Solving Stochastic Inverse Problems Using Sigma-Algebras on Contour Maps

Don Estep

Colorado State University

Research supported by

Defense Threat Reduction Agency (HDTRA1-09-1-0036), Department of Energy (DE-FG02-04ER25620, DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC0001724, DE-SC0005304, INL00120133, DE0000000SC9279), Dynamics Research Corporation (PO672TO001), Idaho National Laboratory (00069249, 00115474), Lawrence Livermore National Laboratory (B573139, B584647, B590495), National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1065046, DMS-1016268, DMS-FRG-1065046, DMS-1228206), National Aeronautics and Space Administration (NNG04GH63G), National Institutes of Health (#R01GM096192), Sandia Corporation (PO299784).

December 10, 2015
Partner in crime

Troy Butler, University of Colorado Denver
Collaborators

Simon Tavener, Colorado State University
Jay Breidt, Colorado State University
Clint Dawson, The University of Texas at Austin
Lindley Graham, The University of Texas at Austin
Jeff Sandelin, Geoeye, Inc.
Joannes Westerink, University of Notre Dame
Tim Wildey, Sandia National Laboratories

Special thanks to
Casey Dietrich, Craig Michoski, and Stewart Stafford at The University of Texas at Austin
Introduction
The inverse problem for model parameters
The physical model

Ingredients

- compact parameter domain $\Lambda \subset \mathbb{R}^n$
- model $M(Y, \lambda)$ with solution $Y = Y(\lambda)$ for $\lambda \in \Lambda$
- quantities of interest $Q(\lambda) = Q(Y(\lambda)) \in \mathbb{R}^d$

We assume that $Q(\lambda)$ is differentiable

The specification of $\Lambda$ is critical and should be determined by physical considerations

The range $\mathcal{D} \subset \mathbb{R}^d$ is the image of $\Lambda$
The inverse problem for parameters

Observations on the output of the map $Q$ are provided:

- This might be imposed, e.g. as part of an optimization problem
- They might be obtained by experimental observation

The goal is to infer properties of the parameters by inverting through the model
Motivating example

Purchasing a batch of milled plates
Motivating example

Response to Heat Source

- Thickness
- Alloy composition
Motivating example

Milling → Batch → Random Samples
Motivating example

Natural variation in the system and experimental/observation error affect the results

What can be determined about the plate thickness and alloy composition?
Inference for a deterministic model

Observation Space

Prediction Space

Solution Space

Observation Functionals

Predication Functionals

Inverse Problem for Parameters

Forward Problem for Predication

Physics Model

Space of Data and Parameters
Formulating the Inverse Problem
Measure theory ideally suited for the treatment of the stochastic inverse problem for a deterministic model

Measure theory is the basis for rigorous probability theory

Measure theory is constructed using a specific approach to the inverse of a map between measurable spaces
A specified domain $X$
A \( \sigma \)-algebra defining the collection of sets whose size can be measured and the operations on those sets.
A procedure for computing measure of sets in $\mathcal{F}_X$
A procedure for computing **measure** of sets in $\mathcal{F}_X$
A procedure for computing measure of sets in $F_X$
A procedure for computing measure of sets in $\mathcal{F}_X$
A procedure for computing measure of sets in $\mathcal{F}_X$
Measure theory ingredients

A procedure for computing measure of sets in $\mathcal{F}_X$
Measure theory ingredients

Approximation of sets is a key element of measure theory
Forward problem 1: Model evaluation

The simplest forward problem:

Given a value of $\lambda$, solve the model to evaluate $Q(\lambda)$
Inverse problem 1: Set-valued solutions

The simplest inverse problem:

Given a value of \( Q \), find all values of \( \lambda \in \Lambda \) with \( Q(\lambda) = Q \)

The solution is generally a set of values
Forward problem 2: Sensitivity analysis

Sensitivity analysis involves the behavior of the model on sets

Given a set $A \subset \Lambda$, evaluate the image $B = Q(A)$
Inverse problem 2: Inverse sensitivity analysis

On what spaces is the inverse map defined?

The natural domain of the inverse map is $D$

The range of the inverse map is a space $\mathcal{L}$ of equivalence classes whose points are sets in $\Lambda$

The range of the inverse map is not $\Lambda$

Properties like well-posedness are posed in $\mathcal{L}$ not in $\Lambda$

The inability to distinguish between representors in a set-valued solution is not ill-posedness
Theorem

$\Lambda$ can be decomposed into a set of smooth lower dimensional contour manifolds representing set-valued solutions

In two dimensions, the manifolds are contour curves
Describing the space of contour manifolds

The contours (set-valued solutions) can be indexed by $\mathcal{D}$

Contours on a map are labeled using the altitude

We require an index set in $\Lambda$ for the inverse problem

**Theorem**

There exists manifolds in $\Lambda$

- transverse to the contours
- intersecting each contour once and only once

We call any such curve a transverse parameterization (TP) and choose one to represent $\mathcal{L}$
Describing the space of contour manifolds

A transverse parameterization for $\mathcal{L}$
The role of a metric and measure

Sensitivity analysis involves concepts such as distances between points, convergence, and sizes of sets

We assume $\Lambda$ and $D$ are equipped with metrics

This induces Borel $\sigma$–algebras $B_\Lambda$ and $B_D$, e.g., starting with open sets

There is a natural associated “volume” measure $\mu_\Lambda$ on $(\Lambda, B_\Lambda)$

The model induces a measure $\mu_D$ on $(D, B_D)$:

$$\mu_D(A) = \mu_\Lambda(Q^{-1}(A)), \quad A \in B_D$$

Induced measure structure is a foundation of measure theory
Measuring volumes using an inverse

Measuring volumes using the model-induced measure

\[ \Lambda \rightarrow Q \rightarrow D \]
Measuring volumes using the model-induced measure
Measuring volumes using an inverse

Measuring volumes using the model-induced measure

Compute measures in \( \Lambda \)
Forward problem 3: Stochastic sensitivity analysis

A probability measure $P_\Lambda$ is given on $(\Lambda, \mathcal{B}_\Lambda)$

This induces a probability measure $P_D$ on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ via

$$P_D(A) = P_\Lambda(Q^{-1}(A)), \quad A \in \mathcal{B}_\mathcal{D}$$

The (forward) computational problem is to approximate $P_D$

The probabilities of arbitrary measurable events can be computed in both $\mathcal{B}_\Lambda$ and $\mathcal{B}_\mathcal{D}$
A probability measure $P_D$ is imposed on $(D, B_D)$

The measurable map between $\mathcal{L}$ and $D$ is 1-1 and onto

This induces a $\sigma$–algebra $B_L$ and a volume measure $\mu_L$ on $(\mathcal{L}, B_L)$

**Theorem**
A probability measure $P_D$ on $(D, B_D)$ corresponds to a unique probability measure $P_L$ on $(\mathcal{L}, B_L)$
Inverse problem 3: Stochastic sensitivity analysis

The density of the inverse probability distribution on $\mathcal{L}$ corresponding to a uniform distribution on $\mathcal{D}$
Natural but not desirable

The inverse solution in $\mathcal{L}$ requires minimal assumptions

But is this what we want?

The physically meaningful space is $\Lambda$

The ideal inferential target is a measure on events in $\Lambda$
Solving the Inverse Problem
Two questions

**Question 1**
How do we extend the solution of the inverse problem to events in $\Lambda$?

**Question 2**
How do we compute probabilities of events in $\Lambda$ in an efficient way?
Approximation of TPs and contour manifolds

Side note: We can approximate key structures

Theorem
A transverse parameterization can be approximated pointwise using a finite number of manifolds

Theorem
Contour manifolds can be approximated using piecewise linear approximations that converge pointwise

Theorem
We can approximate arbitrary events in $\mathcal{F}_C$ and events in $\mathcal{B}_\Lambda$ to any desired accuracy

We avoid explicit approximation of TP and contour manifolds when approximating the inverse probability measures
What about events in $\Lambda$?

The ideal inferential target is a measure on events in $\Lambda$.

But with respect to what $\sigma$-algebra $\mathcal{F}_\Lambda$?

**Theorem**

We can construct a $\sigma$-algebra on $\Lambda$ by combining the induced $\sigma$-algebra $\mathcal{B}_L$ with $\sigma$-algebras $\mathcal{F}_{C_\ell}$ on each contour $C_\ell$, $\ell \in \mathcal{L}$.

There are two natural choices.
The $\sigma$-algebra of contour events

Choice 1: We choose the trivial $\sigma$–algebra $\mathcal{F}_{C_\ell} = \{\emptyset, C_\ell\}$

This yields the $\sigma$–algebra $\mathcal{F}_\Lambda$ of “contour” events in $\Lambda$

Events in $\mathcal{B}_\Lambda$ in the same equivalence class have the same probability
A decomposition of the original $\sigma$-algebra

Choice 2: We construct $\mathcal{B}_{C_\ell}$ by restricting events in $\mathcal{B}_\Lambda$ to $C_\ell$

Theorem
Combining this choice with $\mathcal{B}_L$ yields the original $\sigma-$algebra $\mathcal{B}_\Lambda$!
Structure of measures on $\Lambda$

We “disintegrate” measures with respect to a TP and contours

Let $\pi_{\mathcal{L}}(\lambda)$ map $\lambda \in \Lambda$ to $\ell \in \mathcal{L}$ in the same contour
Let $\pi_{\mathcal{L}}(B)$ map $B \in \mathcal{B}_\Lambda$ to the corresponding event in $\mathcal{B}_{\mathcal{L}}$

Disintegration Theorem
Given measures $\mu_{\Lambda}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and $\mu_{\mathcal{L}}$ on $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$, there is a set of measures $\{\mu_\ell\}$ on $(\Lambda, \mathcal{B}_\Lambda)$ such that

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

giving the disintegration,

$$
\mu_{\Lambda}(A) = \int_{\pi_{\mathcal{L}}(A)} \left( \int_{\pi_{\mathcal{L}}^{-1}(\ell) \cap A} d\mu_\ell(\lambda) \right) d\mu_{\mathcal{L}}(\ell), \quad A \in \mathcal{B}_\Lambda
$$
Structure of measures on $\Lambda$

$\mu_\mathcal{L}$ determined on $(\mathcal{L}, \mathcal{B}_\mathcal{L})$

$\mu_\Lambda$ determined on $(\Lambda, \mathcal{B}_\Lambda)$

Specify $\mu_\ell$ here

$\pi^{-1}_\mathcal{L}(\ell) \cap \Lambda$
Any probability measure on $\Lambda$ disintegrates in terms of a marginal probability $P_L$ on a TP representing $\mathcal{L}$ and conditional probabilities $\{P_\ell\}$ on contours $\{C_\ell\}$ for $\ell \in \mathcal{L}$

$P_\ell$ is the conditional probability for the event $\{Q(\lambda) = Q(\ell)\}$

Recall that the probability $P_L$ is induced by the model after the specification of $P_D$

Theorem
Specifying $P_\ell$ on generalized contours corresponding to $\ell \in \mathcal{L}$ determines a unique probability measure on $(\Lambda, B_\Lambda)$
An Ansatz

**Ansatz:** We assume that a probability measure $P_\ell(\cdot)$ is given on $\pi_\ell^{-1}(\ell)$ for each $\ell \in \mathcal{L}$.

**Theorem**
Under the Ansatz, the stochastic inverse problem has a unique solution on $(\Lambda, \mathcal{B}_\Lambda)$

**Standard Choice for Ansatz:**

$$P_\ell \sim \text{Unif}(\pi_\ell^{-1}(\ell)), \quad \ell \in \mathcal{L}$$
The standard Ansatz

\[ \pi_{\mathcal{L}}^{-1}(\ell) \]

Uniform Density Along Generalized Contours
Solution of the example under the Ansatz

Inverse density on \( \mathcal{L} \)

Inverse density on \( \Lambda \)
Justifications for the standard Ansatz

- Modeling probabilities in the absence of information (Principle of Insufficient Reason)
- It is invariant with respect to choice of TP
- Other measures imply additional geometric structure on the problem, e.g. a Gaussian requires a mean point
- The resulting probability measure inherits key properties of the underlying geometric measure

The approximation schemes can treat any measure in the Ansatz

If the Ansatz is unacceptable, the approximations work with contour events
Theorem

$P_\Lambda$ can be approximated using simple functions $P_L$ and $P_\ell$ are written as densities with respect to the disintegrated volume measures $\mu_L$ and $\mu_\ell$

$P_\Lambda$ is approximated on a partition of $\Lambda$ taken as a subset of the generating sets to $\mathcal{B}_\Lambda$, $\mathcal{B}_L$, and $\{\mathcal{B}_{C_\ell}, \ell \in \mathcal{L}\}$

This results in a direct discretization of the iterated integral in the Disintegration Theorem
Approximations of events in three $\sigma$-algebras

We can approximate events in $\mathcal{B}_\Lambda$, $\mathcal{B}_L$, and $\mathcal{B}_{C_\ell}$ simultaneously.

\[ \pi^{-1}(x_L) \cap A \]

\[ \pi(A) \]

\[ \pi^{-1}(x_L) \]
We have two approaches to approximating events

In both, we use finite collections of a generating set for the $\sigma$–algebras

- Deterministic: use collections of generalized rectangles
- Stochastic: use Voronoi tessalations corresponding to point process samples
Approximations of events
Analysis

In addition to the approximations described above, we have numerical errors in the sample and finite samples to describe measures.

A priori convergence analysis
Complete convergence analysis of the computed probability measure accounting for all sources of discretization error, including stochastic and deterministic (numerical) errors.

A posteriori error analysis
Computable accurate error estimates for the computed probability measure quantifying the effects of all sources of stochastic and deterministic errors.
A model of a chemical reaction:

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

We set \( \lambda = (A, B, IC)^\top \in \Lambda := [0.7, 1.5] \times [2.75, 3.25] \times [1, 2], \)
\( IC = y_{1,0} \)

We set \( Q(\lambda) = \frac{1}{T} \int_0^T (y_1 + y_2) \, dt \)
Example: The Brusselator model

We fix $\lambda_3 = 1.65$ and invert into two parameters

We assume $Q(\lambda)$ is a Beta distribution on $[3.7, 4.0]$

We approximate the solutions to $T = 5$ using a first order method with $\Delta t = 0.1$
Example: The Brusselator model

Using boxes (left) and random Voronoi cells (right)
Example: The Brusselator model

More Voronoi cells (left) then refining $\rho_D$ approximation (right)
Example: Brusselator model

We invert into all three parameters

We assume $Q$ is $N(3.8497, 0.0531)$ on $D := [2.9416, 4.0851]$

We approximate the output distribution using $10^6$ samples in 100 bins

We approximate the inverse using a $15 \times 15 \times 10$ uniform grid

We approximate the solutions to $T = 5$ using a first order method with $\Delta t = 0.1$
Example: Brusselator model

Plots of $P_A(b_j)$ for cells $b_j$

Top row: $IC$ fixed at 1.05, 1.45, 1.65, 1.95
Middle row: $B$ fixed at 2.7667, 2.9, 3.0667, 3.233
Bottom row: $A$ fixed at 0.72667, 0.94, 1.2067, 1.4733
Example: Brusselator model

Plot of region of highest probability

Plot of $P_\Lambda(b_j)$ for cells $b_j$ with $P(b_j) > 0.0001$
Accuracy Study Using an Elliptic Problem
Accuracy study

We “recover” an imposed probability distribution on $\Lambda$

$$
\begin{cases}
-((x^2 e^{-\lambda_1 x} + 0.05)u_x)_x + \lambda_2 u_x \\
= (1-x) \tanh(4(x - \lambda_3)) + \sin(5\pi \lambda_4 x), \quad 0 < x < 1,
\end{cases}
$$

$u(0) = u(1) = 0,$

$\Lambda := [1, 5] \times [0.1, 0.3] \times [0, 1] \times [0, 2]$

$Q_1(\lambda)$ and $Q_2(\lambda)$ are the first and second Fourier coefficients

$Q_3(\lambda)$ is the average spatial value, $Q_4(\lambda)$ is the value at $x = .2$

Discretization

100 equal bins on $\mathcal{D}$, $10^5$ samples to approximate $P_D$

6400 samples to approximate events in $\mathcal{F}_C$ and $B_\Lambda$

Finite element method with 51 elements
We propagate a uniform distribution on $\Lambda$ through the model to obtain $P_D$
Computed measure and error

Slices of the computed inverse probability measure

Slices of the error in the computed inverse measure
High Dimensional Problem: 
Katrina Storm Surge
The inundation of New Orleans from Katrina

Most deaths and property damage from a hurricane result from the storm surge

http://www.nasa.gov/vision/earth/lookingatearth/h2005_katrina.html
Modeling the storm surge

The surge is modeled by the shallow water equations giving water elevation and velocities.

The model depends on 16 parameter fields, including bottom stress, bathymetric depth, and surface stress.

**Advanced Circulation (ADCIRC) Modeling Framework**

- Developed by J. Westerink, R. Luettich, R. Kolar, C. Dawson, et. al.
- Modeling framework for 2D and 3D SWE
- Parallelized for distributed memory, multi-core computers
- Extensive validation for numerous hurricanes
Discretizing: SL15: Bathymetry and Topography
Bathymetry and bottom friction are critical to accurate forecasting of storm surge

We use the Louisiana coastline with hurricane Katrina winds as a particular case study

\[ \mathcal{D} \subset \mathbb{R}^{18} \] defined by observation network of 18 stations recording (noisy values of) maximum surge

\[ \Lambda \subset \mathbb{R}^{30} \] defined by truncated KL expansion of perturbations to near-shore LA bathymetries

The TP is 18 dimensional and the contours are 12 dimensional
The LA coastline seeing the highest maximum surge including near-shore areas with bathymetry values up to 400 m and 18 observation stations marked by asterisks.
Consider variations in a neighborhood around a fixed reference value

One model solution is used to define a reference value

500 samples used to discretize $P_D(Q)$ in $D$

4999 samples to approximate events in $F_C$ and $B_\Lambda$

Computed on Euclid - a 23 node compute cluster with 184 cores at UT Austin

The total wall clock time is about 14 hours and 10 minutes for the inversion
The reference perturbation to log of specified grid bathymetries generating noisy data
Using $P_{\Lambda}$ in high dimensions

Low dimensional projections and marginals are of limited use

We can identify specified induced regions of generalized contours and investigate their structure

Conclusions:

- High probability regions give similar physical conditions compared to the reference parameter
- Low probability regions can give very different physical conditions compared to the reference parameter
- In high probability regions, the locations where conditions differ correspond primarily to hydraulically isolated areas where the observation network is insensitive

Understanding the structure of these regions can aid in the design of optimal observation networks
Maximum variability inside generalized contour

We measure variability inside contour regions by presenting contour plots of differences of fields at different parameter values.

The maximum spatial variability of bathymetries relative to the reference bathymetry inside the region of contours containing the reference bathymetry

Probability of this region is approximately 10.8%
Variability of other highly probable regions

Typical spatial variability of bathymetries relative to the reference bathymetry inside high probability regions
Variability of low probability regions

Typical spatial variability of bathymetries relative to the reference bathymetry inside low probability regions
The condition of the stochastic inverse problem
Solvability of the stochastic inverse problem

How do relations between the quantities of interest affect the solution of the stochastic inverse problem?

**Geometrically distinct QoI**

The component maps of $Q$ are geometrically distinct (GD) if the Jacobian of $Q$ is full rank at every point in $\Lambda$.

**Theorem**

If the component maps of $Q$ are GD, then the generalized contours exist as $n - d$ dimensional manifolds and a TP exists as a $d$ dimensional manifold.
Condition of the numerical solution

The geometry of an event determines the difficulty in computing accurate approximations

Events with sharply acute corners are challenging

Approximation of such events requires more samples
Determinants and parallelepipeds

\( \{v_1, v_2, \cdots, v_d\} \) is a collection of vectors in \( \mathbb{R}^d \)

\( Pa(v_1, \cdots, v_i) \) = parallelepiped defined by the vectors, \( i \leq d \)

\( \mu(Pa(v_1, \cdots, v_i)) \) = \( i \)-dimensional Lebesgue measure (volume)

Set \( V = [v_1 \cdots v_d] \)

Theorem

\[ \det (V) = \mu(Pa(v_1, \cdots, v_d)) \]
Fundamental decomposition of the determinant

**Theorem**
Given vectors \( \{v_1, v_2, \cdots, v_d\} \) in \( \mathbb{R}^d \), there exist vectors \( v_1^\perp, v_1^o \) such that

\[
v_1 = v_1^\perp + v_1^o, \quad v_1^\perp \perp v_1^o, \quad v_1^o \in \text{span} \{v_2, \cdots, v_d\},
\]

\[
det V = \mu(Pa(v_1, v_2, \cdots, v_d)) = |v_1^\perp| \times \mu(Pa(v_2, \cdots, v_d))
\]

The two regions have equal area
Skewness and the number of samples

**Skewness**

\[
Skew(V, v_i) = \frac{|v_i|}{|v_i^⊥|}, \quad Skew(V) = \max Skew(V, v_i)
\]

Large skewness means that the inverse image of a cube in \( \mathcal{D} \) is a highly skewed region in \( \mathcal{L} \)

\( \{x_j\} = \) a net of points in \( \Lambda \)

**Theorem**

The number of sample points needed to compute an accurate inverse solution is proportional to

\[
\max_j \left( Skew(J_{Q|\mathcal{L}}(x_j)) \right)^{d-1}
\]
An inverse problem for Manning’s n

Skewness also affects model predictions

Manning’s n is a parameter field that quantifies the surface roughness that causes bottom stress in coastal hydrodynamics

We invert to determine Manning’s n from wave height measurements at a choice of fixed stations

We consider an idealized inlet with a jetty, sloping bathymetry, and an open ocean boundary at one end

Manning’s n is determined from 9 potential surface types

\[ \Lambda \] is a generalized rectangle defined by physical data
An inverse problem for Manning’s n
Relative skewness of two station pairs

Images of $\Lambda$
Effective support of inverse solutions

Inverse probability densities corresponding to a uniform density in a small box centered at the image of the reference point (black dot)
Poorly conditioned measurements yield predictions of the time of inundation at the indicated points that span a range that is 126% to 794% larger than for the well conditioned measurements.
Conclusion
Applying computational measure theory to the stochastic inverse problem is proving to be a fruitful approach.

Current research includes:
- Full analysis of solution using random sampling
- Adaptive solution of stochastic inverse problems based on a posteriori error estimates
- Evaluation and design of effective experiments
- Functional assimilation in space and time
- Extension to infinite dimensional parameter domains
- Treatment of additional complex multiphysics models


