Published in final edited form as:

Ann Nucl Energy. 2013 February ; 52: . doi:10.1016/j.anucene.2012.05.016.

A numerical method for solving a stochastic inverse problem for parameters

T. Butler^a and D. Estep^{b,*}

^aInstitute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, United States

^bDepartment of Statistics, Colorado State University, Fort Collins, CO 80523, United States

Abstract

We review recent work (Briedt et al., 2011., 2012) on a new approach to the formulation and solution of the stochastic inverse parameter determination problem, i.e. determine the random variation of input parameters to a map that matches specified random variation in the output of the map, and then apply the various aspects of this method to the interesting Brusselator model. In this approach, the problem is formulated as an inverse problem for an integral equation using the Law of Total Probability. The solution method employs two steps: (1) we construct a systematic method for approximating set-valued inverse solutions and (2) we construct a computational approach to compute a measure-theoretic approximation of the probability measure on the input space imparted by the approximate set-valued inverse that solves the inverse problem. In addition to convergence analysis, we carry out an a posteriori error analysis on the computed probability distribution that takes into account all sources of stochastic and deterministic error.

Keywords

A posteriori error analysis; Adjoint problem; Density estimation; Inverse sensitivity analysis; Nonparametric density estimation; Sensitivity analysis

1. Introduction

In recent work (Briedt et al., 2011, 2012), a new approach to the formulation and solution of the stochastic inverse parameter determination problem, i.e. determine the random variation of input parameters to a map that matches specified random variation in the output of the map, is presented. The purpose of this paper is to review the formulation of the stochastic inverse problem, the approximate solution method, and the error analysis, and then apply the results to the interesting Brusselator model. In doing so, we expose some interesting features of the results.

We begin by formulating the model problem. Consider a quantity of interest determined by a linear functional $q(\lambda) = q(y(\lambda))$ applied to the solution $y(\lambda)$ of a model $M(y, \lambda) = 0$, where the parameters and data take values in the parameter space $\Lambda \subset \mathbb{R}^d$ and q takes values in an output space $\mathcal{D}$. Below, we refer to the (forward) map from Λ to $\mathcal{D}$ defined as the composition $M(y, \lambda) = 0 \rightarrow y(\lambda) \rightarrow q(y(\lambda))$.

In the case of interest, M is given as a system of differential equations that determines the solution y as an implicit function of the parameters and data λ . To illustrate the methods and algorithms, we use the well-studied Brusselator model (Prigogine and Lefever, 1968) as an example. If the rate constants in the autocatalytic reaction described by this model are set to unity, then the model reduces to

$$\begin{cases} \dot{y}_1 = A + y_1^2 y_2 - B y_1 - y_1, & t > 0, \\ \dot{y}_2 = B y_1 - y_1^2 y_2, & t > 0, \\ y_1(0) = y_{1,0}, y_2(0) = y_{2,0}. \end{cases} \quad (1.1)$$

Here, A and B represent fixed concentrations of chemicals that drive the chemical reaction, creating temporally varying chemical concentrations y_1 and y_2 . The values of $y_{1,0}$ and $y_{2,0}$ are the initial concentrations of y_1 and y_2 . We take as $\lambda = (A, B, IC)^\top$ to be the model parameters A , B , and the first component of the initial condition $y_{1,0}$, which we denote IC below for ease of reference. The quantity of interest is the sum of the average values of the varying concentrations over $[0, T]$, i.e., $q(\lambda) = (1/T) \int_0^T (y_1 + y_2) dt$.

The forward deterministic model problem is simply to compute a value $q(y(\lambda))$ given values for parameters λ . This is associated with a natural inverse problem, which is given a value $q(y(\lambda))$, determine the corresponding values of λ . We note that this inverse problem is physically motivated and the values of parameters, which characterize the physical system, are often the goal of scientific inference. We write values plural because in the general situation, the parameter space is several dimensions while the output quantity of interest is one dimensional. Hence, the forward map is a many-to-one transformation and the inverse problem has set-valued solutions, with infinite possible values of λ corresponding to a given output value. We illustrate in Fig. 1.1. We note that the model *cannot* distinguish among the possible values in a set-valued inverse. All of those values can produce the observed output.

The indeterminacy of set-valued inverse solutions means the problem of determining a *single* distinguished parameter value leading to the observed output is ill-posed. However, *the inverse problem for a smooth map is well-posed with respect to set-valued inverses*. Consider the case of a smooth map of a two dimensional domain as illustrated in Fig. 1.1. The map behaves smoothly and monotonically with respect to the contours and the contours cannot intersect. However, there are nontrivial difficulties with using set-valued inverse directly, which include: (1) describing the set of set-valued inverses and (2) approximating a particular set-valued inverse.

Next, we describe the stochastic version of the forward and inverse problems. For the forward problem, we assume that the parameters vary randomly in the domain Λ . This means we give a probability measure on Λ that determines the probability of events, or different sets of parameter values. Under smoothness assumptions on the solution of the model M , the output quantity $q(y)$ is now a random variable, again associated with a probability measure. The model that maps the random variable inputs into the random output implicitly defines a map between the probability measure corresponding to the input to the probability measure corresponding to the output. The standard forward stochastic sensitivity problem is to compute this output probability measure. It is often solved with a Monte Carlo approach: We choose values λ of the parameters according to the probability measure on Λ , compute the corresponding values $q(y(\lambda))$, and use the values to approximate either the cumulative density function of the output or employ smoothing to approximate the probability density function.

We distinguish this nonparametric density estimation approach in which we directly study the propagation of the input probability measure to the output from a statistical modeling approach. In the latter, we choose some statistical model, e.g. parameterized distribution or polynomial chaos interpolant, and then attempt to find a best fit of the model or interpolant to the output distribution based on evaluating the model at selected parameter values. The statistical modeling approach is certainly reasonable when the output distribution is close to the model. In our experience, however, even simple problems generally lead to very complicated output distributions that are not easily described in terms of simple models or representations.

The stochastic inverse problem associated to the map is to impose a probability measure on the values of $q(\lambda)$ in $\mathcal{D}$ and then find possible probability measures on the input space Λ that produce this output measure. This is the direct inverse of the forward stochastic sensitivity analysis problem. As with the deterministic version of the inverse problem, there is no unique solution in general. For example, consider the simple linear map $q(\lambda) = \lambda_1 + \lambda_2$, where λ_1, λ_2 are random variables. If we specify that $q(\lambda)$ has a normal $N(0, 2/25)$ distribution, then any bivariate normal density

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \sim N \left(\begin{pmatrix} -\alpha \\ \alpha \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right) \quad \text{with } 2\tau^2(1+\varrho) = \frac{2}{25}, \varrho \in [-1, 1],$$

solves the inverse problem. There are infinitely more (and wilder) solutions, see Briedt et al. (2011).

The simple reason for the infinite number of solutions of the stochastic inverse problem is that the deterministic inverse problem has set-valued solutions. On the other hand, we can compute a unique solution of the stochastic inverse problem as a probability measure on the set of set-valued inverses, provided we can describe the set, see Fig. 1.1. This solution makes it possible to compute the probability of events described as collections of set-valued inverses, however it is not possible to compute the probability of arbitrary events in the parameter space, see Fig. 1.1.

Up to this point, we have described the inverse problem in abstract terms with minimal assumptions on the model formulation and the parameter set. The result is unsatisfactory because the general goal is the ability to compute probabilities of arbitrary events in the parameter set, not a limited set of strange looking events comprised of set-valued inverses (which are themselves difficult to approximate). If we can compute the probabilities of arbitrary events, then we can undertake a number of interesting scientific problems in prediction, engineering design, and optimization. Moreover, if we achieve the desired goal, then we can approximate a unique probability measure on Λ . So clearly, this would require more assumptions on the problem.

Fortunately, in many natural situations, there is considerable information known about the parameters because they are specified as part of creating the model. For example, we certainly know the units of parameter values and we know the relationship between the parameters. We assume that sufficient information is available that we can compute or approximate the volume of an arbitrary (measurable) event in the parameter domain. With this additional structural information, we can formulate the stochastic inverse problem so that we can find a unique probability measure on the parameter domain. Note that the indeterminacy of the inverse problem has not disappeared: if we choose a different volume measure, then we get a different probability measure.

We again distinguish this approach from the problem of inverting a statistical model or representation of the output distribution. In our approach, we are directly inverting the map from the input probability measure to the output probability measure. A major concern with inverting a statistical model of the output is that statistical models are not typically constructed so that their inverses approximate the inverse of the underlying map in general. There is no reason to think the inverses are close in general and it is easy to construct examples in which they are not close.

In Briedt et al. (2011), we formulate this discussion in mathematical terms and then describe an approximate solution method for the stochastic inverse problem that provides the means to approximate the probability of an arbitrary event in the parameter domain. We also show that the approximate solution method converges. We follow in Butler et al. (2012) with an analysis of the numerical errors involved in implementing the approximate solution method.

1.1. Formulation of the inverse problem using the Law of Total Probability

We now formulate the stochastic inverse problem for parameter determination as an inverse problem for an integral equation. The goal is to find a concrete expression for the map between the probability measures of the input and output that is defined implicitly by the physical model.

We assume there is a *parameter volume* measure μ_{Λ} on Λ that determines the volume of sets in Λ . The volume measure is specified as part of the model that defines the map $q(\lambda)$ since the parameters must be explicitly defined in the physical model that determines q . The volume measure depends on the units of measure used for the parameters and also reflects the structural dependency among the parameters, e.g. depending on whether or not μ_{Λ} is a product measure. For example, in the case of three parameters $\lambda = (\lambda_1, \lambda_2, \lambda_3)^{\top}$, we might have three physically independent parameters, and the natural measure is the product Lebesgue measure suitably normalized for units. However, we might have $\lambda_3 = \lambda_1 + \lambda_2$ or λ_2 and λ_3 are both functions of another physical process, in which cases the measure would not be the product Lebesgue. We assume that μ_{Λ} is absolutely continuous with respect to the Lebesgue measure and the volume V of Λ is finite.

Next, we note that the solution of the deterministic model can be expressed in terms of a likelihood function $L(q|\lambda)$ of the output q values given the input parameter values λ , where $L(q|\lambda) = \delta(q - q(\lambda))$ is the unit mass distribution at $q = q(\lambda)$.

In the forward stochastic sensitivity problem, we impose a probability density $\sigma_{\Lambda}(\lambda)$ on the parameter space Λ (normalized with respect to μ_{Λ}). Then the *Law of Total Probability* implies

$$\rho_{\mathcal{D}}(q) = \int_{\Lambda} L(q|\lambda) \sigma_{\Lambda}(\lambda) d\mu_{\Lambda}(\lambda). \quad (1.2)$$

Here, we are putting the Law to the common use of relating an unobserved probability distribution $\rho_{\mathcal{D}}(q)$ to an observed probability density $\sigma_{\Lambda}(\lambda)$. We note that (1.2) is a concrete expression of the Perron–Frobenius map between the input and output measures that is defined implicitly by the physical model.

The stochastic inverse sensitivity analysis problem is the inversion of the integral Eq. (1.2). Namely, we assume that an observed probability density, $\rho_{\mathcal{D}}(q(\lambda))$ is given on the output value $q(\lambda)$ and we seek to compute the corresponding parameter density $\sigma_{\Lambda}(\lambda)$ that yields $\rho_{\mathcal{D}}(q(\lambda))$ via (1.2).

1.2. The approximate solution of the stochastic inverse problem

We note that probability densities are not random. In Briedt et al. (2011), we develop a method for computing approximate probability densities that does not require random sampling. Our approach has two stages:

1. Construct an approximate representation of the set-valued inverse solution of the deterministic model
2. Compute the probability measure on the parameter space that corresponds to the set-valued inverse and the observed output measure using a measure theoretic approximation

These are independently interesting tasks. In addition, we derive a computable a posteriori error bound on the computed distribution that accounts for sampling and discretization errors. We describe these tasks in more detail below.

2. Set-valued inverses

In the case that the forward map q is a linear function of λ , i.e., $q(\lambda) = \gamma^\top \lambda$ for some $\gamma \in \mathbb{R}^d$, then for fixed $q \in q(\Lambda)$ there exists a hyperplane, or linear contour, in Λ such that all points λ on this hyperplane map to q . We can take the set of such hyperplanes as the set of set-valued inverses, see Fig. 2.1. Note, if q is an affine map of λ , i.e., $q(\lambda) = \lambda^\top \gamma + q_0$ for some $q_0 \in \mathbb{R}$, then we use the function above with q replaced by $q - q_0$. In the case of a smooth nonlinear map, the Implicit Function Theorem applied at a point in the domain gives the existence of local set-valued inverse curves or contours in an open neighborhood centered at the point. Applying this local result at each point on a suitably fine mesh in the domain, we can cover the domain with piecewise contours. These generalized contours define natural set-valued inverses of the map, see Fig. 2.1.

In order to describe the set of set-valued inverses, we need a way to index the different set-valued inverses. We can do this using any curve that has the property that it intersects each generalized contour once and only once. We call such a curve a transverse parameterization. We illustrate in Figs. 1.1 and 2.2. In Briedt et al. (2011), we derive a method for computing a transverse parameterization,

Theorem 2.1—If the solution of the model depends smoothly on the parameters, and the set of critical points is a set of measure zero, then there exists a transverse parameterization that can be constructed as a finite set of curves.

The immediate result is that we can solve the stochastic inverse problem uniquely in the set of generalized contours,

Theorem 2.2—If the solution of the model depends smoothly on the parameters, then for a fixed transverse parameterization in Λ , there is a unique distribution on the intersections of the generalized contours with the transverse parameterization corresponding to the probability distribution ρ_D on the output of the map.

We illustrate in Fig. 1.1. Note that once we have computed the distribution on a transverse parameterization, we have used all of the information that can be obtained from solving the model.

2.1. Approximating set-valued inverses

The previous discussion is not directly useful since determining generalized contours amounts to solving the inverse problem exactly. Instead, we devise a way to approximate the

generalized contours. As mentioned above, see Fig. 2.2, the linear case is easy since the generalized contours are hyperplanes.

In the case of a nonlinear map, we can consider the tangent plane approximation to the map in the local neighborhood of a point. We can prove that the generalized contours of the tangent plane approximate the generalized contours of the map, see Fig. 2.2. Finally to build global approximations, we can compute tangent plane approximations centered at points on a suitably fine mesh covering Λ , and then use the corresponding piecewise linear generalized contours to approximate the true generalized contours, see Fig. 2.2.

We can prove

Theorem 2.3—If the points in a sequence of meshes on Λ fill the set, the corresponding generalized linear contours converge point-wise to the generalized contours associated with a map having bounded first and second derivatives.

2.2. Obtaining derivative information from the model

Since this approach to computing an approximate solution of the stochastic inverse problem employs local tangent planes, we require derivatives of the output with respect to the parameters. Since the model defines the solution as an implicit function of the parameters, obtaining the derivative information takes some work. How the derivatives are obtained is immaterial. In Briedt et al. (2011) and Butler et al. (2012), we obtain the necessary derivative information using the adjoint operator to the (linearized) solution operator associated to the model. This is particularly convenient when the quantity of interest is low dimensional. We could also obtain the derivatives by solution of an augmented system obtained by implicit differentiation.

This approach can be written abstractly and applied to many differential equations. To make it concrete, we describe the approach for an initial value problem,

$$\begin{cases} \dot{y} = g(y; \lambda_1), & t > 0, \\ y(0) = \lambda_0, \end{cases} \quad (2.1)$$

where $y \in \mathbb{R}^n$, $g: \mathbb{R}^{n+p} \rightarrow \mathbb{R}^n$ is smooth, and $\lambda = (\lambda_1^\top, \lambda_0^\top)^\top \in \Lambda \subset \mathbb{R}^d (d=p+n)$ are the parameters. We solve (2.1) to calculate a quantity of interest,

$$q(y) = \int_0^T \langle y, \psi \rangle dt. \quad (2.2)$$

We assume that the solution y of (2.1) depends (implicitly) on parameters X in a smooth way and denote solutions of (2.1) as y_λ and the quantity of interest as $q(\lambda)$ to emphasize the implicit dependence of the quantity of interest on the parameters. The smooth dependence of solutions to (2.1) on parameters λ implies the dependence of the quantity of interest on λ is also smooth. We solve the initial value problem at a reference parameter value

$\mu = (\mu_1^\top, \mu_0^\top)^\top$ to obtain (y_μ) . The exact adjoint problem at the reference point is,

$$\begin{cases} -\phi - D_y g(y_\mu; \mu_1)^\top \phi = \psi, & T > t \geq 0, \\ \phi(T) = 0 \end{cases} \quad (2.3)$$

The following theorem (Neckels, 2005) relates the value of $q(\lambda)$ to $q(\mu)$ for λ near μ .

Theorem 2.4—If $g(y; \lambda)$ is twice continuously differentiable with respect to both y and λ and Lipschitz continuous in both y and λ , then the quantity of interest is Fréchet differentiable at (y_μ, μ) with derivative $\nabla q(\mu): \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$\nabla q(\mu)[\lambda] = \langle (\lambda_0 - \mu_0), \phi(0) \rangle + \int_0^T \langle D_{\lambda_1} g(y_\mu; \mu)(\lambda_1 - \mu_1), \phi \rangle dt. \quad (2.4)$$

Additionally, in the absence of numerical error,

$$q(\lambda) \approx q(\mu) + \nabla q(\mu)[\lambda]. \quad (2.5)$$

We use the theorem by numerically solving (2.3) for the adjoint.

We obtain a global piecewise-linear approximation on a partition $\{B_i\}_{i=1}^M$, to obtain

$$\hat{q}(\lambda) := \sum_{i=1}^M (q(\mu_i) + \langle \nabla q(\mu_i), (\lambda - \mu_i) \rangle) \mathbf{1}_{B_i}(\lambda) \quad (2.6)$$

where μ_i is the reference parameter value chosen in cell B_i .

2.3. A numerical example

We apply these ideas to the Brusselator problem (1.1). We allow λ to vary in the 3-dimensional box $\Lambda := [0.7, 1.5] \times [2.75, 3.25] \times [1, 2]$, and assume the underlying volume measure is Lebesgue. The quantity of interest takes on values in the interval $\mathcal{D} := [2.9416, 4.0851]$. To make things interesting, the values for A and B are chosen so that the fixed-point of the system (at $y_1 = A$, $y_2 = B/A$) moves from unstable when $B > 1 + A^2$ to stable when $B < 1 + A^2$. The numerical solution and error estimates are computed using GAASP (Estep et al., 2006) using a first-order discontinuous Galerkin (dG1) method for the forward solve, and a second-order continuous Galerkin (cG2) method for all of the adjoint solves. For all simulations, we set $T = 5$, $\chi_{2,0} = 1$, and we use a time-step of 0.1.

We use a $15 \times 15 \times 10$ grid of uniformly spaced reference parameters at which we solve for $q(\lambda)$ and $\nabla q(\lambda)$ to form the piece-wise tangent plane approximation to the response surface. To visualize this in the plots of Fig. 2.3, we show the variation in the quantity of interest over the (A, B) -plane for $IC = 1.05, 1.45, 1.65$, and 1.95 , the variation in the quantity of interest over the (IC, A) -plane for $B = 2.7667, 2.9, 3.0667$, and 3.233 , and the variation in the quantity of interest over the (IC, B) -plane for $A = 0.72667, 0.94, 1.2067$, and 1.4733 .

We plot a sample of cross-sections of the generalized contours in Fig. 2.4.

3. A computational-measure theoretic inverse

In Briedt et al. (2011), we present an algorithm to approximate the solution of the inverse problem (1.2) that uses a simple function $\sigma_{\Lambda, M}(\lambda)$ computed with respect to a (user chosen) partition $\{b_i\}_{i=1}^{M'}$ of Λ . Simple functions, i.e. piecewise constant functions, are the classic approximation tool in measure theory. Paraphrasing the main result from Briedt et al. (2011),

Theorem 3.1—Given a measurable set $A \subset \Lambda$, we can approximate $P(A)$ using the simple function,

$$\sigma_{\Lambda, M'}(\lambda) = \sum_{k=1}^{M'} \frac{P(b_i)}{\mu_{\Lambda}(b_i)} \mathbf{1}_{b_i}(\lambda). \quad (3.1)$$

The constructive proof yields a computational algorithm that generates a probability $P(b_i)$ for each cell b_i , using only calculations of volumes in Λ . The main steps of the algorithm are based on the following observations.

- The probability of an interval of output data $[q_m, q_M] \subset \mathcal{D}$ is equal to the probability of the region of generalized contours defined by $A = q^{-1}([q_m, q_M])$.
- If $\rho_{\mathcal{D}}(q)$ is constant on $[q_m, q_M]$, then the probability of $b \cap A$ for any event $b \subset \Lambda$ is equal to the probability of A times the ratio of volumes $\mu_{\Lambda}(b \cap A)/\mu_{\Lambda}(A)$.

In words, we determine which contours are contained in an event and the probabilities of those contours (from the output distribution) and then how much of each of those contours are contained inside the event (using the volume measure).

Thus, the algorithm proceeds by first approximating $\rho_{\mathcal{D}}(q)$ with a simple function, which induces regions of contours with probabilities defined by the approximate output density.

Then, the ratios of volumes for each cell in the partition $\{b_i\}_{i=1}^{M'}$ are computed with respect to all the induced regions of contours. From this, we obtain $P(b_i)$ for each cell and obtain (3.1).

3.1. Numerical example

We continue with the Brusselator problem (1.1) and Section 2.3. We assume that the quantity of interest data has a normal distribution with mean 3.8497 and variance 0.0531. To include some effect of finite sampling in our error estimates, we suppose that 10^6 samples of this distribution are used to approximate the density function by binning the samples in 100 equally spaced intervals on $\mathcal{D}$.

We approximate the solution to the inverse probability measure over a $30 \times 30 \times 20$ partition of fine cells $\{b_i\}$ of Λ using the algorithm of Briedt et al. (2011). In Fig. 3.1, we show the probabilities of the fine cells in Λ for various fixed values of IC , B , and A , respectively.

In Fig. 3.2, we show an approximation to the three-dimensional probability measure using a scatter plot where the mid-point of each cell b_i is plotted with a color-value associated with $P(b_i)$, and to indicate the region of highest probability more clearly, we only plot the mid-point of a cell b_i if $P(b_i) > 10^{-3}$.

4. A posteriori error analysis of the computed inverse measure

The main focus of the analysis in Briedt et al. (2011) is on the convergence of the approximate representation to the true representation on a given partition $\{b_i\}_{i=1}^{M'}$ assuming that the map is evaluated exactly. In Butler et al. (2012), we analyze and estimate errors affecting the values $\{P(b_i)\}$ in the representation (3.1) for a fixed partition $\{b_i\}_{i=1}^{M'}$.

In particular, we consider two sources of error that affect the approximation of the representation $\sigma_{\Lambda, M'}(\lambda)$. The first is “statistical error” that arises if the observed probability density $\rho_{\mathcal{D}}(q(\lambda))$ is known only through a finite collection of random samples. This type of

error affects the left-hand side of (1.2). The second source of error arises when we use numerical approximations in the evaluation of the map q , e.g. as happens if q involves solving a differential equation. This means that we use approximate values of q and its gradient to form the approximate representation $q(\lambda) \approx \tilde{q}(\lambda)$. This source of error affects the evaluation of the likelihood function in (1.2).

4.1. General error analysis of distributions

If $F(t)$, $t \in \mathbb{R}^d$, denotes the cumulative distribution function (CDF) on Λ that represents (3.1), then

$$F(t) = P(\{\lambda | \lambda \leq t\}) = P(\lambda \leq t) = \sum_{i=1}^{M'} F_q(q(b_i)) \frac{\int_{b_i \cap \{\lambda \leq t\}} d\mu_\Lambda(\lambda)}{\int_{A_j} d\mu_\Lambda(\lambda)}, \quad (4.1)$$

where $F_q(t)$ denotes the CDF of $q(\lambda)$. This follows from the observation that

$$P(b_i) = F_q(q(b_i)) \frac{\int_{b_i} d\mu_\Lambda(\lambda)}{\int_{A_j} d\mu_\Lambda(\lambda)}, \quad 1 \leq i \leq M', \quad (4.2)$$

where $q(b_i) = \{q(\lambda), \lambda \in b_i\}$. (To simplify the presentation, we assume that b_i is contained in a region of contours A_j induced by the simple function approximation to $\rho_D(q)$, i.e. $A_j = q^{-1}(I_j)$ for some interval $I_j \subset \mathcal{D}$. We can modify the results if this does not hold.)

We consider the use of an approximation $q(\lambda) \approx \tilde{q}(\lambda)$, so that we obtain sample values $\{\tilde{q}(b_i)\}_{i=1}^{M'}$ and we approximate $F_q(t)$ using $\mathcal{N}$ sample values of $\tilde{q}$. The approximate sample CDF $F_{\mathcal{N}}(t)$ is

$$\tilde{F}_{\mathcal{N}}(t) = \sum_{i=1}^{M'} F_{\mathcal{N}}(\tilde{q}(b_i)) \frac{\int_{b_i \cap \{\lambda \leq t\}} d\mu_\Lambda(\lambda)}{\int_{A_j} d\mu_\Lambda(\lambda)}, \quad (4.3)$$

where $F_{\mathcal{N}}(t)$ denotes the sample CDF computed from a finite collection sample values $\{Q_1; \dots; Q_{\mathcal{N}}\}$,

$$F_{\mathcal{N}}(t) = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \mathbf{1}(Q_n \leq t).$$

We calculate probabilities using (4.3) and seek to determine the error $F(t) - \tilde{F}_{\mathcal{N}}(t)$. In Butler et al. (2012), show that,

Theorem 4.1—For any $\varepsilon > 0$,

$$|F(t) - \tilde{F}_{\mathcal{N}}(t)| \leq 2 \left(\frac{\log(\varepsilon^{-1})}{2\mathcal{N}} \right)^{1/2} + L \max_{1 \leq i \leq M'} (\max q(b_i) - \min q(b_i)) \times E \quad (4.4)$$

$$= \varepsilon_S + \varepsilon_D,$$

with probability greater than $1 - \epsilon$. Here, E is the bound on numerical error present in the computed response surface $q(\lambda)$ at any point $\lambda \in \Lambda$, and L is the Lipschitz constant for $F(t)$.

The first term ϵ_S in the bound (4.4) quantifies the effects of using finite samples to approximate F_q while the second term ϵ_D quantifies the effects of errors in the evaluation of q . If no sampling is used to evaluate $\rho_D(q)$ or $F_q(t)$, then (4.4) reduces to

$$|F(t) - \tilde{F}(t)| \leq L \max_{1 \leq i \leq M'} (\max q(b_i) - \min q(b_i)) \times E, \quad (4.5)$$

where we have dropped the subscript $\mathcal{N}$ on F , that is $\tilde{F}(t)$ is the CDF calculated using exact values of $\rho_D(q)$, but approximate values of q . The properties of the bound ϵ_S are explored in Estep et al. (2009a).

In Estep et al. (2009a), when estimating ϵ_D , the a posteriori analysis used to estimate the bound for E only considers the effect of numerical errors in computing the value of the quantity of interest and its derivatives on a fixed partition of the parameter space as described below. Here, we emphasize that we may also include an estimate for the linearization error in using a piecewise linear approximation to the map $q(\lambda)$ on computing $\tilde{F}(t)$. Thus, when E is a bound for the numerical errors in the quantity of interest and derivatives, the bound above describe a bound in the error of the computed CDF compared to the CDF computed using the exact *piecewise-linear* map. Including a bound on the error in the linearization of the map in the E term means that the bound above describes a bound in the error of the computed CDF compared to the CDF computed using the exact *nonlinear* map. Below, we show the effect of the linearization error on the CDF error bound and demonstrate the effectiveness of this CDF error bound.

4.2. An a posteriori estimate for E

Using the bound (4.4) requires an estimate E on the error in the sample values of $q(b_i)$. To obtain this estimate in Butler et al. (2012), we adapt the a posteriori error analysis techniques based on adjoint problems and computable residuals widely used for finite elements (Estep, 1995; Eriksson et al., 1995, 1996; Estep et al., 2000). This has to be extended to treat the inverse problem because we require estimates in both the values of q and ∇q . Of course, derivatives tend to be sensitive to numerical error.

We first summarize the well known result for the errors in values q . To the differential Eq. (2.1), we use the (standard) adjoint problem for estimating the error in approximation $Y \approx y$,

$$\begin{cases} -\dot{\vartheta} = g'(\overline{y}, Y)^\top \phi + \psi(t), & T > t \geq 0, \\ \vartheta(T) = 0, \end{cases} \quad (4.6)$$

where $\overline{g'}(y, Y) = \int_0^1 g'(sy + (1-s)Y) ds$. The standard a posteriori error estimate has the form.

Theorem 4.2

$$\int_0^T (y - Y, \psi) dt = (y - Y(0), \vartheta(0)) + \sum_{n=2}^N ([Y]_{n-1}, \vartheta_{n-1}) - \sum_{n=1}^N \int_{t_{n-1}}^{t_n} (\dot{Y} - g(Y), \vartheta) dt, \quad (4.7)$$

Where $[Y]_{n-1} = Y(t_{n-1}^+) - Y(t_{n-1}^-)$.

The result applies to a large class of standard difference schemes written as either continuous or discontinuous finite element methods, see for example (Estep, 1995; Eriksson et al., 1995, 1996; Estep et al., 2000). A computable estimate of the error in a quantity of interest of (2.1) is obtained by solving (4.6) and computing (4.7). In practice, we typically replace $\overline{g'(y, Y)}$ with $g'(Y)$.

In order to obtain the necessary estimate for E , we must also estimate the error in ∇q . This is considerably more complicated than the situation for q . In Butler et al. (2012), we show how to compute an estimate by solving a set of adjoint problems with different quantities of interest in order to build an estimate for the gradient. We do not state the result here because the notation is very cumbersome. The end result is a complicated version of Theorem 4.2 that provides an estimate in the error of ∇q in each cell b_i . We then maximize the estimates over the cells to obtain the desired bound on E .

4.3. A numerical example

Returning to the Brusselator example in Sections 2.3 and 3.1, the computed $q(\lambda)$ gives $\max_i |M(\max q(b_i) - \min q(b_i))| = 1.63 \times 10^{-1}$. We use the local error estimates on all partitions to bound E globally by 1.97×10^{-4} as described above. The variance of the normal distribution gives a Lipschitz constant on the output distribution bounded by 1.75. Thus, we have, using (4.4) with $\epsilon = 0.05$, that

$$|F(t) - \tilde{F}_{\mathcal{N}}(t)| \leq 1.79 \times 10^{-3},$$

with probability 95%.

Statistical bounds tend to inflate the error bound, so we now verify the deterministic component of the error bound, which is approximately 5.6194×10^{-5} . To focus on the deterministic component, we treat the finite sample distribution of the quantity of interest as the exact distribution, so the error estimate reduces to (4.5).

We then invert through a more accurate approximation of the map $q(\lambda)$ obtained by solving (1.1) using a time step of 0.025. This is one-fourth that of the time step used in the earlier approximations and it results in numerical error estimates that are decreased by, in general, one to two orders of magnitude over all the coarse cells partitioning Λ . We use the same partition of fine cells to approximate the inverse density, and compute the difference in the CDFs at the upper-right corner of every 10th of the fine cells partitioning Λ numbered in a serpentine order starting at the bottom left corner. In Fig. 4.1, we show the histogram of the absolute value of the differences in the CDFs at these 1800 points in Λ . All values are within the error bound.

We now consider the effect of linearization error in the numerically approximated piecewise-linear map. We empirically estimate 0.0105 as a bound for the linearization error, which implies that the bound on the error in the computed CDF from the exact CDF for the nonlinear map is approximately $3.1E-03$. In order to test the effectiveness of this bound, we compute an approximation to $q(\lambda)$ on a refined partition of Λ with the smaller time step of 0.025 used above to reduce numerical and linearization errors by orders of magnitude. Specifically, every cell used to compute the original piecewise linear approximation is divided into eight cells to generate a refined piecewise linear approximation to the map that is computed using the more accurate time mesh, resulting in an approximation with numerical and linearization errors orders of magnitude less than the bound. In Fig. 4.2, we graphically demonstrate the effectiveness of this bound on computations of the error

between the CDFs obtained using the coarser piecewise-linear map and the refined approximation to the map at the upper-right corner of each of the 18,000 cells used to approximate the inverse probability measure.

Acknowledgments

T. Butler's work is supported in part by the Department of Energy (DE-FG02-05ER25699) and the National Science Foundation (DGE-0221595003, MSPA-CSE-0434354). D. Estep's work is supported in part by the Defense Threat Reduction Agency (HDTRA1-09-1-0036), Department of Energy (DE-FG02-04ER 25620, DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC00017 24, DE-SC0005304, INL00120133), Idaho National Laboratory (00069249, 00115474), Lawrence Livermore National Laboratory (B573139, B584647, B590495), National Aeronautics and Space Administration (NNG04GH63G), National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1065046, DMS-1016268, DMS-FRG-1065046), National Institutes of Health (#R01GM096192).

References

- Briedt J, Butler T, Estep D. A computational measure theoretic approach to inverse sensitivity problems I: basic method and analysis. *SIAM J Numer Anal.* 2011; 49:1836–1859. [PubMed: 23637467]
- Butler T, Estep D, Sandelin J. A measure-theoretic computational method for inverse sensitivity problems II: a posteriori error analysis. *SIAM J Numer Anal.* 2012; 50:22–45. [PubMed: 23667271]
- Eriksson, K.; Estep, D.; Hansbo, P.; Johnson, C. *Acta Numer.* Cambridge Univ. Press; Cambridge: 1995. Introduction to adaptive methods for differential equations; p. 105–158.
- Eriksson, K.; Estep, D.; Hansbo, P.; Johnson, C. *Computational Differential Equations.* Cambridge University Press; Cambridge: 1996.
- Estep D. A posteriori error bounds and global error control for approximation of ordinary differential equations. *SIAM J Numer Anal.* 1995; 32(1):1–48.
- Estep D, Larson MG, Williams RD. Estimating the error of numerical solutions of systems of reaction–diffusion equations. *Mem Am Math Soc.* 2000; 146(696):viii–109.
- Estep D, Iqvist AM, Tavener S. Nonparametric density estimation for randomly perturbed elliptic problems I: computational methods, a posteriori analysis, and adaptive error control. *SISC.* 2009a; 31:2935–2959.
- Estep D, Mckeown B, Neckels D, Sandelin J. GAASP: Globally Accurate Adaptive Sensitivity Package. 2006 estep@math.colostate.edu.
- Neckels, D. PhD Thesis. Department of Mathematics, Colorado State University; Fort Collins, CO: 2005. Variational Methods for Uncertainty Quantification. 80523
- Prigogine I, Lefever R. Symmetry breaking instabilities in dissipative systems. *J Chem Phys.* 1968; 48(4):1695–1700.

Further reading

- Butler, T. PhD Thesis. Department of Mathematics, Colorado State University; Fort Collins, CO 80523: 2009. Computational Measure Theoretic Approach to Inverse Sensitivity Analysis: Methods and Analysis.
- Estep D, Holst MJ, Målqvist A. Nonparametric density estimation for randomly perturbed elliptic problems III: convergence, complexity, and generalizations. *J Appl Math Comput.* 2012; 38:367–387.
- Estep D, Iqvist AM, Tavener S. Nonparametric density estimation for randomly perturbed elliptic problems II: applications and adaptive modeling. *J Numer Methods Eng.* 2009b; 80:846–867.
- Estep D, Neckels D. Fast and reliable methods for determining the evolution of uncertain parameters in differential equations. *J Comput Phys.* 2006; 213:530–556.
- Estep D, Neckels D. Fast methods for determining the evolution of uncertain parameters in reaction–diffusion equations. *Comput Methods Appl Mech Eng.* 2007; 196:3967–3979.

Fig. 1.1.

Left: an example of a set-valued inverse. In the case of a smooth two dimensional map, the set-valued inverse is a contour curve. Middle: a probability measure on the output corresponds to a unique probability measure on the set of set-valued inverses. A straight line connecting the point of maximum value to any point of minimum value serves as a transverse parameterization. We show a sample of output values drawn from the output measure and the corresponding set-valued inverses (contours). Right: we can compute the probability of event A which is a union of set-valued inverses (contours). We cannot compute the probability of the simple event B without further information about the underlying geometry of the parameter set.

Fig. 2.1.

Left: in the case of a linear map, the set-valued inverses are hyperplanes. These exist globally over the entire domain. Right: for a nonlinear map, we can use the Implicit Function Theorem on a mesh of points to cover the domain with piecewise contours, or generalized contours.

Fig. 2.2.

Left: for a nonlinear map, we compute a tangent plane approximation near a point. The set-valued inverses of the tangent plane approximate the set-valued inverses of the surface in a sufficiently small neighborhood of the center point. Right: to obtain a set of approximate piecewise linear approximate generalized contours, we construct a piecewise tangent plane approximation. We show a transverse parameterization in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2.3.

Cross-sections of $q(\lambda)$. Top row, from left to right: Varying (A, B) with IC fixed at values 1.05, 1.45, 1.65, 1.95. Middle row, from left to right: Varying (IC, A) with B fixed at values 2.7667, 2.9, 3.0667, 3.233. Bottom row, from left to right: varying (IC, B) with A fixed at values 0.72667, 0.94, 1.2067, 1.4733.

Fig. 2.4.

Approximate cross-sections of generalized contours for the computation in Fig. 2.3. Top row, from left to right: varying (A, B) with IC fixed at values 1.05, 1.45, 1.65, 1.95. Middle row, from left to right: varying (IC, A) with B fixed at values 2.7667, 2.9, 3.0667, 3.233. Bottom row, from left to right: varying (IC, B) with A fixed at values 0.72667, 0.94, 1.2067, 1.4733.

Fig. 3.1.

Plots of $P(b_i)$ for cells b_i in $\mathbf{A}$. Top row, from left to right: intersections with $IC = 1.025, 1.375, 1.725, 1.975$. Middle row, from left to right: Intersections with $B = 2.9083, 3.0417, 3.1417, 3.2417$. Bottom row, from left to right: Intersections with $A = 0.9, 1.06, 1.22, 1.4067$.

Fig. 3.2.
Plot of $P(bi)$ for cells bi in Λ with $P(bi) > 0.0001$.

Fig. 4.1.

Histogram of the absolute value of the differences in the CDFs at 1800 points chosen as the upper-right corner of every 10th cell of the 18,000 fine box partition of Λ . A more accurate numerical map was used as a surrogate for the exact piecewise-linear map to test the deterministic error bound prediction of 5.6194×10^{-5} .

Fig. 4.2.

Left: histogram of the absolute value of the differences in the CDFs of the coarse and refined piecewise linear approximations at the upper-right corner of every one of the 18,000 fine boxes that partition Λ . Right: The points in Λ where the error exceeds $1E-03$ in magnitude, which all occur to the right and above of the region of high probability shown in Fig. 3.2. A more accurate numerical map on a refined partition of Λ was used as a surrogate for the exact nonlinear map to test the deterministic error bound prediction of $3.1E-03$.