Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 213 (2006) 530–556
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Fast and reliable methods for determining the evolution of uncertain parameters in differential equations

D. Estep *, D. Neckels

Department of Mathematics, Colorado State University, 1874 Campus Delivery, Fort Collins, CO 80523, United States

Received 25 February 2005; received in revised form 31 July 2005; accepted 22 August 2005
Available online 8 November 2005

Abstract

A very common problem in science and engineering is the determination of the effects of uncertainty or variation in parameters and data on the output of a deterministic nonlinear operator. For example, such variations may describe the effect of experimental error or may arise as part of a sensitivity analysis of the model. The Monte-Carlo Method is a justifiably popular tool for analyzing such effects. It does, however, require many evaluations of the operator and it is difficult to extract precise information about the accuracy of any particular result. In this paper, we borrow techniques from a posteriori error analysis for finite element methods to compute information about the derivative of an operator with respect to its parameters. These techniques employ the generalized Green's function to describe how variation propagates into the solution around localized points in the parameter space. We show how this derivative information can be used either to create a higher order method or produce an error estimate for information computed from a given representation. In the latter case, this provides the basis for adaptive sampling according to the variation in the *output* values. Both the higher order method and the adaptive sampling method are generally orders of magnitude faster than Monte-Carlo methods in the case that the parameter space is not too high dimensional.

Keywords: A posteriori error estimation; Adaptive sampling; Adaptive error control; Finite element method; Generalized Green's function; Monte-Carlo method; Parameter error; Parameter variation; Reliable sampling; Stochastic system; Sensitivity analysis; Uncertainty quantification; Variational analysis

1. Introduction

In this paper, we study an important and ubiquitous problem in science and engineering that can be described abstractly like this: suppose that a deterministic nonlinear operator $\mathscr{F}$ maps a space of inputs, consisting of data and parameters, into a space of outputs so that to each fixed input state, $\mathscr{F}$ associates a unique output state. In our particular case, $\mathscr{F}$ is the solution operator for a differential equation. We consider the

* Corresponding author. Tel.: +1 970 491 6722; fax: +1 970 491 2161.
E-mail addresses: estep@math.colostate.edu (D. Estep), dcnecke@sandia.gov (D. Neckels).

0021-9991/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2005.08.024

situation in which *functions*, parameterized by some variable, are substituted for the data and parameters in $\mathscr{F}$, rather than individual points. The problem is to describe the resulting output function.

One frame for this problem is uncertainty analysis. In this setting, some of the data and/or the parameters for $\mathscr{F}$ are unknown within a given range and/or subject to random variation, e.g. as a result of experimental or modeling error. The problem is to determine the effect of the uncertainty or variation on the output of the operator. It is natural to weight different scenarios with different probabilities and to consider the input to be a random vector associated with some probability distribution. The output then becomes a random vector associated with a *new* distribution.

To illustrate, we plot the output distribution of the nonlinear function of one parameter $q(\lambda) = \tanh(2\lambda)/\tanh(2)$ for several distributions of the parameter λ on $[-1,1]$ along with q in Fig. 1. The function q acts in a nearly linear fashion on the "narrow" normal distributions with small variance, so the resulting distributions are again approximately normal. For "wide" distributions with large variance, the nonlinear behavior of q leads to increased concentration of values at the boundaries.

Another frame for this problem is sensitivity analysis, in which the goal is to quantify the sensitivity of the output information computed from a model with respect to variations in the input parameters. As well as revealing fundamental stability properties of a model, sensitivity analysis also has important applications. For example, it can be used to estimate the necessary experimental accuracy. Unfortunately, it is usually very difficult to obtain accurate a priori bounds on the effects of variations in parameters. For example, a Gronwall argument is a standard tool for evolutionary problems and it yields bounds that grow exponentially in time regardless of the true behavior.

The Monte-Carlo method is a deservedly popular tool for determining the effects of variation in parameters and data on the output of a nonlinear function. In this approach, a sample of the input space is selected at random according to its distribution, and the model is solved for these parameter values. The resulting collection of output values form a pointwise sample of the output of the function and are used to approximate any desired information, such as the distribution. The Monte-Carlo method converges robustly, is easy to implement, and is relatively immune to the "curse of dimensionality". It is used throughout science and engineering, and it is no exaggeration to state that a good portion of the world's computing resources for science and engineering are used for Monte-Carlo computations.

On the other hand, the Monte-Carlo method generally converges slowly in the sense of requiring a large number of simulations in order to achieve good accuracy. For example, consider the Monte-Carlo approximation $\sum_{i=1}^{n} g(\lambda_i)/n$ to the average value $\int_{\Lambda} g(\lambda)\,\mathrm{d}\lambda/\mathrm{Vol}(\Lambda)$ of a function g over a domain Λ, where $\{\lambda_i\}_{i=1}^{n}$ is a set of points chosen independently at random from Λ. The error is

$$\epsilon(n) = \frac{1}{n}\sum_{i=1}^{n} g(\lambda_i) - \frac{1}{\mathrm{Vol}(\Lambda)}\int_{\Lambda} g(\lambda)\,\mathrm{d}\lambda,$$

which has variance given by

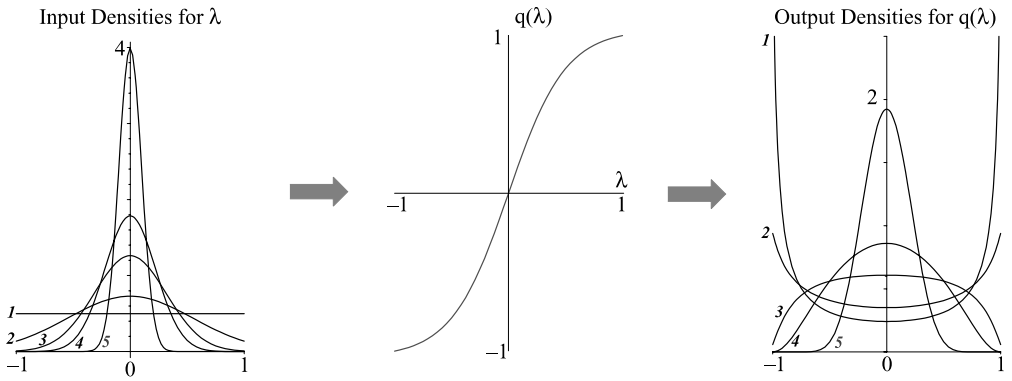


Fig. 1. On the left, we plot normal input distributions for λ ($N(0,\sigma^2)$, $\sigma^2 = \{0.01,0.05,0.1,0.3\}$) as well as the uniform distribution. We plot the function q in the middle. On the right, we plot the output distributions for $q(\lambda)$.

$$\sigma^2 = \frac{1}{\mathrm{Vol}(\Lambda)^2}\int_\Lambda g(\lambda)^2\,\mathrm{d}\lambda - \left(\frac{1}{\mathrm{Vol}(\Lambda)}\int_\Lambda g(\lambda)\,\mathrm{d}\lambda\right)^2.$$

The Law of the Iterated Logarithm implies that with probability 1, for any $C > 1$, all but a finite number of the errors $\epsilon(n)$ are bounded above and below by

$$\pm C\frac{\sigma\sqrt{2\log\log n}}{\sqrt{n}}.$$

To illustrate, we compute the average value of $\lambda_1^2 + \lambda_2^2$ over $[-1,1]\times[-1,1]$ and plot the results in Fig. 2. When the operator $\mathscr{F}$ is expensive or difficult to evaluate, the Monte-Carlo method is expensive to the point of being impractical.

Also, it is difficult to determine how to sample the input space in order to efficiently represent the output distribution when $\mathscr{F}$ is a complicated, nonlinear operator. Simply sampling according to the distribution of the input may not suffice. For example, accurately representing the output corresponding to a *uniform* input distribution efficiently in the example in Fig. 2 requires increased sampling near the boundaries.

In this paper, we present a different approach for ascertaining the effects of variations in parameters and data on a model in the case that the model depends smoothly and deterministically on the inputs. The focus on *deterministic* models, which may be affected by random processes through parameters and data, is worth emphasizing. Such problems have distinct features as compared to models described by stochastic differential equations that can be exploited profitably.

Our approach is based on techniques borrowed from a posteriori error analysis for finite element methods. The goal of a posteriori error analysis is to provide an accurate error estimate of an approximate solution using information obtained from the numerical solution as much as possible (see [1,2]). One avenue [3–5] to a posteriori error analysis uses variational arguments involving the generalized Green's function. The generalized Green's function solves the (linearized) adjoint problem with data specific to the information to be computed from the solution of the original problem and describes how local variation in the model, parameters, and data propagates into the solution. We apply this a posteriori analysis to the problem of determining the effects of variations in parameters and data on a model and show how information obtained from the generalized Green's function can be used either to create a higher order method or produce an error estimate for a given representation. In the latter case, this provides the basis for adaptive sampling. Both the higher order method and the adaptive sampling methods appear to be orders of magnitude faster than Monte-Carlo methods when the dimension of the parameter space is not too large. In the case of high dimensional parameter space, we can use the information with standard techniques to enhance the Monte-Carlo method.

The adjoint problem and generalized Green's function are a powerful tools for the study of differential equations. Our approach to a posteriori analysis is related to applications of the adjoint problem to the analysis of equation sensitivity, optimization, and data assimilation.

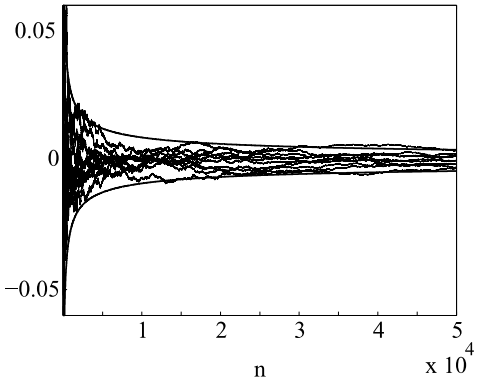


Fig. 2. Using the Monte-Carlo method to approximate the average value of $\lambda_1^2 + \lambda_2^2$ over $[-1,1]\times[-1,1]$. We plot values of the error $\epsilon(n)$ as a function of increasing n for 10 different sequences, along with the bounds given by the Law of the Iterated Logarithm.

The paper is organized like this. We first introduce the new approach in a simple finite-dimensional example, where the ideas are relatively easy to describe. We then extend the approach to ordinary differential equations and discuss its approximation properties. We explain how the approach can be used to create a higher order method or to produce an error estimate for information computed from a representation. The estimate also provides the basis for adaptive sampling. Finally, we present applications to a control problem for the nonlinear Schaefer model of a fish population, the chaotic Lorenz system, and an SIR model of disease with birth/death processes.

1.1. Basic concepts in probability

We use a few useful concepts from probability. We consider the uncertain input $\boldsymbol{\lambda}$ to be random vector on a probability space $(\Omega, \mathscr{B}, P)$, where Ω is a set of outcomes, $\mathscr{B}$ a σ-field of sets in Ω, and P is a probability measure on $\mathscr{B}$. A *random vector* or *variable* $\boldsymbol{X}$ is a measurable function on this probability space with range contained in $\mathbb{R}^d$ or $\mathbb{R}^1$. To each random vector $\boldsymbol{X}$, there is a *distribution measure* $\mu_{\boldsymbol{X}}$ on the Borel subsets of $\mathbb{R}^d$ defined by $\mu_{\boldsymbol{X}}(A) = P(\boldsymbol{X} \in A)$, which gives the probability that $\boldsymbol{X}$ takes values in A. In the case that $\mu_{\mathrm{L}}(A) = 0 \Rightarrow \mu_{\boldsymbol{X}}(A) = 0$, where μ_{L} is the standard Lebesgue measure on $\mathbb{R}^d$, then $\mu_{\boldsymbol{X}}$ is *absolutely continuous* with respect to μ_{L} and there exists a function $\rho : \mathbb{R}^d \to \mathbb{R}$ such that $\mu_{\boldsymbol{X}}(A) = \int_A \rho \, \mathrm{d}\mu_{\mathrm{L}}$. This function is called the *probability density* of $\boldsymbol{X}$.

An example of a density is the *uniform distribution* associated with a set A with positive, finite Lebesgue measure, which has density $\rho(\boldsymbol{s}) = 1/\mu_{\mathrm{L}}(A)$ for all $\boldsymbol{s} \in A$ and is zero otherwise. Another example is the *multivariate normal distribution* with density

$$\rho(\boldsymbol{s}) = \frac{1}{(2\pi)^{p/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu})^{\mathsf{T}} \Sigma^{-1} (\boldsymbol{s}-\boldsymbol{\mu})\right\},$$

where Σ is a symmetric positive-definite matrix and $\boldsymbol{\mu}$ is the mean.

The function

$$F_{\boldsymbol{X}}(\boldsymbol{s}) = P(\boldsymbol{X}_1 \leqslant \boldsymbol{s}_1, \ldots, \boldsymbol{X}_d \leqslant \boldsymbol{s}_d)$$

is called the *cumulative distribution function* of $\boldsymbol{X}$. A random vector always has a cumulative distribution function. In general, $\boldsymbol{X}$ has a density if $F_{\boldsymbol{X}}$ is differentiable, in which case $\rho(\boldsymbol{s}) = \partial_{\boldsymbol{s}_1} \cdots \partial_{\boldsymbol{s}_d} F_{\boldsymbol{X}}$.

Finding the *distribution* of a random vector $\boldsymbol{X}$ means determining the probabilities of $\boldsymbol{X}$ taking various sets of outcomes. We can determine the distribution by computing the distribution measure, the density (if it exists), or the cumulative distribution function of $\boldsymbol{X}$. For numerical reasons, we approximate cumulative distribution functions. For visual reasons, we usually compare densities using plots.

2. A new approach: a finite-dimensional example

We first consider the problem of solving the finite-dimensional nonlinear system of equations,

$$\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{\lambda}) = \boldsymbol{b}, \tag{1}$$

for $\boldsymbol{x} \in \mathbb{R}^n$, where the parameter $\boldsymbol{\lambda}$ is a random vector in $\mathbb{R}^d$ and $\boldsymbol{f} : \mathbb{R}^{n+d} \to \mathbb{R}^n$ is smooth in both variables. In practice, the motivation for solving a nonlinear equation is often to compute a specific piece of information, or a *quantity of interest*, involving the solution. In many cases, this information can be represented as a linear functional of the solution, which is a relatively low-dimensional piece of information that is easier to compute accurately than the entire solution. By the Riesz representation theorem, there is a vector $\boldsymbol{\psi} \in \mathbb{R}^n$ such that $\langle \boldsymbol{x}, \boldsymbol{\psi} \rangle$ yields the quantity of interest, where $\langle , \rangle$ is the Euclidean inner product. For example, $\boldsymbol{\psi} = (1,0,\ldots,0)^{\mathsf{T}}$ yields the first component while $\boldsymbol{\psi} = (1,1,\ldots,1)^{\mathsf{T}}/n$ yields the average of the components.

We begin by assuming that $\boldsymbol{\lambda}$ is distributed closely about some reference value $\boldsymbol{\mu}$ (an assumption that is relaxed later), and we solve the deterministic problem

$$\boldsymbol{f}(\boldsymbol{y}; \boldsymbol{\mu}) = \boldsymbol{b}$$

for $\boldsymbol{y}$. To estimate the quantity of interest corresponding to $\boldsymbol{\psi}$, we use the *generalized Green's vector* $\boldsymbol{\phi}$ that solves the adjoint to the linearized problem,

$$A^{\mathrm{T}}\boldsymbol{\phi} = \boldsymbol{\psi},$$

where $A = D_x\boldsymbol{f}(\boldsymbol{y},\boldsymbol{\mu})$. If, for example, $\boldsymbol{\psi} = (1,0,\ldots,0)^{\mathrm{T}}$, then $\boldsymbol{\phi}$ is directly analogous to a Green's function in differential equations. The standard variational argument yields

$$\langle \boldsymbol{x}, \boldsymbol{\psi}\rangle = \langle \boldsymbol{x}, A^{\mathrm{T}}\boldsymbol{\phi}\rangle = \langle A\boldsymbol{x}, \boldsymbol{\phi}\rangle,$$

which is the analog of the standard representation formula for the Green's function. Computing a Taylor expansion of $\boldsymbol{f}$ around $(\boldsymbol{y},\boldsymbol{\mu})$, we obtain

$$\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\lambda}) = \boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}) + D_x\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{y}) + D_\lambda\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu}) + \boldsymbol{R},$$

where $\boldsymbol{R}$ is a high order remainder. From this, we see that

$$\langle A\boldsymbol{x}\ , \boldsymbol{\phi}\rangle = \langle A\boldsymbol{y}\ , \boldsymbol{\phi}\rangle - \langle D_\lambda\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu}), \boldsymbol{\phi}\rangle - \langle \boldsymbol{R}, \boldsymbol{\phi}\rangle$$

and so

$$\langle \boldsymbol{x}, \boldsymbol{\psi}\rangle = \langle \boldsymbol{y}, \boldsymbol{\psi}\rangle - \langle D_\lambda\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu}), \boldsymbol{\phi}\rangle - \langle \boldsymbol{R}, \boldsymbol{\phi}\rangle.$$

Neglecting the remainder term, we obtain an approximation for the quantity of interest corresponding to $\boldsymbol{\lambda}$ in terms of the quantity of interest at the reference value $\boldsymbol{\mu}$ plus an expression involving the change in $\boldsymbol{f}$ due to the parameter variation, and the generalized Green's vector $\boldsymbol{\phi}$,

$$\langle \boldsymbol{x}, \boldsymbol{\psi}\rangle \approx \langle \boldsymbol{y}, \boldsymbol{\psi}\rangle - \langle D_\lambda\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu}), \boldsymbol{\phi}\rangle. \tag{2}$$

Taking the inner product with the generalized Green's vector $\boldsymbol{\phi}$, which we can view as a discrete convolution, translates the variations in the model into variation in the solution.

If we think of $\boldsymbol{\lambda}-\boldsymbol{\mu}$ as a random vector with a distribution, the expression

$$-\langle D_\lambda\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu})(\boldsymbol{\lambda}-\boldsymbol{\mu}), \boldsymbol{\phi}\rangle$$

yields a random variable with a new distribution. This new random variable is a linear approximation to the true random variable near the reference value. We call using (2) in this way the *Higher Order Parameter Sampling* Method or *HOPS*.

As a first application, we consider:

$$\begin{cases} \lambda_1 x_1^2 + x_2^2 = 1, \\ x_1^2 - \lambda_2 x_2^2 = 1, \end{cases}$$

where $\lambda_1,\lambda_2 > 0$ are the parameters. Solutions $\boldsymbol{x} = (x_1,x_2)$ are intersections of the hyperbola and the ellipse. We concentrate on the solution in the first quadrant. We take λ_1 and λ_2 to be independent with λ_1 uniformly distributed on $\mu_1 \pm 0.1$ and λ_2 normally distributed $N(\mu_2,0.1)$ for some fixed (μ_1,μ_2). We choose the dual data $\boldsymbol{\psi} = (0,1)^{\mathrm{T}}$ so that $\langle \boldsymbol{x},\boldsymbol{\psi}\rangle = x_2$.

We first use two Monte-Carlo computations by solving the system for $n = 1000$ and $n = 10{,}000$ points drawn from the distribution of (λ_1,λ_2). Next, we use HOPS by calculating $\boldsymbol{y} = (\tilde{y}_1,\tilde{y}_2)$ at the mean value $(\mu_1,\mu_2) = (0.5,1)$, numerically solving for the generalized Green's vector, and approximating (2). To compare the results, we compute the value of the HOPS approximation at the 10,000 points used in the Monte-Carlo computation and plot the resulting histograms in Fig. 3. (See Sections 10.1 and 4.1 for details on producing the plots of densities and comparing accuracy of approximate density functions.) The savings in computational effort are extreme since HOPS requires only one solution of the nonlinear system, one solution to the linear adjoint problem, and then a vector dot product for each evaluation of the HOPS model. The Monte-Carlo approach, on the other hand, requires the solution to the nonlinear system for 10,000 points.

In general, linearization around a single point is insufficient to describe the response of the system to variations throughout the parameter space. In this example, near $(\mu_1,\mu_2) = (0.89,1)$ the solution is more sensitive to variations in the parameter. We can see the effect in the degraded accuracy of a HOPS approximation at the reference value, see Fig. 3. A small Monte-Carlo computation with 1000 simulations appears more accurate than the one point linear approximation. In Fig. 4, we plot the norm of the generalized Green's vector against values of λ_1. The increase in the norm as λ_1 approaches 1 indicates the increase in sensitivity of the solution. To

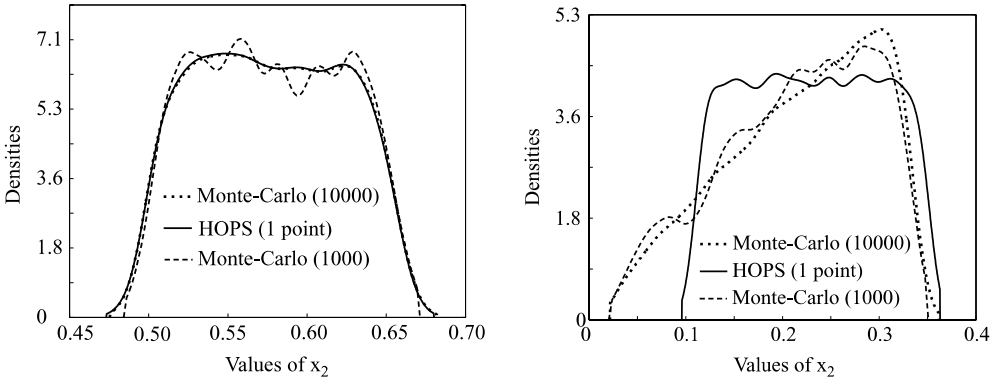


Fig. 3. Comparisons of estimated densities for Monte-Carlo computations using 1000 and 10,000 points along with HOPS one point linear approximations. On the left, $\mu_1 = 0.5$, $\mu_2 = 1$, and on the right, $\mu_1 = 0.89$, $\mu_2 = 1$.

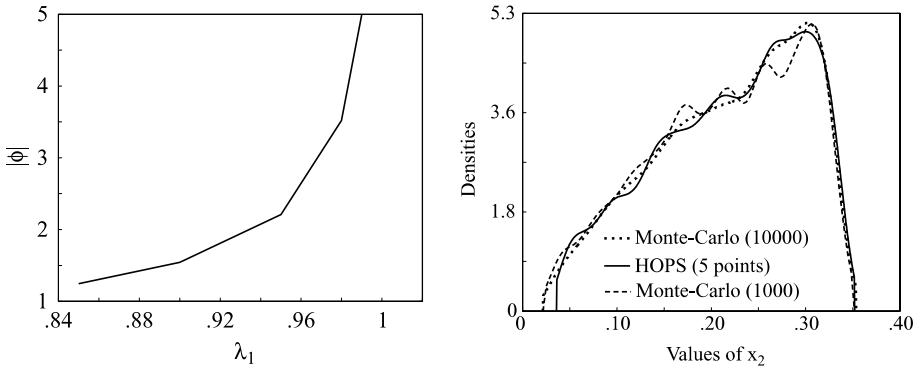


Fig. 4. On the right, we plot $|\phi|$ for λ_1 near 1. On the left, we plot approximate densities for Monte-Carlo computations using 1000 and 10,000 points and a HOPS approximation using five points for $\mu_1 = 0.89$, $\mu_2 = 1$.

address this, we compute HOPS approximations at five points chosen uniformly in the parameter set and combine the results to form a piecewise linear approximation (which we also call a HOPS approximation). The partition for the approximation is taken to be uniformly sized sub-intervals centered at the sample points. The HOPS approximation requires only 10 solutions to yield an approximation with accuracy comparable to the 1000 point Monte-Carlo approximation, see Fig. 4.

3. The HOPS method for an ordinary differential equation

We consider the problem of determining the effects of variations in parameters and data on a quantity of interest computed from the solution of the initial value problem:

$$\begin{cases} \dot{\boldsymbol{x}}(t;\boldsymbol{\lambda}) = \boldsymbol{f}(\boldsymbol{x}(t;\boldsymbol{\lambda});\boldsymbol{\lambda}_1), \quad t > 0, \\ \boldsymbol{x}(0;\boldsymbol{\lambda}) = \boldsymbol{\lambda}_0, \end{cases} \tag{3}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{f}: \mathbb{R}^{n+p} \to \mathbb{R}^n$. The parameter $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_0)^{\mathsf{T}}$ is in $\mathbb{R}^d$ with $d = p + n$, where $\boldsymbol{\lambda}_1 \in \mathbb{R}^p$ represents parameters in the model $\boldsymbol{f}$ and $\boldsymbol{\lambda}_0 \in \mathbb{R}^n$ represents the initial conditions (which we also consider as parameters). We consider $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\omega)$ as a random vector on a probability space $(\Omega, \mathscr{B}, P)$. Under the standard Lipschitz assumption on $\boldsymbol{f}$ (see Theorem 2), $\boldsymbol{x}(t;\boldsymbol{\lambda})$ is a random vector. The set of trajectories indexed by the possible values of ω may be viewed as a collection of random vectors $\boldsymbol{x}(t;\omega) = \boldsymbol{x}(t;\boldsymbol{\lambda}(\omega))$, indexed by t. This is a stochastic process, but this process has more structure than a general stochastic process since increments of the process are characterized deterministically by (3).

The methods discussed in this paper extend to non-autonomous systems directly. They also extend with minor changes to difference equations and partial differential equations. The techniques also can also be extended to the case of time varying coefficients (stochastic processes).

As discussed above, we consider the practical goal of computing a quantity of interest represented as a linear functional of the form

$$q(\omega) = q(\boldsymbol{\lambda}(\omega)) = \int_0^T \langle \boldsymbol{x}(s;\boldsymbol{\lambda}(\omega)), \boldsymbol{\psi}(s)\rangle \, \mathrm{d}s, \tag{4}$$

where $\boldsymbol{\psi}$ is a function of time corresponding to the quantity of interest. Some common choices are:

- $\boldsymbol{\psi}(s) = \delta(s-t)(0,\ldots,1,0,\ldots)^{\mathsf{T}}$, which yields the ith component of $\boldsymbol{x}(t;\omega)$ at time t;
- $\boldsymbol{\psi}(s) = (1,\ldots,1)^{\mathsf{T}}/nT$, which yields the time average over $[0,T]$ of all the components;
- $\boldsymbol{\psi}(s) = (0,\ldots,1,0,\ldots)^{\mathsf{T}}/T$, which yields the time average over $[0,T]$ of a particular component of the solution.

The first goal is to develop a fast method to approximate the cumulative distribution function $F_q(x)$ of q given an arbitrary input distribution for $\boldsymbol{\lambda}$. Since we recover the entire distribution, we can compute the density as well as any desired statistic, e.g. means and moments, or other quantities of the form $E[\mathscr{S}(q(\omega))]$ for measurable functions $\mathscr{S}$. In particular, the plots of densities in this paper are computed using a kernel density estimator on a data set constructed to have the distribution of $F_{\tilde{q}}$, where $\tilde{q}$ is the computed approximation to q. See Section 10.1 for details.

We carry out the analog of the variational argument presented in Section 2 and introduce the generalized Green's function solving the adjoint problem to the linearized equation:

$$\begin{cases} -\dot{\boldsymbol{\phi}}(t) - A^{\mathsf{T}}(t)\boldsymbol{\phi}(t) = \boldsymbol{\psi}(t), \quad T \geqslant t \geqslant 0, \\ \boldsymbol{\phi}(T) = \mathbf{0}, \end{cases} \tag{5}$$

where $A(t) \equiv D_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{y}(t);\boldsymbol{\mu}_1)$ with $\boldsymbol{y}(t)$ denoting a deterministic solution solving the system (3) for the fixed parameter $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\boldsymbol{\mu}_0)^{\mathsf{T}}$.

Arguing as for the finite-dimensional example, see Section 9.1, yields the following representation of the output value.

Theorem 1. *For $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ as above,*

$$q(\boldsymbol{\lambda}) = \int_0^T \langle \boldsymbol{x}, \boldsymbol{\psi}\rangle \, \mathrm{d}s \approx \int_0^T \langle \boldsymbol{y}, \boldsymbol{\psi}\rangle \, \mathrm{d}s + \langle \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0, \boldsymbol{\phi}(0)\rangle + \int_0^T \langle D_{\lambda_1}\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)(\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1), \boldsymbol{\phi}\rangle \, \mathrm{d}s. \tag{6}$$

The last term on the right-hand side describes the effect of variations in the model parameters. Note, that this term suggests that knowledge of $D_{\lambda_1}\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)$ and $\boldsymbol{\phi}$ are necessary to accurately estimate the density of the output corresponding to variation $\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1$ in the input. The second to last term on the right-hand side of (6) describes the effect of variations in the initial conditions.

The nature of this approximation is described precisely in the following theorem.

Theorem 2. *If there is a constant L such that $\boldsymbol{f}$ satisfies the Lipschitz condition,*

$$|\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\lambda}_1) - \boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)| \leqslant L(|\boldsymbol{x}-\boldsymbol{y}| + |\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1|), \tag{7}$$

for all $\boldsymbol{x},\boldsymbol{y} \in \mathscr{K}$ and $\boldsymbol{\mu},\boldsymbol{\lambda} \in \Lambda$, where the solutions to (3) remain in the set $\mathscr{K}$ for all parameters in Λ and $t \in [0,T]$, then the functional (4) of the solution to (3) satisfies

$$\nabla q(\boldsymbol{\mu})[\cdot] = \nabla \int_0^T \langle \boldsymbol{x}(s;\boldsymbol{\lambda}), \boldsymbol{\psi}(s)\rangle \, \mathrm{d}s|_{\boldsymbol{\lambda}=\boldsymbol{\mu}}[\cdot] = \langle [\cdot]_0, \boldsymbol{\phi}(0)\rangle + \int_0^T \langle D_{\lambda_1}\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)[\cdot]_1, \boldsymbol{\phi}\rangle \, \mathrm{d}s, \tag{8}$$

where $[\,\cdot\,]$ appears since this derivative is a linear operator from $\mathbb{R}^d \to \mathbb{R}$, and $\boldsymbol{y}$, $\boldsymbol{\phi}$ are defined as above.

This states that the representation in (6) is a linear approximation to the function q at the parameter value $\boldsymbol{\mu}$. The proof is given in Section 9.2.

The one point HOPS approximation can therefore be written

$$q(\boldsymbol{\lambda}) \approx q(\boldsymbol{\mu}) + \langle \nabla q(\boldsymbol{\mu}), (\boldsymbol{\lambda} - \boldsymbol{\mu}) \rangle.$$

The Lipschitz condition (7) is the natural extension of the standard assumption for a local existence and uniqueness result. For general $\boldsymbol{f}$, this requires the solution set $\mathscr{K}$ to be compact. The necessary restrictions on the parameter set Λ are highly problem dependent. However, in many cases, it appears that at least we have to assume that Λ is compact as well.

For example, consider the model:

$$\begin{cases} x' = x^2, & 0 \leqslant t \leqslant T, \\ x(0) = x_0, \end{cases}$$

where x_0 is the uncertain parameter. The solution is $x(t) = 1/(x_0^{-1} - t)$, which has a finite time blow up at $t = 1/x_0$ whenever $x_0 > 0$. Any investigation of the solution over a time interval $[0,T]$ requires x_0 to be bounded above, otherwise there is no solution on any time interval. Less dramatically, consider computing the mean value of solutions of:

$$\begin{cases} x' = \lambda x, & 0 \leqslant t \leqslant T, \\ x(0) = x_0 \end{cases}$$

with $\lambda < 0$. If $\lambda = N(\mu,\sigma^2)$ where $\mu < 0$, the expected values of the possible solutions is $E[x(t)] = x_0 e^{\mu t} e^{\sigma^2 t^2/2}$, which decays for a short time and then grows exponentially very rapidly! This suggests that a choice of normal perturbation for the parameter is inappropriate if we are interested in the mean value of the solutions.

In general, the differential equation is usually a valid model only for a certain range of parameters and ceases to be a relevant model when the parameters exceed this range. For general problems, the assumption that $\mathscr{K}$ is compact appears to be a natural *minimum* requirement. For specific problems with good dynamical behavior, this assumption can be relaxed.

3.0.1. A scalar example

We consider a simple scalar example:

$$\begin{cases} \dot{x}(t;\lambda) = \lambda x(t;\lambda), & t > 0, \\ x(0) = x_0. \end{cases} \tag{9}$$

Solving for the generalized Green's function corresponding to the reference value μ and the data $\psi(s) = \delta(s - t)$ yields the one point approximation to the functional q,

$$q(\lambda) = x(t;\lambda) \approx y(t;\mu) + (\lambda(\omega) - \mu) \int_0^t e^{\mu(t-s)} y(s;\mu)\, ds, \tag{10}$$

which is valid when the variance of $\lambda(\omega)$ is small. The HOPS approximation has the form $y(t;\lambda) = a(t) + \lambda b(t)$ with

$$a(t) = x_0(1 - \mu t)e^{\mu t}, \quad b(t) = x_0 t e^{\mu t}.$$

It is possible to calculate the associated probability densities exactly and we compare them in Fig. 5.

3.0.2. Multi-point HOPS approximations

Generally, we combine HOPS approximations computed at multiple reference values to obtain an accurate global approximation. We do this in two ways.

In the first approach, we choose a sample $\{\boldsymbol{\mu}_i\}_{i=1}^N$ and then partition the parameter space into a collection of generalized rectangles $\{R_i\}_{i=1}^N$ with $\boldsymbol{\mu}_i \in R_i$ for all i. The corresponding piecewise linear HOPS approximation is defined

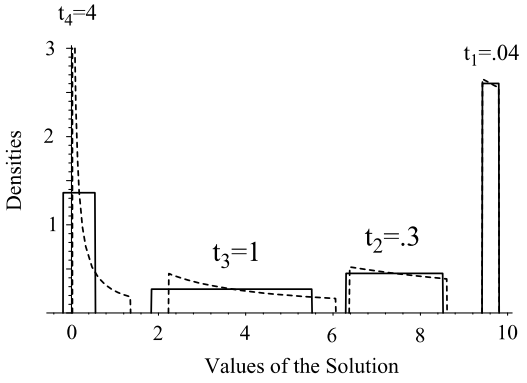


Fig. 5. True (dashed) and approximate (solid) probability densities for (9) with $y_0 = x_0 = 10$, $\lambda(\omega)$ uniformly distributed on $[-3/2,-1/2]$, and $\psi(s) = \delta(s - t_i)$ at times $\{t_i\} = \{0.04,0.3,1,4\}$.

$$q(\boldsymbol{\lambda}) \approx \tilde{q}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} (q(\boldsymbol{\mu}_i) + \langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{\lambda} - \boldsymbol{\mu}_i) \rangle) \chi_{R_i}(\boldsymbol{\lambda}), \tag{11}$$

where χ_{R_i} is 1 for $\boldsymbol{\lambda} \in R_i$ and 0 otherwise. By computing the solution at each $\boldsymbol{\mu}_i$ and applying the approximation to each piece, we expect this piecewise linear approximation to converge in distribution to the variable $q(\boldsymbol{\lambda})$ as the number of sample points increases.

In the second approach, we use a partition of unity. We let $\{\Lambda_i\}_{i=1}^N$ be a finite open cover of the compact parameter space Λ. A *Lipschitz partition of unity* subordinate to $\{\Lambda_i\}$ is a collection of functions $\{\theta_i\}_{i=1}^N$ with the properties:

$$\text{supp}(\theta_i) \subset \bar{\Lambda}_i \text{ and } \theta_i \text{ is differentiable on } \Lambda_i, \quad 1 \leqslant i \leqslant N, \tag{12}$$

$$\|\theta_i\|_{L^\infty(\Lambda_i)} \leqslant C \text{ and } \|\nabla\theta_i\|_{L^\infty(\Lambda_i)} \leqslant C/\text{diam}(\Lambda_i), \quad 1 \leqslant i \leqslant N, \tag{13}$$

$$\theta_i \text{ is continuous on } \Lambda \text{ and } \sum_{i=1}^{N} \theta_i(\boldsymbol{x}) = 1, \quad \boldsymbol{x} \in \Omega, \tag{14}$$

where C is a constant and $\text{diam}(\Lambda_i)$ is the diameter of Λ_i.

Several partitions of unity satisfying (12) and (13) exist. We use the partition of unity suggested for the Modified Shepard's Method by Renka [6]. The support of the partition of unity functions are spheres of radius $\mathscr{R}$ centered at the sample points $\{\boldsymbol{\mu}_i\}_{i=1}^N$, where $\mathscr{R}$ is chosen so that each sphere contains exactly $\mathscr{N}$ of the sample points. The partition of unity functions are defined as

$$\theta_i(\boldsymbol{x}) = \frac{W_i(\boldsymbol{x})}{\sum_{j=1}^N W_j(\boldsymbol{x})}, \quad \text{where } W_i(\boldsymbol{x}) = \left(\frac{\max\{\mathscr{R} - |\boldsymbol{x} - \boldsymbol{\mu}_i|, 0\}}{\mathscr{R}|\boldsymbol{x} - \boldsymbol{\mu}_i|} \right)^2.$$

Assuming that $\boldsymbol{\mu}_i$ is a point inside Λ_i for $i = 1,\ldots,N$, the approximation is defined as

$$q(\boldsymbol{\lambda}) \approx \tilde{q}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} (q(\boldsymbol{\mu}_i) + \langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{\lambda} - \boldsymbol{\mu}_i) \rangle) \theta_i(\boldsymbol{\lambda}). \tag{15}$$

4. Reliably accurate parameter sampling

In HOPS, we attempt to improve computational efficiency. Another important goal is to compute information from an approximation of the output distribution whose accuracy can be reliably quantified. If we write the one point HOPS approximation as

$$q(\boldsymbol{\lambda}) - q(\boldsymbol{\mu}) \approx \langle \nabla q(\boldsymbol{\mu}), (\boldsymbol{\lambda} - \boldsymbol{\mu}) \rangle, \tag{16}$$

we see that the derivative information $\langle \nabla q(\boldsymbol{\mu}),(\boldsymbol{\lambda} - \boldsymbol{\mu}) \rangle$ provides a local estimate of the error that results from using the sample value $q(\boldsymbol{\mu})$ in place of the actual value $q(\boldsymbol{\lambda})$ for $\boldsymbol{\lambda}$ near $\boldsymbol{\mu}$. In this section, we describe several

methods that collectively we call *Reliably Accurate Parameter Sampling* methods, or *RAPS*, which uses this derivative information to provide various a posteriori error estimates and bounds.

4.1. Measuring the error in an approximation of a quantity of interest

We start by discussing ways to measure the error in an approximation $\tilde{q}$ of a quantity of interest q. The approximation $\tilde{q}$ is constructed using some sample $\{\boldsymbol{\mu}_i\}_{i=1}^N$ and the corresponding values $\{q(\boldsymbol{\mu}_i)\}_{i=1}^N$. We form the approximate surface using either a piecewise constant function of the form

$$\bar{q}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} q(\boldsymbol{\mu}_i)\chi_{R_i}(\boldsymbol{\lambda}),$$

where $\{R_i\}$ is some partition consisting of generalized rectangles such that $\boldsymbol{\mu}_i \in R_i$, or using a partition of unity $\{\theta_i\}$ associated to $\{\boldsymbol{\mu}_i\}$ as above to form

$$\bar{q}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} q(\boldsymbol{\mu}_i)\theta_i(\boldsymbol{\lambda}).$$

When $\boldsymbol{\lambda}$ is be a random vector of ω, we obtain $q(\omega) = q(\boldsymbol{\lambda}(\omega)) \approx \tilde{q}(\boldsymbol{\lambda}(\omega))$. To measure the error in this approximation, it is natural to use the expected value the difference of some statistical function of q and $\tilde{q}$, i.e.,

$$|E(\mathscr{S}(q(\omega)) - \mathscr{S}(\tilde{q}(\omega)))|,$$

where, for example, $\mathscr{S}(s) = s$ yields the expected value and $\mathscr{S}(s) = (s - E(q))^2$ yields the variance. In general, we might choose a collection of such statistical functions to evaluate the accuracy.

A more certain way to guarantee that the approximate distribution converges to the true distribution is to control the L^1 norm,

$$E(|q(\omega) - \tilde{q}(\omega)|), \tag{17}$$

since if this tends to zero, then this implies convergence in probability and in distribution [7], and the cumulative distribution function F_q of q can be approximated arbitrarily well by $F_{\tilde{q}}$. By recovering the cumulative distribution function, we can compute any desired statistic, e.g. means and moments, that might be used to characterize the accuracy of the approximate distribution.

On the other extreme, we can require that information computed from $\tilde{q}$ be close pointwise to the corresponding information computed from q. For example, this is the implicit choice when comparing plots of distribution and density functions as often done in this paper. This is also a relevant measure in the context of performing sensitivity analysis.

4.2. Viewing a sample as a piecewise constant approximation

We assume that there is a partition of the compact parameter space Λ into N generalized rectangles $\{R_i\}_{i=1}^N$ where $\boldsymbol{\mu}_i$ is a point inside R_i and $q(\boldsymbol{\mu}_i)$ is calculated for $i = 1,\ldots,N$. This results in a piecewise constant approximation,

$$\tilde{q}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} q(\boldsymbol{\mu}_i)\chi_{R_i}(\boldsymbol{\lambda}). \tag{18}$$

In the case that the sample is generated by a Monte-Carlo computation, this interpretation is different than the standard view of a Monte-Carlo approximation. For example, the Monte-Carlo approximation weights the sample values of the function equally. If we consider the sample values as defining a piecewise constant approximation, the natural approximation to the average value is

$$\sum_{i=1}^{N} g(\omega_i)\frac{\mathrm{Vol}(R_i)}{\mathrm{Vol}(\Omega)},$$

i.e., the sample values are not weighted equally. Nominally, the new interpretation yields a better asymptotic convergence rate, i.e., $O(1/N)$ versus $O(1/\sqrt{N})$, however, the constants in the deterministic error bound grow rapidly as the dimension of the parameter space increases.

We provide evidence below that interpreting a sample as a piecewise constant approximation can be useful when the parameter space is not too high dimensional. Assuming that the error is measured by the expected value of the difference, we can represent it as a sum of "element contributions",

$$E(q(\omega)-\tilde{q}(\omega)) = \int_{\Omega} (q(\boldsymbol{\lambda}(\omega))-\tilde{q}(\boldsymbol{\lambda}(\omega)))\,\mathrm{d}P(\omega) = \sum_i \int_{R_i} (q(\boldsymbol{z})-\tilde{q}(\boldsymbol{z}))\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}),$$

where $\mu_{\boldsymbol{\lambda}}(A) = P(\boldsymbol{\lambda}\in A)$ for Borel sets A. Using a Taylor expansion in each of the R_i as in HOPS, we obtain approximation of the expected value of the error

$$\mathscr{E}^{\mathrm{pc}} = \sum_{i=1}^{N} \int_{R_i} \langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z}-\boldsymbol{\mu}_i)\rangle\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}).$$

Similarly, for the error in a statistical function,

$$\mathscr{E}^{\mathrm{pc}} = \sum_{i=1}^{N} \int_{R_i} \mathscr{S}'(q(\boldsymbol{\mu}_i))\langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z}-\boldsymbol{\mu}_i)\rangle\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}) \tag{19}$$

and in the L^1 norm,

$$\mathscr{E}^{\mathrm{pc}} = \sum_{i=1}^{N} \int_{R_i} |\langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z}-\boldsymbol{\mu}_i)\rangle|\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}). \tag{20}$$

$\mathscr{E}^{\mathrm{pc}}$ is an approximation in the sense that if ∇q is continuous, then a sequence of such approximations computed on partitions with rectangles of decreasing size converges to the true estimate. Unfortunately, this is severely affected by the dimension of the parameter space.

4.2.1. A scalar example

To illustrate this approach, we consider the scalar example (9) with $x_0 = 1$ and λ taken uniformly in $[-0.5,0.5]$. The results for the estimate of the error in the L^1 norm of $q(\boldsymbol{\lambda}) = x(t,\lambda)$ at $T = 10$ are given in Fig. 6. We also plot results for the L^1 norm of the error in the second moment, $E[|q(\boldsymbol{\lambda})^2 - \tilde{q}(\boldsymbol{\lambda})^2|]$, and corresponding estimate at $T = 5$.

4.3. Computing an approximate error function using a partition of unity

In a second approach, we use Renka's partition of unity to create both an approximation from a given sample and an approximate error function. The approximation is defined as

$$\tilde{q}(\lambda) = \sum_{i=1}^{N} q(\boldsymbol{\mu}_i)\theta_i(\lambda), \quad \lambda\in\Lambda. \tag{21}$$

To estimate the expected value of the error, we use (14) to write

$$E(q(\omega)-\tilde{q}(\omega)) = \sum_{i=1}^{N} E((q(\omega)-q(\boldsymbol{\mu}_i))\theta_i(\omega)).$$

Now using (16), we estimate

$$E((q(\omega)-q(\boldsymbol{\mu}_i))\theta_i(\omega)) \approx \int_{\Lambda_i} \langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z}-\boldsymbol{\mu}_i)\rangle\theta_i(\boldsymbol{z})\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}).$$

We obtain the estimate

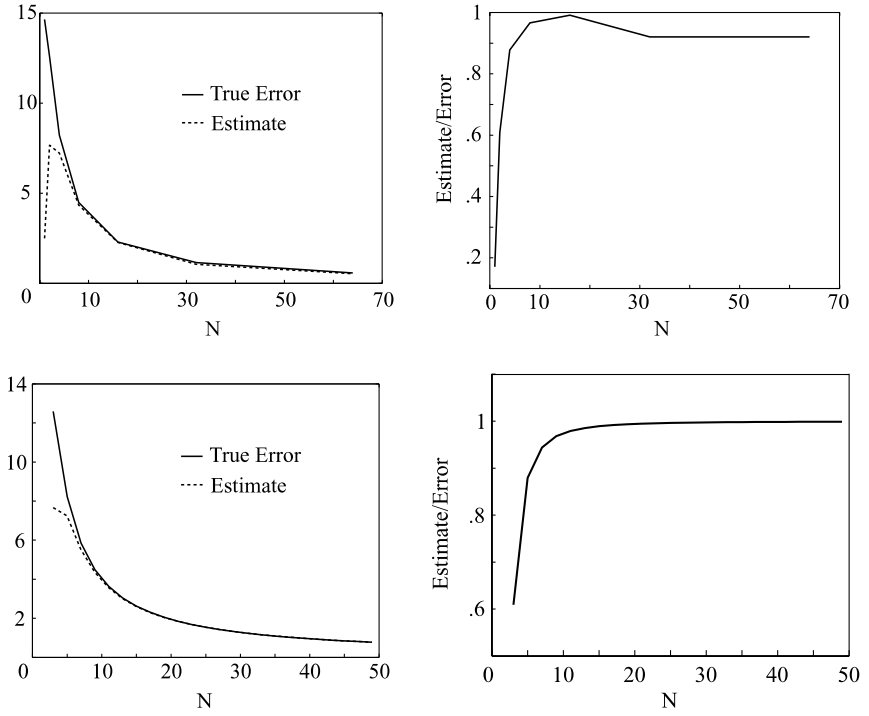


Fig. 6. In the upper left, we plot the error in the L^1 norm and the estimate for the scalar example (9) for a range of sample points N at $T = 10$. In the upper right, we plot the ratio of the estimate to the error. In the lower left, we plot the error and estimate for the L^1 norm of the second moment at $T = 5$. In the lower right, we plot the ratio of the estimate to the error.

$$\mathscr{E}^{\text{pou}} = \sum_{i=1}^{N} \int_{\Lambda_i} \langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z} - \boldsymbol{\mu}_i) \rangle \theta_i(\boldsymbol{z}) \, \mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}).$$

Similarly, we obtain approximations for the error in the expected value of any statistical function,

$$\mathscr{E}^{\text{pou}} = \sum_{i=1}^{N} \int_{\Lambda_i} \langle \mathscr{S}'(q(\boldsymbol{\mu}_i)) \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z} - \boldsymbol{\mu}_i) \rangle \theta_i(\boldsymbol{z}) \, \mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}) \tag{22}$$

and the L^1 norm,

$$\mathscr{E}^{\text{pou}} = \sum_{i=1}^{N} \int_{\Lambda_i} |\langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z} - \boldsymbol{\mu}_i) \rangle \theta_i(\boldsymbol{z})| \, \mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}). \tag{23}$$

Since

$$\int_{\Lambda_i} (\boldsymbol{z} - \boldsymbol{\mu}_i) \theta_i(\boldsymbol{z}) \, \mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}) = \int_{\Omega} (q(\boldsymbol{\lambda}(\omega)) - \boldsymbol{\mu}_i) \theta_i(\boldsymbol{\lambda}(\omega)) \, \mathrm{d}P(\omega) = E((\boldsymbol{\lambda} - \boldsymbol{\mu}_i)\theta_i)$$

in the case of the expected value of a statistical function, we can rewrite (22) as

$$\mathscr{E}^{\text{pou}} = \sum_{i=1}^{N} \langle \mathscr{S}'(q(\boldsymbol{\mu}_i)) \nabla q(\boldsymbol{\mu}_i), E((\boldsymbol{\lambda} - \boldsymbol{\mu}_i)\theta_i) \rangle. \tag{24}$$

4.3.1. A scalar example

To illustrate this approach, we again consider the scalar example (9) with $x_0 = 1$ and λ taken uniformly in $[-0.5, 0.5]$. The results for the estimate of the error in the L^1 norm at $T = 10$ and the estimate of the L^1 norm of the error in the second moment at $T = 5$ are given in Fig. 7.

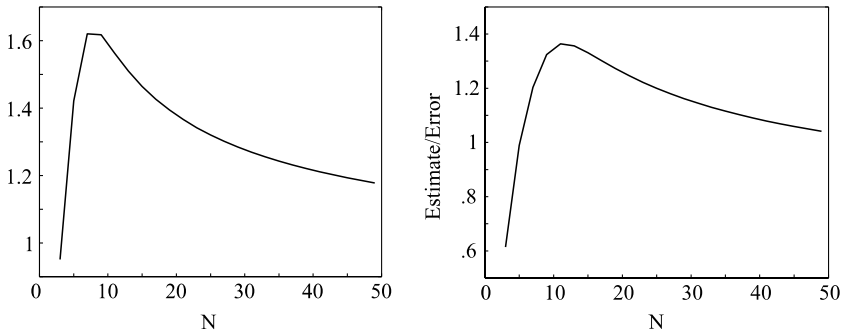


Fig. 7. On the left, we plot ratio of the estimate to the error in the L^1 norm for the scalar example (9) at $T = 10$. On the right, we plot the ratio of the estimate to the error for the L^1 norm of the second moment at $T = 5$. Compare to Fig. 6.

5. Fast adaptive parameter sampling

We observe that obtaining an accurate estimate of the error for a given sample opens up the possibility of optimizing the sample process in order to reach a desired accuracy with minimal computational cost. This can offset the overhead associated with computing the estimate.

In general, constructing an optimal sample set is a complicated and difficult nonlinear problem. In this paper, we borrow ideas from adaptive finite elements to construct two *Fast Adaptive Sampling Procedures*, or FAPS, for finding an approximate solution. The goal of the adaptive strategy is to find a set of sample points that has approximately minimal size yet yields the desired accuracy in the computed information. The adaptive strategy works in an iterative fashion. Given a current set of sample points, we estimate the local contributions to the error for the sample and choose additional sample points in regions in which the contributions to the error are estimated to be largest.

This involves several choices. First, we must choose an error estimate or bound to guide the enrichment decision. Using an optimization approach requires an estimate or bound that can be written as a sum over local contributions with non-negative summands. The estimates (19) and (22) might be used to determine that the computed information is not sufficiently accurate, but they cannot be used for standard adaptive error control because they allow cancellation of errors through the domain. Once the decision to enrich the current sample is made, we use a bound derived from the estimates (20) or (23) for the L^1 norm of the error to choose additional sample points.

Another important choice is a procedure for adding additional sample points to a given sample set. We describe two approaches below.

The choice of the initial set of sample points is third important decision. The standard approach is to use a coarse set in order to increase the possibility of achieving a nearly optimal final sample set. On the other hand, using too coarse an initial sample set can mean that the error estimate fails to recognize significant behavior. Another aspect is the manner in which the initial set is constructed, e.g. the initial sample set might be uniformly spaced or it might be generated by a Monte-Carlo method.

5.1. Adaptive sampling via a standard h-refinement approach

The first approach we describe is analogous to adaptive error control for finite element methods. We use one or more of the estimates (19) and (20) to decide if a given sample set needs to be enriched. If enrichment is needed, we use a bound derived from (20) to choose additional sample points. We add points using a strategy that is closely analogous to the standard h-refinement strategy in adaptive finite element methods. We define $\mathscr{E}_i^{\mathrm{pc}}$ to be the approximate contribution to the error bound from rectangle R_i, i.e.,

$$\mathscr{E}_i^{\mathrm{pc}} = \int_{R_i} |\langle \nabla q(\boldsymbol{\mu}_i), (\boldsymbol{z} - \boldsymbol{\mu}_i)\rangle| \, \mathrm{d}\mu_{\lambda}(\boldsymbol{z}).$$

The adaptive strategy is to refine some fraction of the rectangles on which $\mathscr{E}_i^{\mathrm{pc}}$ is largest. We refine along one dimension at a time since q may be very sensitive to changes in one parameter, but not in others. To measure the contribution to the error bound from each dimension, we define $\mathscr{E}_{i,k}^{\mathrm{pc}}$, $k = 1,\ldots,d$, by

$$\mathscr{E}_{i,k}^{\mathrm{pc}} = \int_{R_i} |\partial_{\lambda_k} q(\boldsymbol{\mu}_i)(z^k - \mu_i^k)|\,\mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}), \tag{25}$$

where $\boldsymbol{z} = (z^1,\ldots,z^d)^{\mathsf{T}}$ and $\boldsymbol{\mu}_i = (\mu_i^1,\ldots,\mu_i^d)^{\mathsf{T}}$ so

$$\mathscr{E}^{\mathrm{pc}} = \sum_{i=1}^{N} \mathscr{E}_i^{\mathrm{pc}} \leqslant \sum_{i=1}^{N}\sum_{k=1}^{d} \mathscr{E}_{i,k}^{\mathrm{pc}}.$$

For a rectangle where $\mathscr{E}_i^{\mathrm{pc}}$ is large enough for refinement, we find the maximum contribution $\mathscr{E}_{i,k}^{\mathrm{pc}}$, $k = 1,\ldots,d$ and we divide the rectangle along this dimension. Since the value at the center of the rectangle is known, we split in thirds along this dimension, compute the values at the centers of the two new rectangles. We iterate until the selection of estimates (19) and (20) used to judge accuracy of the computed information are bounded by TOL.

In the examples below, we use a uniform initial partition of the parameter space, which is reasonable when the parameter space does not have high dimension.

We illustrate the rectangular FAPS in Sections 7.1–7.3 below.

5.2. Adaptive sampling according to a density representing the error

In another approach, we use a partition of unity to generate one or more of the estimates (22) and (23) and construct a density function using the element contributions in the estimate. The integral mean value theorem implies that the approximation $\mathscr{E}^{\mathrm{pou}}$ to the expected value is itself approximated by

$$\mathscr{E}_{\mathrm{loc}}^{\mathrm{pou}}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \langle \nabla q(\boldsymbol{\mu}_i), \boldsymbol{\lambda} - \boldsymbol{\mu}_i\rangle \theta_i(\boldsymbol{\lambda}) f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda})$$

at each point λ, where $f_{\boldsymbol{\lambda}}$ is the density corresponding to $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$. We use the error estimate $|\mathscr{E}_{\mathrm{loc}}^{\mathrm{pou}}|$ as a sampling density to select additional points in such a way that there is increased sampling in regions where this function is relatively large. We can do this either in a deterministic fashion or using a Monte-Carlo approach.

We illustrate this adaptive approach in Section 7.1 below.

5.2.1. Creating a density that indicates regions with insufficient sampling

There are other ways to create a density function. Returning to the local error estimate $|q(\boldsymbol{\lambda}) - q(\boldsymbol{\mu}_i)| \lesssim |\nabla q(\boldsymbol{\mu}_i)||\boldsymbol{\lambda} - \boldsymbol{\mu}_i|$, the values of $q(\boldsymbol{\lambda})$ are approximated by $q(\boldsymbol{\mu}_i)$ to within a tolerance TOL for all $\boldsymbol{\lambda}$ in the ball of radius $r_i = \mathrm{TOL}/|\nabla q(\boldsymbol{\mu}_i)|$ centered at $\boldsymbol{\mu}_i$ with volume $V_i = \pi^{d/2} r_i^d/\Gamma(d/2+1)$. We set $V = \left(\sum_{i=1}^{N} V_i^{-1}\right)^{-1}$, so $\sum_{i=1}^{N} \frac{V}{V_i} = 1$. We interpolate the set of values $\{V/V_i\}$ at the points $\{\boldsymbol{\mu}_i\}$ to obtain a density function that describes the degree to which the function is represented pointwise by its values at $\{\boldsymbol{\mu}_i\}$. Note that the dimension-dependent constants in V_i cancel in the ratio V/V_i. In fact,

$$\frac{V}{V_i} = \frac{|\nabla q(\boldsymbol{\mu}_i)|^d}{\sum_{j=1}^{N} |\nabla q(\boldsymbol{\mu}_j)|^d} = \left(\sum_{j=1}^{N} \left(\frac{|\nabla q(\boldsymbol{\mu}_j)|}{|\nabla q(\boldsymbol{\mu}_i)|}\right)^d\right)^{-1}. \tag{26}$$

As d increases, the difference in the values of the density function corresponding to large and small values of $|\nabla q(\boldsymbol{\mu}_i)|$ is exaggerated. In contrast, simply removing the balls from the parameter space before selecting the additional sample points has decreasing effect on the selection process as the dimension increases.

To illustrate, we consider the function

$$q(\boldsymbol{\lambda}) = \mathrm{e}^{-0.5(\lambda_1^2+\lambda_2^2)} + 0.01\lambda_1^3 + 0.01\lambda_2^2 \tag{27}$$

on $[0,4]\times[0,4]$. We plot q in Fig. 8 for a uniform 21×21 set of sample points. In Fig. 9, we plot the regions in which the function is approximated pointwise to within a tolerance of $\mathrm{TOL} = 0.01$ for three sets of uniformly spaced sample points.

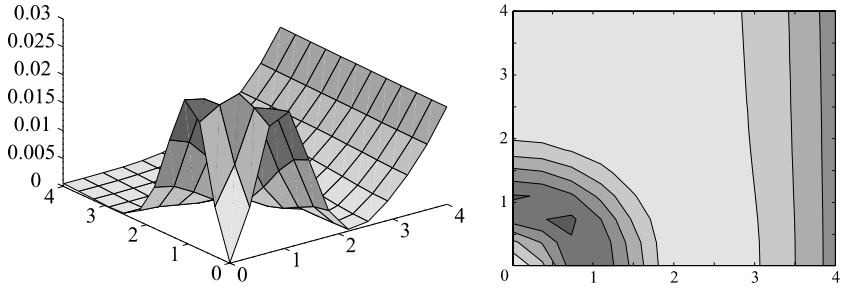


Fig. 8. Surface and contour plots of function q (27) corresponding to TOL = 0.01 and a uniform set of 12×12 sample points. Dark areas indicate regions in which q is not approximated pointwise as well.

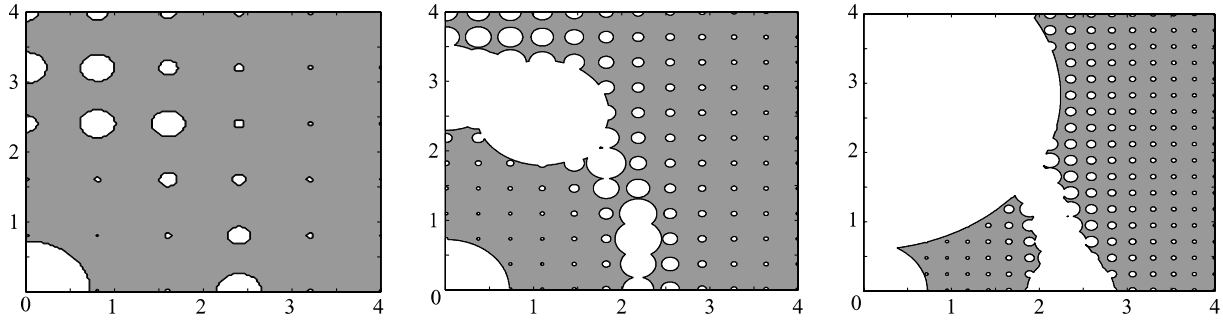


Fig. 9. Plots of the regions in which the function is approximated pointwise to within a tolerance of TOL = 0.01 for three sets of uniformly spaced sample points. From left to right, the sets have 6×6, 12×12, and 18×18 points, respectively. Shaded regions indicate areas where q is not sufficiently well approximated.

This approach extends directly to the estimate (23) after choosing a partition of unity. For example, suppose that Λ_i is a ball of radius r_i centered at $\boldsymbol{\mu}_i$ with volume V_i. We can bound

$$\int_{\Lambda_i} |(\boldsymbol{z} - \boldsymbol{\mu}_i)\theta_i(\boldsymbol{z})| \mathrm{d}\mu_{\boldsymbol{\lambda}}(\boldsymbol{z}) \lesssim C r_i \mu_{\boldsymbol{\lambda}}(\Lambda_i) \approx C r_i f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_i) V_i,$$

where $\boldsymbol{\lambda}_i$ is some point in Λ_i, C is a constant that depends on the construction of θ_i, and

$$|\mathscr{E}^{\mathrm{pou}}| \leqslant \sum_{i=1}^{N} C |\nabla q(\boldsymbol{\mu}_i)| f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_i) r_i V_i.$$

We seek to choose the minimum number of points such that

$$\sum_{i=1}^{N} C |\nabla q(\boldsymbol{\mu}_i)| r_i f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_i) V_i \leqslant \mathrm{TOL}.$$

The standard "Principle of Equidistribution" argument implies that the optimal result is obtained when the summands are equal, which yields a formula for r_i. Defining $V = \left(\sum_{i=1}^{N} V_i^{-1}\right)^{-1}$ as above, we find that in this case,

$$\frac{V}{V_i} = \frac{(|\nabla q(\boldsymbol{\mu}_i)| f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_i))^{d/(d+1)}}{\sum_{j=1}^{N} (|\nabla q(\boldsymbol{\mu}_j)| f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_j))^{d/(d+1)}} = \left(\sum_{j=1}^{N} \left(\frac{|\nabla q(\boldsymbol{\mu}_j)| f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_j)}{|\nabla q(\boldsymbol{\mu}_i)| f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_i)} \right)^{d/(d+1)} \right)^{-1}. \tag{28}$$

Using this, we produce a density function as described in Section 5.2.1. Note that in sharp contrast to (26), there is decreasing effect on the values of the density as d increases. This suggests that the evaluation of q

should not be treated as a purely pointwise interpolation problem when the parameter space has large dimension.

We illustrate this adaptive approach in Section 7.1 below.

6. Application to Monte-Carlo methods

The information about the derivative of the quantity of interest at a reference point that we compute using the generalized Green's function can also be used in context of some standard techniques for improving the Monte-Carlo method. This is important when the parameter space has high dimension, which can have a negative impact on the HOPS and FAPS procedures described above. As an example, we briefly describe applications to the methods of control variates and stratified sampling [8]. Detailed descriptions, analysis, and computations are presented in [9].

6.1. The method of control variates

In the method of control variates, we write the expected value of a statistical quantity

$$E[\mathscr{S}(q(\boldsymbol{\lambda}))] = \int_{\mathbb{R}^p} \mathscr{S}(q(\boldsymbol{z}))f_{\boldsymbol{\lambda}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \int_{\mathbb{R}^p} [\mathscr{S}(q(\boldsymbol{z})) - \mathscr{S}(\tilde{q}(\boldsymbol{z}))]f_{\boldsymbol{\lambda}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} + \int_{\mathbb{R}^p} \mathscr{S}(\tilde{q}(\boldsymbol{z}))f_{\boldsymbol{\lambda}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}, \tag{29}$$

where $\mathscr{S}$ is a statistical function, see Section 4.1. We choose $\tilde{q}$ to be either the FAPS or HOPS approximation to q.

The second integral on the right of (29) is relatively easy to compute since it is the (biased) estimator $E[\mathscr{S}(\tilde{q})]$. We evaluate the first integral on the right of (29) by taking samples $\boldsymbol{\lambda}_i \sim f_{\boldsymbol{\lambda}}$, $i = 1,\ldots,N$, and evaluating

$$\mathrm{d}\mathscr{S} = \frac{1}{N}\sum_{i=1}^{N} \mathscr{S}(q(\boldsymbol{\lambda}_i)) - \mathscr{S}(\tilde{q}(\boldsymbol{\lambda}_i)).$$

The estimator for $E[\mathscr{S}(q)]$ is $E[\mathscr{S}(\tilde{q})] + \mathrm{d}\mathscr{S}$, which is unbiased and has a smaller variance than the original problem when $\tilde{q}$ is a good approximation to q.

6.2. The method of stratified sampling using FAPS

We once again write $E[\mathscr{S}(q)]$ as in (29), and use a stratified Monte-Carlo estimator [8] for $E[\mathscr{S}(q)]$,

$$E[\mathscr{S}(q)] \approx \sum_{i=1}^{r} \frac{p_i}{n_i} \sum_{j=1}^{n_i} \mathscr{S}\left(q(\boldsymbol{y}_j^{\,i})\right),$$

where R_i, $i = 1,\ldots,r$ are a disjoint union of the support of $f_{\boldsymbol{\lambda}}$,

$$p_i = \int_{R_i} f_{\boldsymbol{\lambda}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z},$$

the n_i are integers greater than or equal to 1, and $\boldsymbol{y}_j^i$, $j = 1,\ldots,n_i$ are independent samples drawn from the densities

$$f_i(\boldsymbol{z}) = \frac{1}{p_i} f_{\boldsymbol{\lambda}}(\boldsymbol{z})\chi_{R_i}(\boldsymbol{z}),$$

for $i = 1,\ldots,r$. These samples are created by using the Accept–Reject method for the densities f_i.

We use FAPS in conjunction with this estimator by first allocating N_F samples for partitioning the parameter space according to the L_1 error indicator (20). We choose n_i directly proportional to the quantity $\mathscr{E}_i^{\mathrm{pc}}$ above. This choice utilizes the deterministic information about steep gradients in q, and avoids the bias which is introduced by gridding the parameter space and sampling at box centers.

7. Applications

7.1. The Schaefer model for a harvested fish population

The Schaefer model describes the evolution of a fish population x governed by a logistic mechanism that is harvested at a constant level of effort:

$$\begin{cases} \dot{x}(t;\boldsymbol{\lambda}_1) = \frac{R}{K}(K - x(t;\boldsymbol{\lambda}_1))x(t;\boldsymbol{\lambda}_1) - Hx(t;\boldsymbol{\lambda}_1), & t > 0, \\ x(0;\boldsymbol{\lambda}_1) = x_0, \end{cases} \tag{30}$$

where $\boldsymbol{\lambda}_1 = (R,K,H)^{\mathsf{T}}$. We fix the initial conditions.

First, we consider $H = 0$ and take $\psi(s) = \delta(s - 3)$ and $x_0 = 1$. The norm of the generalized Green's function is large in the area near $R = 0$, see Fig. 10, so we expect that increased sampling will be needed there. In Fig. 11, we plot the final adaptive grid. In Fig. 12, we plot the HOPS and rectangular FAPS approximations.

For the case $H \neq 0$, the system has two fixed points $\tilde{x}_1 = 0$ and $\tilde{x}_2 = K(1 - \frac{H}{R})$. The stability is determined by the derivative $\partial_x f$, which is $R - H$ at $\tilde{x}_1$ and $H - R$ at $\tilde{x}_2$. As $t \to \infty$, the distribution of the solutions split as

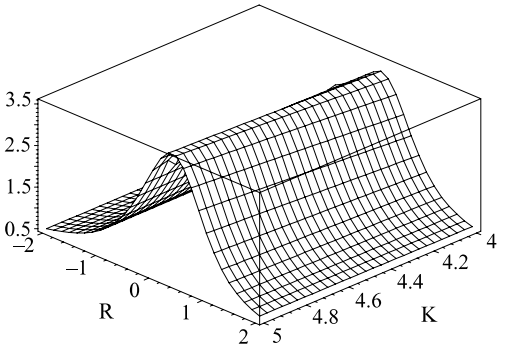
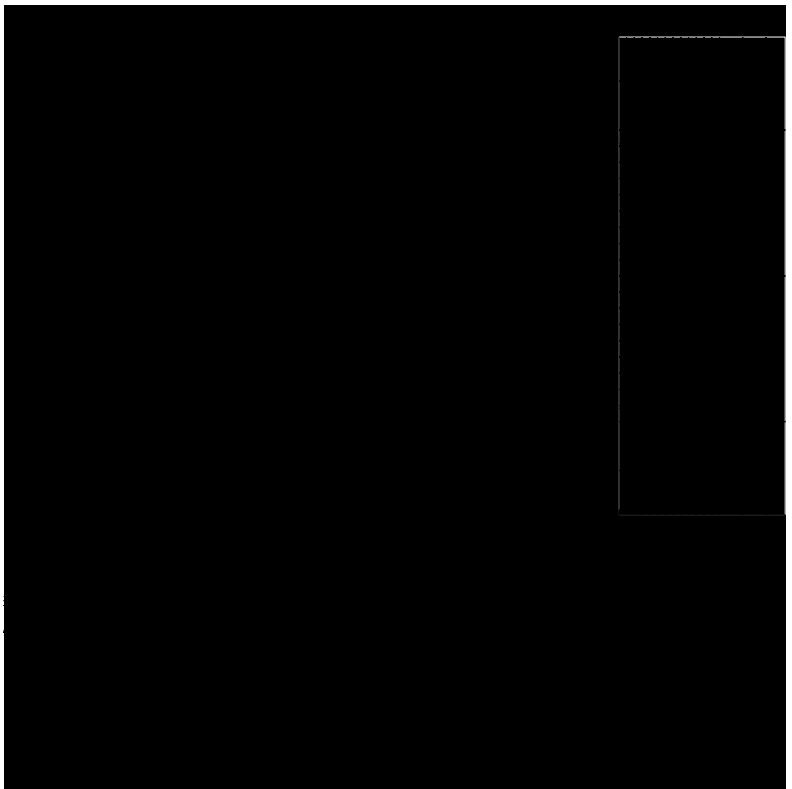
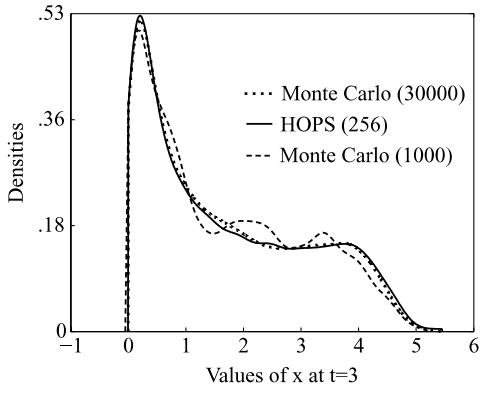
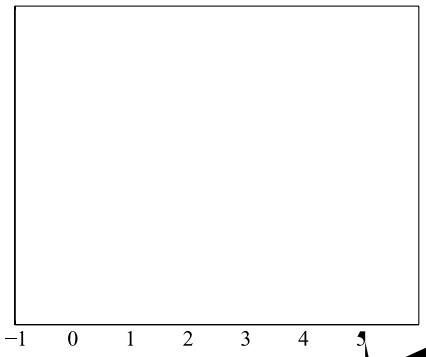


Fig. 10. Norm $\int_0^3 |\phi(s)|\,\mathrm{d}s$ of the Green's function for $\psi(s) = \delta(s - 3)$ and $x_0 = 1$.







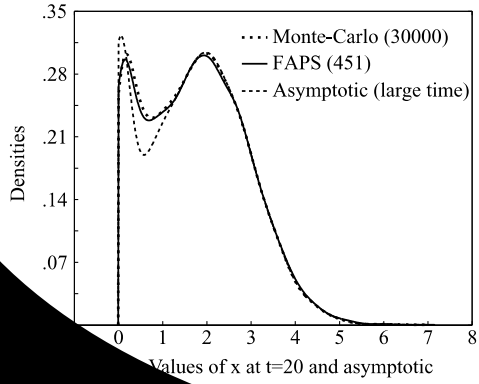
they are attracted to different fixed points. A bifurcation occurs [illegible] unstable for $R < H$ and the opposite when $R > H$. Hence as $t \to \infty$, the me[illegible]t ten[illegible] zero should be the measure of $[R < H]$, and the rest of the solution[illegible] as the [illegible]m variable [illegible] H/R) (their eventual limiting value). The case R[illegible]ce this line h[illegible]asure zero [illegible]rameter distribution. In Fig. 13, the asym[illegible]Carlo and rectang[illegible]APS com[illegible] compared. We notice discrepanci[illegible] which is expected sinc[illegible]stem ap[illegible]m much slower in this ca[illegible]

[illegible]gence properties [illegible]angular FAPS and HOPS in Fig. 14. [illegible]measure [illegible] statistic $\max_{x\in\mathbb{R}}|F_q(x) - F_{\hat{q}}(x)|$ on the cumulative d[illegible]s. For [illegible]ompute the empirical cumulative distribution function

$$[illegible]\ q(\boldsymbol{\lambda}_i) \leqslant x\}$$

[illegible]utions for the quantity of interest $q(\boldsymbol{\lambda}_i)$. Since F_n^{MC} depends [illegible]dom variable. Since our cumulative distribution is deterministi[illegible]mber of Monte-Carlo draws.

[illegible]he exact cumulative distribution function for the quantity of inter[illegible]h results to a massive Monte-Carlo simulation of 70,000 points. The[illegible]S and rectangular FAPS algorithms provide far better approximatio[illegible]



[illegible]ular FAPS, and asymptotic ($t = \infty$) distribu[illegible]

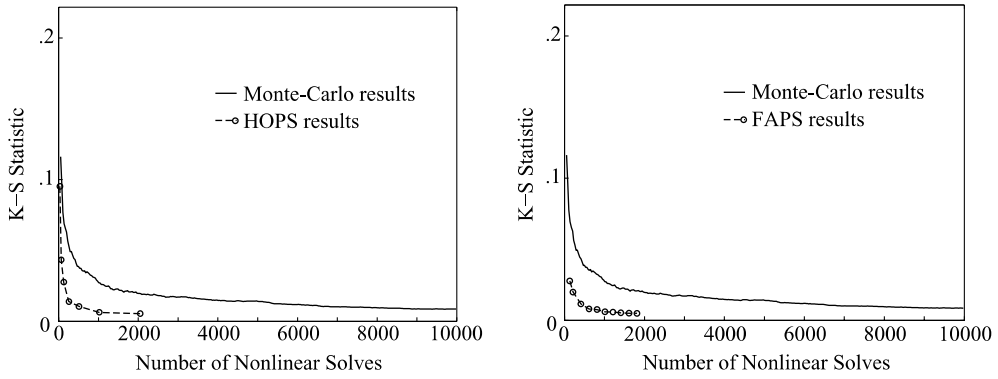


Fig. 14. Comparison of the K–S statistic for the HOPS, rectangular FAPS, and Monte-Carlo methods applied to (30). The error for the Monte-Carlo method is the average over 20 cases. A 70,000 point Monte-Carlo simulation is taken as the reference distribution.

the (K–S) statistic, especially for a small number of samples. The results suggest that simulations of less that 1000 points compare favorably with 10,000 point Monte-Carlo simulation.

We conclude by presenting some results for the partition of unity FAPS applied to (30) with $H = 0$, $x_0 = 1$, and $T = 3$. We first consider sampling according to the density described in Section 5.2, which represents the error. In Fig. 15, we plot the true error $(q(\boldsymbol{\lambda}) - \tilde{q}(\boldsymbol{\lambda}))f_{\lambda}(\boldsymbol{\lambda})$ and the approximation $\mathscr{E}^{\text{pou}}_{\text{loc}}(\boldsymbol{\lambda})$ computed using 20 sample points. To choose additional sample points according to the computed density, we employ a Markov Chain Monte-Carlo method [10] that creates a large sample of points with distribution $|\mathscr{E}^{\text{pou}}_{\text{loc}}|$, from which we draw a few points for evaluation. In Fig. 16, we plot approximate densities computed using FAPS.

7.2. The chaotic Lorenz model

We investigate the Lorenz equations:

$$\begin{cases} \dot{x}_1 = \sigma(x_2 - x_1), \\ \dot{x}_2 = rx_1 - x_2 - x_1x_3, \\ \dot{x}_3 = x_1x_2 - bx_3, \end{cases} \tag{31}$$

where we fix the parameters σ,r,b at standard values believed to yield chaotic behavior as well as the initial values $x_{2,0} = 0$, $x_{3,0} = 24$, but take the initial value for x_1 uniformly distributed in $[-2,2]$. We follow the value of the x_1 coordinate of the solution for a sequence of times, so that the functional is $q(\boldsymbol{\lambda}) = q(x_{1,0}) = x_1(t;x_{1,0})$.

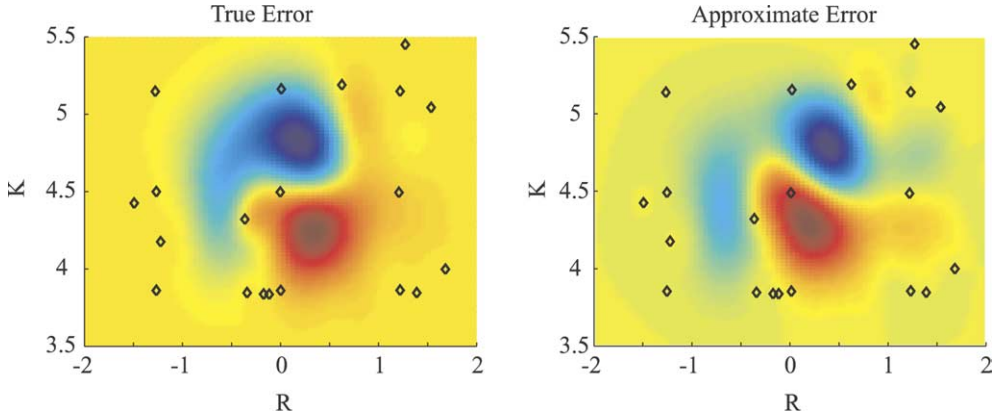


Fig. 15. On the left, we plot the true error. On the right, we plot the partition of unity FAPS approximation $\mathscr{E}^{\text{pou}}_{\text{loc}}(\boldsymbol{\lambda})$ computed from 20 sample points using the density described in Section 5.2. The sample points are indicated with circles.

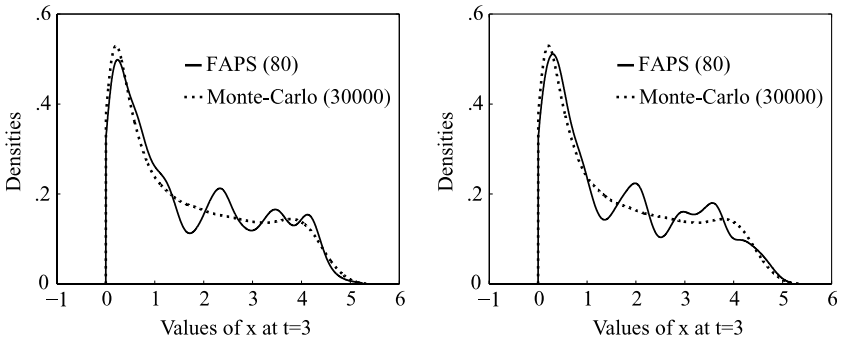


Fig. 16. We plot the approximate densities computed from a partition of unity FAPS using a sample of 80 points along with a Monte-Carlo computation with 30,000 points. On the left, we show the density described in Section 5.2, and on the right, we use the density described in Section 5.2.1.

All trajectories approach an attractor and the initial distribution is eventually spread onto this attractor, see Fig. 17. However, since the problem is chaotic, solutions that start close to each other eventually diverge, see Fig. 17. There is a separatix manifold [11] originating out of the x_3-axis that separates trajectories in the sense that close nearby solutions passing by this separatix rapidly move to neighborhoods of different hyperbolic

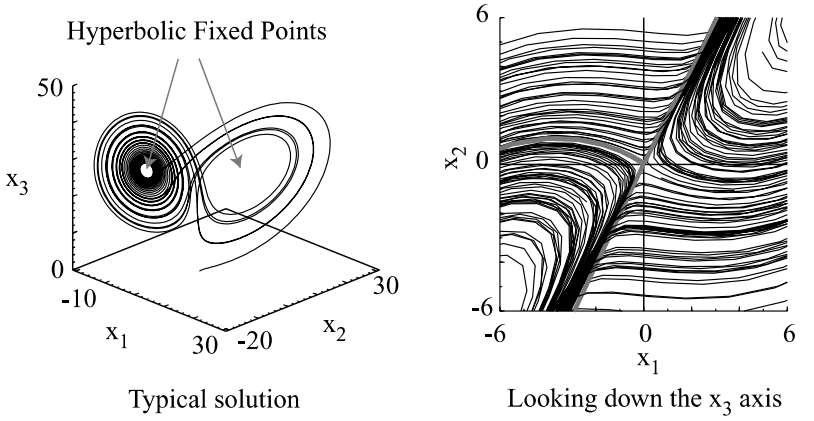


Fig. 17. On the left, we plot a typical solution of the Lorenz equations. On the right, we plot many solutions from the perspective of looking down the x_3-axis. The separatix manifold lies between the two solutions plotted with thick grey lines. The hyperbolic fixed points lie off the lower left and upper right corners.

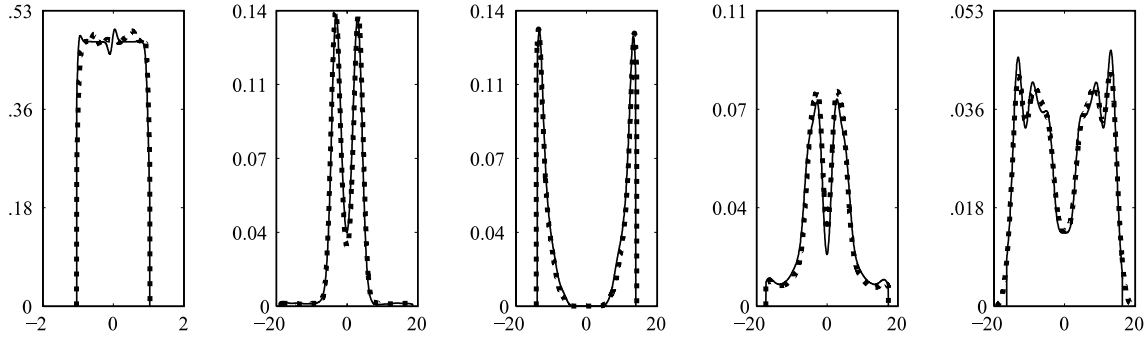


Fig. 18. Rectangular FAPS-computed densities of $q(x_{1,0}) = x_1(t;x_{1,0})$ compared with a 40,000 point Monte-Carlo computation at times 0.1, 1, 2, 3, 5 left to right. TOL = 0.01 resulted in 64, 94, 134, 240, 336 simulations, respectively.

Description	Name	Reference	Perturbation
Recovery rate	a_R	0.2	± 0.1
Natural growth rate	r_n	1	± 0.2
Carrying capacity	k	100	± 5
Probability of inheriting resistance	p_R	0.1	± 0.01
Natural death rate	d_n	0.2	± 0.1
Contraction rate	r_I	0.2	± 0.1
Death rate from disease	r_I	1	± 0.2

Fig. 20. Parameter reference values and ranges for SIR model. We take the parameters to be uniformly distributed in the ranges given.

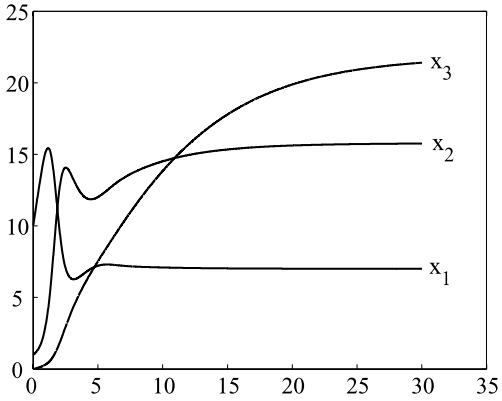


Fig. 21. Solution of the SIR model for $x_{1,0} = 10$, $x_{2,0} = 1$, $x_{3,0} = 0$ and mean parameters given in Fig. 20.

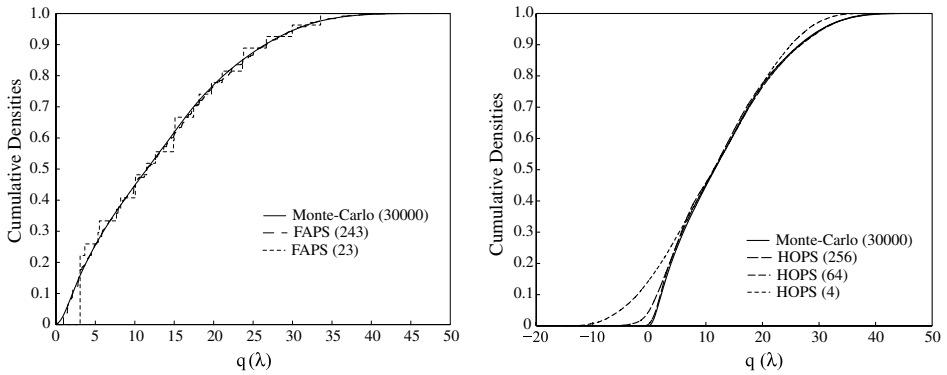


Fig. 22. On the left, we plot some HOPS approximations to the cumulative density. On the right, we plot rectangular FAPS approximations. Results are compared to a 30,000 point Monte-Carlo simulation.

$$q(\boldsymbol{\lambda}) = \int_0^T x_2(s;\boldsymbol{\lambda})\,\mathrm{d}s,$$

where $\boldsymbol{\lambda}$ denotes the vector of parameters in Fig. 20. We show the computed densities in Fig. 22. The HOPS procedure produces a reasonable approximation for a very small number of points. FAPS also does a very good job, as shown in Fig. 23.

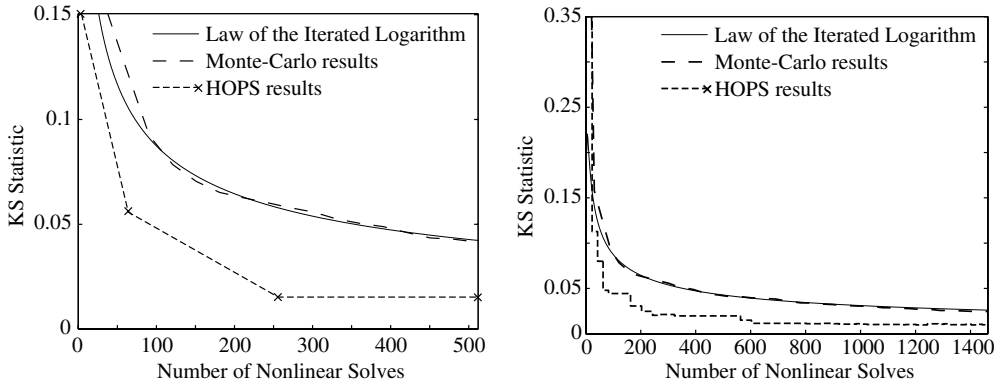


Fig. 23. A comparison of the K–S statistic for FAPS and HOPS for the SIR model. We compare against the mean Monte-Carlo behavior and the bound from the *Law of the Iterated Logarithm*. Left: *HOPS*, Right: *FAPS*.

8. Conclusion

In this paper, we describe a new class of methods for ascertaining the effects of variation and uncertainty in data and parameters on the output of a deterministic, smooth model. These methods are based on the introduction of the generalized Green's function and a variational analysis. The information yielded by the generalized Green's function is used both to create a higher order method and to produce an computational error estimate for a given pointwise representation. In the latter case, the estimate provides the basis for adaptive sampling. Both the higher order method and the adaptive sampling methods can be orders of magnitude faster than Monte-Carlo methods when the number of parameters is not too high.

The HOPS method provides a description of the output variation that is higher order than the traditional Monte-Carlo approach, gives no estimate on the accuracy. In the RAPS method, we show how the information yielded by the generalized Green's function can be used to generate various error estimates. We also describe how the error estimate can be used to create several adaptive sampling procedures called FAPS. We provide a number of examples of RAPS and FAPS computations that suggest the error estimates are accurate and adaptive sampling can lead to significant gains in computational efficiency. All of these methods are closely related and may be used in combination with each other in both Monte-Carlo and deterministic computations. The derivative information we compute using the generalized Green's function may also be used in combination with standard techniques for augmenting Monte-Carlo computations.

9. Proofs

9.1. Proof of Theorem 1

We consider a perturbation $\boldsymbol{\lambda} = \boldsymbol{\mu} + \boldsymbol{h}$ along with the corresponding solution $\boldsymbol{x}(t) = \boldsymbol{x}(t;\boldsymbol{\lambda})$ of (3). We let $\boldsymbol{y}(t)$ and $\boldsymbol{\phi}(t)$ be as defined in the text and we $\boldsymbol{e}(t) \equiv \boldsymbol{x}(t) - \boldsymbol{y}(t)$. We have the relation

$$\begin{aligned}\int_0^T \langle \boldsymbol{e}, \boldsymbol{\psi}\rangle \,\mathrm{d}s &= \int_0^T \langle \boldsymbol{e}, -\dot{\boldsymbol{\phi}} - A^{\mathrm{T}}\boldsymbol{\phi}\rangle \,\mathrm{d}s = \langle \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0, \boldsymbol{\phi}(0)\rangle + \int_0^T \langle \dot{\boldsymbol{e}} - A\boldsymbol{e}, \boldsymbol{\phi}\rangle \,\mathrm{d}s \\ &= \langle \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0, \boldsymbol{\phi}(0)\rangle + \int_0^T \langle \dot{\boldsymbol{x}} - A\boldsymbol{x}, \boldsymbol{\phi}\rangle \,\mathrm{d}s - \int_0^T \langle \dot{\boldsymbol{y}} - A\boldsymbol{y}, \boldsymbol{\phi}\rangle \,\mathrm{d}s,\end{aligned}$$

since $\boldsymbol{\phi}(T) = \boldsymbol{e}(0) = \boldsymbol{0}$. Expanding $\boldsymbol{f}$ in a Taylor approximation,

$$\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\lambda}_1) = \boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1) + D_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{y}) + D_{\boldsymbol{\lambda}_1}\boldsymbol{f}(\boldsymbol{y};\boldsymbol{\mu}_1)(\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1) + \boldsymbol{R}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\lambda}_1,\boldsymbol{\mu}_1),$$

where $|\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\lambda}_1,\boldsymbol{\mu}_1)|/(|\boldsymbol{x}-\boldsymbol{y}| + |\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1|) \to \boldsymbol{0}$ as $(|\boldsymbol{x}-\boldsymbol{y}| + |\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1|) \to \boldsymbol{0}$. Using the above, we get

$$\int_0^T \langle \boldsymbol{x}, \boldsymbol{\psi} \rangle \, \mathrm{d}s = \int_0^T \langle \boldsymbol{y}, \boldsymbol{\psi} \rangle \, \mathrm{d}s + \langle \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0, \boldsymbol{\phi}(0) \rangle + \int_0^T \langle D_{\lambda_1} \boldsymbol{f}(\boldsymbol{y}; \boldsymbol{\mu}_1)(\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1), \boldsymbol{\phi} \rangle \, \mathrm{d}s + \int_0^T \langle \boldsymbol{R}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\lambda}_1, \boldsymbol{\mu}_1), \boldsymbol{\phi} \rangle \, \mathrm{d}s. \tag{33}$$

We interpret the second term as the effect of varying the initial condition and the third as the effect of varying the parameters in the model itself. Note that even when $\boldsymbol{f}$ is linear in $\boldsymbol{x}$ and in $\boldsymbol{\lambda}_1$, $\boldsymbol{R}$ is typically not zero since $\boldsymbol{f}$ is not linear in their combination.

9.2. Proof of Theorem 2

We use the notation of the proof in Section 9.2 and apply the definition of the derivative with $\boldsymbol{x} = \boldsymbol{x}(s;\boldsymbol{\mu} + \boldsymbol{h})$ to get

$$\begin{aligned} v(\boldsymbol{h}) &= \left| \int_0^T \langle \boldsymbol{x}, \boldsymbol{\psi} \rangle \, \mathrm{d}s - \int_0^T \langle \boldsymbol{y}, \boldsymbol{\psi} \rangle \, \mathrm{d}s - \left(\langle \boldsymbol{h}_0, \boldsymbol{\phi}(0) \rangle + \int_0^T \langle D_{\lambda_1} \boldsymbol{f}(\boldsymbol{y}; \boldsymbol{\mu}_1) \boldsymbol{h}_1, \boldsymbol{\phi} \rangle \, \mathrm{d}s \right) \right| \\ &\leqslant \int_0^T |\boldsymbol{R}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu}_1 + \boldsymbol{h}_1, \boldsymbol{\mu}_1)||\boldsymbol{\phi}| \, \mathrm{d}s. \end{aligned}$$

Writing the solution to (3) as

$$\boldsymbol{x}(t; \boldsymbol{\lambda}) = \boldsymbol{\lambda}_0 + \int_0^t \boldsymbol{f}(\boldsymbol{x}(s; \boldsymbol{\lambda}); \boldsymbol{\lambda}_1) \, \mathrm{d}s$$

and using the Lipschitz property of $\boldsymbol{f}$, we get

$$|\boldsymbol{x}(t) - \boldsymbol{y}(t)| \leqslant |\boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0| + \int_0^t |\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{\mu}_1 + \boldsymbol{h}_1) - \boldsymbol{f}(\boldsymbol{y}; \boldsymbol{\mu}_1)| \, \mathrm{d}s \leqslant L \int_0^t |\boldsymbol{x} - \boldsymbol{y}| \, \mathrm{d}s + (LT + 1)|\boldsymbol{h}|,$$

where ($|\boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0| \leqslant |\boldsymbol{h}|$). By Gronwall's lemma [12],

$$|\boldsymbol{x}(t) - \boldsymbol{y}(t)| \leqslant (LT + 1)|\boldsymbol{h}|\mathrm{e}^{Lt}.$$

We choose δ so ($|\boldsymbol{x} - \boldsymbol{y}| + |\boldsymbol{\lambda}_1 - \boldsymbol{\mu}_1|) \leqslant \delta$ implies $\boldsymbol{R}(\boldsymbol{x}, \boldsymbol{y};\ \boldsymbol{\mu}_1 + \boldsymbol{h}_1, \boldsymbol{\mu}_1) \leqslant \epsilon(|\boldsymbol{x} - \boldsymbol{y}| + |\boldsymbol{h}_1|) \leqslant \epsilon(|\boldsymbol{x} - \boldsymbol{y}| + |\boldsymbol{h}|)$. For $\boldsymbol{h}$ sufficiently small ($|\boldsymbol{h}| \leqslant \delta_1$),

$$|\boldsymbol{x}(s) - \boldsymbol{y}(s)| + |\boldsymbol{h}| \leqslant (LT + 1)|\boldsymbol{h}|\mathrm{e}^{LT} + |\boldsymbol{h}| \leqslant \delta.$$

For $|\boldsymbol{h}| \leqslant \delta_1$,

$$|v(\boldsymbol{h})| \leqslant \int_0^T |\boldsymbol{R}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu}_1 + \boldsymbol{h}_1, \boldsymbol{\mu}_1)||\boldsymbol{\phi}| \, \mathrm{d}s \leqslant \epsilon \left[((LT + 1)\mathrm{e}^{LT} + 1) \int_0^T |\boldsymbol{\phi}| \, \mathrm{d}s \right] |\boldsymbol{h}|$$

and so $|v(\boldsymbol{h})|/|\boldsymbol{h}| \to 0$ as $|\boldsymbol{h}| \to 0$. This proves the result.

10. Details of implementation

The implementation of the proposed algorithms requires a number of numerical approximations and this section contains some of the relevant details.

10.1. Obtaining density plots

At the end of the RAPS and FAPS procedures, the piecewise constant function $\tilde{q}(\boldsymbol{\lambda})$ leads immediately to an approximation of $F_{\tilde{q}}(x) = \sum_{q(\boldsymbol{\mu}_i) \leqslant x} \mu_{\boldsymbol{\lambda}}(R_i)$. At the end of the HOPS procedure, the piecewise linear function $\tilde{q}(\boldsymbol{\lambda})$ may either be converted to a piecewise constant by applying the FAPS procedure to this linear approximation, or be converted into a sample of points by applying $\tilde{q}$ to a sample of points from $\boldsymbol{\lambda}$. In any case, the function

$$F_{\tilde{q}}(x) = \frac{1}{N}\{\text{Number of points} \leqslant x\}$$

naturally approximates F_q.

The cumulative distribution functions created by the FAPS and HOPS algorithms are not differentiable, and so these approximating distributions do not have associated densities. To approximate f_q, we take any density $k(x)$, define $k_\zeta(x) = k(x/\zeta)/\zeta$, $K_\zeta = \int_{-\infty}^{x} k_\zeta(x)\,\mathrm{d}x$, and

$$K_\zeta * F_{\tilde{q}}(x) = \int_{-\infty}^{\infty} K_\zeta(x-y)\,\mathrm{d}F_{\tilde{q}}(y).$$

This gives an approximation $K_\zeta * F_{\tilde{q}}$ to $F_{\tilde{q}}$ that is differentiable and

$$(K_\zeta * F_{\tilde{q}})'(x) = k_\zeta * F_{\tilde{q}}(x) = \int_{-\infty}^{\infty} k_\zeta(x-y)\,\mathrm{d}F_{\tilde{q}}(y) = \frac{1}{\zeta}\sum_{i=1}^{N} k\left(\frac{x-q(\boldsymbol{\mu}_i)}{\zeta}\right)\boldsymbol{\mu}_{\boldsymbol{\lambda}}(R_i)$$

in the case $F_{\tilde{q}}$ is computed from the piecewise constant approximation, or

$$(K_\zeta * F_{\tilde{q}})'(x) = \frac{1}{\zeta N}\sum_{i=1}^{N} k\left(\frac{x-q(\boldsymbol{\mu}_i)}{\zeta}\right)$$

when the cumulative distribution function is constructed from a data set of N points. A typical choice of k is the standard $N(0,1)$ density function.

As $\zeta \to 0$ and $F_{\tilde{q}} \to F_q$, formally

$$\begin{aligned}|K_\zeta * F_{\tilde{q}}(x) - F_q(x)| &= \int_{-\infty}^{\infty} \big(F_{\tilde{q}}(x-y) - F_q(x-y)\big)k_\zeta(y)\,\mathrm{d}y + \int_{-\infty}^{\infty} \big(F_q(x-y) - F_q(x)\big)k_\zeta(y)\,\mathrm{d}y \\ &\to 0 + \delta(y-0)\big(F_q(x-y) - F(x)\big) = 0,\end{aligned}$$

and so $K_\zeta * F_{\tilde{q}}$ is a differentiable cumulative distribution function that approximates F_q. Choosing the "bandwidth" parameter ζ is somewhat subjective, and is a balance between increasing the variance of the approximating density on the one hand, or the bias on the other [13].

10.2. Computing error integrals

The FAPS method requires the evaluation of the errors $\mathscr{E}_{i,k}^{\mathrm{pc}}$ (25) with respect to the measure $\mu_{\boldsymbol{\lambda}}$. There are at least two ways to do this, depending on how the distribution of $\boldsymbol{\lambda}$ is represented numerically.

In the first method a list of M data points representing the distribution of $\boldsymbol{\lambda}$ is read in. The measure $\mu_{\boldsymbol{\lambda}}$ is approximated by the discrete measure that assigns

$$\mu_{\boldsymbol{\lambda}}(A) = \frac{1}{M}(\text{Number of points in } A).$$

The approximate error then becomes

$$\mathscr{E}_{i,k}^{\mathrm{pc}} \approx \frac{1}{M}\sum_{z_j \in R_i} |\partial_{\lambda_k} q(\boldsymbol{\mu}_i)(z_j^k - \mu_i^k)|,$$

where the $z_j = (z_j^1, \ldots, z_j^d)^{\mathsf{T}}$ are the points from the sample that lie in R_i.

If instead we choose to represent the distribution analytically, where the distribution has a density $\rho_{\boldsymbol{\lambda}}(\boldsymbol{z})$, the integral becomes

$$\mathscr{E}_{i,k}^{\mathrm{pc}} = \int_{R_i} |\partial_{\lambda_k} q(\boldsymbol{\mu}_i)(z^k - \mu_i^k)|\rho_{\boldsymbol{\lambda}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z},$$

which may be computed using standard quadrature rules. In the special case that the components of $\boldsymbol{\lambda}$ are independent, then the density may be factored $\rho_{\boldsymbol{\lambda}}(\boldsymbol{z}) = \prod_{i=1}^{d}\rho_{\lambda_i}(z^i)$ and so the error becomes

$$\mathscr{E}_{i,k}^{\mathrm{pc}} = \int_{R_i} |\partial_{\lambda_k} q(\boldsymbol{\mu}_i)(z^k - \mu_i^k)\rho_{\lambda_1}(z^1)\cdots\rho_{\lambda_p}(z^d)\,\mathrm{d}\boldsymbol{z} = \frac{\mu_{\boldsymbol{\lambda}}(R_i)}{b}\int_{R_i^k} |\partial_{\lambda_k} q(\boldsymbol{\mu}_i)(z^k - \mu_i^k)|\rho_{\lambda_k}(z^k)\,\mathrm{d}z^k,$$

where R_i^k is the interval along the kth side of R_i and $b = \int_{R_i^k} \rho_{\lambda_k} \, \mathrm{d}z^k$. This integral for large d is much simpler to compute, since it involves only two 1-dimensional integrals as opposed to an integral over $\mathbb{R}^d$. In particular, the curse of dimensionality is avoided in this case.

For a non-independent density, Markov-Chain Monte-Carlo methods may be used to compute the integral.

10.3. Computation of gradients

In practice, we want to numerically approximate the derivative in (8) and apply the derivative to points $\boldsymbol{\lambda} - \boldsymbol{\mu}$. It is necessary to rework the integral (second) term in the derivative, since we wish to avoid calculating an integral for many points in the distribution of $\boldsymbol{\lambda}$. We may rewrite the derivative with respect to the $\boldsymbol{\lambda}_1$ as

$$\partial_{\lambda_1} q(\boldsymbol{\mu})[\cdot] = \left\langle [\cdot], \int_0^T D_{\lambda_1} \boldsymbol{f}(\boldsymbol{y}; \boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\phi} \, \mathrm{d}s \right\rangle,$$

so that we do not have to integrate random variables to calculate the derivative. Rather, we calculate the derivative and perform a dot product.

We also note that for $i \leqslant p$,

$$\partial_{\lambda_i} q(\boldsymbol{\mu}) = \partial_{\boldsymbol{\lambda}} q(\boldsymbol{\mu}) \boldsymbol{e}_i = \sum_{j=1}^{n} \int_0^T \partial_{\lambda_i} f_j(\boldsymbol{y}; \boldsymbol{\mu}_1) \phi_j \, \mathrm{d}s, \tag{34}$$

where $\boldsymbol{f} = (f_1, \ldots, f_n)^{\mathrm{T}}$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^{\mathrm{T}}$. Once we have numerically solved for $\boldsymbol{y}$ and $\boldsymbol{\phi}$, the integrals may be computed on the time grid by any quadrature rule. For instance, if we solve the forward and adjoint problems on the mesh $t_0 = 0 < t_1 < \cdots < t_m = T$,

$$\int_0^T \partial_{\lambda_i} f_j(\boldsymbol{y}\,; \boldsymbol{\mu}_1) \phi_j \, \mathrm{d}s \approx \sum_{k=0}^{m-1} \frac{1}{2} (\partial_{\lambda_i} f_j(\boldsymbol{y}(t_{k+1}); \boldsymbol{\mu}_1) \phi_j(t_{k+1}) + \partial_{\lambda_i} f_j(\boldsymbol{y}(t_k); \boldsymbol{\mu}_1) \phi_j(t_k))(t_{j+1} - t_j) \tag{35}$$

is the trapezoidal rule.

Acknowledgement

The research activities of D. Estep are partially supported by the Department of Energy through Grant DE-FG02-04ER25620 and DE-FG02-05ER25699, the National Aeronautics and Space Administration through Grant NNG04GH63G, the National Science Foundation through Grants DMS-0107832, DGE-0221595003, and MSPA-CSE-0434354, the Sandia Corporation through Contract No. PO299784, and the United States Department of Agriculture through Contract No. 58-5402-3-306.

The research activities of D. Neckels are partially supported by the National Science Foundation through Grants DGE-0221595003 and MSPA-CSE-0434354.

References

[1] K. Eriksson, D. Estep, P. Hansbo, C. Johnson, Computational Differential Equations, Cambridge University Press, New York, 1996.
[2] K. Eriksson, D. Estep, P. Hansbo, C. Johnson, Introduction to adaptive methods for differential equations, Acta Numerica (1995) 105–158.
[3] D. Estep, A posteriori error bounds and global error control for approximations of ordinary differential equations, SIAM J. Numer. Anal. 32 (1995) 1–48.
[4] D.J. Estep, M.G. Larson, R.D. Williams, Estimating the error of numerical solutions of systems of reaction–diffusion equations, Mem. Amer. Math. Soc. 146 (696) (2000) viii+109.
[5] D. Estep, M. Holst, M. Larson, Generalized Green's functions and the effective domain of influence, SIAM J. Sci. Comput. 26 (2005) 1314–1339.
[6] R. Renka, Multivariate interpolation of large sets of scattered data, ACM Trans. Math. Software 14 (1988) 139–148.
[7] P. Billingsley, Probability and measure, Wiley Series in Probability and Mathematical Statistics, third ed., Wiley, New York, 1995 (a Wiley-Interscience Publication).
[8] G.S. Fishman, Monte-Carlo, Springer Series in Operations Research, Springer-Verlag, New York, 1996, concepts, algorithms, and applications.

[9] D. Estep, D. Neckels, Applying adjoint techniques to the methods of stratified sampling and control variates in Monte-Carlo computations (in preparation).
[10] J.M. Hammersley, D.C. Handscomb, Monte-Carlo Methods, Methuen & Co. Ltd., London, 1965.
[11] D. Estep, C. Johnson, The computability of the Lorenz system, Math. Models Meth. Appl. Sci. 8 (1998) 1277–1305.
[12] L. Perko, Differential equations and dynamical systems, third ed.Texts in Applied Mathematics, vol. 7, Springer-Verlag, New York, 2001.
[13] W.R. Pestman, Mathematical Statistics, Walter de Gruyter, New York, 1998.