



Published in final edited form as:

SIAM J Numer Anal. 2011 September 1; 49(5): 1836–1859. doi:10.1137/100785946.

A MEASURE-THEORETIC COMPUTATIONAL METHOD FOR INVERSE SENSITIVITY PROBLEMS I: METHOD AND ANALYSIS

J. Breidt[†], T. Butler[‡], and D. Estep[†]

J. Breidt: jbreidt@stat.colostate.edu; T. Butler: tbutler@ices.utexas.edu; D. Estep: estep@math.colostate.edu

[†]Department of Statistics, Colorado State University, Fort Collins, CO 80523

[‡]Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712

Abstract

We consider the inverse sensitivity analysis problem of quantifying the uncertainty of inputs to a deterministic map given specified uncertainty in a linear functional of the output of the map. This is a version of the model calibration or parameter estimation problem for a deterministic map. We assume that the uncertainty in the quantity of interest is represented by a random variable with a given distribution, and we use the law of total probability to express the inverse problem for the corresponding probability measure on the input space. Assuming that the map from the input space to the quantity of interest is smooth, we solve the generally ill-posed inverse problem by using the implicit function theorem to derive a method for approximating the set-valued inverse that provides an approximate quotient space representation of the input space. We then derive an efficient 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.

Keywords

adjoint problem; density estimation; inverse sensitivity analysis; model calibration; nonparametric density estimation; parameter estimation; sensitivity analysis; set-valued inverse

1. Introduction

We develop and analyze a numerical method to solve the inverse sensitivity analysis problem: Given a specified variation and/or uncertainty in the output of a smooth map, determine variations in the input parameters that produce the observed uncertainty. We formulate this inverse problem using probability to describe variation by assuming that the inputs and outputs are random variables. This inverse problem has an abstract interpretation in which the density is imposed on the output in order to observe the consequences for the inputs. It also has an experimental interpretation in which the model output matches observed values of an experiment and the imposed density is associated with the experimental data, i.e., reflecting the uncertainty in the data or arising as a consequence of experimental error.

To motivate this inverse sensitivity analysis problem, consider the situation of a manufacturer who will purchase a large number of metal plates of a given alloy and thickness that are to be used subsequently in a high temperature environment. In order to ensure the plates maintain integrity, the manufacturer specifies that a given heat load must be distributed quasi-uniformly after ten minutes of exposure, with some conditions on how much the temperature may vary through the plate. The plates are milled with variations in

the purity of the alloy and the thickness of the plates, both of which affect the heat distribution under load. To check a batch of plates to see if it meets the requirements, the manufacturer tests the heat specification on a random sample of plates drawn from the batch. The random selection of samples, the variation in plate properties, and measurement error combined lead to a description of the test results as a random variable. After delivery, the manufacturer decides that knowing the statistics on the size of the plates and the composition of the alloy would be useful. The heat equation models the heat distribution under a given load once the conductivity determined by the alloy composition and the thickness of the plates are specified. The inverse sensitivity problem is to determine the distribution on the space of parameters consisting of the thickness and alloy purity from the distribution of the results of the heat experiments on the plates.

The probabilistic inverse problem can be described more precisely as follows. Given

- a model $\mathcal{M}(Y, \lambda)$ with solution $Y = G(\lambda)$ depending on parameters and data λ in parameter space $\Lambda \subset \mathbb{R}^d$,
- a linear functional $q(\lambda) = q(Y(\lambda))$ taking values in an output space \mathcal{D}
- an *observed* probability density $\rho_{\mathcal{D}}(q(\lambda)) = \rho_{\mathcal{D}}(q(Y(\lambda)))$ on the output value $q(\lambda)$, determine
- a probability density $\sigma_{\Lambda}(\lambda)$ on the parameter space Λ that produces the observed density.

We assume the model $\mathcal{M}(Y, \lambda)$ depends smoothly on the inputs, so the map $q(\lambda)$ is implicitly a smooth and *deterministic* function of λ .

There are several important issues associated with this problem. In general, the parameter space is multidimensional while there is a single observation (or a low dimensional set of observations at most). So, the inverse problem is ill-posed in the sense that the inverse solution of the deterministic model is set-valued. Under the assumption of a smooth model, we address this issue by constructing a systematic method for approximating set-valued inverses. Second, we are particularly interested in models that are complicated and/or expensive to evaluate, e.g., requiring the solution of a differential equation, so that the map to the output is determined implicitly. We address this issue by using adjoint operators [22, 20, 6, 21, 23, 12, 13, 9, 10, 7, 11] to compute the required derivative information. Third, while probability densities describe random variables, the densities themselves are not random. Common approaches to approximating probability densities often use a random representation obtained by some variation of Monte Carlo sampling [14, 17, 18]; however, this is not a requirement. In particular, the approach described in this paper is not stochastic, rather it is based on the simple approximation commonly used in measure theory.

In this paper, we present the basic method and analysis of a measure-theoretic computational approach for the probabilistic inverse sensitivity analysis problem. In [4], we present a numerical analysis of the discretization error that arises when evaluating the model by numerical solution and using a finite number of random samples to represent the distribution on the output quantity. In [5], we discuss the problem of dealing with multiple quantities of interest, which has application to data assimilation and “cascaded” uncertainty in operator decomposition solution of multiphysics problems.

This paper is structured as follows. In section 2, we formulate the probabilistic inverse problem that we solve and discuss the relation to a Bayesian inverse problem. In section 3.1, we deal with the set-valued nature of the inverse problem by introducing a theory of generalized contours and explain how the generalized contours can be approximated. In section 3.2, we develop a computational measure theoretic method for approximating the

inverse parameter distribution using approximate generalized contours. In section 4, we apply the method to a variety of problems. Finally, section 5 summarizes the work.

2. Formulation of the probabilistic inverse problem

The inverse problem we study is the direct inversion of the forward stochastic sensitivity analysis problem for a deterministic model. We consider a deterministic operator $q(\lambda)$ that maps values in a parameter space Λ to an output space \mathcal{D} . We assume there is a *parameter volume* measure μ_Λ on Λ that determines the volume of sets in Λ . 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. 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 . We assume that μ_Λ is absolutely continuous with respect to the Lebesgue measure and the volume V of Λ is finite.

We first describe the forward stochastic sensitivity analysis for the deterministic map $q(\lambda)$. We assume that a probability density $\sigma_\Lambda(\lambda)$ is specified on the parameter space Λ . This density distinguishes the probability of different events in Λ , i.e., the probability of an event A in Λ , by which we mean a measurable set of values, is computed via

$$P(A) = \int_A \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda).$$

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 $q = q(\lambda)$. This implies the fundamental relationship

$$\text{Law of Total Probability } \rho_{\mathcal{D}}(q|A) = \frac{\int_A L(q|\lambda) \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda)}{\int_A \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda)}. \quad (2.1)$$

This is a Fredholm integral equation of the first kind that determines a conditional probability density $\rho_{\mathcal{D}}(q|A)$ on the output given that the parameters come from A . Thus, we may determine the conditional probability of event $B \subset \mathcal{D}$ as

$$P(B|A) = \int_B \rho_{\mathcal{D}}(q|A) d\mu_{\mathcal{D}}(q) = \frac{\int_B \int_A L(q|\lambda) \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda) d\mu_{\mathcal{D}}(q)}{\int_A \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda)}.$$

For forward sensitivity analysis it is common to take $A = \Lambda$ so that $P(B|A) = P(B)$, and we arrive at the common form for the law of total probability given by

$$\rho_{\mathcal{D}}(q) = \int_\Lambda L(q|\lambda) \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda). \quad (2.2)$$

This describes an analogue of a Perron–Frobenius map where the deterministic map $q(\lambda)$ defines a transformation of the density $\sigma_\Lambda(\lambda)$ to $\rho_{\mathcal{D}}(q)$. This forward sensitivity analysis problem is often solved using a Monte Carlo approach: Random parameter sample values λ are drawn from the distribution σ_Λ on the parameter space; corresponding values of $q(\lambda)$ are computed; and these values are binned to produce an approximate probability distribution on the output.

The stochastic inverse sensitivity analysis problem that we study is the inversion of the *Law of Total Probability* (2.2).

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 (2.2).

It is important to note that what we seek for the solution of the inverse problem is the actual probability density that can be used to compute the probability of events in the parameter space Λ . In other words, we seek to compute the inverse of the analogue of the Perron–Frobenius map between the densities on the input and output spaces. The purpose of this paper is to describe a method for solving the inverse problem by providing a way to approximate the probability of an arbitrary event in the input space. This can be used subsequently to generate an approximation of the inverse density and/or to compute any desired statistical moments of the inverse density.

We emphasize the fundamental role of the underlying parameter volume measure μ_{Λ} in defining the solution of the inverse problem. In particular, the a priori specification of μ_{Λ} imposes the structure of the measure on Λ , e.g., whether the measure on Λ is a product measure or not. In general, there are many combinations of σ_{Λ} and μ_{Λ} that can yield a given observed density on the output.

We provide a simple illustration of the inverse problem using the map

$$q(\lambda) = \lambda_1 + \lambda_2,$$

where λ_1, λ_2 are random variables. For the inverse problem, we specify that $q(\lambda)$ has a $\mathcal{N}(0, 2/25)$ distribution and seek to determine the parameter distribution $\sigma_{\Lambda}(\lambda)$ that yields the specified output density. This output distribution can be generated by choosing λ_1, λ_2 to be independent identically distributed $\mathcal{N}(0, 1/25)$ random variables; see Figure 2.1. As well, we could choose any bivariate normal density

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

If we find a distribution on Λ that generates $q(\lambda)$ according to a $\mathcal{N}(0, 2/25)$ distribution, then we accept this as a solution to the inverse problem. The choice of the underlying parameter volume measure μ_{Λ} is critical to this task. In Figures 2.1–2.3, we show five different probability densities $\sigma_{\Lambda}(\lambda)$ that yield the identical $\mathcal{N}(0, 2/25)$ density on $q(\lambda)$. Each of the five different densities correspond to five different underlying volume distributions μ_{Λ} as shown.

The specification of μ_{Λ} has to do with how measurements in Λ are carried out and the relationships between the parameters. As noted, the volume measure should be specified as part of defining the model. In many situations involving deterministic models, the product Lebesgue measure appropriately scaled to account for units is the natural choice. But, this is not always the case. Continuing the motivating problem, as a first approximation, we might consider the thickness and alloy composition to be physically independent parameters and impose a product measure on the space formed by the two variables using independent normalized Lebesgue measures. A more realistic description will take into account the fact that the thickness of the plates indirectly depends on the alloy composition during the milling process. We can model the milling process to determine the thickness as an indirect

function of the physically independent variables of pressure in the milling process and the alloy composition. The measure on the space consisting of the thickness and alloy composition is then determined by propagating the product measure imposed on the independent alloy composition and pressure variables through the milling model. The resulting measure on the space consisting of the alloy composition and thickness will not be a product measure.

The plots of inverse densities given in Figures 2.2–2.3 also illustrate the important point that injecting probability into the inverse problem by itself does not reduce the ill-posedness, even after specifying the parameter volume measure. The consequence of ill-posedness on the stochastic inverse problem is illustrated by the complex measure structure of the inverse probability densities in the plots. For example, these densities are not product measures. In general, it is not possible to determine densities for the individual parameters without further information. We can determine only a measure on the entire parameter space.

Comparison to a Bayesian inverse problem

There is another natural inverse problem associated with the *Law of Total Probability* (2.1) that is important in the case of a *general* likelihood function $L(q|\lambda)$, not necessarily arising from a deterministic map. Namely, we may use Bayes' theorem to invert the likelihood function to obtain the “posterior density” $p(\lambda|q)$ given the “prior density” σ_{Λ} on the input space Λ and a “data density” $\rho_{\mathcal{D}}$ on the output space \mathcal{D} . We emphasize that the solution of this Bayesian inverse problem is a conditional distribution. This is very natural when the map from the input to output space has been modeled statistically by specifying $L(q|\lambda)$ given information about the statistical properties of the input parameters and output quantity, e.g., when the map is derived empirically, rather than from physical principles.

This Bayesian inverse problem is at the heart of Bayesian inference [26, 1, 19, 18]. In this approach, the inferential target is a single, unknown parameter (or parameter vector) λ . We are given data in the form of observations q_1, \dots, q_n for which a typical assumption is conditional independence,

$$p(q_1, \dots, q_n|\lambda) \sim \prod_{i=1}^n p(q_i|\lambda) \quad (2.3)$$

where $\{p(q_i|\lambda)\}$ are conditional probability densities with respect to some appropriate measure, and are specified up to the value of λ . The right-hand side of (2.3) is the likelihood of the observations given the parameter. We are also given a prior distribution on λ that gives a probabilistic description of the uncertainty about the values of λ before any data are observed. This prior distribution is exactly $\sigma_{\Lambda}(\lambda)$ in the notation used above. Bayesian inference then proceeds by using Bayes' theorem to compute the posterior conditional distribution of λ given the observations q_1, \dots, q_n :

$$p(\lambda|q_1, \dots, q_n) \propto p(q_1, \dots, q_n|\lambda)\sigma_{\Lambda}(\lambda) = \prod_{i=1}^n p(q_i|\lambda)\sigma_{\Lambda}(\lambda). \quad (2.4)$$

We could adopt a Bayesian approach to solve the inverse problem we study by modeling $\sigma_{\Lambda}(\lambda)$ parametrically as $\sigma_{\Lambda}(\lambda|\theta)$ in terms of new (lower-dimensional) parameters θ . This is known as a *mixture* or *hierarchical* model. In Bayesian terminology, $\sigma_{\Lambda}(\lambda|\theta)$ is the prior while a new distribution σ_{θ} describing θ is the hyperprior. Assuming that the hyperprior is specified, we then compute the posterior distribution on θ given “data” from $\rho_{\mathcal{D}}(q(\lambda))$. Any desired inferences about the distribution of λ given θ can then be obtained from the

posterior. The difficulty with this approach is specifying a reasonable conditional model, which is difficult to verify empirically.

The inverse problem solved in this paper shares some characteristics with the Bayesian inverse problem, but has fundamental differences as well. In the Bayesian problem, the inferential target is the *parameter* λ , and σ_{Λ} is *given* as prior information. The likelihood $L(q|\lambda)$ typically involves a nontrivial stochastic structure and is *not deterministic*.

By contrast, in the inverse problem we solve the inferential target is the *distribution* σ_{Λ} , which is *not given* as the prior. Further, our likelihood $L(q|\lambda)$ is given by a *deterministic* map, which completely determines the set-valued inverse.

The choice of inverse problem to solve depends completely on the available information. In the case of a deterministic physics-based model, the unknowns and quantities subject to uncertainty are the data and parameter values that are input into the model and the observations that are supposed to match model output while the likelihood function determined by the map is completely trivial in a statistical/probabilistic sense. Based on the law of total probability, the inverse problem we solve is the direct inverse of the probabilistic forward sensitivity problem for a deterministic model.

3. Solving the inverse problem

As noted above, while probability densities describe the random nature of a random variable, the densities themselves are not random. While a common approach to compute a discrete approximation of a probability density employs random sampling, this is not necessary. In this paper, we describe a method for computing approximate probability densities that does not require random sampling. Our approach breaks the solution down into two stages:

1. Construct an approximate representation of the set-valued inverse solution of the deterministic model.
2. Use measure-theoretic computational methods to approximate the probability density (measure) structure on the parameter space that corresponds to the set-valued inverse and the observed output density.

These are independently interesting tasks.

We present a brief overview before providing the details. Under the assumption of a smooth map, if we are given a fixed output value $\bar{q} \in \mathcal{D}$ then the implicit function theorem guarantees the existence of a $(d-1)$ -dimensional manifold in Λ that is mapped to \bar{q} . Motivation comes from the two-dimensional case, $\lambda = (\lambda_1, \lambda_2)$, where the manifolds are contours of the surface $q(\lambda_1, \lambda_2)$ (left-hand illustration in Figure 3.1). Every point in Λ lies on a unique contour, so we may consider Λ as a set described by its contours. The set of (generalized) contours is an equivalence class in the input space, i.e., a quotient space representation of the input space. In Λ , there exists 1-dimensional curves transverse to the contours that intersect each contour once and only once (right-hand illustration in Figure 3.1). We can take one of these curves as the index for the set of contours. There is a bijection between the points on an index curve and the points in the range of the output $q(\Lambda)$. Therefore, any measure posed on the range of the output imposes a measure on the index curve. Thus, the intersections of the contours with the index curve is a random variable with a distribution uniquely defined by the distribution of the output $p_{\mathcal{D}}(q(\lambda))$ (left-hand illustration in Figure 3.2). In other words, there exists a unique solution to the inverse sensitivity analysis problem in the set of the contours.

However, determining the set of contours analytically is infeasible in practice. In [23], the forward sensitivity analysis problem defined by (2.2), where a given density $\sigma_{\Lambda}(\lambda)$ is propagated through the output surface $q(\lambda)$, is solved using a piecewise-linear tangent plane approximation to the output surface. This requires computations involving only inner products, which is cheap compared to the full model evaluation cost of $q(\lambda)$ for each new value of λ . The derivatives of $q(\lambda)$ are computed implicitly using adjoint methods. Motivated by this approach, we use a piecewise-linear tangent plane approximation to the output surface $q(\lambda)$ to construct approximate contours and an approximate index set.

The next step is to determine the probability density on the parameter set that corresponds to the distribution on the transverse parameterization of the space of approximate contours. In order to assign a probability to a measurable set in Λ , we first recognize that such a set is defined by the contours it contains and the amount of each contour it contains (right-hand illustration in Figure 3.2). The parameter volume measure μ_{Λ} specified on Λ quantifies the amount of each contour contained in any given set. Combining the results of the generalized contours with such a measure, the monotone convergence theorem, and additivity properties of measures, we develop an algorithm to estimate the probability of any measurable set in Λ . This algorithm employs a piecewise constant approximation of measures that is commonly used in measure theory. This yields a direct computational method to approximate $\sigma_{\Lambda}(\lambda)$.

In the next two sections, we provide details of the two ingredients of the approximate solution method.

Remark 3.1. Many solution methods for both statistical and deterministic inverse problems deal with ill-posedness by introducing some form of regularization, either directly or reposing the inverse problem as an optimization problem. Such methods avoid the need to deal with set-valued inverse solutions.

Remark 3.2. There are cases of interest, e.g., a parameter domain that contains a bifurcation point, for which the described method cannot be used in a straightforward fashion. We note that while an approach based on random sampling may be applied nominally to such a problem, the interpretation of the results is still problematic.

Remark 3.3. While the solution method for the inverse problem proposed here relies on derivatives of a quantity of interest, it is not dependent on how those derivatives are computed. Instead of an adjoint-based approach, the derivatives might be computed using (deterministic) forward sensitivity analysis that computes the derivatives directly along with the solution of the model. Yet another approach, presented, e.g., in [27], employs a stochastic spectral method to obtain a polynomial representation of $q(\lambda)$, which is then used to compute gradients.

3.1. Determining the inverse of the deterministic model using generalized contours

We consider a finite dimensional map q from the space of parameters to the output defined implicitly by solving a finite dimensional nonlinear system of equations,

$$f(x;\lambda)=b, \quad (3.1)$$

where $x \in \mathbb{R}^n$, parameter $\lambda \in \Lambda \subset \mathbb{R}^d$ (assuming that Λ is compact) is a random vector, and $f: \mathbb{R}^{n+d} \rightarrow \mathbb{R}^n$ is assumed smooth in both variables. The goal is to compute a quantity of interest $q(\lambda) = q(x(\lambda)) = \langle x, \psi \rangle$, described as a linear functional of the solution $x(\lambda)$. If x depends smoothly on λ , then the dependence of q on λ is also smooth.

This problem applies in particular to differential equations that depend on a finite set of parameters. For differential equations, we require the same assumptions as the standard existence and uniqueness theorems to guarantee the smoothness of $q(\lambda)$. This is discussed in more detail in the second part of this paper [4].

For any $\bar{q} \in q(\Lambda)$, we define $\check{q}(\lambda) := q(\lambda) - \bar{q}$. By assumption, $\check{q}(\lambda) : \mathbb{R}^d \rightarrow \mathbb{R}$ is continuously differentiable and there exists $\bar{\lambda} \in \Lambda$ such that $q(\bar{\lambda}) = \bar{q}$, which implies that $\check{q}(\bar{\lambda}) = 0$. We are mainly interested in the case where the quantity of interest varies as the parameters vary, so we assume that $\lambda_d \check{q}(\bar{\lambda}) \neq 0$, i.e. there is at least one nontrivial partial derivative. We may relax the restriction of $\lambda_d \check{q}(\bar{\lambda}) \neq 0$ for a finite number of points in Λ , where $q(\lambda)$ possibly attains a local extreme value and ignore this set of points when considering the generalized contours.

By the implicit function theorem, there exists an open set $U_{\bar{\lambda}} \subset \Lambda^{d-1}$, where $\Lambda^{d-1} := \{\lambda^{d-1} := (\lambda_1, \dots, \lambda_{d-1}) | \lambda = (\lambda_1, \dots, \lambda_d) \in \Lambda\}$, containing λ^{d-1} , an open set $V_{\bar{\lambda}} \subset \Lambda_d$ where $\Lambda_d := \{\lambda_d | \lambda \in \Lambda\}$, and a differentiable function $g_{\bar{\lambda}} : U_{\bar{\lambda}} \rightarrow V_{\bar{\lambda}}$ such that

$$\{(\lambda^{d-1}, g_{\bar{\lambda}}(\lambda^{d-1}))\} = \{\lambda | q(\lambda) = \bar{q}\} \cap (U_{\bar{\lambda}} \times V_{\bar{\lambda}}). \quad (3.2)$$

Since the implicit function theorem is a local result, there may be additional points in Λ that map to \bar{q} , but are not contained in the set defined by (3.2). Thus, given $\bar{q} \in q(\Lambda)$, we choose a collection of sets $\{U_{\bar{\lambda}} \times V_{\bar{\lambda}}\} = \cup_{\alpha \in A} \{U_{\bar{\lambda}_\alpha} \times V_{\bar{\lambda}_\alpha}\}$, where $\cup_{\alpha \in A} \{\bar{\lambda}_\alpha\}$ is the set of all $\lambda \in \Lambda$ such that $q(\lambda) = \bar{q}$. Then using the same notation as in (3.2), the function $g_{\bar{\lambda}}(\lambda^{d-1})$ might be piecewise defined. The set in (3.2) is a $(d-1)$ -dimensional manifold that is a natural inverse of $q(\lambda)$ given \bar{q} . We call this set the *generalized contour*.

Theorem 3.1. *If we choose distinct $\bar{q}_1, \bar{q}_2 \in q(\Lambda)$, then the generalized contours for \bar{q}_1 and \bar{q}_2 are unique and do not intersect.*

Proof. The nonintersection property follows immediately from the fact that $q(\lambda)$ is a function. Uniqueness follows immediately from the choice $\{U_{\bar{\lambda}} \times V_{\bar{\lambda}}\} = \cup_{\alpha \in A} \{U_{\bar{\lambda}_\alpha} \times V_{\bar{\lambda}_\alpha}\}$, where $\cup_{\alpha \in A} \{\bar{\lambda}_\alpha\}$ is the set of all $\lambda \in \Lambda$ such that $q(\lambda) = \bar{q}$ for a given value of $\bar{q} \in q(\Lambda)$.

In two dimensions, the generalized contours are simply contours of the surface $q(\lambda_1, \lambda_2)$. We denote a generalized contour for a specific quantity of interest \bar{q} as $q^{-1}(\bar{q})$. Since $q(\lambda)$ is smooth and Λ is compact, $q(\Lambda)$ defines a compact interval of real numbers, $I_q := [q_m, q_M] = q(\Lambda)$, where q_m and q_M are the absolute minimum and absolute maximum of $q(\lambda)$, respectively. We redefine $q(\Lambda)$ to be the open interval (q_m, q_M) , which we also denote by I_q .

We next prove that there exists (possibly discontinuous) 1-dimensional curves that are transverse to the generalized contours that can be used to *index* the family of generalized contours. We call any curve that has the property that it intersects each generalized contour once and only once a *transverse parameterization* (TP).

We give a constructive proof that is a useful algorithm. The algorithm produces discontinuous curves in Λ in general.

Theorem 3.2. *Suppose f is smooth in (3.1) and $q(\lambda)$ is a linear functional of the solution to (3.1). There exists a transverse parameterization for the set of generalized contours.*

Proof. We construct the transverse curve from a finite number of connected curves. We fix $\epsilon > 0$ and $\epsilon > \delta > 0$, and set $I_{q,\epsilon} = [q_m + \epsilon, q_M - \epsilon]$. If Λ is compact, then the existence of

transverse curves is guaranteed by the smoothness of $q(\lambda)$. To construct a curve, we begin at a point $\gamma_M \in \Lambda$ such that $q(\gamma_M) = q_M - \delta$, and follow the direction of the negative gradient until the curve either intersects the boundary or a minimum or saddle is reached, and denote that point γ_m . From smoothness, exactly one contour for each value of $q(\lambda)$ between $(q(\gamma_m), q(\gamma_M))$ is intersected by this curve. If $(q(\gamma_m), q(\gamma_M))$ does not completely cover $I_{q,\varepsilon}$, then we select a point $\tau_m \in \Lambda$ such that $q(\tau_m) = q_m + \delta$, and follow the direction of the gradient until the curve either intersects the boundary or a maximum or saddle is reached, and denote this point τ_M . We now check if $(q(\gamma_m), q(\gamma_M)) \cup (q(\tau_m), q(\tau_M))$ covers $I_{q,\varepsilon}$. If so, then we eliminate any part of the second curve that gives an overlap with contours intersected by the first. Otherwise, we continue to create this curve as above trying to cover the output interval defined by $(q(\tau_M), q(\gamma_m))$. This process produces a countable number of connected curves whose union forms a (possibly discontinuous) transverse curve through the generalized contours that corresponds to a countable open cover of $I_{q,\varepsilon}$, which is compact. Hence, there is a finite subcover of $I_{q,\varepsilon}$, which implies that the transverse parameterization can be constructed from a finite number of curves.

In practice, we construct the transverse curve to the generalized contours of I_q by initially following the first two steps above with $\varepsilon = 0$, i.e., locate $\gamma_M \in \Lambda$ such that $q(\gamma_M) = q_M$ and $\tau_m \in \Lambda$ such that $q(\tau_m) = q_m$ and construct the pieces of the transverse curve by following the negative and positive directions of the gradient, respectively. If we now take ε to be half the minimum of $q(\gamma_M) - q(\gamma_m)$ and $q(\tau_M) - q(\tau_m)$, then following the steps above, we construct a curve transverse to all the contours of I_q in a finite number of steps.

3.1.1. Approximating the set of generalized contours—Suppose that q is a linear function of λ , i.e., $q(\lambda) = \gamma^\top \lambda$ for some $\gamma \in \mathbb{R}^d$ (recall $\Lambda \subset \mathbb{R}^d$). Then for fixed $\bar{q} \in q(\Lambda)$ we have (with the same conventions as above) $U_{\bar{\lambda}}$, $V_{\bar{\lambda}}$, and $g_{\bar{\lambda}} : U_{\bar{\lambda}} \rightarrow V_{\bar{\lambda}}$ such that $\{(\lambda^{d-1}, g_{\bar{\lambda}}(\lambda^{d-1}))\}$ is the generalized contour. In this case, we write the function $g_{\bar{\lambda}}(\lambda^{d-1}) = (\bar{q} - (\gamma^{d-1})^\top (\lambda^{d-1})) / \gamma_d$ explicitly. The generalized contour above is a $(d-1)$ -dimensional hyperplane, and we refer to this as a *generalized linear contour*.

We approximate generalized contours locally by generalized linear contours, and approximate a generalized contour by a generalized piecewise-linear contour. We use generalized piecewise-linear contours computed from a piecewise-linear tangent plane approximation to $q(\lambda)$. If q is an affine map of λ , i.e., $q(\lambda) = \gamma^\top \lambda + q_0$ for some $q_0 \in \mathbb{R}$, then we use the function above with \bar{q} replaced by $\bar{q} - q_0$.

We obtain derivative information required to compute the tangent plane approximations implicitly by introducing the adjoint operator. This approach is very useful when the forward map is complicated to evaluate, e.g., involving the solution of a differential equation. But, the derivative information can be obtained by any convenient method.

Local linearization of the linear functional: The goal is to approximate the map $q(\lambda)$ with a piecewise-linear map $\hat{q}(\lambda)$ since it is possible to calculate the generalized contours for this approximate map.

Theorem 3.3. *The generalized linear contours converge pointwise to the true contours locally in Λ .*

Proof. Suppose we choose a reference parameter value $\lambda = \mu$ at which to solve

$$f(x; \lambda) = b$$

exactly. Call this reference solution y . Then according to Taylor's theorem,

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

where $\mathcal{R} \sim \mathcal{O}(\|x - y\|^2 + \|\lambda - \mu\|^2)$, for $|\alpha| = 2$. Here $D_x f$ and $D_\lambda f$ denote the derivatives of f with respect to x and λ , respectively.

In order to compute the tangent plane approximation efficiently, we use the *generalized Green's vector* ϕ that solves the adjoint to the linearized problem

$$A^\top \phi = \psi, \quad (3.3)$$

where $A = D_x f(y; \mu)$. Recall that $q(\lambda) = \langle x, \psi \rangle$, so by substitution of the above and standard linear algebra we arrive at

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

Neglecting the higher order term leads to an approximation of q by an affine map \hat{q} . If we denote the generalized contour of q given \bar{q} by $\{(\lambda^{d-1}, g_{\bar{\lambda}}(\lambda^{d-1}))\}$ and the generalized linear contour of \hat{q} given \bar{q} by $\{(\lambda^{d-1}, \hat{g}_{\bar{\lambda}}(\lambda^{d-1}))\}$, then at any $\lambda^{d-1} \in U_{\bar{\lambda}}$,

$$\left[g_{\bar{\lambda}}(\lambda^{d-1}) - \hat{g}_{\bar{\lambda}}(\lambda^{d-1}) \right] \left[\phi^\top \partial_{\lambda d} f(y, \mu) \right] = - \langle \mathcal{R}, \phi \rangle. \quad (3.4)$$

By assumption, $\lambda_d q(\lambda) = \phi^\top \lambda_d f(y, \mu) = 0$, so we rewrite (3.4) as

$$\left[g_{\bar{\lambda}}(\lambda^{d-1}) - \hat{g}_{\bar{\lambda}}(\lambda^{d-1}) \right] = C \langle \mathcal{R}, \phi \rangle,$$

where $C^{-1} = -\phi^\top \lambda_d f(y, \mu)$, is a nonzero constant determined entirely by the reference point (y, μ) . Thus, if we define

$$\|U_{\bar{\lambda}}\| = \sup_{\lambda \in U_{\bar{\lambda}}} \|\lambda - \bar{\mu}\|_2,$$

where $\|\cdot\|_2$ denotes the standard Euclidean norm, then as $\|U_{\bar{\lambda}}\| \rightarrow 0$, $\|\mathcal{R}\|_2 \rightarrow 0$ which implies that $|g_{\bar{\lambda}}(\lambda^{d-1}) - \hat{g}_{\bar{\lambda}}(\lambda^{d-1})| \rightarrow 0$.

Global linearization of the linear functional: We extend the local linearization technique to obtain a global piecewise-linear approximation of the linear functional over all of Λ . We first define a partition of *cells* $\{B_i\}_{i=1}^M$ of Λ . The geometry is immaterial, as long as we can integrate constant functions over the cells. We apply the local linearization technique described above for each cell, and defining

$$1_{B_i}(\lambda) := \begin{cases} 1 & \text{if } \lambda \in B_i, \\ 0 & \text{if } \lambda \notin B_i, \end{cases}$$

we obtain a global piecewise-linear approximation $\hat{q}(\lambda)$ to $q(\lambda)$ defined by

$$\widehat{q}(\lambda) := \sum_{i=1}^M (q(\mu_i) + \langle \nabla q(\mu_i), (\lambda - \mu_i) \rangle) \mathbf{1}_{B_i}(\lambda), \quad (3.5)$$

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

Theorem 3.4. *As $\|B_i\| \rightarrow 0$ (or as $M \rightarrow \infty$ when the number of sample points are distributed uniformly), the generalized linear contour converges pointwise to the generalized contour.*

Proof. For the finite system of nonlinear equations, we have

$$\nabla q(\mu_i) = \phi_i^\top D_\lambda f(y_i; \mu_i),$$

where ϕ_i solves the linearized adjoint problem using the reference point (y_i, μ_i) . If we let $\langle \mathcal{R}_i, \phi_i \rangle$ denote the higher-order terms neglected in the linearization of $q(\lambda)$ in cell B_i , then we can write the error of the piecewise-linear approximation, $e(\lambda) = \widehat{q}(\lambda) - q(\lambda)$, as

$$e(\lambda) = - \sum_{i=1}^M \langle \mathcal{R}_i, \phi_i \rangle \mathbf{1}_{B_i}(\lambda).$$

The generalized linear contour of \widehat{q} given \widehat{q} is a collection of hyperplanes in Λ . Using the same notation as above,

$$|g_{\widehat{\lambda}}(\lambda^{d-1}) - \widehat{g}_{\widehat{\lambda}}(\lambda^{d-1})| \leq C \sum_{i=1}^M |\langle \mathcal{R}_i, \phi_i \rangle|, \quad C^{-1} = \min_i \{|\phi_i^\top \partial_{\lambda d} f(y_i, \mu_i)|\}.$$

This yields the convergence result.

The transverse parameterization (TP) for the generalized linear contours is constructed using \widehat{q} in the same way as described in the proof of Theorem 3.2. Since \widehat{q} is a piecewise-linear surface, the resulting TP is a piecewise-linear curve in Λ .

Examples: We illustrate the convergence of generalized linear contours to true contours in the two examples below.

In the first example, we suppose that $q(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 \exp[-(\lambda_1^2 + 1.25 \lambda_2^2 - 1)]$ over $[0, 2] \times [0, 2]$. We approximate q over a uniform partition $\{B_i\}$ of $[0, 2] \times [0, 2]$ into squares, and we linearize around the midpoint of each B_i to form \widehat{q} in (3.5). We plot various contour curves and two TP's on each plot. The results are summarized in Figure 3.3.

For a second example, we suppose $q(\lambda_1, \lambda_2) = \exp[\cos(\lambda_1) + \sin(\lambda_2)]$ on $[-2\pi - 0.1, 2\pi + 0.1]^2$. We proceed as above to obtain the numerical results summarized in Figure 3.4.

3.2. Computing the parameter probability density

We now explain how to use the unique solution to the inverse problem in the space of generalized contours to compute an approximation of the probability density σ_Λ on Λ . We first observe if $I = [q_1, q_2] \subset \mathcal{D}$ is an event with probability $P(I) = P(q(\lambda) \in I)$, then this corresponds to a measurable set in Λ that is defined as the set of all contours obtained by $q^{-1}(I)$. From the basic assumptions of smoothness and the nonintersecting property of the

contours, the set of all contours is a set in Λ that is contained between the two contours defined by $q^{-1}(q_1)$ and $q^{-1}(q_2)$ (or possibly one of these contours and the boundary of Λ). We assign this set the probability $P(I)$. It follows immediately that we can define the inverse into the set of generalized contours for a given distribution of $q(\lambda)$ uniquely.

Theorem 3.5. *Suppose f is smooth in (3.1) and $q(\lambda)$ is a linear functional of the solution to (3.1). If $q(\lambda)$ is a random variable with distribution $F_q(q(\lambda))$, then for a fixed TP in Λ , the distribution of the intersections of the generalized contours on the TP, which is a random variable, is unique.*

The probability of a measurable set in Λ is determined by the contours the set contains and the amount of each contour the set contains and the probabilities of those contours. The parameter volume measure μ_Λ determines the contours a given set contains and the amount of each contour the set contains.

3.2.1. Computational measure theory—The method we develop for computing an approximate probability distribution is based on constructions used in measure theory.

Theorem 3.6. *Given a measurable set $A \subset \Lambda$, we can approximate $P(A)$ using a simple function approximation to $\sigma_\Lambda(\lambda)$, which requires only calculations of volumes in Λ*

The constructive proof below parallels Algorithm 1 for approximating the probability of a measurable set $A \subset \Lambda$.

Proof. For λ restricted between any two contours induced by a subinterval of a partition of \mathcal{D} as in Algorithm 1, $q(\lambda)$ is approximately a uniformly distributed random variable. Suppose

that $\{q_j\}_{j=0}^N$ is a partition of \mathcal{D} such that $q_0 < q_1 < \dots < q_N$, and if $E_j = [q_{j-1}, q_j]$, then $\mathcal{D} = \cup_j E_j$. Let $A_j = \{\lambda \mid q(\lambda) \in E_j\}$. We assume that $\Lambda = \cup_j A_j$. The probability of A_j is given by

$$P(A_j) = \int_{A_j} \sigma_\Lambda(\lambda) d\mu_\Lambda(\lambda).$$

We can compute this probability because of the 1-1 correspondence between the contours and output values, i.e., $P(A_j) = P(E_j) = \int_{E_j} \rho_{\mathcal{D}}(q) d\mu_{\mathcal{D}}(q)$. Therefore, we have a simple function approximation to $\sigma_\Lambda(\lambda)$ given by

$$\sigma_\Lambda(\lambda) \approx \sigma_{\Lambda,N}(\lambda) = \sum_{j=1}^N \frac{P(A_j)}{\mu_\Lambda(A_j)} 1_{A_j}(\lambda),$$

Algorithm 1

Approximate Parameter Probability Distribution Method

Fix simple function approximation, $\rho_{\mathcal{D}}^{(M)}(q)$, to $\rho_{\mathcal{D}}(q)$ that induces a partition $\cup_{i=1}^{N(M)} [q_{i-1}, q_i]$ of \mathcal{D} where for each $i = 1, \dots, N(M)$, $\rho_{\mathcal{D}}^{(M)}(q)$ is constant on each subinterval $[q_{i-1}, q_i]$

$\cup_{i=1}^{N(M)} [q_{i-1}, q_i]$ induces a partition of Λ by generalized contours and $\{A_j\}_{j=1}^{N(M)}$ denotes this partition

Let P_j denote probability of A_j given by $\int_{[q_{j-1}, q_j]} \rho_{\mathcal{D}}^{(M)}(d) d\mu_{\mathcal{D}}(q)$

Partition Λ with small cells $\{b_i\}_{i=1}^{M'}$

for $i = 1, \dots, M'$ **do**

for $j = 1, \dots, M$ **do**

 Calculate ratio of volume of $b_i \cap A_j$ to volume of A_j , store in matrix V_{ij}

end for

 Set $P(b_i)$ equal to $\sum_{j=1}^{N(M)} V_{ij} P_j$

end for

Given event $A \subset \Lambda$, estimate $P(A)$ using

- inner sums, i.e., sum of $P(b_i)$ for all $i \in I_C \subset \{1, \dots, M'\}$ such that $b_i \subset A$,
- outer sums, i.e., sum of $P(b_i)$ for all $i \in I_C \subset \{1, \dots, M'\}$ such that $b_i \cap A \neq \emptyset$,
- average of inner and outer sums, or
- $\int_A \sigma_{\Lambda, M'}(\lambda) d\mu_{\Lambda}(\lambda)$, where $\sigma_{\Lambda, M'}(\lambda) = \sum_{i=1}^{M'} P(b_i) \mathbf{1}_{b_i}(\lambda)$.

Given event $A \subset \Lambda$, we use the law of total probability to write

$$P(A) = \sum_{j=1}^N P(A|A_j)P(A_j).$$

Using the above simple function approximation to the parameter density, we have

$$P(A|A_j) = \frac{P(A \cap A_j)}{P(A_j)} = \frac{\int_{A \cap A_j} d\mu_{\Lambda}(\lambda)}{\int_{A_j} d\mu_{\Lambda}(\lambda)} = \frac{\mu_{\Lambda}(A \cap A_j)}{\mu_{\Lambda}(A_j)}.$$

Hence, the probability $P(\lambda \in A | q(\lambda) \in E_j) = P(A|A_j)$ can be calculated from the volume measure on model space since it depends only on measurable sets in Λ if we use the approximation $q(\lambda) \sim \mathcal{U}(E_j)$ for $\lambda \in A_j$. The value is the ratio of volume of $A \cap A_j$ to the volume of A_j . Since the density on data space is a nonnegative integrable function, there

exists a sequence of simple functions $\{\rho_{\mathcal{D}}^{(M)}(q)\}_{M=1}^{\infty}$ with

$$\rho_{\mathcal{D}}^{(M)}(q) = \sum_{k=1}^{2^{2M}+1} \frac{k-1}{2^M} \mathbf{1}_{I_{M,k}}(\rho_{\mathcal{D}}(q)),$$

and $I_{M,k} = [(k-1)/2^M, k/2^M]$. We first observe that the partition $\{I_{M,k}\}$ induces a partition $\{E_{M,k}\}$ of \mathcal{D} . Also, we observe that $\rho_{\mathcal{D}}^{(M)}(q) \rightarrow \rho_{\mathcal{D}}(q)$ in L^1 as $M \rightarrow \infty$ by the monotone convergence theorem, and for any measurable set $E \subset \mathcal{D}$

$$\int_E \rho_{\mathcal{D}}^{(M)}(q) d\mu_{\mathcal{D}}(q) = \sum_{k=1}^{2^{2M}+1} \frac{k-1}{2^M} \mu_{\mathcal{D}}(E_{M,k} \cap E) \rightarrow P_{\mathcal{D}}(E) \text{ as } M \rightarrow \infty.$$

Thus, we can approximate the value of $P(A|A_j)$ by the ratio of volume of $A \cap A_j$ to the volume of A_j obtained from the volume measure on model space if the induced partitions $\{A_j\}$ come from a sufficiently fine partition $\{E_j\}$ of data space so that the distribution of $q(\lambda)$ for $\lambda \in A_j$ is approximated by $\mathcal{U}E_j$.

Since $P(A) = \sup\{P(K) : K \subset A, K \text{ compact}\}$ and $P(A) = \inf\{P(U) : A \subset U, U \text{ open}\}$, we can estimate $P(A)$ using the inner and outer sums described by Algorithm 1.

Remark 3.4. If the set A has not (yet) been specified, we may still carry out the first part of Algorithm 1 to obtain a *discretized* approximation of the measure P on model space.

Remark 3.5. The set of cells $\{b_i\}_{i=1}^{M'}$ in Algorithm 1 is introduced purely for computational purposes and is not necessary to the approximation of $P(A)$. We choose $\{b_i\}_{i=1}^{M'}$ in order to approximate $P(A)$, for *any* event $A \subset \Lambda$, without carrying out the calculations in the nested loops of Algorithm 1 for each new event. If we are interested only in one event, $A \subset \Lambda$, then we might skip the step of partitioning Λ by $\{b_i\}_{i=1}^{M'}$ and replace the step in the nested loop by the following: Calculate ratio of volume of $A \cap A_j$ to volume of A_j , store in vector V_j . We may then approximate $\sum_{j=1}^{N(M)} V_j P_j$.

Remark 3.6. Note that as we refine the partition $\{E_j\}$ on the data space, which in turn refines the partition $\{A_j\}$ on model space, we should consider refining the mesh that defines the partition $\{b_i\}$ on model space. The reason is that we assign a probability $P(b_i)$ to each cell b_i that in essence reapproximates the simple function approximation,

$$\sigma_{\Lambda}(\lambda) \approx \sigma_{\Lambda, N}(\lambda) = \sum_{j=1}^N \frac{P(A_j)}{\mu_{\Lambda}(A_j)} \mathbf{1}_{A_j}(\lambda),$$

by the new simple function

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

If the partition $\{b_i\}$ remains fixed as the approximation of $\rho_{\mathcal{D}}(q)$ by simple functions is refined by the partition $\{E_j\}$, then the representation of $\sigma_{\Lambda}(\lambda)$ as a simple function converges with respect to the fixed $\{b_i\}$. When choosing $\{b_i\}$, we should consider that a cell b_i might be large relative to the A_j that it intersects, i.e., b_i might intersect many A_j . When this is the case, estimating the probability over b_i by a constant $P(b_i)$ might not be an appropriate approximation. In general, it is not computationally demanding to estimate an appropriate size of the b_i .

Observations on simple function approximations: The use of simple function approximations of a probability density is sufficiently unusual in the context of stochastic analysis of differential equations as to justify comment. Simple function approximations form the basis for classic measure theory because they yield several benefits, including

- Simple function approximations are widely applicable under minimal assumptions on the density being approximated. As the examples below suggest, probability densities solving inverse problems appear to be highly complex in general.

- The convergence analysis for simple function approximations is also widely applicable. This contrasts with sampling techniques such as Markov chain Monte Carlo methods whose convergence properties are stochastic and can be highly sensitive to properties of the problem.

Though we have not exploited the fact in this paper, simple function approximations also offer significant benefits for stochastic sensitivity analysis of differential equations [12, 13, 9, 10, 7]. In particular, combining a simple function approximation with sensitivity derivatives of a quantity of interest with respect to parameters provides both a natural dimension reduction mechanism and the basis for adaptive sampling.

Of course, a significant issue with simple function approximations is the nominal dependence of accuracy on the dimension of the parameter space. This may be a consequence of the common approach of using hyper-rectangular cell discretizations of the underlying space combined with the unfortunate growth in diagonal dimension of hyper-rectangles as dimension increases, though we report on some inconclusive results of using radial basis functions in [12]. In our experience, the effects of dimension are nominal up to dimensions of 8–10, and we have effectively used the piecewise constant approximations to dimensions of order 15–18. We note that this is *effective* dimension. By exploiting dimension reduction, the nominal dimension of the parameter space may be higher.

4. Examples

We apply the new method to solve inverse problems associated with a variety of maps. We first consider three constrained geometric optimization problems. We then discuss examples involving a nonlinear ordinary differential equation and a nonlinear elliptic partial differential equation with two parameters. Finally, we discuss the determination of regions with high probability.

In the following examples, we have chosen the uniform Lebesgue measure for the parameter volume measure and often impose a normal distribution on the output quantity of interest. The first choice is made because it is commonly the (implicit) default, e.g., in Bayesian inference. The imposition of a normal distribution on the output is also a common choice. In our examples, it serves the purpose of illustrating the complex nature of the inverse probability measure that results even when a normal distribution has been imposed on the output. However, we emphasize that neither of these choices are important in terms of implementing the numerical solution method, which is readily applied for any distributions.

4.1. A 2-dimensional nonlinear function

We consider the map determined implicitly as the solution of the finite-dimensional nonlinear system of equations given by

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

where λ_1 and λ_2 are the parameters. Geometrically, solutions $x = (x_1, x_2)^T$ to the system represent intersections of the hyperbola and ellipse. The quantity of interest is the second component of the solution in the first-quadrant, i.e., $q(\lambda) = q(x(\lambda)) = x_2 = \langle x, \psi \rangle$, where $\psi = (0, 1)^T$. According to (3.3), the adjoint problem is

$$\begin{pmatrix} 2\mu_1 y_1 & 2y_1 \\ 2y_2 & -2\mu_2 y_2 \end{pmatrix} \phi = \psi,$$

where $\mu = (\mu_1, \mu_2)^T$ and $y = (y_1, y_2)^T$ are the reference parameter and reference solution for the forward problem.

In order to create an interesting example, we choose

$\Lambda = [.79, .99] \times [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}]$ based on a sensitivity analysis of the forward problem in [23]. We use six-uniformly spaced mesh points in both the λ_1 and λ_2 directions of Λ to create cells $\{B_i\}_{i=1}^{25}$ that partition Λ . We use the centroid of each cell as the reference parameter $\mu_i = (\mu_{1,i}, \mu_{2,i})^T$ in that cell and solve the forward problem to obtain reference solutions $y_i = (y_{1,i}, y_{2,i})^T$ at these points, and then solve for the generalized Green's vector $\phi_i = (\phi_{1,i}, \phi_{2,i})^T$ at the reference point (μ_i, y_i) . According to (3.5), we obtain a global piecewise-linear approximation \tilde{q} to q defined as

$$\tilde{q}(\lambda) := \sum_{i=1}^{25} \left(y_{2,i} + (\lambda - \mu_i)^T \begin{pmatrix} y_{1,i}^2 & 0 \\ 0 & -y_{2,i}^2 \end{pmatrix} \phi_i \right) \mathbf{1}_{B_i}(\lambda).$$

We assume that the output data is a random variable with normal distribution on the data space defined by $\tilde{q}(\Lambda)$ (Figure 4.1). We assume μ_Λ is the Lebesgue measure. We implement Algorithm 1 to calculate $P(b_j)$ for small cells for each fine partition of Λ and determine the probabilities of events $A \subset \Lambda$. We plot the results in Figure 4.2.

4.1.1. A three-parameter geometric constrained optimization problem—The map to be inverted is determined by minimizing the distance to the point $(1, -1, 1)$ among points constrained to lie on the surface $g = 4$, where

$$g(x_1, x_2, x_3; \lambda_1, \lambda_2, \lambda_3) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2.$$

Geometrically, the parameters determine the shape of the ellipsoid that defines the constraint. Using the method of Lagrange multipliers we set up a system of nonlinear equations with four state variables and three parameters. We take the quantity of interest as the first state variable, which geometrically is interpreted as the first spatial coordinate in the solution to the constrained minimization problem. We set $\Lambda = [.35, .65] \times [.28, .52] \times [.42, .78]$ and construct a piecewise-linear approximation using 125 points in Λ . We assume a normal distribution on $q(\lambda)$ and taking the underlying parameter volume measure μ_Λ to be a normalized Lebesgue measure. We use 3375 small cells $\{B_i\}$ in Algorithm 1. We plot the probabilities at the midpoint of each cell with the color of the point determined by the probability of the small cell in Figures 4.3–4.4.

4.1.2. A four-parameter geometric constrained optimization problem—The map to be inverted is determined by minimizing the distance to the point $(5, 5, 5)$ among points constrained to lie on the intersection of the surfaces $g = 1$ and $h = 0$, where

$$\begin{aligned} g(x_1, x_2, x_3; \lambda_1, \lambda_2) &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + x_3^2, \\ h(x_1, x_2, x_3; \lambda_3, \lambda_4) &= \lambda_3 x_1 + \lambda_4 x_2 + x_3. \end{aligned}$$

Geometrically, $g = 1$ defines a hyperboloid of one sheet and $h = 0$ defines a plane through the origin, and the intersection of the two constraints is a closed curve. Using the method of Lagrange multipliers we set up a system of nonlinear equations with five state variables and four parameters. We take the quantity of interest as the first state variable, which geometrically is interpreted as the first spatial coordinate in the solution to the constrained minimization problem. We set $\Lambda = [1.4, 2.6] \times [0.7, 1.3] \times [1.4, 2.6] \times [0.35, 0.65]$ and construct a piecewise-linear approximation using 750 points in Λ . We assume a normal distribution on $q(\lambda)$ and take μ_Λ to be a normalized Lebesgue measure. We use 60750 small cells $\{b_j\}$ in Algorithm 1. Displaying a 4-dimensional distribution is problematic. We plot “snapshots” of the approximated probability density for three fixed λ_4 values in Figure 4.5.

4.1.3. A two-parameter ordinary differential equation—We now study the nonlinear ordinary differential equation

$$\begin{cases} \dot{x} = \lambda_1 \sin(\lambda_2 x), & 0 < t \leq T, \\ x(0) = 1. \end{cases}$$

The linear functionals (quantities of interest, $q(\lambda)$) we study take the form

$$q(\lambda) = \langle x(t), \psi(t) \rangle = \int_0^T \langle x(s; \lambda), \psi(s) \rangle ds,$$

and we take the quantity of interest to be the average value of $x(t)$ over the time interval $[0, 2]$. Thus, we set $\psi(t) = \mathbf{1}_{[0,2]}(t)/2$, and the generalized Green’s function $\phi(t)$ solves the adjoint problem,

$$\begin{cases} -\dot{\phi}(t) - A^\top(t)\phi(t) = \psi(t), & T > t \geq 0, \\ \phi(T) = \psi(T), \end{cases}$$

where $A(t) := f'(y(t; \mu))$ is the Jacobian of $f = \lambda_1 \sin(\lambda_2 x)$ evaluated at $y(t; \mu)$, μ is a reference parameter, and $y(t; \mu)$ is the solution to (4.1.3) for this reference parameter. Compare this to (3.3). Using substitution, integration by parts, and Taylor’s theorem, we arrive at a linear approximation to $q(\lambda)$ for parameters near μ , and analogous to the finite dimensional case, we obtain a global piecewise-linear approximation to $q(\lambda)$ over $\Lambda = [0.8, 1.2] \times [0.1, 0.1]$ shown in Figure 4.6.

Remark 4.1. There can be substantial error in the reference solutions and gradients used when applying the method to differential equations whose solutions must be approximated numerically, and we study the effect of these errors in the second paper [4].

4.1.4. A two-parameter elliptic partial differential equation—We now study a nonlinear elliptic partial differential equation

$$\begin{cases} -\Delta u = \lambda_1(u - \lambda_2)^2, & (x, y) \in \Omega = [0, 1] \times [0, 1], \\ u = 0, & (x, y) \in \partial\Omega. \end{cases}$$

The quantities of interest, $q(\lambda)$, take the form

$$q(\lambda) = \langle u, \psi \rangle = \int_{\Omega} u(x, y) \psi(x, y) dx dy,$$

and we take the quantity of interest to be the average value of u over Ω . Thus, we set $\psi(x, y) = 1$, and the generalized Green's function $\phi(t)$ solves the adjoint problem,

$$\begin{cases} -\Delta\phi - A^T\phi = \psi, & (x, y) \in \Omega, \\ \phi = 0, & (x, y) \in \partial\Omega, \end{cases}$$

where $A := f'(w(x, y; \mu); \mu)$ is the Jacobian of $f = \lambda_1 \exp(\lambda_2 u)$ evaluated at $w(x, y; \mu)$, μ is a reference parameter, and $w(x, y; \mu)$ is the solution to (4.1.4) for this reference parameter. Using substitution, the weak form of (4.1.4), and Taylor's theorem, we arrive at a linear approximation to $q(\lambda)$ for parameters near μ , and just as with the previous examples, we obtain a global piecewise-linear approximation to $q(\lambda)$ over $\Lambda = [.95, 1.05] \times [-.1, .1]$ using Algorithm 1. We show the results in Figure 4.7.

4.2. Determining regions of high probability

The new method can be applied to find regions of high probability. Consider $q(\lambda) = \lambda_1 + \lambda_2$, where $\Lambda = [0, 1] \times [0, 1]$. Figure 4.8 shows the generalized contours for 500 samples of $q(\lambda)$ taken from a $N(0, 2/25)$ distribution along with the TP and the intersections of contours on the TP. Where the contours intersect the TP most densely corresponds to a region of high probability in the space of contours.

We can locate regions of high probability by sorting through the probability of the fine cells $\{b_i\}$. We can rank order these cells and determine any cells of high probability. We can also determine regions of neighboring cells that all have relatively high probability. We illustrate using the four-parameter geometric constrained optimization problem in section 4.1.2. In Table 1, we list the ten small cells with highest probability. If we let the events $\{b_i\}$ become small, under a smoothness assumption, the probabilities of these events are related to the maximum-likelihood estimate.

5. Conclusion

We consider the probabilistic inverse sensitivity analysis problem: Given a specified uncertainty in the output of a map, determine variations in the parameters that produce the observed uncertainty. We formulate this inverse problem using the law of total probability. We describe and analyze a method for computing the approximate probability density that solves the inverse problem and does not require random sampling. Our approach breaks the solution down into two stages:

1. Construct an approximate representation of the set-valued inverse solution of the ill-posed deterministic inverse problem.
2. Approximate the density on the parameter space that corresponds to the set-valued inverse and the observed output density using a simple function representation.

We illustrate the method and several features using a variety of examples.

In [4] we present numerical analysis of discretization error, e.g., in evaluating the model by numerical solution and in finite sampling. In [5], we discuss the problem of dealing with multiple quantities of interest, which has application to data assimilation and "cascaded" uncertainty in operator decomposition solution of multiphysics problems.

Acknowledgments

The first author's work was supported in part by the National Aeronautics and Space Administration, Earth Sciences Division (#NNX08AK08G) the National Science Foundation (SES-0922142), and the Joint NSF/NIGMS Initiative to Support Research in the Area of Mathematical Biology (#R01GM096192). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health. The last author's work was supported in part by the Defense Threat Reduction Agency (HDTRA1-09-1-0036), Department of Energy (DE-FG02-04ER25620, DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC0001724), Lawrence Livermore National Laboratory (B573139, B584647), the National Aeronautics and Space Administration (NNG04GH63G), the National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559), Idaho National Laboratory (00069249), NSF/NIGMS (#R01GM096192), and the Sandia Corporation (PO299784).

This author's work was supported in part by the Department of Energy (DE-FG02-05ER25699) and the National Science Foundation (DGE-0221595003, MSPACSE-0434354).

REFERENCES

1. Bernardo JM. Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B.* 1979; 41:113–147.
2. Billingsley, P. *Probability and Measure*. 3rd ed.. New York: John Wiley & Sons; 1995.
3. Butler, T. Ph.D. thesis. Fort Collins, CO: Department of Mathematics, Colorado State University; 2009. *Computational Measure Theoretic Approach to Inverse Sensitivity Analysis: Methods and Analysis*.
4. Butler T, Estep D. A measure-theoretic computational method for inverse sensitivity problems II: A posterior error analysis. *SIAM J. Numer. Anal.* submitted.
5. Butler T, Estep D. A measure-theoretic computational method for inverse sensitivity problems III: Multiple output quantities of interest. in preparation, 2010.
6. Cacuci, D. *Sensitivity and Uncertainty Analysis: Theory*. Vol. Vol. I. Boca Raton, FL: Chapman & Hall/CRC; 1997.
7. Estep D, Holst MJ, Målqvist A. Nonparametric density estimation for randomly perturbed elliptic problems III: Convergence, complexity, and generalizations. *J. Appl. Math. Comput.* to appear.
8. Estep D, Larson MG, Williams RD. Estimating the error of numerical solutions of systems of reaction-diffusion equations. *Mem. Amer. Math. Soc.* 2000; 146 pp. viii+109.
9. Estep D, Målqvist A, Tavener S. Nonparametric density estimation for randomly perturbed elliptic problems. I: Computational methods, a posteriori analysis, and adaptive error control. *SIAM J. Sci. Comput.* 2009; 31:2935–2959.
10. Estep D, Målqvist A, Tavener S. Nonparametric density estimation for randomly perturbed elliptic problems. II: Applications and adaptive modeling. *Internat. J. Numer. Methods Engrg.* 2009; 80:846–867.
11. Estep, D.; Mckeown, B.; Neckels, D.; Sandelin, J. 2006. GAASP: Globally Accurate Adaptive Sensitivity Package. write to estep@math.colostate.edu for information
12. 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.
13. Estep D, Neckels D. Fast methods for determining the evolution of uncertain parameters in reaction-diffusion equations. *Comput. Methods Appl. Mech. Engrg.* 2007; 196:3967–3979.
14. Huelsenbeck JP, Larget B, Miller RE, Ronquist F. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 2002; 51:673–688. [PubMed: 12396583]
15. Folland, G. *Real Analysis*. 2nd ed.. New York: John Wiley & Sons, *Modern Techniques and their Applications*; 1999.
16. Gentle, JE. *Random Number Generation and Monte Carlo Methods*. 2nd ed.. New York: Springer; 2003.
17. Gilks, WR.; Richardson, S.; Spiegelhalter, DJ. *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: CRC Press; 1995.
18. Kaipio, J.; Somersalo, E. *Statistical and Computational Inverse Problems*. New York: Springer-Verlag; 2005.

19. Knill, DC.; Richards, W. Perception as Bayesian Inference. Cambridge, UK: Cambridge University Press; 1996.
20. Lanczos, C. Linear Differential Operators. Mineola, NY: Dover Publications; 1997.
21. Marchuk, GI. Adjoint Equations and Analysis of Complex Systems. Dordrecht: Kluwer Academic Publishers; 1995.
22. Marchuk, GI.; Agoshkov, VI.; Shutyaev, VP. Adjoint Equations and Perturbation Algorithms in Nonlinear Problems. Boca Raton, FL: CRC Press; 1996.
23. Neckels, D. Ph.D. thesis. Fort Collins, CO: Department of Mathematics, Colorado State University; 2005. Variational Methods for Uncertainty Quantification.
24. Robert, CP.; Casella, G. Monte Carlo Statistical Methods. New York: Springer-Verlag; 2004.
25. Sandelin, J. Ph.D. thesis. Fort Collins, CO: Department of Mathematics, Colorado State University; 2006. Global Estimate and Control of Model, Numerical, and Parameter Error.
26. Tarantola, A. Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia: SIAM; 2005.
27. Zabaras N, Ganapathysubramanian B. A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *J. Comput. Phys.* 2008; 227:4697–4735.

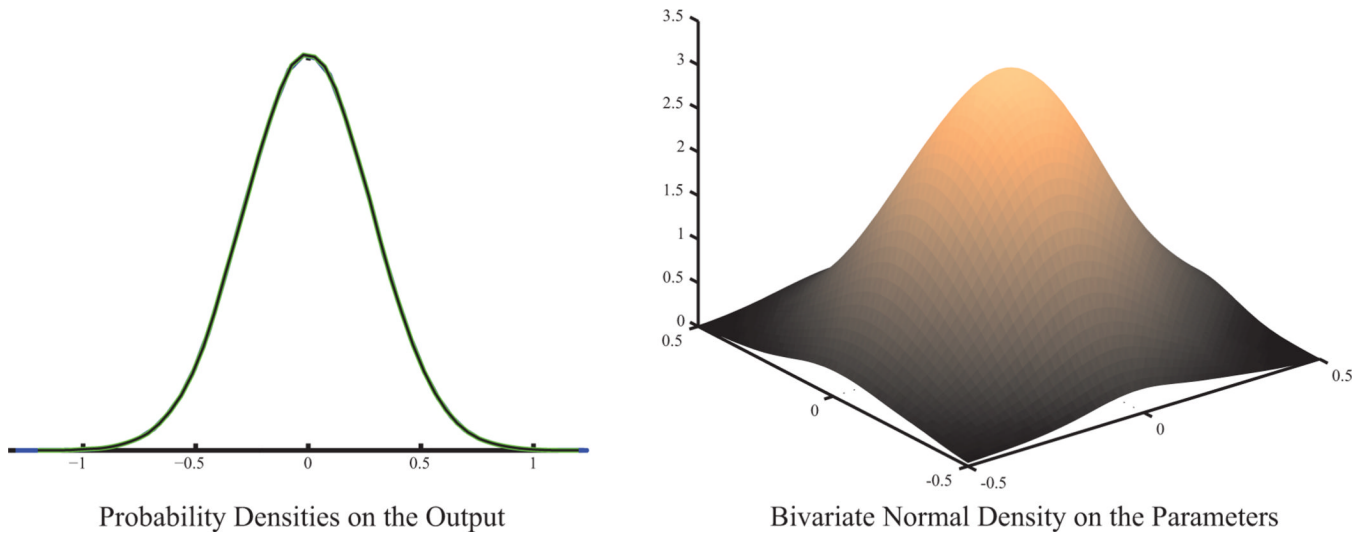


Fig. 2.1. Left: The $N(0, 2/25)$ distribution imposed on the output $\lambda_1 + \lambda_2$. Right: The joint distribution of two independent $N(0, 1/25)$ parameters λ_1 and λ_2 . Summing these variables is one way to compute the imposed normal on the output quantity. Figures 2.2–2.3 show alternatives.

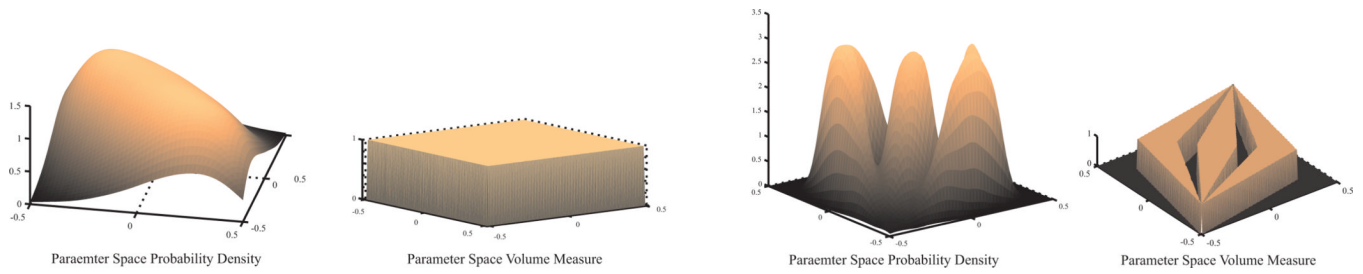


Fig. 2.2. The joint distributions of parameters (λ_1, λ_2) sampled with respect to the density $\rho_{\Lambda}(\lambda)$ and the corresponding volume measure presented in pairs of plots. Left two plots: The volume measure is uniform Lebesgue on Λ . Right two plots: The volume measure is uniform Lebesgue a set with three distinct parts.

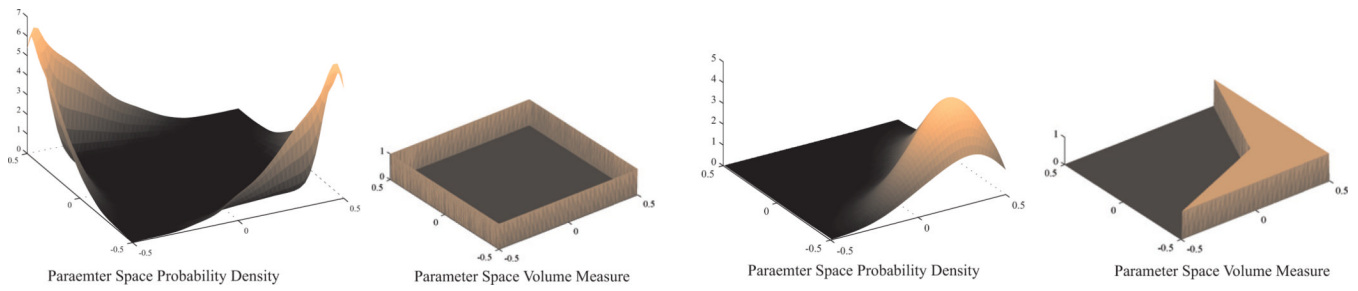


Fig. 2.3. The joint distributions of parameters (λ_1, λ_2) sampled with respect to the density $\rho_{\Lambda}(\lambda)$ and the corresponding volume measure presented in pairs of plots. Left two plots: The volume measure is uniform Lebesgue on the boundary. Right two plots: The volume measure is uniform Lebesgue on a nonconvex interior set.

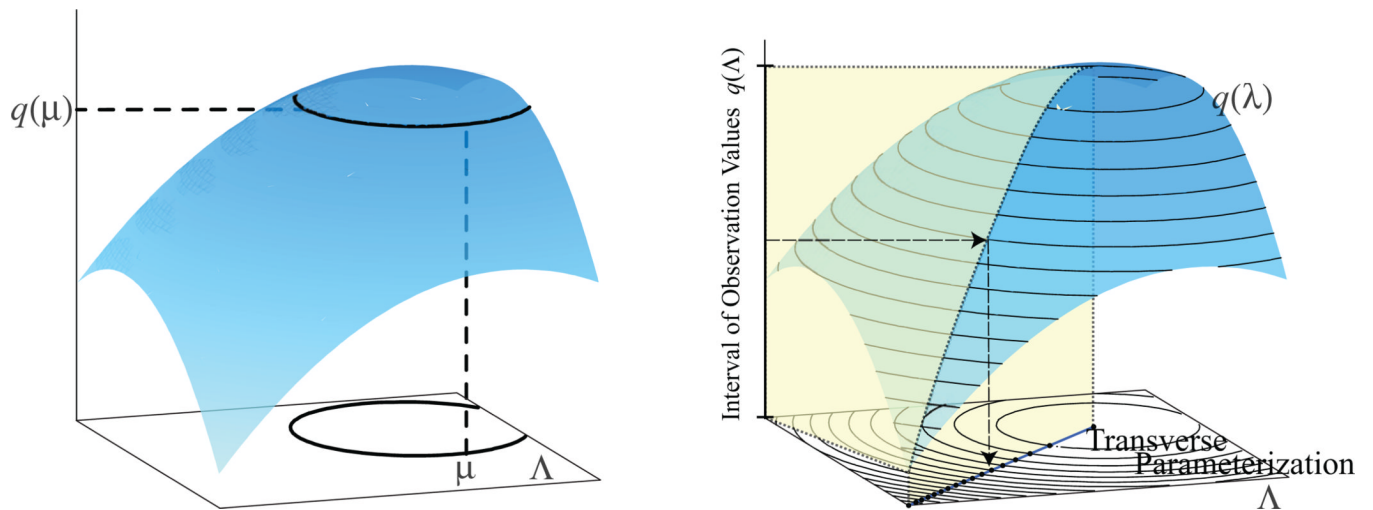
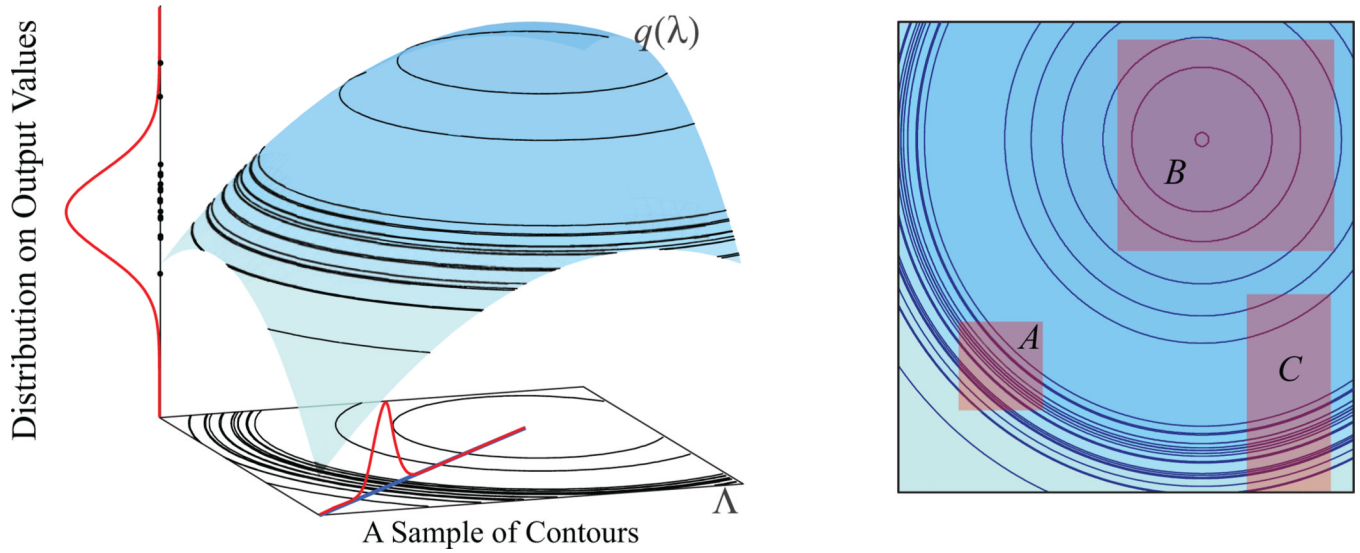


Fig. 3.1.

Left: Each observation value corresponds to a unique contour curve. Right: On the horizontal plane, we show a transverse parameterization. Each point on the transverse parameterization corresponds to a unique contour curve, so the transverse parameterization acts as an index for the space of contour curves. There is a unique map from the points in the interval containing the observed output values to the points on the transverse parameterization.

**Fig. 3.2.**

Left: We show a probability distribution imposed on the output values. A sample of output values drawn from this distribution corresponds to a unique sample of contour curves. Right: Plotted is a sample of contour lines in parameter space corresponding to a specified distribution on the output observation values along with three events. We specify the Lebesgue measure as the parameter volume measure. Event B has relatively low probability because while it has relatively large area, the probability of the contours is relatively low (visible because the density is sparse). Event A has intermediate probability because while the area of event A is relatively small, A contains contours with relatively high probability (which is visible because of the dense sample of contours). The probability of event C is largest because it contains the same high probability contours as A but has larger area.

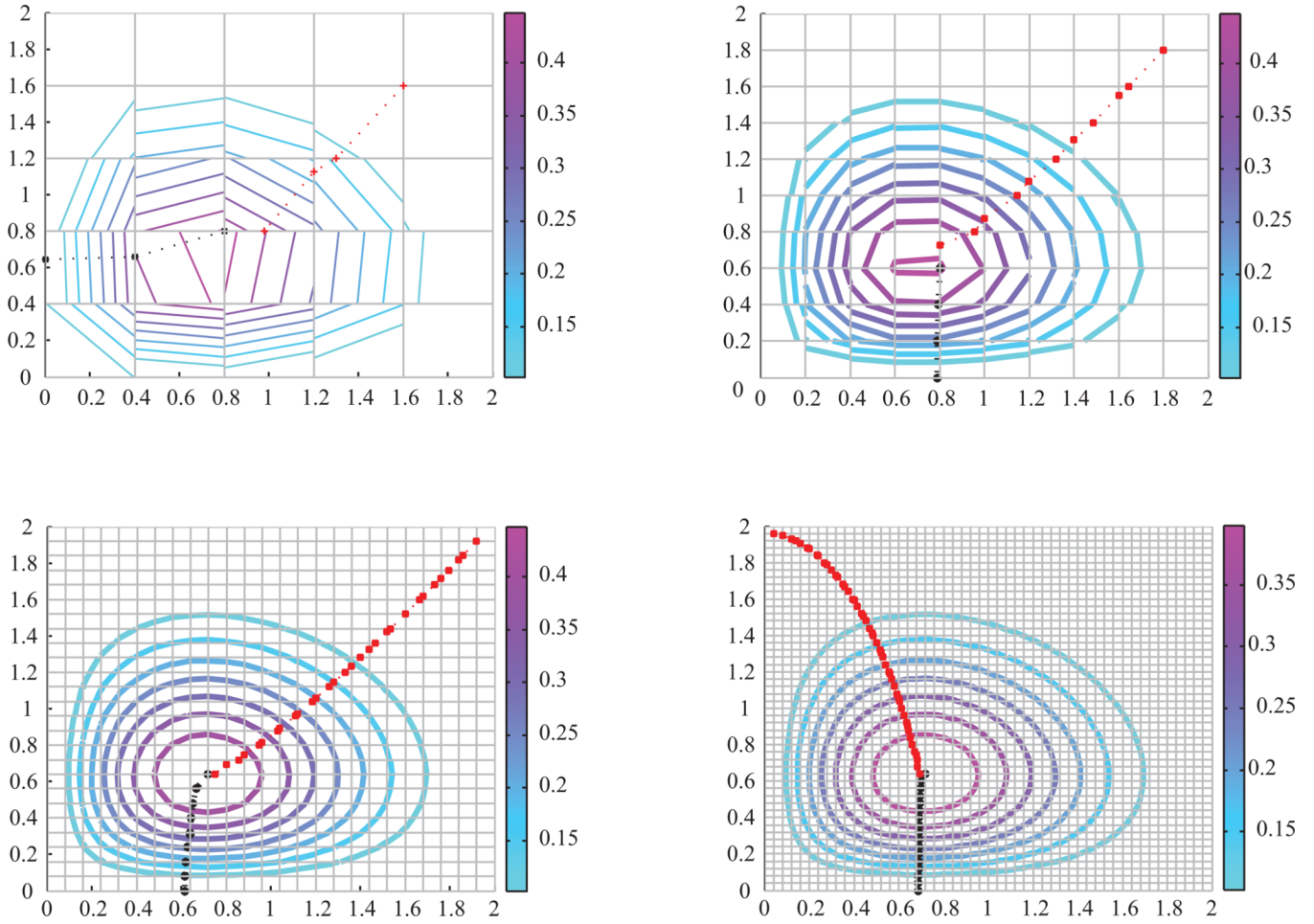


Fig. 3.3. Contours of \hat{q} using 5×5 cells (top left), 10×10 cells (top right), 25×25 cells (bottom left), and 50×50 cells (bottom right). The TP is created using the algorithm outlined in the proof of its existence and is denoted by the circle-dotted and plus-dotted lines. The circle-dotted line is constructed from the maximum of $q(\lambda)$ and follows the negative direction of the gradient of $q(\lambda)$, and the plus-dotted line is constructed from the minimum of $q(\lambda)$ and follows the direction of the gradient.

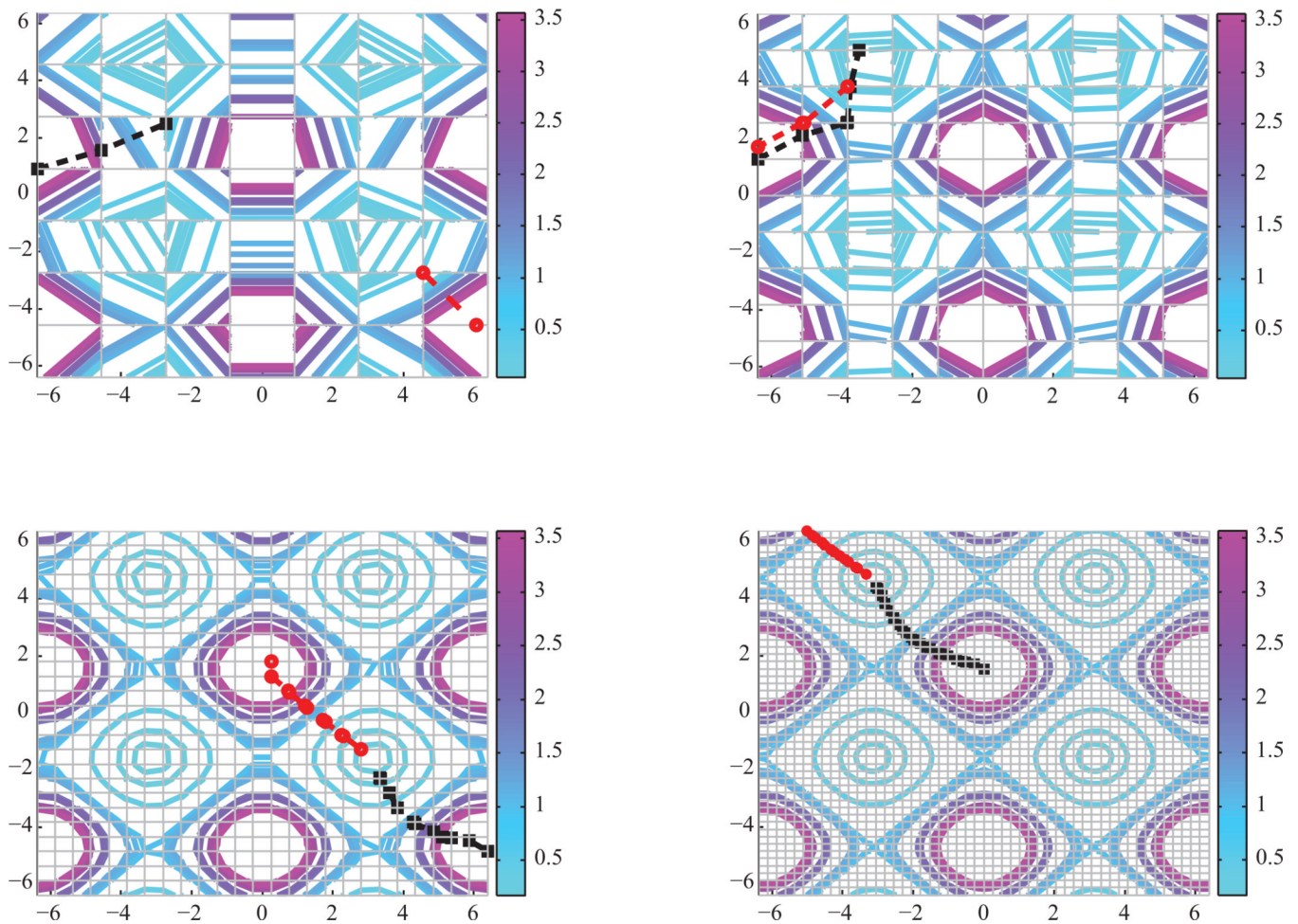


Fig. 3.4. Contours of \hat{q} using 7×7 cells (top left), 10×10 cells (top right), 25×25 cells (bottom left), and 50×50 cells (bottom right). The TP is created using the algorithm outlined in the proof of its existence and is denoted by the square-dotted and circle-dotted lines. The square-dotted line is constructed from the maximum of $q(\lambda)$ and follows the negative direction of the gradient of $q(\lambda)$, and the circle-dotted line is constructed from the minimum of $q(\lambda)$ and follows the direction of the gradient.

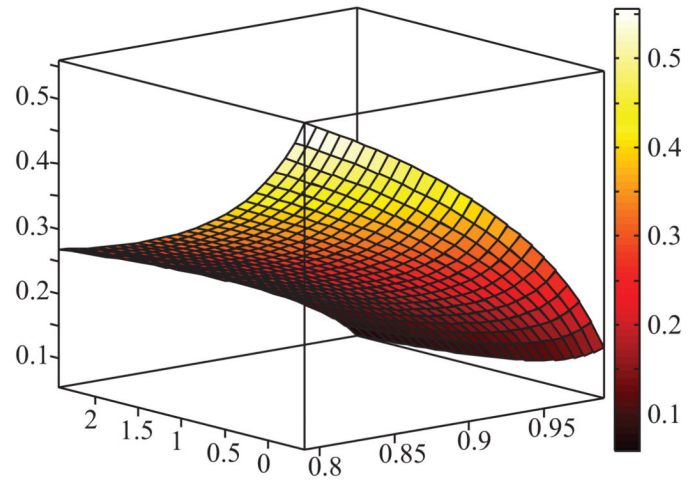
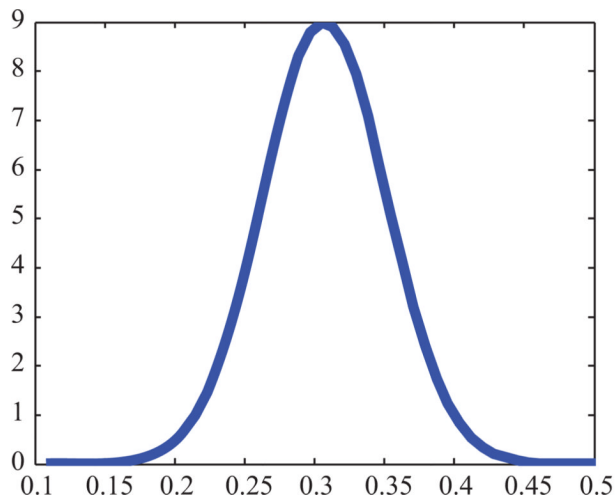


Fig. 4.1.
Left: Uncertainty of output is modeled as a random variable with a normal distribution.
Right: A plot of the map $q : \Lambda \rightarrow \mathbb{R}$.

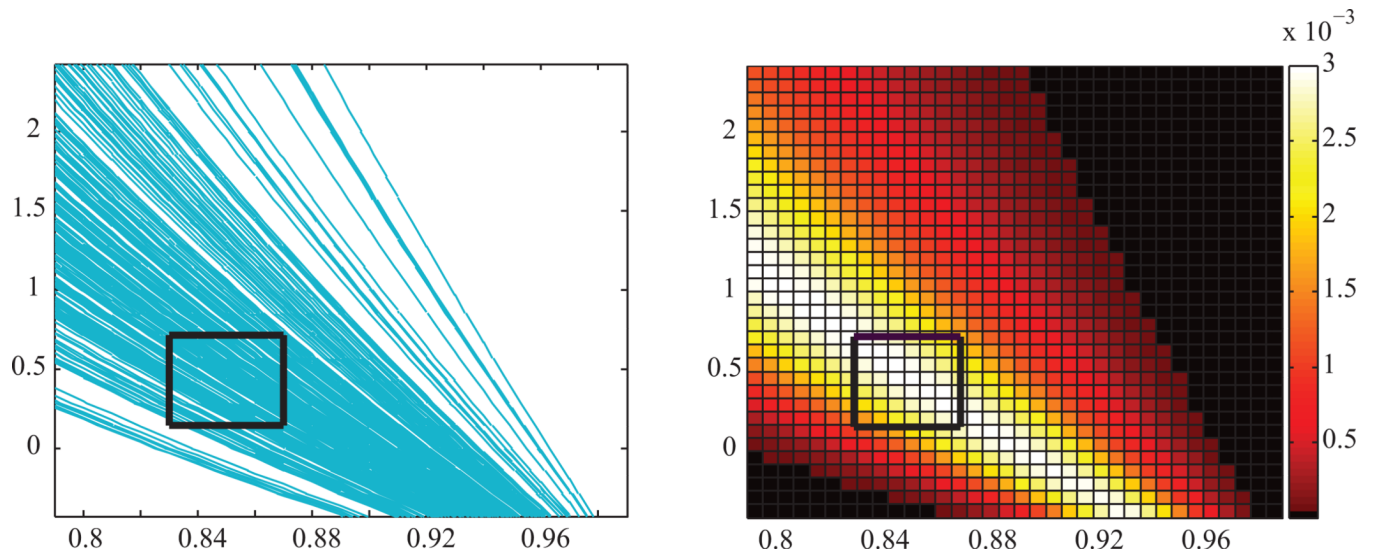


Fig. 4.2.

Illustration of an application of Algorithm 1. Left: We determine which contours are contained in an event $A \subset \mathbf{A}$ and how much of each contour is inside the event. Right: We estimate the probabilities of small cells contained in the event and use an inner and outer estimate to obtain an approximation of the probability of the event A .

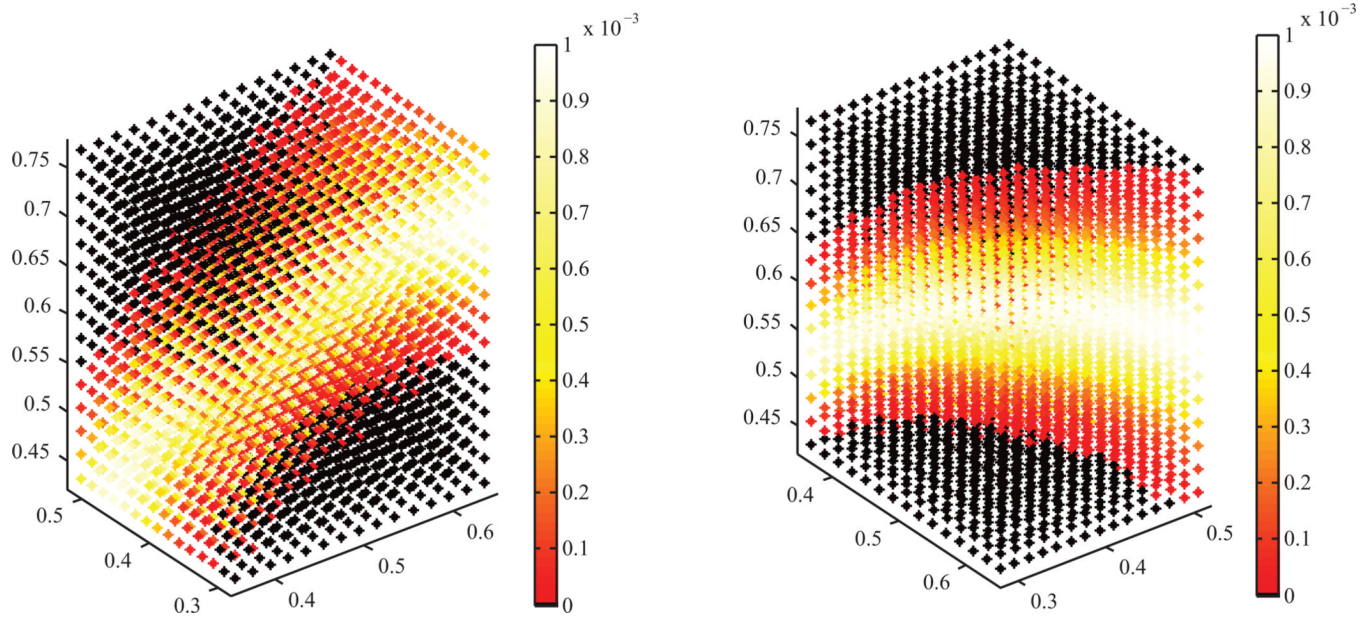


Fig. 4.3. We use $15 \times 15 \times 15$ small cells in Algorithm 1. We plot the approximate distribution from several angles. Left: A 3-dimensional view. Right: The same 3-dimensional view rotated 90 degrees clockwise.

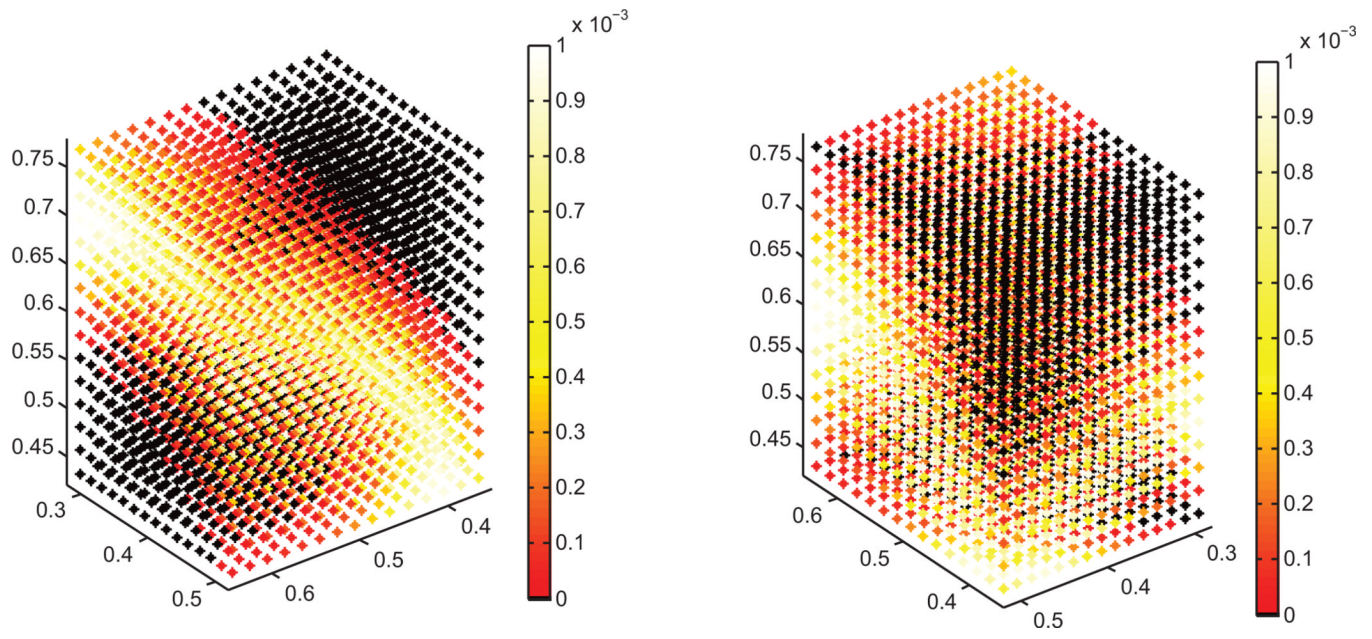
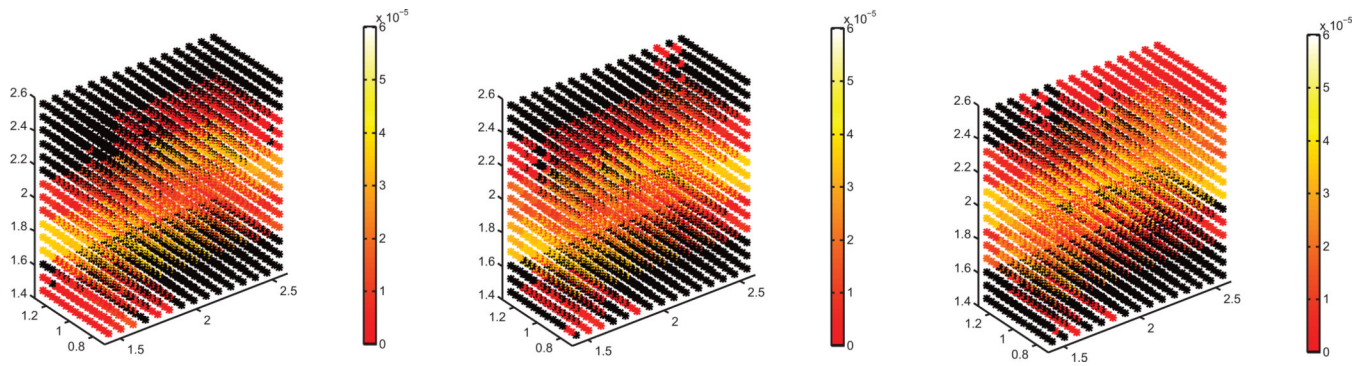
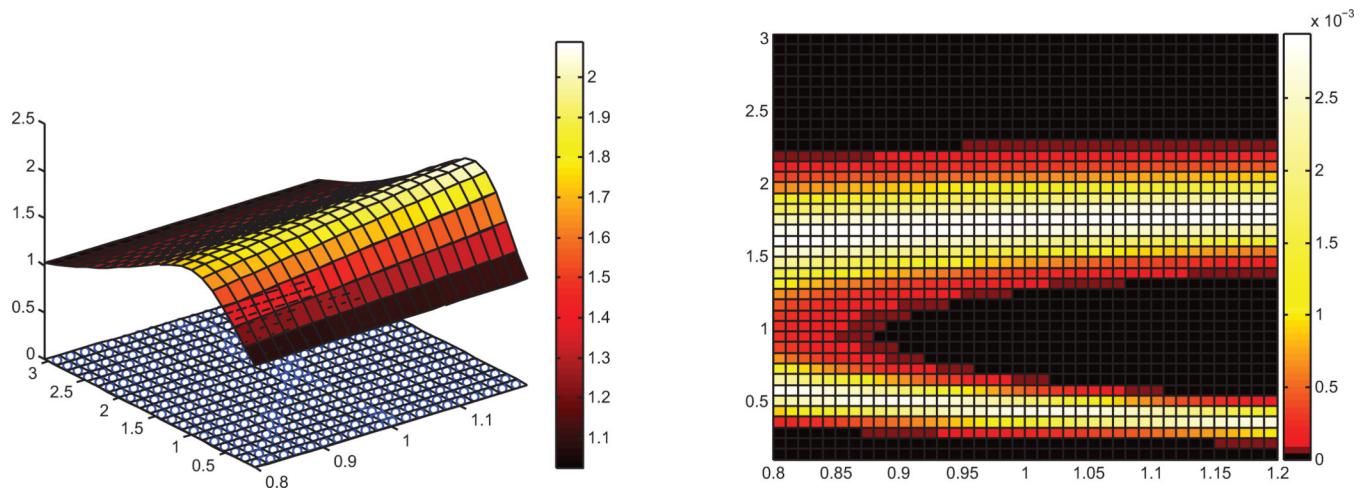


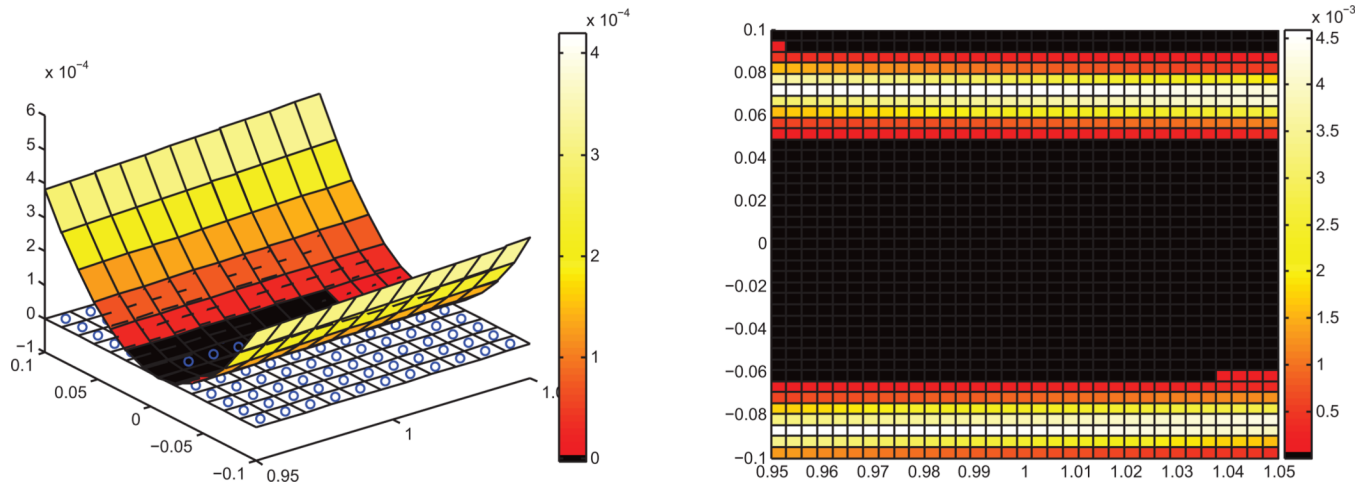
Fig. 4.4. We use $15 \times 15 \times 15$ small cells in Algorithm 1. We plot the approximate distribution from several angles. Left: The original 3-dimensional view rotated 180 degrees clockwise. Right: The original 3-dimensional view rotated 270 degrees clockwise.

**Fig. 4.5.**

We use $15 \times 15 \times 15 \times 18$ small cells in Algorithm 1. We plot “snapshots” of the approximate probability distribution for three values of the fourth parameter. Left: The fourth parameter is set at its minimum value. Middle: The fourth parameter is set at its midpoint value. Right: The fourth parameter is set at its maximum value. Notice how the probabilities vary in space as we vary the fourth parameter.

**Fig. 4.6.**

Left: The global piecewise-linear approximation to $q(\lambda)$ obtained using Algorithm 1. The cells in Λ illustrate the coarse discretization of this space for the forward problem of obtaining a piecewise-linear approximation and the circles in each cell indicate the reference parameter used to linearize $q(\lambda)$ in that cell. We assume a normal distribution for $q(\lambda)$ and use a grid of 40×40 small cells.

**Fig. 4.7.**

Left: Global piecewise-linear approximation to $q(\lambda)$ obtained using Algorithm 1. We used a 11×13 grid of coarse cells to discretize Λ and used the midpoint of each cell as the reference parameter in that cell. We assume a normal distribution of $q(\lambda)$ and we use a 33×39 grid of small cells.

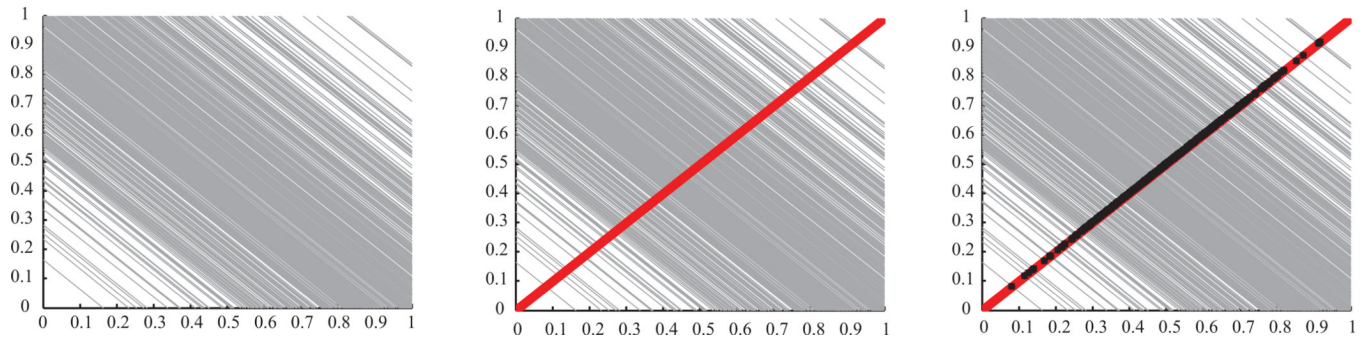


Fig. 4.8.

Left: Generalized contours from 500 samples of $q(\lambda) = \lambda_1 + \lambda_2$ generated from a $N(0, 2/25)$ distribution. Middle: The TP intersects each contour once and goes from the minimum of $q(\lambda)$ in the lower left corner to the maximum of $q(\lambda)$ in the upper-right corner of the plot. Right: Intersections of contours on the TP are marked with a star and can be used to index the inverses and determine a unique distribution of the contours on the TP using any consistent indexing scheme.

Table 1

We indicate the location of the ten cells with the highest probabilities for the example in section 4.1.2. The first column gives the probability and the second column gives the dimensions and location of the cells. There are clearly two distinct regions for events with relatively high probability. In general, one can use this information to determine where the largest regions of highest probability are located in a high-dimensional parameter space.

$P(b_i)$ order 10^{-4}	b_i location
0.600381927	$[2.44, 2.52] \times [1.22, 1.26] \times [2.04, 2.12] \times [0.4, 0.4167]$
0.600446977	$[2.36, 2.44] \times [1.06, 1.1] \times [1.96, 2.04] \times [0.4333, 0.45]$
0.600462420	$[2.44, 2.52] \times [1.18, 1.22] \times [2.04, 2.12] \times [0.4333, 0.45]$
0.600465732	$[2.36, 2.44] \times [0.98, 1.02] \times [2.04, 2.12] \times [0.4167, 0.4333]$
0.600470136	$[2.36, 2.44] \times [1.06, 1.1] \times [1.96, 2.04] \times [0.4167, 0.4333]$
0.600474821	$[2.36, 2.44] \times [1.26, 1.3] \times [1.96, 2.04] \times [0.4167, 0.4333]$
0.600501752	$[2.36, 2.44] \times [0.98, 1.02] \times [2.04, 2.12] \times [0.4333, 0.45]$
0.600463048	$[1.4, 1.48] \times [1.18, 1.22] \times [1.64, 1.72] \times [0.3833, 0.4]$
0.600464252	$[1.4, 1.48] \times [1.18, 1.22] \times [1.64, 1.72] \times [0.35, 0.3667]$
0.600468545	$[1.4, 1.48] \times [1.18, 1.22] \times [1.64, 1.72] \times [0.3667, 0.3833]$