SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 2, pp. 174–202

A Measure-Theoretic Computational Method for Inverse Sensitivity Problems III: Multiple Quantities of Interest*

T. Butler†, D. Estep‡, S. Tavener§, C. Dawson¶, and J. J. Westerink‖

Abstract. We consider inverse problems for a deterministic model in which the dimension of the output quantities of interest computed from the model is smaller than the dimension of the input quantities for the model. In this case, the inverse problem admits set-valued solutions (equivalence classes of solutions). We devise a method for approximating a representation of the set-valued solutions in the parameter domain. We then consider a stochastic version of the inverse problem in which a probability distribution on the output quantities is specified. We construct a measure-theoretic formulation of the stochastic inverse problem, then develop the existence and structure of the solution using measure theory and the disintegration theorem. We also develop and analyze an approximate solution method for the stochastic inverse problem based on measure-theoretic techniques. We demonstrate the numerical implementation of the theory on a high-dimensional storm surge application where simulated noisy surge data from Hurricane Katrina is used to determine the spatially variable bathymetry fields of highest probability.

Key words. approximation of measures, approximation of events, disintegration of measure, inverse sensitivity analysis, measure theory, probability measures, stochastic inverse problems, set-valued solutions, storm surge

AMS subject classifications. 60H30, 60H35, 65M32

DOI. 10.1137/130930406

1. Introduction.

Computing information about input parameters for a mathematical model of a physical system based on observations on model output quantities is a critical

*Received by the editors July 23, 2013; accepted for publication February 14, 2014; published electronically May 8, 2014. The work of the second and third authors was supported in part by Department of Energy grants DE-FG02-04ER25620 and INL00120133, Idaho National Laboratory grants 00069249 and 00115474, and Lawrence Livermore National Laboratory grant B590495. The work of the first, second, and third authors was also supported in part by National Science Foundation grant DMS-1228206.

http://www.siam.org/journals/juq/2/93040.html

†Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO 80217 (butler.troy.d@gmail.com).

‡Department of Statistics, Colorado State University, Fort Collins, CO 80523 (estep@stat.colostate.edu). The work of this author was supported in part by the Defense Threat Reduction Agency (HDTRA1-09-1-0036), the Department of Energy (DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC0001724, DE-SC0005304, DE0000000SC9279), Lawrence Livermore National Laboratory (B573139, B584647), the National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1065046, DMS-1016268, DMS-FRG-1065046), and the National Institutes of Health (R01GM096192).

§Department of Mathematics, Colorado State University, Fort Collins, CO 80523 (tavener@math.colostate.edu).

¶Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Austin, TX 78712 (clint@ices.utexas.edu). The work of this author was supported in part by the National Science Foundation (DMS-1228243).

‖Department of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame, South Bend, IN 46556 (jjw@nd.edu). The work of this author was supported in part by the National Science Foundation (DMS-1228212).

component of scientific inference and engineering design. The solution of a model induces a map from the space of input parameters and data (referred to as parameters in this paper) to the output quantities of interest computed from the solution of the model. In many situations, these quantities of interest are related to the experimentally observable aspects of the modeled system. In that case, an important problem is to determine possible parameter values that match the observations on the quantities given by the output of the model. Important in its own right, information about parameter values can also be used for predicting unobserved behavior of the system and determining parameter values that lead to optimal behavior.

This parameter determination problem is nominally a deterministic inverse problem. In many situations, however, the observational data on the model outputs is affected by error and uncertainty that are described in stochastic terms. The result is a type of stochastic inverse problem: If we describe the stochastic quantities in terms of the associated probability distributions, then the inverse problem is to determine information about the probability distribution(s) on the input parameter space that match the distributions associated with observations on the output. For example, given the shallow-water equations, a wind model for Hurricane Katrina, and noisy data corresponding to the recorded maximum surge at various points of interest (e.g., at levees near the city of New Orleans), an inverse problem of practical importance is to determine the probability measure on model parameters corresponding to bathymetry along the coast of Louisiana.

A significant complication arises in the common situation in which the model induces a noninvertible solution operator. Under typical assumptions on the model, the solution of the deterministic inverse problem for a single output value consists of a set of points in the parameter domain (as opposed to a single point). The solution of the inverse problem on the range of the solution operator yields a space of equivalence classes of set-valued solutions. Contour maps of elevation present a familiar example of such solutions. Set-valued solutions significantly complicate the approximate solution of the deterministic and stochastic inverse problems.

This paper continues the formulation and solution of stochastic inverse problems begun in [2, 5]. Our approach can be described as carrying out measure-theoretic computations in the space of set-valued solutions made concrete by explicit approximation of points in that space. In [2], we constructed and analyzed a method based on approximating set-valued solutions in the case of a single quantity of interest. In [5], we performed a full numerical analysis, including a priori convergence and a posteriori estimation analysis, for the approach in [2]. In this paper, we treat the additional geometric complications that arise from multiple quantities of interest, using an altered approach based on approximating events rather than the manifolds defining the set-valued solutions. We also provide a fuller measure-theoretic development of the stochastic inverse problem.

In section 2, we begin with a first principles "measure-theoretic" description of the stochastic inverse problems considered in this paper. The motivation is to be absolutely clear about the inverse problem we solve, since there are several valid inverse problems associated with uncertainty quantification for a physics-based model. Next we consider the solution of stochastic inverse problems, which is broken down into two aspects:

1. describe an approximate constructive representation of the set-valued inverse solution of the deterministic model;

2. approximate the probability density (measure) on the parameter space that corresponds to the set-valued inverse and the observed output density.

The first aspect is described in section 3. The second aspect is described in section 4. In that section, we also comment briefly on the relation to other inverse problems and solution approaches. Finally, we present the main numerical examples in section 5.

2. Formulation of stochastic inverse problems. Our starting point is a physical description expressed as a deterministic model $M(Y,\lambda)$ that determines a solution $Y(\lambda)$ that is an implicit function of parameters and data λ in a compact[1] set $\mathbf{\Lambda}_0 \subset \mathbb{R}^n$. The domain $\mathbf{\Lambda}_0$ is the largest physically meaningful domain of parameter values. The goal of solving the model is to compute a set of "geometrically distinct" quantities of interest, $\{Q_i(Y)\}_{i=1}^m$, where Q_i is a functional valued in $\mathbb{R}$. This induces a map $Q(\lambda) = (Q_1(\lambda),\ldots,Q_m(\lambda))^\top = (Q_1(Y(\lambda)),\ldots,Q_m(Y(\lambda)))^\top$ from $\mathbf{\Lambda}_0$ into its range $\mathcal{D}_0$. We give a precise definition of "geometrically distinct" in section 3. In the case of linear maps, this is simply linear independence. Our interest lies in the case that $m < n$. In practice, it is often the situation that the goal of the mathematical modeling is to compute a relatively small set of quantities of interest. Moreover, it is often the situation that only a handful of quantities are experimentally observable.

We describe a sequence of forward problems and corresponding inverse problems in measure-theoretic terms. For other approaches to stochastic inverse problems involving measure theory, see Banks and Bihari [1] and Tarantola [15].

Forward problems for a deterministic, physics-based model. We first describe a sequence of "forward" problems for a deterministic, physics-based model. The lowest level of forward problem is simply to evaluate the map Q for a particular parameter value λ. Of course, depending on the model, the approximate solution of the model and issues such as convergence and accuracy of the approximate solution may be very complex. We do not address these issues here; see [5].

The next level of forward problem is fundamental to the deterministic sensitivity analysis of the model. There are several kinds of sensitivity analysis, e.g., computation of derivatives, dynamical behavior, etc. However, all examples are concerned with describing how the output of the model behaves as the parameters are varied over sets of values. We define a compact set $\mathbf{\Lambda} \subset \mathbf{\Lambda}_0$ that forms the domain of consideration for the map Q, and we use $\mathcal{D} = Q(\mathbf{\Lambda}) \subset \mathcal{D}_0$ to denote the corresponding range of the map.

The sensitivity analysis of the behavior on sets requires additional structure on the parameter space that allows evaluation of the distance between points, e.g., so that we can discuss convergence of sequences, limit points, set topology, and distance between a point and a set. Hence, we assume that $\mathbf{\Lambda}$ and $\mathcal{D}$ are metric spaces and that the map Q is locally differentiable with respect to the metrics on $\mathbf{\Lambda}$ and $\mathcal{D}$. This is very often the case under typical assumptions on the model.

In turn, this makes $\mathbf{\Lambda}$ into a measure space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ using the Borel σ-algebra $\mathcal{B}_{\mathbf{\Lambda}}$ and "volume" (Lebesgue) measure $\mu_{\mathbf{\Lambda}}$. Assuming the Borel σ-algebra $\mathcal{B}_{\mathcal{D}}$ on $\mathcal{D}$, $\mu_{\mathbf{\Lambda}}$ induces a

[1] We explicitly use the compactness assumption in our solution methodology. The method can be extended to deal with noncompact $\mathbf{\Lambda}$, but additional considerations are required. We note that parameters are often limited to compact sets for physical reasons.

volume measure $\mu_{\mathcal{D}}$ on $\mathcal{D}$. This "push-forward" measure is defined for measurable $A \in \mathcal{B}_{\mathcal{D}}$,

$$\mu_{\mathcal{D}}(A) = \int_A d\mu_{\mathcal{D}} = \int_{Q^{-1}(A)} d\mu_{\mathbf{\Lambda}} = \mu_{\mathbf{\Lambda}}(Q^{-1}(A)),$$

yielding the measure space $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$. Our definition of the inverse problem and computation of its solution depend critically on these volume measures. We emphasize that these measures are not probability measures. They are introduced naturally as part of the construction of the mathematical model and are related to the physical meaning of the parameters.

The third forward problem is the stochastic sensitivity analysis problem. This builds on the second forward problem. We assume that a probability measure $P_{\mathbf{\Lambda}}$ is given on $\mathbf{\Lambda}$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$, yielding the probability space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, P_{\mathbf{\Lambda}})$. Hence, the computation of the probability of an *arbitrary* event $E \in \mathcal{B}_{\mathbf{\Lambda}}$ can be written in terms of an integral involving a probability density $\rho_{\mathbf{\Lambda}}$ with respect to $\mu_{\mathbf{\Lambda}}$,

$$P_{\mathbf{\Lambda}}(E) = \int_E \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}}.$$

The goal of the stochastic sensitivity analysis is to compute the probability density $\rho_{\mathcal{D}}$ with respect to $\mu_{\mathcal{D}}$ induced by $\rho_{\mathbf{\Lambda}}$; i.e., for event $A \in \mathcal{B}_{\mathcal{D}}$,

$$P_{\mathcal{D}}(A) = \int_A \rho_{\mathcal{D}} d\mu_{\mathcal{D}} = \int_{Q^{-1}(A)} \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}} = P_{\mathbf{\Lambda}}(Q^{-1}(A)).$$

This is a familiar problem in uncertainty quantification, whose solution can be obtained via a classic Monte Carlo method. Namely, we draw samples from $\mathbf{\Lambda}$ according to the probability measure $\rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}}$, evaluate the map for each sample, bin the results, apply smoothing, and thus obtain an approximation to $\rho_{\mathcal{D}}$.

Inverse problems for a deterministic, physics-based model. We next present a sequence of inverse problems directly corresponding to the forward problems described above. The first problem is to compute the set of input values corresponding to a single given value for the map Q in $\mathcal{D}$. Note that because Q is an "n-to-m" map with $n > m$, each value in the range of Q corresponds to a set of values in $\mathbf{\Lambda}$. For example, if Q maps $\mathbf{\Lambda} \subset \mathbb{R}^2$ to $\mathcal{D} \subset \mathbb{R}^1$, then the inverse set corresponding to a given value is a contour curve; see Figure 1. In general, we call an inverse set a generalized contour. By the assumption that Q is locally differentiable, the implicit function theorem guarantees that the generalized contours exist as locally smooth manifolds (section 3).

Corresponding to the second forward problem, we next consider the inverse problem for Q on a given range $\mathcal{D}$. We assume that $\mathcal{D}$ is the proper range of Q over $\mathbf{\Lambda}$, and so we can decompose $\mathbf{\Lambda}$ into a union of generalized contours corresponding to—and indexed by—the points in $\mathcal{D}$. This is an equivalence class decomposition of $\mathbf{\Lambda}$, where two points are equivalent if they are in the same generalized contour. In the case of inverting a two-to-one dimensional map, this decomposition is a contour map; see Figure 1. We denote the space of equivalence classes in $\mathbf{\Lambda}$ by $\mathcal{L}$, so that each point in $\mathcal{L}$ corresponds to a set of points in $\mathbf{\Lambda}$. The inverse of Q defines a one-to-one, onto map between $\mathcal{D}$ and $\mathcal{L}$. The inverse image of a set in $\mathcal{D}$ is a set of points in $\mathcal{L}$ corresponding to a collection of set-valued inverses in $\mathbf{\Lambda}$; see Figure 1.

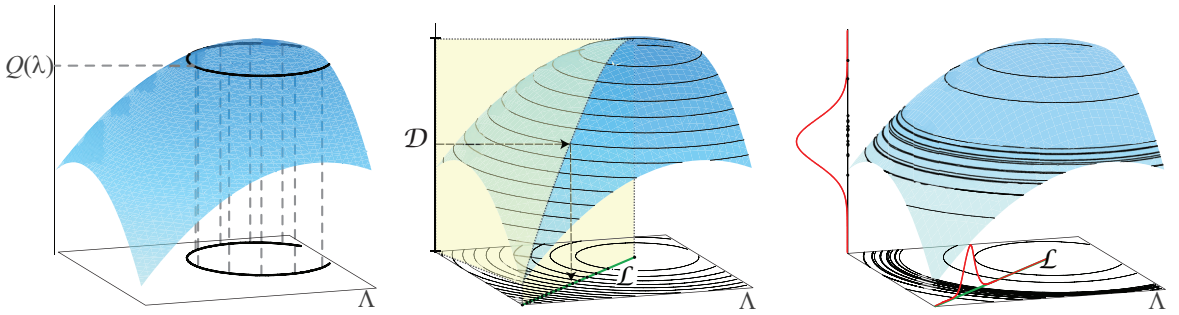

Figure 1. *Illustrations of several inverse problems for an inverted quadratic "bowl." Left: The set-valued inverse of a single output value. Middle: There is an invertible map between the set of equivalence classes $\mathcal{L}$ and the range $\mathcal{D}$. We show a partition of $\mathbf{\Lambda}$ by a set of contour equivalence classes. Right: A probability distribution on the range $\mathcal{D}$ gives a unique inverse distribution on the set of equivalence classes $\mathcal{L}$. We show contours corresponding to a sample drawn on the distribution on $\mathcal{D}$. Figures are adapted from* [2].

Dealing with set-valued solutions requires adjustment of common conceptions because notions such as continuity and well-posedness must be phrased with respect to points in $\mathcal{L}$, not with respect to points in $\mathbf{\Lambda}$. For example, the condition or degree of near-ill-posedness of the inverse map is determined by how well the map distinguishes different equivalence classes from each other. Different representers from the same equivalence class cannot be distinguished by the solution of the inverse problem, but this is not ill-posedness in the context of set-valued solutions. While this is a significant restriction, practical experience with contour maps demonstrates that the space of generalized contours is nonetheless very useful.

Finally, a σ-algebra $\mathcal{B}_{\mathcal{L}}$ on $\mathcal{L}$ can be generated using inverse images of a collection of Borel sets in $\mathcal{B}_{\mathcal{D}}$, and the volume measure $\mu_{\mathcal{D}}$ induces a volume measure on $\mathcal{L}$,

$$\mu_{\mathcal{L}}(A) = \int_A d\mu_{\mathcal{L}} = \int_{Q(A)} d\mu_{\mathcal{D}},$$

for $A \in \mathcal{B}_{\mathcal{L}}$. We obtain a measure space $(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, \mu_{\mathcal{L}})$.

Next, we introduce a stochastic version of the previous inverse problem. Namely, we assume that a probability measure $P_{\mathcal{D}}$ is given on $\mathcal{D}$ as a probability density $\rho_{\mathcal{D}}$ with respect to $\mu_{\mathcal{D}}$. The inverse problem is to compute the induced probability measure $P_{\mathcal{L}}$ described as a probability density $\rho_{\mathcal{L}}$ on $\mathcal{L}$ with respect to $\mu_{\mathcal{L}}$,

$$P_{\mathcal{L}}(A) = \int_A \rho_{\mathcal{L}} d\mu_{\mathcal{L}} = \int_{Q(A)} \rho_{\mathcal{D}} d\mu_{\mathcal{D}} = P_{\mathcal{D}}(Q(A)),$$

for $A \in \mathcal{B}_{\mathcal{L}}$. In other words, we carry out probability computations in a space $(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, P_{\mathcal{L}})$ whose points consist of generalized contours. We illustrate in Figure 1.

Under our assumptions, the equivalence relation on $\mathbf{\Lambda}$ determined by Q^{-1} is a measurable map from $\mathbf{\Lambda}$ to $\mathcal{L}$. The induced σ-algebra $\mathcal{C}_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ can be generated from the set of equivalence classes of a set of generating events for $\mathcal{B}_{\mathcal{L}}$. We note that $\mathcal{C}_{\mathbf{\Lambda}}$ is a proper subset of $\mathcal{B}_{\mathbf{\Lambda}}$ in general, and we call events in $\mathcal{C}_{\mathbf{\Lambda}}$ "contour events"; see Figure 2. The induced probability measure $P_{\mathbf{\Lambda},\mathcal{C}_{\mathbf{\Lambda}}}$ is defined as $P_{\mathbf{\Lambda},\mathcal{C}_{\mathbf{\Lambda}}}(A) = P_{\mathcal{L}}(\mathcal{E}_A) = P_{\mathcal{D}}(Q(A))$, where $\mathcal{E}_A$ is the event in $\mathcal{L}$ corresponding to the equivalence class of A.

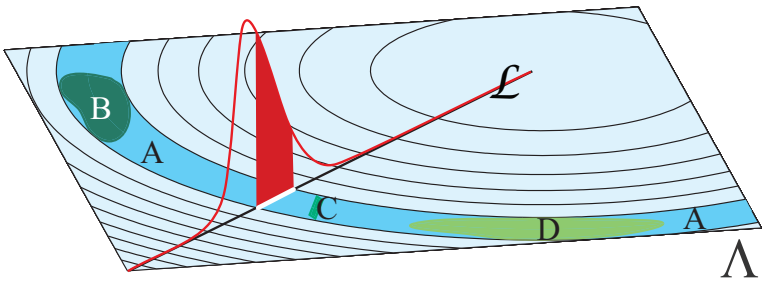

Figure 2. *Illustration of the solution of the inverse problem in* $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}})$ *following Figure 1. Event A is a "contour event" in* $\mathcal{C}_{\mathbf{\Lambda}}$ *that is in the equivalence class corresponding to the event in* $\mathcal{L}$ *marked in white. The probability of A is the probability of the corresponding event in* $\mathcal{L}$*. The probabilities of events B, C, and D, which are all in the same equivalence class as A, are equal to the probability of A.*

In the second and third inverse problems, we have used all the information that is possible to obtain from inversion of the map Q. Yet, the third inverse problem does not match the third forward problem. In the third forward problem, we can distinguish the individual probabilities of distinct events in the original σ-algebra $\mathcal{B}_{\mathbf{\Lambda}}$. However, in the third inverse problem, we can only distinguish the individual probabilities of distinct events in the contour σ-algebra $\mathcal{C}_{\mathbf{\Lambda}}$. Distinct events in $\mathcal{B}_{\mathbf{\Lambda}}$ that are in the same equivalence class have the same probability regardless of differences in size, location, and shape; see Figure 2.

However, the events in the original σ-algebra $\mathcal{B}_{\mathbf{\Lambda}}$ have important physical meaning in the context of the model. Different events in $\mathcal{B}_{\mathbf{\Lambda}}$ corresponding to the same equivalence class in $\mathcal{L}$ may represent significantly different physical conditions. Hence, applications in scientific prediction and engineering design naturally involve distinguishing between distinct events in $\mathcal{B}_{\mathbf{\Lambda}}$ as much as possible. Thus, the ideal inferential target is a probability measure with respect to $\mathcal{B}_{\mathbf{\Lambda}}$, not merely with respect to $\mathcal{C}_{\mathbf{\Lambda}}$. But this clearly requires more than the information that can be obtained by inverting Q. Indeed, in general there is no unique solution of this problem without further assumption; see [2].

Roughly speaking, the disintegration theorem discussed in section 4.1 guarantees that *any* measure on $\mathbf{\Lambda}$ can be decomposed into a form involving measures in $\mathcal{L}$ and along each generalized contour corresponding to points in $\mathcal{L}$, in a manner somewhat analogous to a product decomposition. The measures on $\mathcal{L}$ and on each generalized contour are restrictions of measures on $\mathbf{\Lambda}$ that make sense on these sets of $\mu_{\mathbf{\Lambda}}$-measure zero. Considering an application to a probability measure on $\mathbf{\Lambda}$ solving the stochastic inverse problem, the component of the probability measure on $\mathcal{L}$ is determined. Thus, if we prescribe measures along each generalized contour, we obtain a unique solution of the stochastic inverse problem.

The measures along the generalized contours *cannot* be obtained by observation on Q. This motivates the adoption of an ansatz in which the probability measures along generalized contours are specified. The ansatz is stated precisely in section 4.1. With the adoption of the ansatz, there is a unique solution of the stochastic inverse problem from $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$ to $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$. We then show how to approximate this solution using measure-theoretic techniques. If the ansatz is unacceptable, then the inverse problem can be solved only in $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$, i.e., with respect to contour events. Our approximation technique can be modified to approximate this solution as well.

3. Set-valued solutions of deterministic inverse problems. We first consider the underlying problem of defining a solution to the deterministic inverse problem of mapping an n-dimensional set $\Lambda \subset \mathbb{R}^n$ into an $(m \leq n)$-dimensional space $\mathcal{D} \subset \mathbb{R}^m$. Following [2], the set-valued inverses of $Q(\lambda)$ are called *generalized contours*, and any indexing manifold of the generalized contours in $\mathbf{\Lambda}$, i.e., a representation of $\mathcal{L}$ in $\mathbf{\Lambda}$, is called a *transverse parameterization*.

3.1. Linear maps. It is enlightening to first consider the case of a linear map Q defined by m linearly independent linear maps $Q_i(\lambda) = \sum_{j=1}^{n} a_{ij}\lambda_j$ acting on $\mathbf{\Lambda} \subset \mathbb{R}^n$. In this case, we can describe $\mathcal{L}$ and the generalized contours corresponding to points in $\mathcal{L}$ precisely.

Let $A : \mathbf{\Lambda} \to \mathcal{D}$ denote the $m \times n$ matrix $[A]_{ij} = a_{ij}$, so $Q(\lambda) = A\lambda$ is the vector-valued map with components $\mathbf{Q}_i^\top \lambda$. Let $\mathbf{u}_i = \mathbf{Q}_i / \|\mathbf{Q}_i\|$, so that $Q_i(\lambda) = \|\mathbf{Q}_i\| \, \mathbf{u}_i^\top \lambda$, where $\|\cdot\|$ is the usual Euclidean norm. We denote the null space of A by $\mathcal{N}$ and note that $\mathcal{N} = (\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\})^\perp$. We complete $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ to obtain an orthonormal basis $\{\mathbf{u}_{m+1}, \ldots, \mathbf{u}_n\}$ for $\mathcal{N}$. The dual of the quotient space $\mathcal{L} = \Lambda/\mathcal{N}$ can be identified with $\mathcal{N}^\perp$ and is isomorphic to $\mathcal{D}$ by the first isomorphism theorem. For any vector $\lambda \in \Lambda$, let $\tilde{\lambda}$ denote the unique representation of λ using the basis of direction vectors defined by $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n\}$.

Thus, we have written the desired representation $\mathbf{\Lambda} \cong \mathcal{L} \times \mathcal{N}$. If we choose any generalized contour from $\mathcal{L}$ and select a fixed representer from this equivalence class, then the m-dimensional hyperplane defined by span $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ passing through this fixed representer intersects each of the unique equivalence classes of $\mathcal{L}$ once and only once.

Next, we compute the representation of Q on $\mathcal{L} \times \mathcal{N}$. There exists a dual basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ of $n \times 1$ vectors such that $\mathbf{b}_i^\top \mathbf{u}_j = 1$ when $i = j$, and $\mathbf{b}_i^\top \mathbf{u}_j = 0$ when $i \neq j$. Given $\tilde{\lambda}$, the action of A on $\mathbf{\Lambda} \cong \mathcal{L} \times \mathcal{N}$ is defined by left multiplication by the matrix

$$\begin{pmatrix} B^\top A^\top & 0_{m \times (n-m)} \end{pmatrix}, \tag{3.1}$$

where B is the $n \times m$ matrix with ith column the ith dual basis vector $\mathbf{b}_i$. The $m \times m$ matrix $B^\top A^\top$ is the diagonal matrix with ith diagonal given by $\|\mathbf{Q}_i\|$. The ratio of the largest and smallest diagonal entries of $B^\top A^\top$ determines the condition number of the map from $\mathcal{L}$ to $\mathcal{D}$.

Example 1. *Let $\mathbf{\Lambda} = [0,1]^3$ and consider $\mathcal{D} \subset \mathbb{R}^2$ defined by vector-valued map $Q(\lambda) = \begin{pmatrix} Q_1(\lambda) & Q_2(\lambda) \end{pmatrix}^\top$ for two linearly independent linear component maps $Q_1(\lambda) \approx 0.506\lambda_1 + 0.253\lambda_2 + 0.085\lambda_3$ and $Q_2(\lambda) \approx 0.463\lambda_1 + 0.918\lambda_2 + 0.496\lambda_3$. The left and middle plots of Figure 3 show the planes in $\mathbf{\Lambda}$ equal to the generalized contours and lines in $\mathbf{\Lambda}$ equal to the transverse parameterizations for the component maps $Q_1(\lambda)$ and $Q_2(\lambda)$, respectively. The right plot of Figure 3 shows the generalized contours in $\mathbf{\Lambda}$, which are the lines defined by intersections of the planes defining the generalized contours of the component maps, and the transverse parameterization, which is the plane in $\mathbf{\Lambda}$ with normal vector defined by the cross product of the vectors parallel to the lines defining the transverse parameterizations for the component maps.*

Some standard presentations of inverse problems for linear maps present the discussion in the context of a map between spaces of equal dimensions, which makes it easier to carry out regularization by adding an invertible operator. We can use the matrix representation (3.1) of Q to define an invertible map $\tilde{Q}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ that is equivalent to Q. We let $\gamma \in \mathcal{L}$ and $\eta \in \mathcal{N}$ denote the m- and $(n-m)$-dimensional vectors such that $\tilde{\lambda} = \begin{pmatrix} \gamma^\top & \eta^\top \end{pmatrix}^\top$, and let

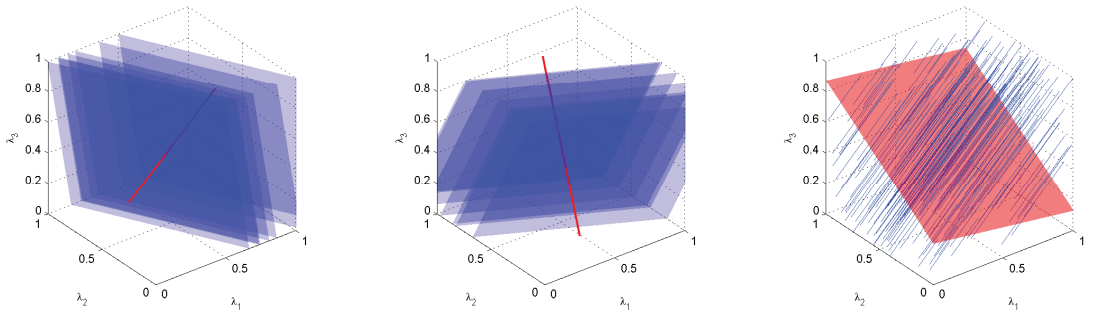

Figure 3. *Left: The planes in* $\mathbf{\Lambda}$ *are generalized contours of* $Q_1(\lambda)$*, and the line indexing these planes is a transverse parameterization. Middle: The planes in* $\mathbf{\Lambda}$ *are generalized contours of* $Q_2(\lambda)$*, and the line indexing these planes is a transverse parameterization. Right: The lines in* $\mathbf{\Lambda}$ *are generalized contours of* $Q(\lambda)$*, and the plane indexing these lines is a transverse parameterization.*

$\tilde{Q}(\gamma, \eta) := \tilde{Q}(\tilde{\lambda}) = \tilde{A}\tilde{\lambda}$ be the bijection between $\mathbf{\Lambda} \cong \mathcal{L} \times \mathcal{N}$ and $\mathcal{D} \times \mathcal{N}$ with matrix

$$\tilde{A} := \begin{pmatrix} B^\top A^\top & 0_{m\times(n-m)} \\ 0_{(n-m)\times m} & I_{(n-m)\times(n-m)} \end{pmatrix}. \tag{3.2}$$

For fixed components $d_1, \ldots, d_m$ and arbitrary parameter values $d_{m+1}, \ldots, d_n$ in $\mathbb{R}^{n-m}$, the solution to the square system $\tilde{A}\tilde{\lambda} = \mathbf{d}$ defines the same affine hyperplane as the generalized contour corresponding to $Q^{-1}(d_1, \ldots, d_m)$.

Any invertible matrix $C_{(n-m)\times(n-m)}$ can be used in (3.2) instead of $I_{(n-m)\times(n-m)}$. It is a modeling choice to use the identity to propagate vectors in the null space $\mathcal{N}$ justified by simplicity. If having to make such a modeling choice is unacceptable, then reverting to (3.1) is the alternative.

3.2. Nonlinear maps. We return to the general case of a nonlinear map Q.

3.2.1. Brief review for $m = 1$. In [2], when $m = 1$, we prove that any two distinct generalized contours associated with distinct output values are unique nonintersecting manifolds. Additionally, we show that there exists a piecewise-smooth one-dimensional curve that intersects each generalized contour once and only once, that this curve can be constructed by a finite number of connected curves, and that this curve can be used as the transverse parameterization representing $\mathcal{L}$ in $\mathbf{\Lambda}$.

3.2.2. The case $m > 1$. We prove the existence of both the generalized contours and the transverse parameterization for the case of multiple quantities of interest. The extension from $m = 1$ requires characterization of the specification of "multiple quantities of interest."

Definition 3.1. *We say that the component maps of* m*-dimensional piecewise-smooth vector-valued map* $Q(\lambda)$ *are* geometrically distinct *(GD) if the Jacobian of* Q *has full rank at every point in* $\mathbf{\Lambda}$*.*

Example 2. *Consider a model* M *with input parameters* x *and* y*. Suppose*

$$\begin{pmatrix} z \\ \theta \end{pmatrix} = \begin{pmatrix} x^2 + y^2 \\ \textit{angle of } (x, y) \textit{ with the } x - \textit{axis} \end{pmatrix}$$

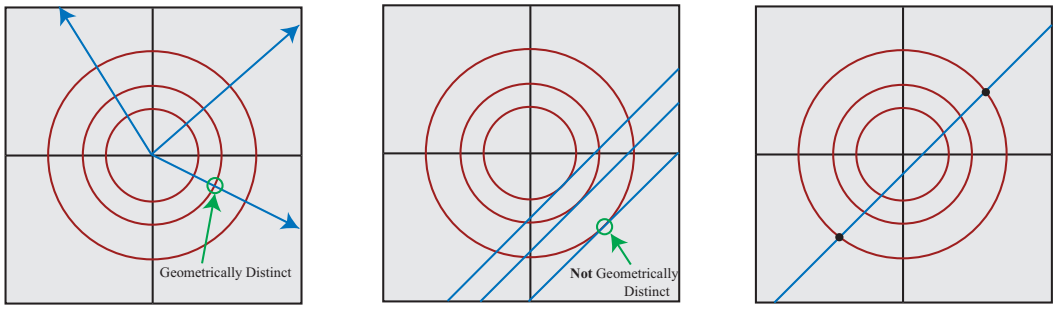

Figure 4. *Left: The generalized contours for GD Q. Middle: The generalized contours for non-GD $\tilde{Q}$. Right: Inverting both components of $\tilde{Q}$ gives answers consisting of pairs of points almost everywhere.*

and $Q := (z, \theta)^\top$. The generalized contours of the first component of Q are circles, while the generalized contours for the second component of Q are rays from the origin; see Figure 4. The two sets of generalized contours are GD. If we invert from both components of Q simultaneously, then the solutions consist of individual points since the map is 1–1 and onto.

Now suppose

$$\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} x^2 + y^2 \\ y - x \end{pmatrix},$$

with $\tilde{Q} := (z, w)^\top$. The generalized contours for the first component are again circles, while the generalized contours for the second component are lines with slope 1; see Figure 4. Given a generalized contour of either component, we can find generalized contours of the other component that are tangent at an intersection point, thus violating the condition for $\tilde{Q}$ being GD. The consequence is that if we invert using both components of $\tilde{Q}$ simultaneously, the solutions are equivalence classes consisting of two points almost everywhere.

Henceforth, we assume that $Q : \mathbf{\Lambda} \to \mathcal{D}$ is a locally differentiable nonlinear vector-valued map with GD component maps. We adopt the following notational conventions:

- points in $\mathcal{D}$ are denoted by q,
- superscripts such as $q^{(1)}$ and $q^{(2)}$ differentiate between points in $\mathcal{D}$, and
- subscripts denote a specific coordinate; e.g., $q_2^{(3)}$ denotes the second coordinate of the third point.

Theorem 3.1. *If $q^{(1)}, q^{(2)} \in \mathcal{D}$ are distinct points, then there exist unique collections of $(n-m)$-dimensional manifolds defining the two generalized contours for $q^{(1)}$ and $q^{(2)}$ that are unique and do not intersect.*

Proof. The multivariate implicit function theorem guarantees existence of locally smooth $(n-m)$-dimensional manifolds whose unions define the separate generalized contours in $\mathbf{\Lambda}$. While these separate generalized contours associated with distinct output points can be defined by unions of disconnected $(n-m)$-dimensional manifolds, they cannot intersect, as this would contradict Q being a function. ∎

It remains to show that there exists a transverse parameterization given by an m-dimensional manifold indexing the $(n-m)$-dimensional generalized contours of the vector-valued map $Q(\lambda)$. We present this as an independently interesting result.

Theorem 3.2. *There exists an m-dimensional manifold defining a transverse parameteriza-*

tion for the set of generalized contours. Moreover, a piecewise-linear m-dimensional transverse parameterization can be constructed in a finite number of steps.[2]

Proof. The generalized contours for $Q(\lambda)$ are defined by intersections of m distinct $(n-1)$-dimensional generalized contours from the m component maps indexed by points in $\mathcal{D}$. The goal is to construct an alternative representation of $\mathcal{L}$ consisting of points in $\mathbf{\Lambda}$. The assumption of local differentiability of $Q(\lambda)$ implies that for almost every $\lambda \in \mathbf{\Lambda}$ the Jacobian of $Q(\lambda)$ is computable. The kth row of the Jacobian corresponds to the normal vector of the $(n-1)$-dimensional generalized contour for the kth component $Q_k(\lambda)$. For fixed $q \in \mathcal{D}$, choose any $\lambda \in \{\lambda \,|\, Q(\lambda) = q\}$ such that the Jacobian is full rank. Such a point is guaranteed to exist by the GD assumption. Let $\nu_{\lambda,q}$ denote the generalized cross-product of the m rows of the Jacobian of $Q(\lambda)$ evaluated at the choice of λ. The point λ and the vector $\nu_{\lambda,q}$ define a hyperplane through λ that *locally intersects distinct* generalized contours since it is spanned by the normal vectors to the m distinct $(n-1)$-dimensional generalized contours of each of the component maps. We may repeat this procedure at a countable number of points in $\mathbf{\Lambda}$ to define a piecewise-linear m-dimensional manifold in $\mathbf{\Lambda}$ such that evaluation of $Q(\lambda)$ at points on this manifold defines an open cover of $\mathcal{D}$. Since $\mathcal{D}$ is compact, there is a finite subcover. This implies that we can construct a piecewise-linear m-dimensional manifold in a finite number of steps that intersects once and only once each of the unique $(n-m)$-dimensional manifolds defining the unique generalized contours of $Q(\lambda)$. Any such m-dimensional piecewise-defined manifold is a transverse parameterization. ∎

Locally, the set-valued inverses and the transverse parameterization can be approximated by linear manifolds. We may then use a local change of coordinates, a.e., as in the linear case, to describe the local linear approximations to the $(n-m)$-dimensional manifolds defining the generalized contours as $\mathbf{\Lambda} \cong \mathcal{L} \times \mathcal{N}$.

Example 3. *Let $\mathbf{\Lambda} = [0.05, 1]^3$ and $Q(\lambda) = (\, Q_1(\lambda)\ Q_2(\lambda)\,)^\top$, where $Q_1(\lambda) = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$ and $Q_2(\lambda) = \arccos(\lambda_3/\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2})$. Figure 5(left) shows the set $\mathcal{D} := Q(\Lambda) \subset \mathbb{R}^2$ of all possible output data. The generalized contours for $Q_1(\lambda)$ are the intersections of spherical shells centered at the origin with the set $\mathbf{\Lambda}$, and any straight line segment from $(0.05, 0.05, 0.05)$ to $(1, 1, 1)$ is a transverse parameterization. The generalized contours for $Q_2(\lambda)$ are the intersections of cones centered at the origin with $\mathbf{\Lambda}$, and any straight line segment from $(1, 1, 0.05)$ to $(0.05, 0.05, 1)$ is a transverse parameterization. The generalized contours for $Q(\lambda)$ are circles of various radii centered at the origin at various heights in planes perpendicular to the λ_3-axis. A transverse parameterization may be described by the half-plane through $\mathbf{\Lambda}$, as shown in Figure 5(right). Thus, the value of λ_3 is uniquely determined by any $Q(\lambda)$, whereas the values of λ_1 and λ_2 can only be determined as being upon a circle of fixed radius in the horizontal plane defined by λ_3 constant.*

A piecewise-linear approximation to Q and a transverse parameterization can be constructed systematically using interpolation on a triangulation of $\mathbf{\Lambda}$ (see [2] for an alternative approach). We note that this approximation is *not* used in the algorithms for solving the stochastic inverse problem discussed below.

Theorem 3.3. *Suppose that Q is locally differentiable with GD component maps. There*

[2] If $\mathbf{\Lambda}$ is not compact but $\mu_\Lambda(\Lambda)$ is finite, then there is no guarantee that the number of steps is finite, and the proof is no longer constructive.

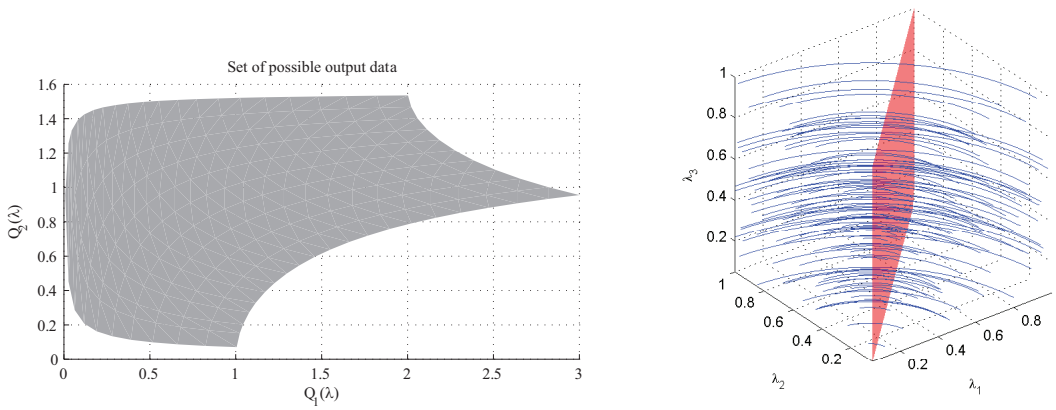

Figure 5. *Left: The set $\mathcal{D}$ of possible output data. Right: The quarter circles of various radii and vertical height represent a random collection of unique generalized contours for the nonlinear vector-valued map $Q(\lambda)$, and the transverse parameterization that puts into* 1–1 *correspondence these set-valued inverses and the points in $\mathcal{D}$ is plotted as the half-plane.*

exists a sequence of piecewise-linear approximations that converge pointwise to a manifold in $\mathbf{\Lambda}$ defining a transverse parameterization.

Proof. Given a triangulation of $\mathbf{\Lambda}$, we evaluate Q at the vertices of each generalized triangle to construct a piecewise-linear interpolant $\tilde{Q}$. The evaluation of $\tilde{Q}$ on the vertices of the input generalized triangles defines both a cover and triangulation of $\tilde{\mathcal{D}} := \tilde{Q}(\mathbf{\Lambda})$. Since we assume $m < n$, each generalized m-dimensional triangle in $\tilde{\mathcal{D}}$ is in 1-1 correspondence with m vertices of an n-dimensional generalized triangle in $\mathbf{\Lambda}$ that shares a single "face" of this n-dimensional generalized triangle by construction. In other words, the triangulation of $\tilde{\mathcal{D}}$ defines a triangulation of the *faces* of the input generalized triangles into unions of m-dimensional triangles. Thus, by construction of the cover of $\tilde{\mathcal{D}}$, there exists a finite index map between the generalized triangles in $\tilde{\mathcal{D}}$ and a triangulation of the faces of generalized triangles in $\mathbf{\Lambda}$. We compute a transverse parameterization for the map $\tilde{Q}$ in the following way: (1) we choose a finite nonintersecting subcover of $\tilde{\mathcal{D}}$ from the above triangulation, and (2) for each generalized triangle in $\tilde{\mathcal{D}}$ we use the index map to choose a specific m-dimensional triangle on a single face of the associated input generalized triangle. By construction of the local interpolant, the union of such m-dimensional triangles in $\mathbf{\Lambda}$ forms a linear approximation to an m-dimensional manifold that indexes the local linear approximations to the $(n-m)$-dimensional generalized contours that are contained in the input generalized triangles. Since Q is assumed to be locally differentiable, for any $\epsilon > 0$, there exists a triangulation of $\mathbf{\Lambda}$ such that

$$\max\{|Q(\lambda) - \tilde{Q}(\lambda)| : \lambda \in \mathbf{\Lambda}\} < \epsilon. \tag{3.3}$$

We define a sequence of triangulations of $\mathbf{\Lambda}$ and piecewise linear approximations $\tilde{Q}$ satisfying (3.3) with ϵ_k, where $\{\epsilon_k\}$ is a sequence such that $\epsilon_k \downarrow 0$. This sequence of piecewise-linear $(n-m)$-dimensional approximations converges pointwise to the generalized contours, and subsequently the corresponding sequence of piecewise-linear m-dimensional manifolds converges

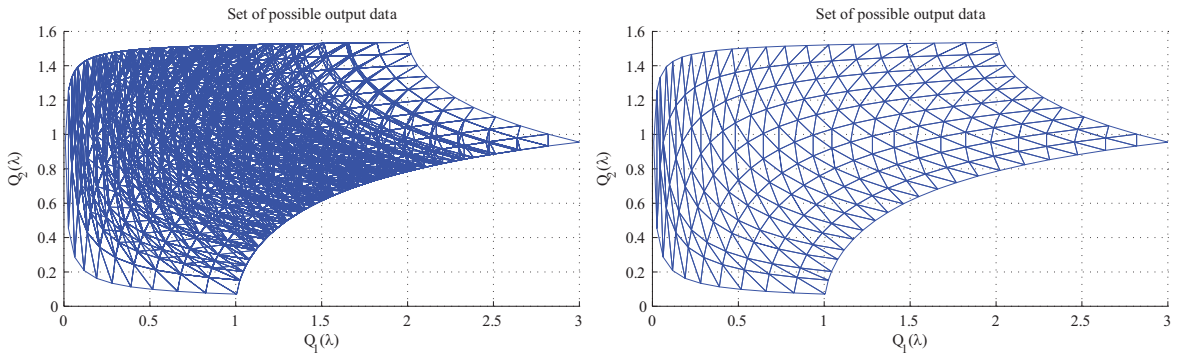

Figure 6. *Left: Sets that cover $\mathcal{D}$ obtained by propagation of each triangular input element used in the coarse piecewise-linear interpolant approximating the vector-valued map. Right: A finite subcover of $\mathcal{D}$ used to define the global approximation to the transverse parameterization that indexes the generalized contours.*

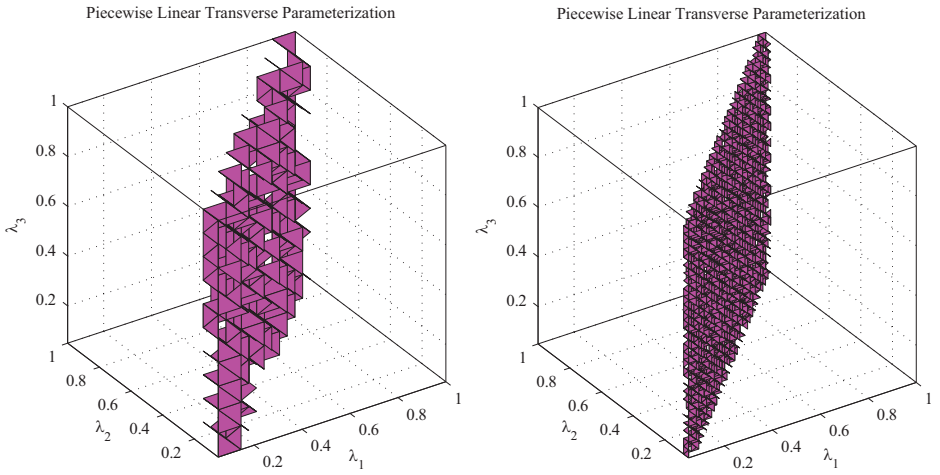

Figure 7. *Left: The approximation to the transverse parameterization computed using the finite subcover shown in Figure 6. Right: A refined approximation to the transverse parameterization computed using a finite subcover of $\mathcal{D}$ obtained from the refined piecewise-linear interpolant to $Q(\lambda)$.*

to a transverse parameterization. ■

Example 4. *To illustrate the convergence of approximate piecewise-linear transverse parameterizations to the transverse parameterization shown in Figure 5(right), we show approximations at two resolutions. We compute two piecewise-linear interpolants approximating Q, using first $6,000$ and then $162,000$ tetrahedral elements to define a triangulation of $\mathbf{\Lambda}$. Figure 6(left) shows the cover of $\mathcal{D}$ defined from each local interpolant across each element of the coarser approximation. The right plot in Figure 6 shows one particular subcover of $\mathcal{D}$ used to define the transverse parameterization for this coarse approximation. The transverse parameterizations computed using the coarse and fine piecewise-linear interpolants are shown in Figure 7.*

4. The stochastic inverse problem.

We now consider the solution of the stochastic inverse problem. The computation of a probability measure on $\mathbf{\Lambda}$ corresponding to a probability

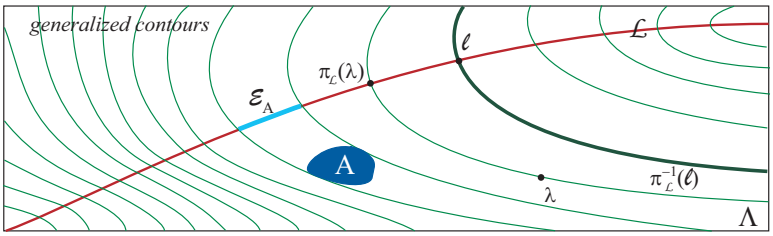

Figure 8. *Notation for solution of the stochastic inverse problem.*

measure on $\mathcal{D}$ depends on the geometric structure imbued by the generalized contours of Q and the volume measures on $\mathbf{\Lambda}$ and $\mathcal{D}$. In section 4.1, we prove the existence of a solution to the stochastic inverse problem. In section 4.2, we discuss an approximation of the solution.

4.1. Existence, uniqueness, and structure of the solution. Recall that in section 2 we described how the inverse of the map Q induces a probability space $(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, P_{\mathcal{L}})$ from a probability space $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, P_{\mathcal{D}})$, where $\mathcal{L}$ is the space of equivalence classes corresponding to Q^{-1}. Theorem 3.2 implies that $\mathcal{L}$ can be described by a transverse parameterization in $\mathbf{\Lambda}$, so we abuse notation to let $\mathcal{L}$ denote any such transverse parameterization. We define the equivalence map $\pi_{\mathcal{L}} : \mathbf{\Lambda} \to \mathcal{L}$ by $\pi_{\mathcal{L}}(A) = \mathcal{E}_A$ for $A \in \mathcal{B}_{\mathbf{\Lambda}}$, where recall that $\mathcal{E}_A$ denotes the event in $\mathcal{B}_{\mathcal{L}}$ corresponding to event $A \in \mathcal{B}_{\mathbf{\Lambda}}$. We illustrate the notation in Figure 8.

The next step is to use the disintegration theorem to decompose *any* measure on the measurable space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ into a combination of measures on $\mathcal{L}$ and on the generalized contours corresponding to points in $\mathcal{L}$.

The disintegration theorem. Note that the assumption that Q is locally differentiable implies that it is a measurable map. Adapting the disintegration theorem [8, 7], we have the following.

Theorem 4.1 (the disintegration theorem). *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps, and assume that Ψ is a measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$. There is a family of measures $\{\Psi_\ell\}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ defined for almost every $\ell \in \mathcal{L}$ such that*

$$\Psi_\ell(\lambda) = 0, \quad \lambda \in \mathbf{\Lambda} \setminus \pi_{\mathcal{L}}^{-1}(\ell), \quad \text{a.e. } \ell \in \mathcal{L},$$

i.e., $\Psi_\ell(A) = \Psi_\ell(\pi_{\mathcal{L}}^{-1}(\ell) \cap A)$ for all $A \in \mathcal{B}_{\mathbf{\Lambda}}$, and which gives the following disintegration for Ψ:

$$\Psi(A) = \int_{\mathcal{E}_A} \Psi_\ell(A)\, d\mu_{\mathcal{L}}(\ell) = \int_{\mathcal{E}_A} \left(\int_{\pi_{\mathcal{L}}^{-1}(\ell) \cap A} d\Psi_\ell(\lambda) \right) d\mu_{\mathcal{L}}(\ell) \tag{4.1}$$

for $A \in \mathcal{B}_{\mathbf{\Lambda}}$.

Proof. The proof follows from the disintegration theorem in [7] (see Appendix A) after identifying

$$\int_{Q(A)} \Psi_{Q^{-1}(q) \cap \mathcal{L}}(A)\, d\mu_{\mathcal{D}}(q) = \int_{\mathcal{E}_A} \Psi_\ell(A)\, d\mu_{\mathcal{L}}(\ell).$$

We note that the events in the σ-algebras on $\mathcal{L}$ and $\pi_{\mathcal{L}}^{-1}(\ell)$, $\ell \in \mathcal{L}$, can be obtained by restriction of events in $\mathcal{B}_{\mathbf{\Lambda}}$ to $\mathcal{L}$ and $\pi_{\mathcal{L}}^{-1}(\ell)$, respectively. ∎

The disintegration theorem is extremely useful in a variety of contexts. For example, in product measure spaces, the disintegration theorem can be applied to the coordinate projection maps to obtain a disintegration of measure familiar from Fubini's theorem. The disintegration theorem also justifies the restriction of a measure to the zero-measure boundaries of many types of domains for evaluation of integrals in theorems such as Green's theorem, the divergence theorem, and Stokes' theorem.

We can use Theorem 4.1 to disintegrate the volume measure $\mu_{\mathbf{\Lambda}}$, which in turn allows definition of the solution of the stochastic inverse problem. We set $\mu_{\mathcal{N}}(\ell;\cdot) = \Psi_\ell(\cdot)$ in the case that $\Psi = \mu_{\mathbf{\Lambda}}$. We obtain the following result.

Corollary 4.1 (disintegration of the volume measure). *For any $A \in \mathcal{B}_{\mathbf{\Lambda}}$,*

$$\mu_{\mathbf{\Lambda}}(A) = \int_{\mathcal{E}_A} \mu_{\mathcal{N}}(\ell; A)\, d\mu_{\mathcal{L}}(\ell) = \int_{\mathcal{E}_A} \int_{\pi_{\mathcal{L}}^{-1}(\ell)\cap A} d_\lambda \mu_{\mathcal{N}}(\ell;\lambda)\, d\mu_{\mathcal{L}}(\ell).$$

The solution of stochastic inverse problems. We use Corollary 4.1 to define a solution of the stochastic inverse problem from $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ to $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ in terms of probability densities with respect to the measures $\mu_{\mathcal{N}}$ and $\mu_{\mathcal{L}}$. We begin by discussing the solution of the stochastic inverse problem from $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ to $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$.

Theorem 4.2. *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps. Given a probability measure $P_{\mathcal{D}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ that is absolutely continuous with respect to $\mu_{\mathcal{D}}$, there exists a unique probability measure $P_{\mathcal{L}}$ on $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$ that is absolutely continuous with respect to $\mu_{\mathcal{L}}$.*

Proof. The map Q is a bijection between $\mathcal{B}_{\mathcal{D}}$ and $\mathcal{B}_{\mathcal{L}}$, and setting $P_{\mathcal{L}}(A) = P_{\mathcal{D}}(Q(A))$ for $A \in \mathcal{B}_{\mathcal{L}}$ defines the solution to the stochastic inverse problem between $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ and $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$. If $\mu_{\mathcal{L}}(A) = 0$ for $A \in \mathcal{B}_{\mathcal{L}}$, then by construction $\mu_{\mathcal{D}}(Q(A)) = 0$ and $P_{\mathcal{D}}(Q(A)) = P_{\mathcal{L}}(A) = 0$. So, $P_{\mathcal{L}}$ is absolutely continuous with respect to $\mu_{\mathcal{L}}$. Thus, there exists density $\rho_{\mathcal{L}}$ such that

$$P_{\mathcal{L}}(A) = \int_A \rho_{\mathcal{L}}(\ell) d\mu_{\mathcal{L}}(\ell) = \int_A \rho_{\mathcal{L}} d\mu_{\mathcal{L}}, \quad A \in \mathcal{B}_{\mathcal{L}}.$$

This solution is unique in $L_1(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, \mu_{\mathcal{L}})$ since any other probability measure on $\mathcal{L}$ that solves the inverse problem and is absolutely continuous with respect to $\mu_{\mathcal{L}}$ must have a density that is equal to $\rho_{\mathcal{L}}$ a.e. ∎

This result gives a solution of the stochastic inverse problem from $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ to $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}})$ in terms of a probability density, where recall that $\mathcal{C}_{\mathbf{\Lambda}} \subset \mathcal{B}_{\mathbf{\Lambda}}$ is the σ-algebra of contour events on $\mathbf{\Lambda}$ generated by the set of equivalence classes corresponding to a generating set for $\mathcal{B}_{\mathcal{L}}$.

Theorem 4.3. *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps. Given a probability measure $P_{\mathcal{D}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ that is absolutely continuous with respect to $\mu_{\mathcal{D}}$, there exists a unique probability measure $P_{\mathbf{\Lambda},\mathcal{C}_{\mathbf{\Lambda}}}$ on $(\Lambda, \mathcal{C}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$.*

Proof. The equivalence relation determined by $\pi_{\mathcal{L}}$ defines a measurable map from $\mathbf{\Lambda}$ to $\mathcal{L}$ so that, for arbitrary $A \in \mathcal{C}_{\mathbf{\Lambda}}$, $P_{\mathbf{\Lambda},\mathcal{C}_{\mathbf{\Lambda}}}(A) := P_{\mathcal{L}}(\mathcal{E}_A)$ is well defined. The conclusion follows from Theorem 4.2. ∎

However, we seek a solution of the stochastic inverse problem from $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ to $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$. To define the solution, we use a disintegration of any probability measure imposed on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ given by Theorem 4.1 and Theorem 4.2.

Theorem 4.4 (disintegration of a probability measure). *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps, and assume that $P_{\mathbf{\Lambda}}$ is a probability measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$. There is a family of probability measures $\{P_\ell\}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ defined for almost every $\ell \in \mathcal{L}$ such that $P_\ell(\lambda) = 0$, $\lambda \in \mathbf{\Lambda} \setminus \pi_{\mathcal{L}}^{-1}(\ell)$, a.e. $\ell \in \mathcal{L}$, and*

$$P_{\mathbf{\Lambda}}(A) = \int_{\mathcal{E}_A} P_\ell(A)\, \rho_{\mathcal{L}}(\ell)\, d\mu_{\mathcal{L}}(\ell) = \int_{\mathcal{E}_A} \left(\int_{\pi_{\mathcal{L}}^{-1}(\ell) \cap A} dP_\ell(\lambda) \right) \rho_{\mathcal{L}}(\ell)\, d\mu_{\mathcal{L}}(\ell) \tag{4.2}$$

for $A \in \mathcal{B}_{\mathbf{\Lambda}}$.

Note that P_ℓ is the conditional probability on $\mathbf{\Lambda}$ for the event defined by $\{\lambda \,|\, Q(\lambda) = Q(\ell)\}$.

We see that the disintegration theorem combined with the geometric structure induced by the set-valued inverse of Q imports a significant amount of structure on any probability measure $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$. Indeed, $P_{\mathbf{\Lambda}}$ is uniquely determined by specifying P_ℓ valued on generalized contours $\pi_{\mathcal{L}}^{-1}(\ell)$ for $\ell \in \mathcal{L}$.

We now state the following.

Ansatz: We assume that for $\ell \in \mathcal{L}$ a probability density $\rho_{\mathcal{N},}(\ell; \cdot)$ is given on the generalized contour corresponding to ℓ such that

$$dP_\ell(\lambda) = \rho_{\mathcal{N}}(\ell; \lambda)\, d_\lambda \mu_{\mathcal{N}}(\ell; \lambda), \quad \lambda \in \pi_{\mathcal{L}}^{-1}(\ell) \cap A,\ A \in \mathcal{B}_{\mathbf{\Lambda}}. \tag{4.3}$$

This yields the next result.

Theorem 4.5. *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps. Assume that $P_{\mathbf{\Lambda}}$ is a probability measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and that the ansatz holds. There exists a unique probability measure $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$, and*

$$P_{\mathbf{\Lambda}}(A) = \int_{\mathcal{E}_A} \left(\int_{\pi_{\mathcal{L}}^{-1}(\ell) \cap A} \rho_{\mathcal{N}}(\ell; \lambda)\, d_\lambda \mu_{\mathcal{N}}(\ell; \lambda) \right) \rho_{\mathcal{L}}(\ell)\, d\mu_{\mathcal{L}}(\ell) \tag{4.4}$$

for $A \in \mathcal{B}_{\mathbf{\Lambda}}$.

The approximation method described in section 4.2 can be used for any choice of density $\rho_{\mathcal{N}}$. However, we prefer a specific choice of density which amounts to assuming that a "non-preferential" weighting determined by the volume measure is used to compute probabilities of events inside of a contour event. Namely, we use the next formulation.

Standard choice for ansatz:

$$\rho_{\mathcal{N}}(\ell; \cdot) = \frac{1}{\mu_{\mathcal{N}}(\ell; \pi_{\mathcal{L}}^{-1}(\ell))}, \quad \ell \in \mathcal{L}. \tag{4.5}$$

In the case of a product Lebesgue volume measure on $\mathcal{B}_{\mathbf{\Lambda}}$, our choice for the ansatz implies a uniform density along generalized contours.

This choice is related to a fundamental question about modeling probability. Consider a compact metric space with a volume measure. In the absence of any specific information about probabilities of events in the metric measure space, is there a natural way to impose

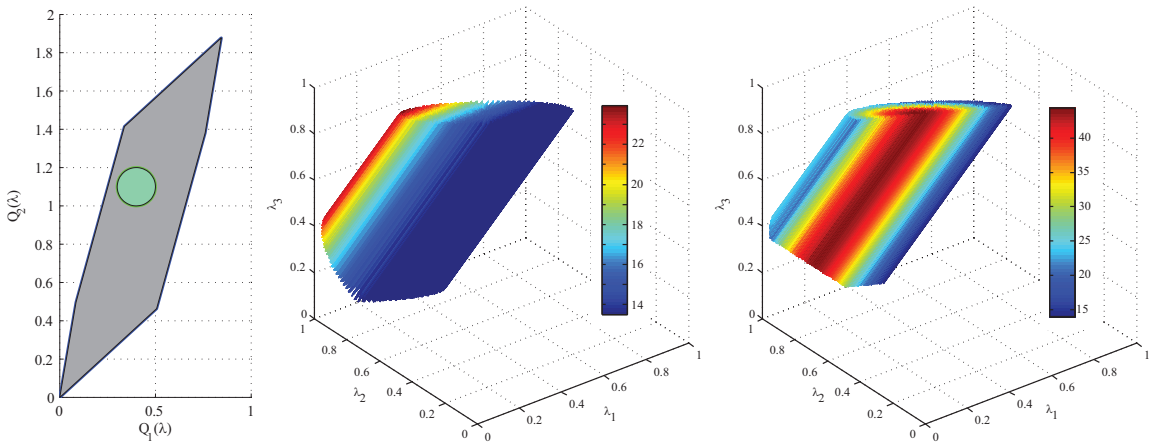

Figure 9. *Left: Plot of $\mathcal{D} = Q(\mathbf{\Lambda})$ indicating the set of possible output values. The disc indicates a set of possible measurement data I. Middle: Plot of the inverse density $\rho_{\mathbf{\Lambda}}$ corresponding to a uniform density on I. Right: Cross section of the inverse density $\rho_{\mathbf{\Lambda}}$ corresponding to a truncated normal density on I.*

a probability measure? Simplicity argues for apportioning probability to events according to relative volume (measure) size. Compare to the discussion in section 3.1 dealing with set-valued solutions of a linear map. We are applying this model choice in the generalized contours. An important characteristic of this choice is that it respects some geometric properties of the volume measure such as translation invariance and isotropy, e.g., that hold for a product Lebesgue volume measure. Moreover, the same ansatz applies to any Q that may be considered and applied without modification if the dimension of Q is changed, e.g., because data becomes available on an additional quantity of interest. While any choice of density represents a priori knowledge about the probability structure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that cannot come from observations on the map Q, other choices involve assuming additional structure beyond that of (4.5).

Example 5. *We consider the linear map in Example 1. We assume that the real data point q is some point in the disc I of radius* 0.1 *in $\mathcal{D}$ centered at* $(0.4, 1.1)$*; see Figure 9(left). The support of the inverse is a "truncated" cylinder (middle plot of Figure 9). For the ansatz, we set the conditional density along each generalized contour to be the reciprocal of its Euclidean length in domain $\mathbf{\Lambda}$. Since the lengths of the generalized contours vary, the resulting density of $P_{\mathbf{\Lambda}}$ is* not *uniform.*

In the first computation, we assume that the data point q is distributed uniformly in I. This gives the probability measure on $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}})$ defined by the density $\rho_{\mathcal{L}}(\ell) = \frac{1}{2\pi(0.1)^2}\mathbf{1}_{Q(\ell)\in I}(\ell)$. In Figure 9(middle), we plot the corresponding approximate inverse density $\rho_{\mathbf{\Lambda}}$ for a large sampling of generalized contours.

In the second computation, we assume that q is distributed according to a truncated multivariate normal distribution centered in $I \subset \mathcal{D}$. This gives

$$\rho_{\mathcal{L}}(\ell) = \nu \exp((Q(\ell) - q_0)\Sigma^{-1}(Q(\ell) - q_0))\mathbf{1}_{Q(\ell)\in I}(\ell),$$

where $\Sigma = .005\mathbf{I}_{2\times 2}$, q_0 is the center of disc I, and ν is a normalizing constant. Figure 9(right) shows the corresponding approximate inverse density $\rho_{\mathbf{\Lambda}}$.

We emphasize once more that our numerical approach for approximating the inverse density applies to any assumption for the ansatz.

4.2. Approximations of solutions. We turn to the approximation of the solutions of stochastic inverse problems given by Theorems 4.2–4.5. There are several approximation issues that need to be addressed in any practical computation involving $P_{\mathbf{\Lambda}}$ arising from the abstract theoretical development and numerical implementation. The fundamental approximation issues are

- approximation of events in $\mathcal{B}_{\mathcal{D}}$,
- approximation of events in $\mathcal{B}_{\mathbf{\Lambda}}$,
- approximation of events in $\mathcal{C}_{\mathbf{\Lambda}}$.

The numerical approximation issues are

- approximation of $P_{\mathcal{D}}$ and subsequently $P_{\mathbf{\Lambda}}$ on partitions of $\mathcal{D}$ and $\mathbf{\Lambda}$,
- error in numerical evaluation of the model.

In [5], we carry out a priori convergence analysis and a posteriori estimation analysis to deal with both numerical approximation issues in the case of a single quantity of interest for a somewhat different but related approximation. Aspects of that analysis are very similar for multiple quantities of interest and the approximation in this paper. Hence, we assume that the model solution can be evaluated exactly.

We approximate events in $\mathcal{D}$ and $\mathbf{\Lambda}$ by taking finite collections of generating events for the σ-algebras $\mathcal{B}_{\mathcal{D}}$ and $\mathcal{B}_{\mathbf{\Lambda}}$ that satisfy a certain approximation property. We let $\{I_i\}$ and $\{b_j\}$ denote sequences of Borel sets generating $\mathcal{B}_{\mathcal{D}}$ and $\mathcal{B}_{\mathbf{\Lambda}}$, respectively, and let $\{I_i\}_{i=1}^{M}$ and $\{b_j\}_{j=1}^{N}$ denote a finite partition of $\mathcal{D}$ and $\mathbf{\Lambda}$ taken from these sequences. We assume that any event in $\mathcal{B}_{\mathcal{D}}$ (respectively, $\mathcal{B}_{\mathbf{\Lambda}}$) can be approximated in the corresponding measure $\mu_{\mathcal{D}}$ (respectively, $\mu_{\mathbf{\Lambda}}$) by such finite partitions. Since $\mathcal{C}_{\mathbf{\Lambda}} \subset \mathcal{B}_{\mathbf{\Lambda}}$, unions of elements in $\{b_j\}_{j=1}^{N}$ simultaneously approximate events in $\mathcal{C}_{\mathbf{\Lambda}}$ and $\mathcal{B}_{\mathbf{\Lambda}}$.

We employ two different types of generating sets. Following [2], we may use collections of generalized "rectangles." For typical metrics on $\mathbb{R}^n$ we can approximate events using such rectangular cells (e.g., see Theorem 2.41 on p. 70 of [10]). However, the computational performance of this approach is affected negatively as the dimension increases. We also employ Voronoi cell tesselations based on a point process sampling of the domain [14]. This approach works well in high dimensions.

We let $\rho_{\mathcal{D},M}$ and $\rho_{\mathbf{\Lambda},N}$ denote simple function approximations to the densities of $P_{\mathcal{D}}$ and $P_{\mathbf{\Lambda}}$ on the partitions. We use these measure-theoretic approximations to prove the following.

Theorem 4.6. *Suppose that $Q : \mathbf{\Lambda} \to \mathcal{D}$ is sufficiently smooth with GD component maps. Given probability measure $P_{\mathcal{D}}$ that is absolutely continuous with respect to $\mu_{\mathcal{D}}$ and event $A \in \mathcal{B}_{\mathbf{\Lambda}}$, there exists a sequence of approximations $P_N(A)$ using simple function approximations to $\rho_{\mathbf{\Lambda},N}$ and $\rho_{\mathcal{D},M}$ requiring only calculations of volumes in $\mathbf{\Lambda}$ that converges to $P(A)$ as $N, M \to \infty$.*

Proof. By assumption, the Radon–Nikodym density $\rho_{\mathcal{D}} \in L_1(\mathcal{D})$ has a piecewise-continuous representation. There exists a partition $\{\tilde{I}_i\}_{i=1}^{\tilde{M}}$ of $\mathcal{D}$ such that $\rho_{\mathcal{D}}$ is continuous on each $\tilde{I}_i$. On any $\tilde{I}_i$ where $\rho_{\mathcal{D}}$ is bounded on the smallest compact set containing $\tilde{I}_i$, we define $g(q) := \sup\{\rho_{\mathcal{D}}(q) : q \in \tilde{I}_i\}$. On the remaining $\tilde{I}_i$, we use the fact that $\rho_{\mathcal{D}}(q) \in L_1(\mathcal{D})$, which implies there exists an $\epsilon_i > 0$ such that $\rho_{\mathcal{D}}(q) < g(q)$, with $g(q) := \nu_i \|q - q^{(i)}\|_2^{-(m-\epsilon)}$ on $\tilde{I}_i$,

where m is the dimension of $\mathcal{D}$, ν_i is a constant, $q^{(i)}$ is a fixed point in $\tilde{I}_i$, and $\|\cdot\|_2$ is the usual Euclidean norm. Since $\mu_{\mathbf{\Lambda}}(\mathbf{\Lambda})$ is finite, $\mu_{\mathcal{D}}(\mathcal{D}) < \infty$ and $g \in L_1(\mathcal{D})$.

Now consider any refinement $\{I_i\}_{i=1}^{M}$ of $\{\tilde{I}_i\}_{i=1}^{\tilde{M}}$, and define the simple function approximation,

$$\rho_{\mathcal{D},M} = \sum_{i=1}^{M} p_i \mathbf{1}_{I_i}(q), \quad p_i = \int_{I_i} \rho_{\mathcal{D}}\, d\mu_{\mathcal{D}}.$$

The function $g \in L_1(\mathcal{D})$ satisfies the inequality $\rho_{\mathcal{D},M} \le g$ for all M. Moreover, $\int \rho_{\mathcal{D}}\, d\mu_{\mathcal{D}} = \int \rho_{\mathcal{D},M}\, d\mu_{\mathcal{D}}$ for all M. Finally, if we consider a sequence of such refinements of $\{\tilde{I}_i\}_{i=1}^{\tilde{M}}$ corresponding to increasing M, the corresponding simple function approximations converge to $\rho_{\mathcal{D}}$ pointwise.

For any refinement $\{I_i\}_{i=1}^{M}$ of $\{\tilde{I}_i\}_{i=1}^{\tilde{M}}$, we let $A_i := Q^{-1}(I_i) \in \mathcal{B}_{\mathcal{L}}$ denote the set of induced generalized contours. By construction, $\{A_i\}_{i=1}^{M}$ defines a partition of $\mathcal{L}$ where $\mu_{\mathcal{L}}(A_i) > 0$ for each i, and $\mu_{\mathcal{L}}(A_i \cap A_j) = 0$ for any $i \neq j$ since any intersection between A_i and A_j can occur only on their boundaries, which define a set of zero $\mu_{\mathcal{L}}$-measure. Furthermore, the interior of any A_i contains all of the unique generalized contours associated with the interior of I_i. By the equivalence relation, we can also identify each A_i as a contour event in $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}})$. From Theorem 4.3, it follows that

$$\rho_{\mathcal{L},M}(C) \approx \sum_{i=1}^{M} p_i \mathbf{1}_{A_i}(C), \qquad C \in \mathcal{C}_{\mathbf{\Lambda}},$$

defines an approximation to the unique solution to the stochastic inverse problem in $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}})$.

To build the approximation of events in $\mathbf{\Lambda}$, we identify each b_j as a member of a particular contour event A_i using, for example, the value of Q at the barycentric center of each b_j or the expected value of Q over the image of b_j in $\mathcal{D}$, and setting $P(b_j) = P(A_i)\mu_{\mathbf{\Lambda}}(b_j)/\mu_{\mathbf{\Lambda}}(A_i)$ for $b_j \subset A_i$. This results in the simple function approximation

$$\rho_{\mathbf{\Lambda}} \approx \rho_{\mathbf{\Lambda},N} := \sum_{j=1}^{N} P(b_j) \mathbf{1}_{\mathbf{\Lambda}}(\lambda). \tag{4.6}$$

For a fixed M, by construction, $\rho_{\mathbf{\Lambda},N} \to \rho_{\mathbf{\Lambda}}$ pointwise and there exists a function $h \in L_1(\mathbf{\Lambda})$ such that $\rho_{\mathbf{\Lambda},N} \le h$ for all N.

To prove convergence, we apply the Lebesgue dominated convergence theorem to a sequence of such approximations $\rho_{\mathcal{D},M}$ and $\rho_{\mathbf{\Lambda},N}$ for increasing M and N. ■

We summarize this stage of the approximation in Algorithm 1. To simplify notation, we let V denote the *volume matrix* with ij-component set equal to the volume of $A_i \cap b_j$. We can use V to compute ratios of volumes in a simple way. Specifically, the ratio of volume of b_j in A_i is given by $V_{ij}/\sum_j V_{ij}$, and the probability of b_j is approximated by $\sum_i p_i \big(V_{ij}/\sum_j V_{ij}\big)$.

Once the probabilities of cells $\{b_j\}$ have been approximated in Algorithm 1, we may estimate $P(A)$ for arbitrary event $A \in \mathcal{B}_{\mathbf{\Lambda}}$ in various ways, e.g., using one of the following:

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

Algorithm 1. Approximation to inverse density.

Generate approximating sets $\{I_i\}_{i=1}^M$ and $\{b_j\}_{j=1}^N$.
Fix and normalize the simple function approximation $\rho_{\mathcal{D},M} = \sum_{i=1}^M p_i \mathbf{1}_{I_i}(q)$.
Let $\{A_i\}_{i=1}^M \subset \Lambda$ denote the induced regions of generalized contours partitioning $\mathbf{\Lambda}$.
for $j = 1, \ldots, N$ **do**
  for $i = 1, \ldots, M$ **do**
    Compute $\mu_\Lambda(A_i \cap b_j)$ and store as ij-component in matrix V.
  end for
end for
for $j = 1, \ldots, N$ **do**
  Set $P(b_j)$ to $\sum_{i=1}^M p_i (V_{ij} / \sum_{j=1}^N V_{ij})$.
end for

Note that in Algorithm 1 it is straightforward to handle the case of $m = n$ GD quantities of interest due to the use of simple function approximations. In this case, the "volume measure" on the generalized contours is a point mass distribution, and the volume measure associated with the transverse parameterization/measurable data is simply μ_Λ. This implies that for a given I_i, the corresponding A_i defines a set of positive volume measure with respect to μ_Λ that consists of points in $\mathbf{\Lambda}$ defining the generalized contours.

Example 6. *We consider the nonlinear vector-valued map from Example 3. We assume that there is a nominal true value of* $Q(\lambda) = q = (0.75, 0.80)$*, while measured values are drawn from a truncated normal distribution* $\tilde{N}(\mu_Q, \Sigma_Q)$ *on* $\mathcal{D}$ *with*

$$\mu_Q = \begin{pmatrix} 0.75 \\ 0.80 \end{pmatrix} \quad \textit{and} \quad \Sigma_Q = \begin{pmatrix} 0.025 & 0 \\ 0 & 0.01 \end{pmatrix},$$

where $\mathcal{D}$ *is shown in Figure 5(left). The value of* λ_3 *is uniquely determined by the vector-valued map for any fixed output, whereas the values of* λ_1 *and* λ_2 *lie on the generalized contour consisting of the circle of fixed radius in the horizontal plane of the known* λ_3*-value. We apply Algorithm 1 using a* 20×20 *uniform partition* $\{I_i\}$ *of* $\mathcal{D}$ *and* 50^3 *uniformly sized cells* $\{b_j\}$ *to partition* $\mathbf{\Lambda}$*. We show the two-dimensional marginals of* $\rho_{\mathbf{\Lambda}}(\lambda)$ *in Figure 10. In Figure 11, we plot the two-dimensional marginals of* $\rho_{\mathbf{\Lambda}}(\lambda)$ *computed using observations on only one of the components to the vector-valued quantity of interest. As the dimension of the generalized contours decreases, corresponding to an increase in the number of quantities of interest, a smaller portion of the volume receives a larger portion of the probability; compare the colorbars in Figures 10 and 11.*

We conclude by describing the alterations to Algorithm 1 required to treat a different density in the ansatz. We assume that a family of probabilities $\{P_\ell\}$ is given on the generalized contours corresponding to points in $\mathcal{L}$, as described in Theorem 4.5. We define a simple function approximation to the conditional densities along the regions of generalized contours by

$$\sum_{i=1}^M \sum_{k=1}^{K_i} u_{ik} \mathbf{1}_{\mathcal{L}_{A_i \cap \varepsilon_k}}(\lambda). \tag{4.7}$$

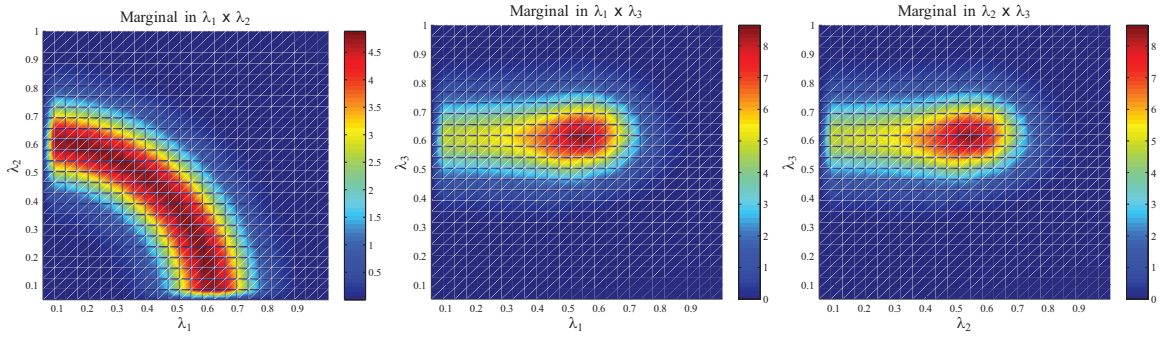

Figure 10. *Plots of marginal densities of $\rho_{\mathbf{\Lambda}}(\lambda)$ in the $\lambda_1\lambda_2$ (left), $\lambda_1\lambda_3$ (middle), and $\lambda_2\lambda_3$ (right) domains.*

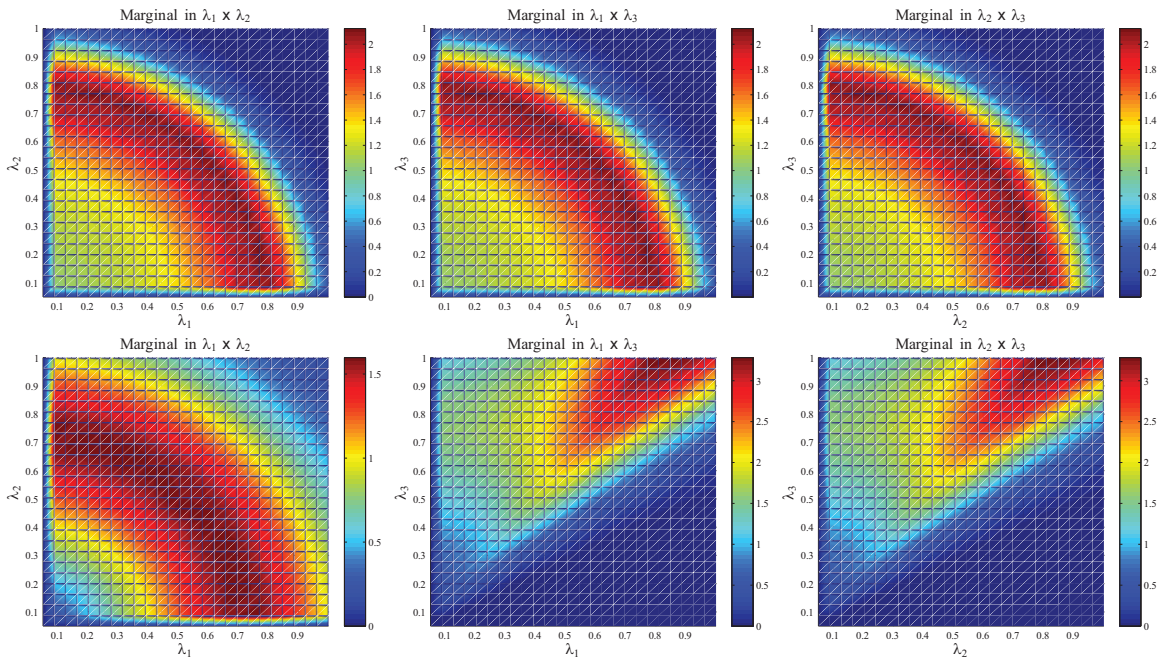

Figure 11. *Top: Plots of marginal densities of $\rho_{\mathbf{\Lambda}}(\lambda)$ in the $\lambda_1\lambda_2$ (left), $\lambda_1\lambda_3$ (middle), and $\lambda_2\lambda_3$ (right) domains when inverting the marginal density on $Q_1(\lambda) = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$. Bottom: Plots of marginal densities of $\rho_{\mathbf{\Lambda}}(\lambda)$ in the $\lambda_1\lambda_2$ (left), $\lambda_1\lambda_3$ (middle), and $\lambda_2\lambda_3$ (right) domains when inverting the marginal density on $Q_2(\lambda) = \arccos(\lambda_3/\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2})$.*

Here, $\{\mathcal{E}_k\}_{k=1}^{K_i}$ denotes a partition of a region of induced generalized contours A_i such that for a fixed A_i the function $\sum_{k=1}^{K_i} u_{ik}\mathbf{1}_{\mathcal{L}_{A_i}\cap\mathcal{E}_k}(\lambda)$ defines the simple function approximation to P_ℓ for $\ell \in \mathcal{E}_{A_i}$. In other words, the conditional densities defined by P_ℓ are used to weight the volumes of $A_i \cap \mathcal{E}_k$. Now, the ansatz will substitute for $\mu_{\mathbf{\Lambda}}$ the approximation of (4.7) in the computation of a simple function using the partition $\{b_j\}_{j=1}^{N}$. The ratios of volume are computed with respect to the partition defined by $\cup_i \{\mathcal{E}_k\}_{k=1}^{K_i}$ and are weighted by u_{ik}. Use of (4.7) in defining $P(b_j)$ leads to the proportion of the conditionally weighted volume in $\mathcal{E}_k$

that comes from b_j given by

$$\sum_{k=1}^{K_i} u_{ik} \left(\frac{V_{jk}^{(i)}}{\sum_{j=1}^{N} V_{jk}^{(i)}} \right). \tag{4.8}$$

In other words, for a fixed i, (4.8) defines a new approximation to the conditional densities defined on the region of induced generalized contours denoted A_i with respect to the cells $\{b_j\}$ used to approximate A_i.

The final approximation and convergence follow as above. We summarize the approximation to the unique density $\rho_{\mathbf{\Lambda}}$ in Algorithm 2 below.

Algorithm 2. Approximation to exact parameter probability distribution.

Generate approximating sets $\{I_i\}_{i=1}^{M}$ and $\{b_j\}_{j=1}^{N}$.
Fix and normalize the simple function approximation $\rho_{\mathcal{D},M} = \sum_{i=1}^{M} p_i \mathbf{1}_{I_i}(q)$.
Let $\{A_i\}_{i=1}^{M} \subset \Lambda$ denote the induced region of generalized contours partitioning Λ.
Fix the simple function approximations $\sum_{k=1}^{K_i} u_{ik} \mathbf{1}_{A_i \cap \mathcal{E}_k}(\lambda)$ to conditional densities defined by P_ℓ on each A_i.
for $j = 1, \ldots, N$ **do**
  for $i = 1, \ldots, M$ **do**
    for $k = 1, \ldots, K_i$ **do**
      Compute $\mu_\Lambda(b_j \cap \mathcal{E}_k)$ and store in the jk component of matrix $V^{(i)}$.
    end for
  end for
end for
for $j = 1, \ldots, N$ **do**
  Set $P(b_j)$ to $\sum_{i=1}^{M} p_i \left[\sum_{k=1}^{K_i} u_{ik} \left(V_{jk}^{(i)} / \sum_{j=1}^{N} V_{jk}^{(i)} \right) \right]$.
end for

4.3. Relation to other inverse problems and solution methods.

We briefly comment on the relation between the stochastic inverse problem we solve and two other inverse problems and solution methods.

A statistical inverse problem for an unknown process. A classic statistics inverse problem deals with the situation in which the underlying process is unknown except for some experimental data. The experimental data that can be gathered is limited: it is described in terms of finite collections of measurements on specified quantities of interest at a specified group of measurement sites. The underlying process often includes natural stochastic variation, while the experimental observations are subject to errors that behave stochastically, which means that repeated experiments generally yield different data. The goal of statistical analysis is to create and analyze a statistical model that can be used to make inferences about process behavior that is unobserved, e.g., interpolating at locations other than the measurement sites. The fact that having observations at only a fixed set of sites limits the range of scales in behavior that can be described, and the desire to avoid overfitting to a particular data set, both

provide motivation for imposing smoothness or regularization assumptions in the construction of a statistical model and in solving the inverse problem for parameters in a statistical model.

The situation is significantly different in the case that the underlying process is determined by a known deterministic model, which can be queried to any desired resolution given sufficient computational resources. For one thing, uncertainty about process behavior on any scale can be eliminated by increasing the number and/or arbitrarily changing the location of measurement/computational sites. For another, repeated evaluations of the model process yield the same set of data values. The only possible source of stochastic variation in the problem comes if there is variation in the input data, yielding variation in the output data. This raises significant theoretical difficulties with the straightforward application of statistical modeling and analysis techniques.

Regularization in deterministic inverse problems. The field of deterministic inverse problems is dominated by methods based on regularization. In this approach, the model is regularized, e.g., by adding an invertible operator to the original model operator in a linear case, to obtain a new model that has a well-posed inverse in the original input domain. In the best of cases, the solution produced by the altered model converges to a representative element of a set-valued inverse of the original model in the limit of vanishing regularization.

Regularization destroys the geometry of the generalized contours, making it impossible to solve the stochastic inverse problem into $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ that we have posed. This represents a serious loss of information about the model behavior. By way of analogy, when hiking backcountry throughout a mountainous region, a full contour map of elevations is required. Knowing approximate elevations along some particular trail or in a small neighborhood of a particular point is useless.

5. Numerical examples.

We present two examples illustrating aspects of the formulation and solution of the stochastic inverse problem.

Accuracy study: Elliptic equation.

The first example presents a numeric demonstration of the approximation properties of the solution algorithm. There are multiple ways to evaluate the accuracy of an approximate solution of the stochastic inverse problem. Our interest is measuring accuracy of the computed probability distribution on the parameter space, which is of primary physical importance in a number of applications. Of course, given information about the parameter input values, we may also then make predictions about the output of the model corresponding to that input information. In this case, we evaluate accuracy by considering the predicted behavior. However, the second approach is relatively weak given that existence of set-valued solutions.

So, we design a test in which we attempt to "recover" a given distribution on $\mathbf{\Lambda}$ that produces a distribution on $\mathcal{D}$. The model consists of an elliptic differential equation,

$$\begin{cases} -((x^2 e^{-\lambda_1 x} + 0.05)u')' + \lambda_2 u' = (1-x)\tanh(4(x-\lambda_3)) + \sin(5\pi\lambda_4 x), & 0 < x < 1, \\ u(0) = u(1) = 0, \end{cases} \tag{5.1}$$

where λ_1 affects the diffusion, λ_2 affects convection, and λ_3 and λ_4 determine properties of the forcing data. We assume that $\mathbf{\Lambda} := [1,5] \times [0.1, 0.3] \times [0,1] \times [0,2]$. We numerically solve (5.1) with a second order finite element method with 51 uniform elements partitioning $[0,1]$

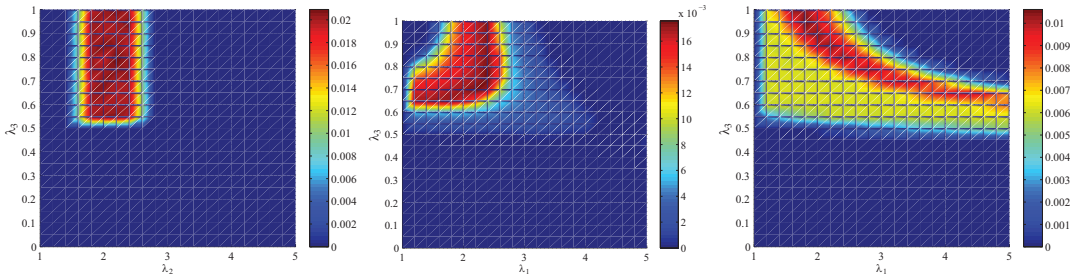

Figure 12. *Left: Exact probability of each cell b_i from a 20×20 grid of cells. Middle and right: The approximate probabilities computed when using only the first Fourier coefficient and spatial average of $u(x;\lambda)$, respectively.*

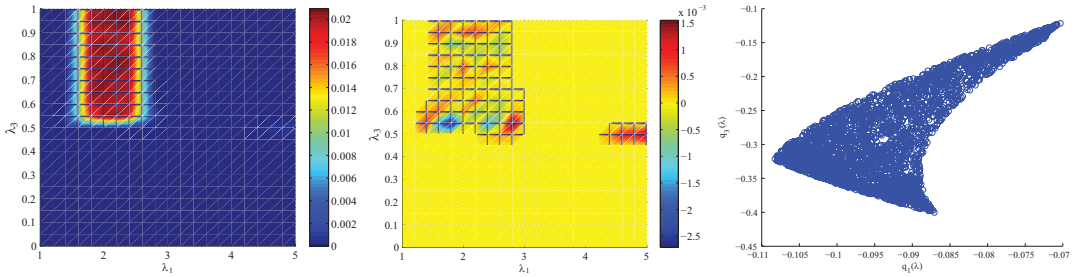

Figure 13. *Left: Approximate probability of each b_i from the 20×20 grid of cells computed using approximations to generalized contours from the 2×1 vector-valued map $Q(\lambda)$. Middle: Error in the left-hand plot of probabilities from exact probabilities of these cells. Observe the order of magnitude change in the colorbar of the right plot from the other two. Right: A scatterplot of 3000 i.i.d. samples from $\rho_{\mathcal{D}}$ defining associated generalized contours used in computations of probabilities here and in the middle and right plots of Figure 12. The propagation of uniform densities leads to a complex distribution.*

employing lumped mass quadrature, which provides sufficient accuracy for all solutions so as to avoid biasing results by numerical error.

We consider four quantities of interest: $Q_1(\lambda)$ and $Q_2(\lambda)$ denote the first and second Fourier coefficients of $u(x;\lambda)$, respectively; $Q_3(\lambda)$ denotes the average spatial value of $u(x;\lambda)$; and $Q_4(\lambda) = u(0.2;\lambda)$. In all cases, we choose a uniform density in the input parameters to propagate forward to obtain a density on the chosen quantities of interest that is then inverted back to the parameters. The resulting output distributions are quite complex; see, for example, Figure 13.

A two-parameter case. We first fix $\lambda_2 = 0.1$ and $\lambda_4 = 1$ and let $\lambda_1 \sim \mathcal{U}([1.5, 2.5])$ and $\lambda_3 \sim \mathcal{U}([0.5, 1])$. The resulting density $\rho_{\mathbf{\Lambda}}$ has small support relative to the domain $\mathbf{\Lambda}$; see Figure 12(left). As mentioned, the corresponding probability measure on $\mathcal{D}$ is quite complex; see Figure 13.

We first invert using observations on only one of $Q_1(\lambda)$ or $Q_3(\lambda)$ by applying Algorithm 1 with 100 equally spaced bins $\{I_i\}$ discretizing $\mathcal{D}$ and 20×20 equally sized cells $\{b_i\}$ to partition $\mathbf{\Lambda}$. We generate the approximate $\rho_{\mathcal{D}}$ using 10^5 samples, while we use 6400 solutions of (5.1) in the point sample-based approach [6] to approximate the inverse distribution. We see that the resulting approximations to $\rho_{\mathbf{\Lambda}}$ have supports that contain the support of the

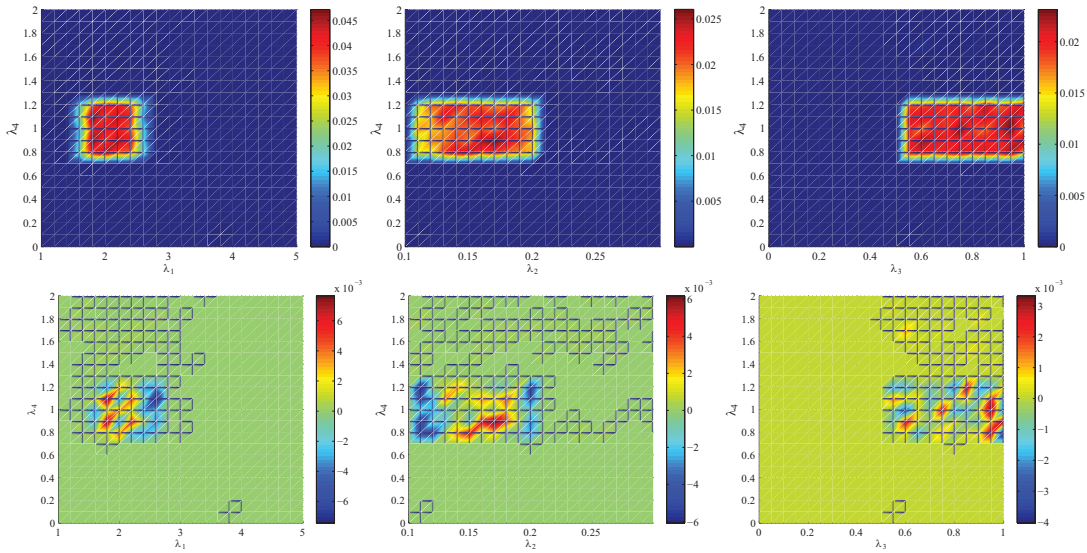

Figure 14. *Top: Approximate probabilities of* 20×20 *cells in* $\lambda_1\lambda_4$ *(left),* $\lambda_2\lambda_4$ *(middle), and* $\lambda_3\lambda_4$ *(right) subdomains computed from approximate marginal to computed inverse density. Bottom: Errors in probabilities of the top-row plots in the same planes computed by differencing with the exact inverse density.*

original density, but are larger; see Figure 12. This is an effect of the set-valued inverses in this case.

Next, we invert using observations on $Q = (Q_1, Q_3)^\top$ using the same solutions to (5.1) and cells $\{b_i\}$ partitioning $\mathbf{\Lambda}$. The resulting approximate density is close to the exact density $\rho_{\mathbf{\Lambda}}$. The error in computation of the probabilities of each of the cells $\{b_i\}$ is shown in the center plot of Figure 13.

A four-parameter case. We now let all four parameters vary stochastically with $\lambda_1 \sim \mathcal{U}([1.5, 2.5])$, $\lambda_3 \sim \mathcal{U}([0.5, 1])$, $\lambda_2 \sim \mathcal{U}([0.1, 0.2])$, and $\lambda_4 \sim \mathcal{U}([0.7, 1.2])$ to generate a probability measure on $\mathcal{D}$ by forward propagation. We use $Q = (Q_1, Q_2, Q_3, Q_4)^\top$ to invert the observed density on $\mathcal{D}$ into $\mathbf{\Lambda}$. The observed density $\rho_{\mathcal{D}}$ is discretized in the four-dimensional data space $\mathcal{D}$ in a way analogous to the two-dimensional case using Monte Carlo integration to approximate the probabilities of the four-dimensional Voronoi cells partitioning $\mathcal{D}$ implicitly defined by the pointwise values of $Q(\lambda)$ at each solution to (5.1). We use exactly the same number of solutions to (5.1) as before and bin the same number of i.i.d. samples of $\rho_{\mathcal{D}}$. To illustrate the approximate density, we plot some of the representative approximate probabilities on grids of 20×20 cells partitioning each two-dimensional configuration of $\mathbf{\Lambda}$ along with the errors in Figure 14. Comparing the error plots of Figure 14 to the error plot of Figure 13, observe that the same order of accuracy is maintained as before, despite doubling the dimension of the stochastic parameter space and maintaining the same number of solutions as in (5.1).

Extensions to higher dimensions: Storm surge and bathymetry. Mathematical models of storm surge are described by the shallow water equations (SWE) derived from the depth-averaged incompressible Navier–Stokes equations (see, e.g., [16]) and solve for water

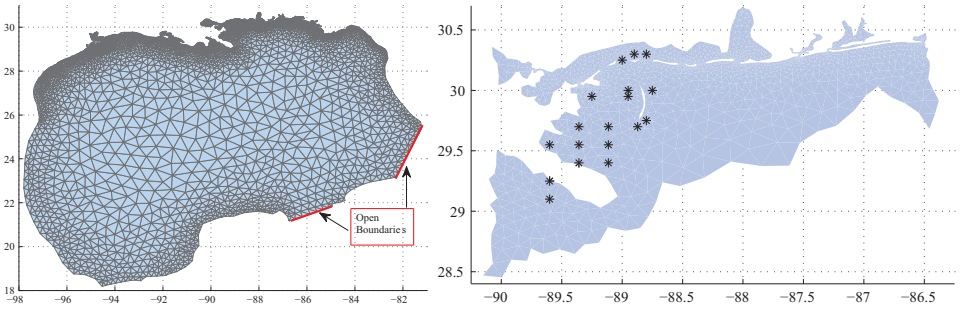

Figure 15. *Left: The mesh for the Gulf of Mexico used in all simulations, consisting of* 8,006 *nodes and* 14,269 *elements; figure modified from* [4]. *Right: The Louisiana coastline subdomain seeing the highest surge, including near-shore areas with mean bathymetry values of* 100 *m and* 18 *observation stations marked by asterisks.*

elevation (or water depth) and velocities. Common quantities of interest are maximum water elevations in the spatial domain. For storm surge applications, these models are solved on complex physical domains. These models are forced primarily by tides, winds, and waves, with wind stresses being the dominant forcing during a hurricane. We use the state-of-the-art advanced circulation (ADCIRC) model [13], which discretizes the SWE using finite element methods defined on unstructured meshes in space and finite difference schemes in time. For a description of the ADCIRC model and its application to storm surge, see [3, 9]. We use a dynamic Holland wind model [12] to generate a parametric wind field of Hurricane Katrina and tidal forcing on the open boundary of the domain (see Figure 15) for all simulations in this study. The atmospheric hurricane data (central pressure, maximum wind speed, radius of maximum winds) and the best track data are available on the NOAA archive (ftp://ftp.tpc.ncep.noaa.gov/atcf/archive/).

Accurate knowledge of the bottom topography, coastal elevation, and friction characteristics, both under water and on potentially inundated land, are needed but not always available. In this study, the uncertain model parameter is the bathymetry in a subdomain around the southeastern Louisiana coastline (see Figure 15). The given grid defines a bathymetry field as nodal attributes. We consider this given bathymetry field as the "mean field" and use a 30-term truncated Karhunen–Loève expansion (KLE) to represent the field around the logarithm of the mean [11]. The coefficients of the KLE define uncertain parameters $\lambda \in \mathbf{\Lambda} \subset \mathbb{R}^{30}$, with $\Lambda \subset \mathbb{R}^{30}$ described below. The KLE was computed using a Gaussian kernel with correlations and variances that vary smoothly with the mean bathymetric values, so there is increased variability in the perturbations near shore where there is increased model sensitivity to changes in bathymetry.

We use a single perturbation to define the target "original" bathymetry field (denoted $\tilde{\lambda}$) and generate maximum surge data at 18 simulated buoy observation stations defining $\mathcal{D} \subset \mathbb{R}^{18}$ (see Figure 15). To define $\rho_{\mathcal{D}}$ we use an additive noise model with independent Gaussian distributions and standard deviations given as functions of the recorded surge. We assume that measurement error is unbiased with a 95% confidence interval of length approximately 8 cm for 4 m of surge. Because the dimension of $\mathcal{D}$ is 18, the induced regions of generalized contours are defined by unions of 12-dimensional manifolds.

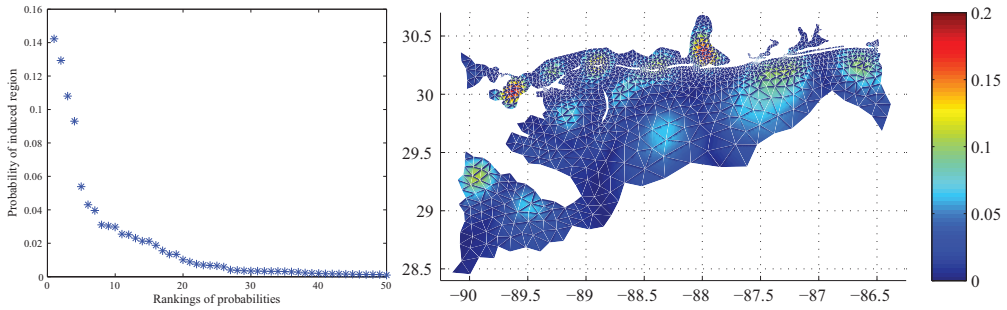

Figure 16. *Left: Probabilities of the highest* 50 *induced regions of* 12*-dimensional generalized contours. Right: The magnitude of the maximum amount of relative spatial variability of bathymetry* within *the high probability induced region containing the original bathymetry field.*

The eigenfunctions computed from the KLE are orthonormal L_2 functions on the spatial domain. This implies that any L_2 bound on the distance of the perturbations to the log of the original bathymetry imposes an l_1 bound on the coefficients λ and subsequently defines the domain $\Lambda \subset \mathbb{R}^{30}$. For any specific coefficient λ_i in the KLE, we allow $\max |\lambda_i - \tilde{\lambda}_i| = 0.75$ and $\max \|\lambda - \tilde{\lambda}\|_{l_1} = 3$. Since we use the KLE to define perturbations to the log of the original bathymetry, this means that any single perturbed bathymetry field can have pointwise values ranging from approximately $1/20$ to 20 times the original bathymetric values.

We compute a total of 5000 simulations. The first simulation is used to define the "perfect" noise-free data, and the remaining 4999 simulations correspond to 4999 uniformly sampled parameters from $\mathbf{\Lambda}$. The simulations were executed in parallel on the cluster Euclid at the University of Texas at Austin, with each simulation using eight processors. The total wall time for the experiment including reading/writing all output/input files in MATLAB and submitting jobs to the queue was under 14 hours and 10 minutes.

We use simulated data from 500 of the 4999 simulations to define (implicitly) a partitioning of $\mathcal{D}$ and compute a simple function approximation to $\rho_{\mathcal{D}}$ on this partition using Monte Carlo integration. Specifically, we "bin" 10^5 i.i.d. samples of $\rho_{\mathcal{D}}$ within this partition. We follow the remaining steps of Algorithm 1 to compute the probabilities of the induced regions of generalized contours approximated by the 4999 samples of $\mathbf{\Lambda}$.

This example raises the difficult problem of how to present the solution of a stochastic inverse problem in high dimensions. Once the dimension is larger than 4 or 5, plots of marginals become relatively meaningless. It becomes necessary to present important characteristics of the inverse distribution.

For example, a key issue is determining the regions of highest probability, given the observations on the output of the model. With the partition of $\rho_{\mathcal{D}}$ as defined above, there are four induced regions of generalized contours with probabilities ranging from just under 10% to approximately 14%, and these are significantly larger than the probabilities of all other induced regions (see Figure 16(left)). The induced region of generalized contours with the third largest probability (of approximately 10.8%) contains the original nominal "true" bathymetry field.

As one test of accuracy, we note that various refinements/coarsenings of the approximation to $\rho_{\mathcal{D}}$ (i.e., using more or fewer samples to discretize $\mathcal{D}$) lead to different values of the

probability of the induced region of generalized contours that contains the original nominal "true" bathymetry field, but the ranking of the approximate contour event that contains this value is consistently in the three highest-probability contour events.

Another significant issue for solutions of stochastic inverse problems is the degree of variability of physical conditions within a given contour event and between two contour events of significantly different probabilities. For example, if there is a low degree of variability within a given contour event, it is reasonable to argue that the choice of a particular representative element to represent points in the event is immaterial. To quantify the amount of variability within or between induced regions of generalized contours, we show plots of the magnitude of spatial variability of the corresponding perturbed bathymetry fields relative to the original bathymetry corresponding to the variability of points in $\Lambda \subset \mathbb{R}^{30}$. In Figure 16(right) we show the magnitude of the *maximum* amount of relative spatial variability *within* the induced region of generalized contours containing the original value. This variability was determined by maximizing the distance of points within the induced region from the original value. We observe for the bathymetry field defined by this point within the induced region that over the majority of the physical domain there is less than a 10% difference in bathymetric values relative to the original values. There exist two relatively small and isolated areas along the northern part of the domain with bathymetries differing between 15 and 20% relative to the original values. These two areas are close to the coastline and located roughly 40 km away from the northernmost elevation stations used in the simulation. This suggests that the addition of GD quantities of interest in these areas can improve results further, and this will be the topic of future study.

In the top two plots of Figure 17, we show the magnitudes of the relative spatial variability *between* the region containing the original value and the nearby highest probability induced regions of generalized contours. In the bottom two plots of Figure 17, we show the magnitudes of the relative spatial variability between the region containing the original value and low probability induced regions of generalized contours. We observe that there are significant increases in spatial variability relative to the original value in low probability regions compared to high probability regions. Generally, samples from the most probable regions of generalized contours provide estimates closer to the original bathymetric field compared to samples from less probable regions of generalized contours.

Appendix A. The disintegration theorem. A development of the disintegration theorem for probability measures is given in [8]. We provide a summary closely following the presentation in [7], but updated for our notation. Below, Q is a locally differentiable map with GD components between measurable spaces $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ and $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, Ψ is a σ-finite measure on $\mathcal{B}_{\mathbf{\Lambda}}$, and $\mu_{\mathcal{D}}$ is a σ-finite measure on $\mathcal{B}_{\mathcal{D}}$. A disintegration is defined in [7] as the following.

Definition A.1. *The measure Ψ on the measurable space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ has a disintegration $\{\Psi_q\}$ with respect to Q and $\mu_{\mathcal{D}}$, or a $(Q, \mu_{\mathcal{D}})$-disintegration, if*

(i) *Ψ_q is a σ-finite measure on $\mathcal{B}_{\mathbf{\Lambda}}$ concentrated on $\{Q = q\}$, that is, $\Psi_q\{Q \neq q\} = 0$, for $\mu_{\mathcal{D}}$-almost all q,*

and, for each nonnegative measurable f on $\mathbf{\Lambda}$,

(ii) *$q \mapsto \Psi_q f$ is measurable;*

(iii) *$\Psi f = \mu_{\mathcal{D}}^q(\Psi_q f)$.*

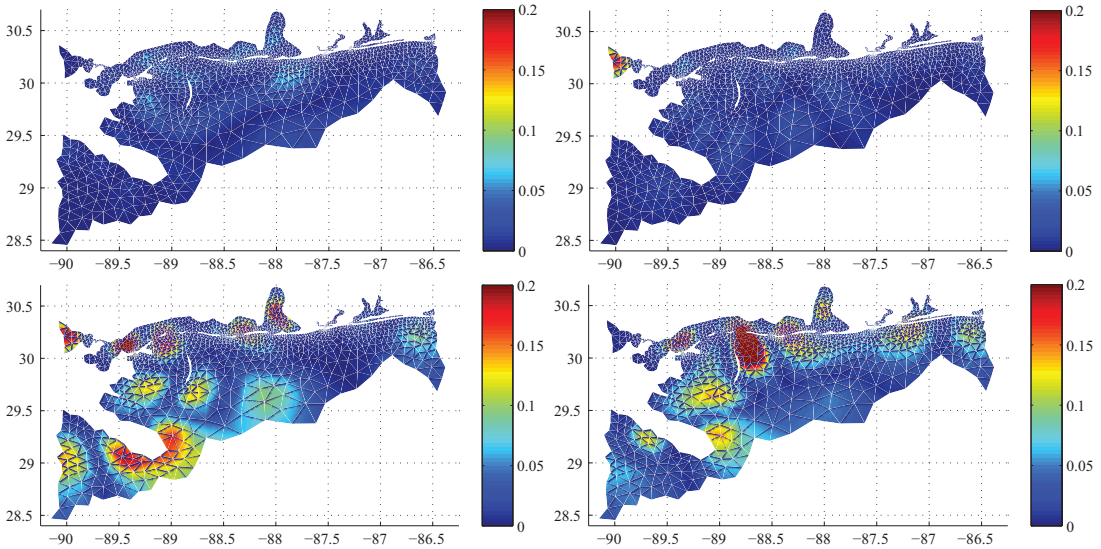

Figure 17. *The top two plots show the magnitude of the relative spatial variability of bathymetry* between *the nearby highest probability induced regions of generalized contours and the region containing the original bathymetry field. The bottom two plots show the magnitude of the relative spatial variability of bathymetry* between *induced regions of generalized contours with low probability and the region containing the original bathymetry field.*

We call $\{\Psi_q\}$ *the disintegrating measures.*

From [7], we have the following existence theorem.

Theorem A.1. *Let* Ψ *be a* σ*-finite Radon measure on a metric space* $\mathbf{\Lambda}$*. If* $\mathcal{B}_{\mathcal{D}}$ *is countable generated and contains all the singleton sets* $\{q\}$*, then* Ψ *has a* $(Q, \mu_{\mathcal{D}})$*-disintegration. The* Ψ_q *measures are uniquely determined up to an almost sure equivalence.*

The disintegrating measures are only probabilities, i.e., conditional probabilities, if we can standardize the disintegrating measures. Thus, if $0 < \mu_{\mathcal{D}}(\mathcal{D}) < \infty$, then we can always make the disintegrating measures into probabilities. This is a fact exploited by the ansatz. It is also used in [7] to prove the following.

Theorem A.2. *Let* Ψ *have a* $(Q, \mu_{\mathcal{D}})$*-disintegration* $\{\Psi_q\}$*, with* σ*-finite* Ψ *and* $\mu_{\mathcal{D}}$*.*

(i) *The image measure* $Q\Psi$ *is absolutely continuous with respect to* $\mu_{\mathcal{D}}$ *with density* $\Psi_q\mathbf{\Lambda}$*.*

(ii) *The measures* $\{\Psi_q\}$ *are finite for* $\mu_{\mathcal{D}}$*-almost all* q *if and only if* $Q\Psi$ *is* σ*-finite.*

(iii) *The measures* $\{\Psi_q\}$ *are probabilities for* $\mu_{\mathcal{D}}$*-almost all* q *if and only if* $\mu_{\mathcal{D}} = Q\Psi$*.*

(iv) *If* $Q\Psi$ *is* σ*-finite, then* $(Q\Psi)\{\Psi_q\mathbf{\Lambda} = 0\} = 0$ *and* $(Q\Psi)\{\Psi_q\mathbf{\Lambda} = \infty\} = 0$*. For* $Q\Psi$*-almost all* q*, the measures*

$$\tilde{\Psi}_q(\cdot) = \frac{\Psi_q(\cdot)}{\Psi_q\mathbf{\Lambda}}\{0 < \Psi_q\mathbf{\Lambda} < \infty\}$$

are probabilities that give a Q*-disintegration of* Ψ*.*

REFERENCES

[1] H.T. Banks and K.L. Bihari, *Modelling and estimating uncertainty in parameter estimation*, Inverse Problems, 17 (2001), pp. 95ff.

[2] J. Breidt, T. Butler, and D. Estep, *A measure-theoretic computational method for inverse sensitivity problems* I: *Method and analysis*, SIAM J. Numer. Anal., 49 (2011), pp. 1836–1859.

[3] S. Bunya, J.C. Dietrich, J.J. Westerink, B.A. Ebersole, J.M. Smith, J.H. Atkinson, R. Jensen, D.T. Resio, R.A. Luettich, C. Dawson, V.J. Cardone, A.T. Cox, M.D. Powell, H.J. Westerink, and H.J. Roberts, *A high resolution coupled riverine flow, tide, wind, wind wave and storm surge model for southern Louisiana and Mississippi: Part* I—*Model development and validation*, Monthly Weather Rev., 138 (2010), pp. 345–377.

[4] T. Butler, M.U. Altaf, C. Dawson, I. Hoteit, X. Luo, and T. Mayo, *Data assimilation within the advanced circulation (ADCIRC) modeling framework for hurricane storm surge forecasting*, Monthly Weather Rev., 140 (2012), pp. 2215–2231.

[5] T. Butler, D. Estep, and J. Sandelin, *A computational measure theoretic approach to inverse sensitivity problems* II: *A posteriori error analysis*, SIAM J. Numer. Anal., 50 (2012), pp. 22–45.

[6] T. Butler, D. Estep, S. Tavener, T. Wildey C. Dawson, and L. Graham, *Solving Stochastic Inverse Problems Using Sigma-Algebras on Contour Maps*, manuscript, University of Colorado Denver, 2013.

[7] J.T. Change and D. Pollard, *Conditioning as disintegration*, Stat. Neerl., 51 (1997), pp. 287–317.

[8] C. Dellacherie and P.A. Meyer, *Probabilities and Potential*, North–Holland, Amsterdam, 1978.

[9] J.C. Dietrich, S. Bunya, J.J. Westerink, B.A. Ebersole, J.M. Smith, J.H. Atkinson, R. Jensen, D.T. Resio, R.A. Luettich, C. Dawson, V.J. Cardone, A.T. Cox, M.D. Powell, H.J. Westerink, and H.J. Roberts, *A high resolution coupled riverine flow, tide, wind, wind wave and storm surge model for southern Louisiana and Mississippi: Part* II—*Synoptic description and analysis of Hurricanes Katrina and Rita*, Monthly Weather Rev., 138 (2010), pp. 378–404.

[10] G.B. Folland, *Real Analysis Modern Techniques and Their Applications*, 2nd ed., John Wiley & Sons, New York, 2008.

[11] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, rev. ed., Dover, New York, 2012.

[12] G.J. Holland, *An analytic model of the wind and pressure profiles in hurricanes*, Monthly Weather Rev., 108 (1980), pp. 1212–1218.

[13] R.A. Luettich and J.J. Westerink, *ADCIRC: A Parallel Advanced Circulation Model for Oceanic, Coastal and Estuarine Waters*, 2005; users' manual for version 47 available at http://www.adcirc.org/document47/ADCIRC_title_page.html; code available from jason.fleming@seahorsecoastal.com.

[14] D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, John Wiley & Sons, New York, 1999.

[15] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.

[16] C.B. Vreugdenhil, *Numerical Methods for Shallow Water Flow*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.