developed in [8, 10, 14]). The application of the Prohorov metric to optimal design problems formulated as (15) is explained more fully in [12]: briefly, define the Prohorov metric ρ on the space $\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K)$, and consider the metric space $(\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K), \rho)$. Since $\Omega_Z \times \Omega_C^K$ is compact, $(\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K), \rho)$ is also compact. Therefore an optimal distribution $\hat{P} = \hat{P}^{L,K}$ in $\tilde{\mathbb{P}}(\Omega_Z \times \Omega_C^K)$. Under appropriate assumptions one can guarantee consistency [14] of the resulting estimator $\hat{P}^{L,K}$. More precisely, using this optimal distribution $\hat{P} = \hat{P}^{L,K}$ in the corresponding weighted least squares criterion $J_{\text{WLS}}^{L,K}(\boldsymbol{\theta}; \mathbf{y})$ gives rise to an estimate $\hat{\boldsymbol{\theta}}^{L,K}$ for $\boldsymbol{\theta}_0$. Following the consistency [27] arguments found in [14, Sections 3.3], one can argue $\hat{\boldsymbol{\theta}}^{L,K} \rightarrow \boldsymbol{\theta}_0$ as $L \rightarrow \infty$, $K \rightarrow m$ (see Theorem 3.3.1 and its extension to weighted least squares in Section 3.3.1 of [14]). An essential assumption is on the sampling locations associated with \mathcal{Z} which require that as $L \rightarrow \infty$, the sampling points $z_{i,j} = (t_i, x_j)$ “fill up” the region Ω_Z – see condition (A5) of Section 3.3 in

[14]. This is in some sense equivalent to $\hat{P}^{L,K} \rightarrow P_0 = \Delta_{\theta_0}$ in the Prohorov metric (i.e., in distribution) as $L \rightarrow \infty, K \rightarrow m$.

The formulation of the cost functional (15) may take many forms. We outline the use of traditional optimal design methods, D-optimal, E-optimal, or SE-optimal design criteria, to determine the form of \mathcal{J} . Each of these design criteria are functions of the inverse of the FIM (assumed hereafter to be invertible) defined in (14).

In D-optimal design, the cost functional is written

$$\mathcal{J}_D(F) = \det((F(\mathcal{Z}, \mathcal{S}, \theta_0))^{-1}) = \frac{1}{\det(F(\mathcal{Z}, \mathcal{S}, \theta_0))}.$$

By minimizing \mathcal{J}_D , we minimize the volume of the confidence interval ellipsoid describing the uncertainty in our parameter estimates. Since F is symmetric and positive semi-definite, $\mathcal{J}_D(F) \geq 0$. Additionally, since F is assumed invertible, $\mathcal{J}_D(F) \neq 0$, therefore, $\mathcal{J}_D : \mathbb{R}^{\kappa_\theta \times \kappa_\theta} \rightarrow (0, \infty)$.

In E-optimal design, the cost functional \mathcal{J}_E is the largest eigenvalue of $(F(\mathcal{Z}, \mathcal{S}, \theta_0))^{-1}$, or equivalently

$$\mathcal{J}_E(F) = \max \left\{ \frac{1}{\text{eig}(F(\mathcal{Z}, \mathcal{S}, \theta_0))} \right\}.$$

To obtain a smaller standard error, we must reduce the length of the principal axis of the confidence interval ellipsoid. Since F is positive definite, all eigenvalues are therefore positive. Thus $\mathcal{J}_E : \mathbb{R}^{\kappa_\theta \times \kappa_\theta} \rightarrow (0, \infty)$.

In SE-optimal design, the cost functional \mathcal{J}_{SE} is a sum of the elements on the diagonal of $(F(\mathcal{Z}, \mathcal{S}, \theta_0))^{-1}$ weighted by the respective parameter values [12, 13], written

$$\mathcal{J}_{SE}(F) = \sum_{i=1}^{\kappa_\theta} \frac{(F(\mathcal{Z}, \mathcal{S}, \theta_0))_{i,i}^{-1}}{\theta_{0,i}^2}.$$

Thus in SE-optimal design, the goal is to minimize the standard deviation of the parameters, normalized by the true parameter values. As the diagonal elements of F^{-1} are all positive and all parameters are assumed non-zero in $\theta \in \mathbb{R}^{\kappa_\theta}$, $\mathcal{J}_{SE} : \mathbb{R}^{\kappa_\theta \times \kappa_\theta} \rightarrow (0, \infty)$.

In [13], it is shown that the D-, E-, and SE-optimal design criteria select different time grids and in general yield different standard errors. As we might expect these design cost functionals will also generally choose different observation variables (maps) [15] in order to minimize different aspects of the confidence interval ellipsoid.

3 Algorithmic Considerations

We complete our outline of design methods with a very brief discussion of algorithmic issues. In most choice-of-variable optimal design problems, one does not have a continuum of measurement possibilities; rather, there are $K^* \leq m$ possible variable observation maps \mathcal{C} . Denote this set as $\Omega_C^{K^*} \subset \mathbb{R}^{K^* \times m}$. While we may still use the Prohorov metric framework to guarantee existence and convergence of (15), we have a stronger result first proposed in [6] that is useful in numerical implementation. Because for a given K , $K \leq K^*$, Ω_C^K is finite, all bounded discrete distributions made from the elements of Ω_C^K have the form

$$P_S^K = \sum_{k=1}^K \Delta_{\mathcal{C}_k}.$$

Moreover, the set $\tilde{\mathbb{P}}_2(\Omega_C^K)$ of all discrete distributions that use K sampling methods is also finite. For a fixed distribution of sampling points $P_{\mathcal{Z}}^L$, we may compute using (14) the set of all possible FIM $F(\mathcal{Z}, \mathcal{S}, \theta)$ that could be formulated from $P_S^K \in \tilde{\mathbb{P}}_2(\Omega_C^K)$. By the properties of matrix multiplication and addition, this set is also finite. Then the functional (15) applied to all F in the set produces a finite set contained in \mathbb{R}^+ . Because this set is finite, it is well-ordered by the relation \leq and therefore has a minimal element. Therefore, for any distribution of sampling points $P_{\mathcal{Z}}^L$, we may find at least one solution $\hat{P}_S^K \in \tilde{\mathbb{P}}_2(\Omega_C^K)$. Moreover, \hat{P}_S^K may be determined by a search over all matrices $\mathcal{S} = \text{column}(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K)$ formed by K elements having support in Ω_C^K .

Due to the computational demands of performing nonlinear optimization for L sampling points and K observation maps (for a total of $L + K$ dimensions), we solve the coupled set of equations

$$\hat{\mathcal{S}} = \arg \min_{\{\mathcal{S} | P_{\mathcal{S}}^K \in \mathbb{P}_2(\Omega_C^K)\}} \mathcal{J} \left(F(\hat{\mathcal{Z}}, \mathcal{S}, \boldsymbol{\theta}_0) \right) \quad (16)$$

$$\hat{\mathcal{Z}} = \arg \min_{\{\mathcal{Z} | P_{\mathcal{Z}}^L \in \mathbb{P}_1(\Omega_Z)\}} \mathcal{J} \left(F(\mathcal{Z}, \hat{\mathcal{S}}, \boldsymbol{\theta}_0) \right), \quad (17)$$

where $\mathcal{S} \in \mathbb{R}^{K \times m}$ represents a set of K sampling maps and $\mathcal{Z} = \{z_l\}_{l=1}^L$, $z_l \in \Omega_Z$, is an ordered set of L sampling points. These equations are solved iteratively as

$$\hat{\mathcal{S}}_i = \arg \min_{\{\mathcal{S} | P_{\mathcal{S}}^K \in \mathbb{P}_2(\Omega_C^K)\}} \mathcal{J} \left(F(\hat{\mathcal{Z}}_{i-1}, \mathcal{S}, \boldsymbol{\theta}_0) \right) \quad (18)$$

$$\hat{\mathcal{Z}}_i = \arg \min_{\{\mathcal{Z} | P_{\mathcal{Z}}^L \in \mathbb{P}_1(\Omega_Z)\}} \mathcal{J} \left(F(\mathcal{Z}, \hat{\mathcal{S}}_i, \boldsymbol{\theta}_0) \right), \quad (19)$$

where \mathcal{J} is the D-, E-, or SE-optimal design criterion. We begin by solving for $\hat{\mathcal{S}}_1$ where $\hat{\mathcal{Z}}_0$ is specified by the user. The system (18)-(19) is solved until $\left| \mathcal{J} \left(F(\hat{\mathcal{Z}}_i, \hat{\mathcal{S}}_i, \boldsymbol{\theta}_0) \right) - \mathcal{J} \left(F(\hat{\mathcal{Z}}_{i-1}, \hat{\mathcal{S}}_{i-1}, \boldsymbol{\theta}_0) \right) \right| < \epsilon$ or until $\hat{\mathcal{S}}_i = \hat{\mathcal{S}}_{i-1}$. For each iteration, (18) is solved using a global search over all possible \mathcal{S} as explained above. Since the sensitivity equations cannot be easily solved for in the models chosen here and in [15, 16, 17, 18] to illustrate this method, one can use a modified version of *tssolve.m* [5], which implements the *myAD* package developed in [23]. Solving (19) requires using a nonlinear constrained optimization algorithm. While Matlab's *fmincon* is a natural choice for such problems, as reported in [13], it does not perform well in this situation. Instead, we recommend the optimization tool *SolvOpt* developed by Kuntsevich and Kappel [24]. Once either of the convergence requirements are met and $\hat{\mathcal{S}}$ and $\hat{\mathcal{Z}}$ are determined, one can compute standard errors using the asymptotic theory described in [14].

Acknowledgements

This research was supported in part by Grant Number NIAID R01AI071915-10 from the National Institute of Allergy and Infectious Diseases, and in part by the Air Force Office of Scientific Research under grant number AFOSR FA9550-12-1-0188.

References

- [1] B.M. Adams, *Non-parametric Parameter Estimation and Clinical Data Fitting with a Model of HIV Infection*, PhD Thesis, NC State Univ., 2005.
- [2] B.M. Adams, H.T. Banks, M. Davidian, H. Kwon, H. T. Tran, S.N. Wynne, and E.S. Rosenberg, HIV dynamics: Modeling, data analysis, and optimal treatment protocols, CRSC-TR04-05, February 2004; *J. Comp. Appl. Math.*, **184** (2005), 10–49.
- [3] B.M. Adams, H.T. Banks, M. Davidian and E.S. Rosenberg, Model fitting and prediction with HIV treatment interruption data, CRSC-TR05-40, October, 2005; *Bulletin of Math. Biology*, **69** (2007), 563–584.
- [4] A.C. Atkinson and R.A. Bailey, One hundred years of the design of experiments on and off the pages of *Biometrika*, *Biometrika*, **88** (2001), 53–97.
- [5] A. Attarian, *tssolve.m*, Retrieved August 2011, from <http://www4.ncsu.edu/~arattari/>.
- [6] M. Avery, H.T. Banks, K. Basu, Y. Cheng, E. Eager, S. Khasawinah, L. Potter, and K.L. Rehm, Experimental design and inverse problems in plant biological modeling, CRSC-TR11-12, October, 2011; *J. Inverse and Ill-posed Problems*, DOI 10.1515/jiip-2012-0208.
- [7] A.C. Atkinson and A.N. Donev, *Optimum Experimental Designs*, Oxford University Press, New York, 1992.

- [8] H.T. Banks, *A Functional Analysis Framework for Modeling, Estimation and Control in Science and Engineering*, Chapman and Hall/CRC Press, Boca Raton, FL, 2012.
- [9] H.T. Banks, R. Baraldi, K. Cross, C. McChesney, L. Poag, E. Thorpe, and K.B. Flores, Uncertainty quantification in modeling HIV viral mechanics, CRSC-TR13-16, N. C. State University, Raleigh, NC, December, 2013; *Mathematical Biosciences and Engineering*, submitted.
- [10] H.T. Banks and K.L. Bihari, Modeling and estimating uncertainty in parameter estimation, CRSC-TR99-40, December, 1999; *Inverse Problems*, **17** (2001), 95–111.
- [11] H.T. Banks, A. Cintrón-Arias, and F. Kappel, Parameter selection methods in inverse problem formulation, CRSC-TR10-03, N.C. State University, February, 2010, Revised, November, 2010; in *Mathematical Modeling and Validation in Physiology: Application to the Cardiovascular and Respiratory Systems*, (J. J. Batzel, M. Bachar, and F. Kappel, eds.), pp. 43 - 73, Lecture Notes in Mathematics Vol. 2064, Springer-Verlag, Berlin 2013.
- [12] H.T. Banks, S. Dediu, S.L. Ernstberger and F. Kappel, Generalized sensitivities and optimal experimental design, CRSC-TR08-12, September, 2008, (Revised), November, 2009; *J. Inverse and Ill-posed Problems*, **18** (2010), 25–83.
- [13] H.T. Banks, K. Holm and F. Kappel, Comparison of optimal design methods in inverse problems, CRSC-TR10-11, July 2010; *Inverse Problems*, **27** (2011), 075002.
- [14] H.T. Banks, S. Hu and W.C. Thompson, *Modeling and Inverse Problems in the Presence of Uncertainty*, Taylor/Francis-CRC Press, Boca Raton, FL, 2014.
- [15] H.T. Banks and K.L. Rehm, Experimental design for vector output systems, CRSC-TR12-11, April, 2012; *Inverse Problems in Sci. and Engr.*, (2013), 1–34. DOI: 10.1080/17415977.2013.797973
- [16] H.T. Banks and K.L. Rehm, Experimental design for distributed parameter vector systems, CRSC-TR12-17, August, 2012; *Applied Mathematics Letters*, **26** (2013), 10–14; <http://dx.doi.org/10.1016/j.aml.2012.08.003>.
- [17] H.T. Banks, D. Rubio, N. Saintier and M.I. Tropicovsky, Optimal design techniques for distributed parameter systems, CRSC-TR13-01, January, 2013; *Proceedings 2013 SIAM Conference on Control Theory*, SIAM, 83–90.
- [18] H.T. Banks, D. Rubio, N. Saintier and M.I. Tropicovsky, Optimal electrode positions for the inverse problem of EEG in a simplified model in 3D, *Proceedings MACI 2013: Fourth Conference on Applied, Computational and Industrial Mathematics*, May 15-17, 2013, Buenos Aires, AR, **4** (2013), 521–524.
- [19] H.T. Banks, D. Rubio, N. Saintier and M.I. Tropicovsky, Optimal design for parameter estimation in EEG problems in a 3D multilayered domain, CRSC-TR14-02, N. C. State University, Raleigh, NC, March, 2014; *Mathematical Biosciences and Engineering*, submitted.
- [20] M.P.F. Berger and W.K. Wong (Editors), *Applied Optimal Designs*, John Wiley & Sons, Chichester, UK, 2005.
- [21] V.V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York and London, 1972.
- [22] V.V. Fedorov and P. Hackel, *Model-Oriented Design of Experiments*, Springer-Verlag, New York, NY, 1997.
- [23] M. Fink, myAD, Retrieved August 2011, from <http://www.mathworks.com/matlabcentral/fileexchange/15235-automatic-differentiation-for-matlab>.
- [24] A. Kuntsevich and F. Kappel, SolvOpt, Retrieved December 2009, from <http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/>.
- [25] W. Müller and M. Stehlik, Issues in the optimal design of computer simulation experiments, *Appl. Stochastic Models in Business and Industry*, **25** (2009), 163–177.

- [26] M. Patan and B. Bogacka, Optimum experimental designs for dynamic systems in the presence of correlated errors, *Computational Statistics and Data Analysis*, **51** (2007), 5644–5661.
- [27] G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, John Wiley & Sons, New York, NY, 1989.
- [28] D. Ucinski and A.C. Atkinson, Experimental design for time-dependent models with correlated observations, *Studies in Nonlinear Dynamics and Econometrics*, **8(2)** (2004), Article 13: The Berkeley Electronic Press.
- [29] X.-G. Zhu, E. de Sturler, and S. P. Long, Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm, *Plant Physiology*, **145** (2007), 513–526.