The Use of Adjoints for Error Estimation and Uncertainty Quantification

Donald Estep

University Interdisciplinary Research Scholar
Department of Statistics
Department of Mathematics
Center for Interdisciplinary Mathematics and Statistics
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), Lawrence Livermore National Laboratory (B573139, B584647, B590495), the National Aeronautics and Space Administration (NNG04GH63G), the National Institutes of Health (5R01GM096192), the National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1016268, DMS-FRG-1065046), and Idaho National Laboratory (00069249, 00115474)

6th International Conference on Automatic Differentiation, 2012
The work I am describing represents two decades of research by students, postdocs, and collaborators.

I mention in particular Simon Tavener at Colorado State University.

The collaborators include people at a number of government labs and agencies and several companies.
Uncertainty Quantification

Uncertainty quantification deals with the effects of error, uncertainty, and variation on model predictions

Sources of uncertainty include

- Unknown aspects of the physical processes
- Numerical solution errors
- Uncertainty and experimental error in data and parameters
- Natural stochastic variation
- Indefiniteness of solution of inverse problems
The Role of Derivative Information

Mathematical models typically present locally smooth behavior with respect to inputs, discretization, and perturbation.

Variation, errors, and uncertainty in the inputs and model evaluation affect the output.
The Role of Derivative Information

Mathematical models typically present locally smooth behavior with respect to inputs, discretization, and perturbation.

Derivatives of the model can provide significant information about the effects of variation and error.
The Role of Adjoint Operators

However, derivatives of an operator can be difficult to compute.

We use adjoint operators to compute sensitivity and derivatives in an efficient manner.
Functionals, Duality, and Adjoints
We focus on computing particular information from a model expressed in terms of functionals of the model solution.

A bounded linear functional $\ell$ is a continuous linear map from a vector space $X$ to the reals $\mathbb{R}$.

Presumably, it is easier to compute an accurate functional value than a model solution that is accurate everywhere.

Often, we require only a small set of functional values.

The Fourier coefficients of a function are linear functionals. Applications typically use a truncated Fourier series.
The dual space $X^*$ of a normed vector space $X$ is the vector space of continuous linear functionals on $X$.

The dual space is the collection of “reasonable” snapshots.

$X^*$ is a vector space with dual norm defined for $\ell \in X^*$ as

$$||\ell||_* = \sup_{x \in X, \|x\| = 1} |\ell(x)|$$
Motivation for the Adjoint Operator

Let $L$ be a continuous linear map between normed vector spaces $X$, $Y$ with dual spaces $X^*$, $Y^*$

The goal is to compute a linear functional value of the output

$$\ell(L(x))$$

Some important questions:

- Can we find a way to compute the functional value efficiently?
- What is the error in the functional value if approximations are involved?
- Given information of $\ell(L(x))$, what can we say about $x$?
- What happens if $x$ or $\ell(L(x))$ is stochastic?
Definition of the Adjoint Operator

The adjoint is defined “implicitly”

For each \( y^* \in Y^* \) there is an \( x^* \in X^* \) defined by

\[
x^*(x) = y^*(L(x))
\]

sample of \( x \) in \( X \) = sample of image \( L(x) \) of \( x \) in \( Y \)

The adjoint map \( L^* : Y^* \to X^* \) satisfies the bilinear identity

\[
(L^*(y^*))(x) = y^*(L(x)), \quad x \in X, \ y^* \in Y^*
\]

If \( L \) has matrix \( \mathbf{L} \) then \( L^* \) has matrix \( \mathbf{L}^\top \)
The bilinear identity does not define a unique adjoint to a nonlinear operator in general.

There are multiple ways to define a valid adjoint for a nonlinear operator.

The correct choice depends on the purpose of the analysis.

A common approach uses the adjoint of a linearization of the operator.

It is important to consider an appropriate linearization point.
In the context of Sobolev spaces, the adjoint of the differential operator is obtained by a succession of integration by parts in the weak formulation to move all derivatives onto the test function.

Boundary terms involving functions and derivatives arise during integration by parts and must be resolved.

The formal adjoint neglects the boundary terms.

It is important to consider boundary conditions in practice.

There are several useful ways to define an adjoint to a nonlinear differential equation.
Example of a Formal Adjoint Operator

For the model initial value problem

\[ L(u) = \frac{\partial u}{\partial t} - \nabla \cdot a \nabla u + b \cdot \nabla u + cu, \quad 0 < t \leq T \]

The adjoint to the model problem is given by

\[ L^*(v) = -\frac{\partial v}{\partial t} - \nabla \cdot a \nabla v - \text{div}(bv) + cv = 0, \quad T > t \geq 0 \]
The Fundamental Role of the Adjoint

There is a close connection between the stability properties of an operator and its adjoint.

The singular values of a matrix $L$ are the eigenvalues of the square, symmetric transformations $L^*L$ or $LL^*$.

The solvability of $L(y) = b$ is closely related to the solvability of $L^*(\phi) = \psi$.

In a Hilbert space, the existence of sufficiently many solutions of $L^*(\phi) = 0$ implies that $L(y) = b$ can have at most one solution.
Suppose we wish to compute a functional $\ell(x)$ of the solution $x \in \mathbb{R}^n$ of a $n \times n$ system

$$Ax = b$$

For a linear functional $\ell(\cdot) = (\cdot, \psi)$, the adjoint problem is

$$A^* \phi = \psi$$

Variational analysis yields the representation formula

$$\ell(x) = (x, \psi) = (x, A^* \phi) = (Ax, \phi) = (b, \phi)$$

We can compute the effect of variations in the data by computing one adjoint solution and taking inner products

This generalizes the method of Greens functions
A Second Computational Benefit

If \( A = A(\lambda) \) and \( b = b(\lambda) \) depend on a parameter \( \lambda \), differentiation gives

\[
A \frac{\partial x}{\partial \lambda} + A'(\lambda)x = \frac{\partial b}{\partial \lambda}
\]

where \( A' = \partial A/\partial \lambda \)

\[
\frac{\partial \ell(x)}{\partial \lambda} = \left( \frac{\partial x}{\partial \lambda}, \psi \right) = \left( \frac{\partial x}{\partial \lambda}, A^* \phi \right) = \left( A \frac{\partial x}{\partial \lambda}, \phi \right) = \left( A \frac{\partial b}{\partial \lambda} - A'(\lambda)x, \phi \right)
\]
Using Duality and Adjoints to Estimate the Error in Numerical Solutions
We solve a differential equation

\[ L(u) = f \]

for a quantity of interest given by a linear functional

\[ (u, \psi) \]

\( \psi \) determines the quantity of interest

\( (\ , \ ) \) denotes the \( L^2 \) inner product over space and time

We compute a numerical approximation

\[ U \approx u \]

The local resolution of the approximation is measured using the residual

\[ R(U) = L(U) - f \]
Abstract A *Posteriori* Error Analysis

We solve a **linear adjoint problem**

\[(DL)^* \phi = \psi\]

We obtain the **error representation**

\[(U - u, \psi) = (R(U), \mathcal{W}(\phi))\]

for a suitable **adjoint weight** \(\mathcal{W}(\phi)\)

Scaling residuals by the adjoint solution accounts for the **accumulation, cancellation, and propagation of local contributions** to the error

The stability information is specific to the quantity of interest
Targeting Error in a Quantity of Interest is Important

Different quantities of interest can have much different stability properties

Solutions of the chaotic Lorenz system and a plot of the distance between the solutions versus time
Targeting Error in a Quantity of Interest is Important

The rate that errors grow depends strongly on the information being computed.

If we take the quantity of interest to be the average distance over time, the sensitivity is much different.

We compare to an ensemble average of 100 accurate solutions computed using time step .0001 for 15 time units.

<table>
<thead>
<tr>
<th>End Time</th>
<th>Coarse Solutions</th>
<th>Fine Solutions</th>
<th>Ensemble Ave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ave</td>
<td>Var</td>
<td>Ave</td>
</tr>
<tr>
<td>20</td>
<td>27.620</td>
<td>52.011</td>
<td>27.622</td>
</tr>
<tr>
<td>80</td>
<td>26.470</td>
<td>79.461</td>
<td>26.467</td>
</tr>
<tr>
<td>320</td>
<td>26.3</td>
<td>83.7</td>
<td>26.3</td>
</tr>
</tbody>
</table>
Let $U$ be a Galerkin finite element approximation for

$$-\nabla \cdot a \nabla u + b \cdot \nabla u + cu = f \quad \text{on } \Omega$$

Formal adjoint problem:

$$-\nabla \cdot a \nabla \phi - \text{div}(b \phi) + c \phi = \psi \quad \text{on } \Omega$$

Estimate:

$$(e, \psi) = (a \nabla U, \nabla (I - \pi_h) \phi) + (b \cdot \nabla U + cU - f, (I - \pi_h) \phi)$$

$\pi_h$ is a projection into the space of $U$
An Estimate for an Elliptic Problem

\[
\begin{cases}
-\Delta u = 200\pi^2 \sin(10\pi x) \sin(10\pi y), & (x, y) \in \Omega = [0, 1] \times [0, 1], \\
u = 0, & (x, y) \in \partial\Omega
\end{cases}
\]

The solution is \( u = \sin(10\pi x) \sin(10\pi y) \)
Consider a convection-diffusion problem with strong convection

\[
\begin{align*}
-\nabla \cdot \left( (0.05 + \tanh(10(x - 5)^2 + 10(y - 1)^2)) \nabla u \right) \\
+ \begin{pmatrix} -100 \\ 0 \end{pmatrix} \cdot \nabla u &= 1, \quad (x, y) \in \Omega = [0, 10] \times [0, 2], \\
u &= 0, \quad (x, y) \in \partial\Omega
\end{align*}
\]
Effect of Stability on a Stationary Problem

Final meshes for an average error of 4% 24,000 elements versus 3500 elements

Adjoint solutions are plotted on the right
Analysis for Multiphysics Systems

There has been a breakthrough in the *a posteriori* error analysis for multiscale, multiphysics systems.

Analysis includes:

1. Defining appropriate adjoint operators and quantifying complex stability properties
2. Estimating the effects of processing information and errors passed between components
3. Estimating the effects of using a finite number of iterations in the solution
4. Devising new ways to mitigate error in a complex multiphysics simulation
Forward Stochastic Sensitivity Analysis
The forward stochastic sensitivity analysis problem

An abstract forward stochastic sensitivity analysis problem

Given

- parameter space $\Lambda$
- model $M(Y, \lambda)$ with solution $Y$ that is an implicit, smooth function of $\lambda \in \Lambda$
- linear functional $q(\lambda) = q(Y(\lambda))$
- observed probability distribution on the parameter space $\Lambda$

Determine the probability distribution on the output values $q(\lambda)$ in the output space $\mathcal{D}$
With the output distribution in hand, we can compute probabilities of events.
The Effect of Numerical Error

Monte Carlo approach:

Draw samples of $\lambda$, compute corresponding values of $q(\lambda)$, and approximate the output distribution, e.g. by binning

Monte Carlo sample solutions can exhibit a variety of behaviors

The numerical error of samples is expected to vary significantly

This can severely bias Monte Carlo results
The chaotic Lorenz problem

\[
\begin{align*}
\dot{u}_1 &= -10u_1 + 10u_2, \\
\dot{u}_2 &= \lambda u_1 - u_2 - u_1u_3, \\
\dot{u}_3 &= -\frac{8}{3}u_3 + u_1u_2, \\
0 &< t,
\end{align*}
\]

\[
\begin{align*}
u_1(0) &= -6.9742, \\
u_2(0) &= -7.008, \\
u_3(0) &= 25.1377
\end{align*}
\]

We vary \( \lambda \sim \text{Unif}[25, 31] \)

The errors using uniform time steps vary significantly

We compare 1000 point Monte-Carlo sampling with both a fixed time step computation and an adaptive computation with error smaller than \( 10^{-5} \)
Numerical Error May Affect a Computed Distribution

Distributions for fixed and adapted time steps
Note the difference in the horizontal scales!
Compute the probability distribution for quantity of interest $Q(U)$ where $U$ solves (a.s.)

$$\begin{cases} -\nabla \cdot A \nabla U = f, & x \in \Omega, \\ U = 0, & x \text{ in } \partial \Omega, \end{cases}$$

It is important to estimate the effects of the various discretization steps.
There are two approximations used for the computed distribution

- (Deterministic) We compute numerical solutions
- (Stochastic) We compute only a finite number of sample values

We estimate the error in the approximate cumulative distribution of the quantity of interest using adjoints

The analysis estimates each source of error accurately

We devise an extended adaptive algorithm that efficiently balances all sources of error
Stochastic Elliptic Problem

We use a $9 \times 9$ uniform partition of $[0, 1] \times [0, 1]$

Here $a$ is .1 in the interior “cross”, 1 elsewhere, $A$ is uniformly distributed in $\pm 10\%$ of the value of $a$

We ask for 15% error with probability 0.95.

Initial: $h = 1/18$, $\mathcal{I} = 40$, $\mathcal{P} = 1$, and $\mathcal{N} = 60$, and $\Delta \mathcal{I} = 0.3\mathcal{I}$, $\Delta \mathcal{P} = 1$
Stochastic Elliptic Problem

Error in the approximate CDFs:

The tolerance was reached after three iterations with $h = 1/54$, $I = 160$, $P = 3$, and $N = 240$.
The tolerance was reached after three iterations with $h = 1/54$, $\mathcal{I} = 160$, $\mathcal{P} = 3$, and $\mathcal{N} = 240$.
Consider an initial value problem

\[
\begin{cases}
\dot{x}(t; \lambda) = f(x(t; \lambda); \lambda_1), & t > 0, \\
x(0; \lambda) = \lambda_0
\end{cases}
\]

with \( x \in \mathbb{R}^n \) and \( f : \mathbb{R}^{n+d} \rightarrow \mathbb{R}^n \)

The random parameter \( \lambda = (\lambda_1, \lambda_0)^\top \in \mathbb{R}^{d+n} \)

Quantity of interest

\[
q(\lambda) = \int_0^T (x(s; \lambda), \psi(s)) \, ds,
\]
The adjoint problem is

$$\begin{cases} 
- \dot{\phi}(t) - (D_x f(y(t); \mu_1))^*(t) \phi(t) = \psi(t), & T \geq t \geq 0, \\
\phi(T) = 0,
\end{cases}$$

$y(t)$ denotes the deterministic solution with the reference parameter value $\mu = (\mu_1, \mu_0)^*$

**Theorem**

$$\nabla q(\mu)(\lambda - \mu) \approx (\lambda_0 - \mu_0, \phi(0)) + \int_0^T (D_{\lambda_1} f(y; \mu_1)(\lambda_1 - \mu_1), \phi) \, ds$$
We approximate the output distribution using a piecewise constant approximation

\[ \tilde{q}(\lambda) = \sum_{i=1}^{N} q(\mu_i) \chi_{R_i}(\lambda), \]

corresponding to a sample \( \{\mu_i\}_{i=1}^{N} \) and partition of generalized rectangles \( \{R_i\}_{i=1}^{N} \) with \( \mu_i \in R_i \) for all \( i \)

\( \chi_{R_i} \) is 1 for \( \lambda \in R_i \) and 0 otherwise

We have the error estimate

\[ q(\lambda) - q(\mu) \approx \nabla q(\mu)(\lambda - \mu) \]

which we evaluate using the adjoint expression
We use the error estimate to adaptively guide the sampling process.

The refinement process works iteratively:

- For a given sample, we estimate the local contributions to the error using the estimate.
- We add new sample points in regions and in the dimensions that contribute the most.

This provides automated localized dimension reduction.
A Predator Prey Example

We model a prey $u_1$ with a logistic birth/death process consumed by predator $u_2$

\[
\begin{align*}
\partial_t u_1 - \delta \Delta u_1 &= \lambda_4 u_1 (1 - \frac{u_1}{\lambda_5}) - \lambda_6 u_2 h(u_1; \lambda_2), \\
\partial_t u_2 - \delta \Delta u_2 &= \lambda_1 u_2 h(u_1; \lambda_2) - \lambda_3 u_2, \\
\partial_n u_1 &= \partial_n u_2 = 0, \\
u_1 &= u_{1,0}, u_2 = u_{2,0},
\end{align*}
\]

$\Omega \times (0, T], \partial \Omega \times (0, T], \Omega \times \{0\}$

The (Holling II) functional response $h(u) = h(u; \lambda_2)$ satisfies

- $h(0) = 0$
- $\lim_{u \to \infty} h(u) = 1$
- $h$ is strictly increasing
We assume a (truncated) normal distribution in the region

<table>
<thead>
<tr>
<th>Description</th>
<th>Name</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>encounter gain</td>
<td>$\mu_1$</td>
<td>[.5, 1.5]</td>
</tr>
<tr>
<td>response gain</td>
<td>$\mu_2$</td>
<td>[5.05, 15.15]</td>
</tr>
<tr>
<td>predator death rate</td>
<td>$\mu_3$</td>
<td>[.5, 1.5]</td>
</tr>
<tr>
<td>prey growth rate</td>
<td>$\mu_4$</td>
<td>[2.5, 7.5]</td>
</tr>
<tr>
<td>prey carrying capacity</td>
<td>$\mu_5$</td>
<td>[.5, 1.5]</td>
</tr>
<tr>
<td>encounter loss</td>
<td>$\mu_6$</td>
<td>[.5, 1.5]</td>
</tr>
</tbody>
</table>

We use the $L^1$ norm of the prey population at $t = 10$ as the quantity of interest ($\psi = \delta(t - 10)(0, 1)^T$)

We use a 12400 point Monte-Carlo computation as a reference
Evolution of the Solution

We show a few snapshots in time of the predator (left) and prey (right). Warmer colors mean higher density.
Evolution of the Solution

We show a few snapshots in time of the predator (left) and prey (right). Warmer colors mean higher density.
Evolution of the Solution

We show a few snapshots in time of the predator (left) and prey (right). Warmer colors mean higher density.
Evolution of the Solution

We show a few snapshots in time of the predator (left) and prey (right). Warmer colors mean higher density.
Evolution of the Solution

We show a few snapshots in time of the predator (left) and prey (right). Warmer colors mean higher density.
Note that each “point” requires solving the PDE.
FAPS sample points are added in directions with the largest gradient components.
Inverse Stochastic Sensitivity Analysis
An abstract probabilistic inverse problem

Given

- parameter space $\Lambda$
- model $M(Y, \lambda)$ with solution $Y$ that is an implicit, smooth function of $\lambda \in \Lambda$
- linear functional $q(\lambda) = q(Y(\lambda))$
- observed probability distribution on the output values $q(\lambda)$ in the output space $\mathcal{D}$

Determine a probability distribution on the parameter space $\Lambda$ that produces the observed distribution on $\mathcal{D}$
What is a solution of the inverse problem?

The solution is a probability distribution that can be used to compute probabilities of events in $\Lambda$.

This provides the capability of answering questions relevant to experimental design, engineering design, optimization, policy formulation, risk assessment, etc.
Issues to be addressed

The solution of the stochastic inverse problem requires resolution of several issues:

- Nonuniqueness in inverse problems
- Approximation of a (potentially complicated) probability measure
- Approximation of (potentially complicated) events in $\Lambda$ by collections of smaller events of simple construction
- Error arising from numerical evaluation of the model
- Error arising from finite sampling
Nonuniqueness and Set-valued inverses

The solution of an inverse problem is set-valued, comprised of the set of all values that map to a given output value.

Contours
Given the smoothness of the output map, the Implicit Function Theorem guarantees existence of lower dimensional manifolds in (potentially small) neighborhoods of the parameter set $\Lambda$, where points on a contour curve all map to the same output value.

In two dimensions, the manifolds are contour curves.
Set-valued inverses

$q(\mu)$
Generalized contours

Globally, we choose a collection of sets \( \{ U_{\lambda_\alpha} \times V_{\lambda_\alpha} \} \) and local representations \( \{ g_{\lambda_\alpha} \} \) corresponding to a collection of points \( \{ \bar{\lambda}_\alpha \} \) giving the value \( \bar{q} \).

This defines \((d - 1)\)-dimensional piecewise manifolds we call generalized contours.

These manifolds are the natural set-valued inverses of \( q(\lambda) \).
Generalized contours

$q(\lambda)$
How do we describe the set of set-valued inverses?

There exists 1-dimensional curves
- transverse to the contours
- intersecting each contour once and only once

We can use any transverse curve to parameterize the contours

**Theorem**

A transverse parameterization (TP) can be constructed numerically from a finite number of connected curves
Parameterizing the set of generalized contours

Interval of Values $q(\Lambda)$

Transverse Parameterization
Inversion into the index set of the generalize contours is well-posed:

**Theorem**
If \( q(\lambda) \) has distribution \( F_q(q(\lambda)) \), then the corresponding distribution of the points on the TP (intersections with the generalized contours) is unique.
Inverting into the set of set-valued solutions
Approximations of generalized contours

To exploit this, we have to approximate the generalized contours.

Locally, we use adjoints to determine local linear approximations to generalized contours.

Globally, we use a piecewise-linear tangent plane approximation to $q(\lambda)$ to compute piecewise-linear approximations of generalized contours.

**Theorem**
The generalized linear contours converge pointwise to the true contours in $\Lambda$. 
Approximations of generalized contours

The local and global approximations
Suppose

\[ q(\lambda_1, \lambda_2) = \exp[\cos(\lambda_1) + \sin(\lambda_2)] \]

We take \( \Lambda = [-2\pi - 0.1, 2\pi + 0.1] \times [-2\pi - 0.1, 2\pi + 0.1] \)

Form piecewise-linear approximation of \( q \) to form \( \hat{q} \)

We plot various contour curves and a TP on each plot for various approximations \( \hat{q} \)

The convergence is fast and in some sense monotonic.
Convergence of contours - example

7 \times 7 and 10 \times 10 grids
Convergence of contours - example

25 × 25 and 50 × 50 grids
The Brusselator model

A model of a chemical reaction:

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

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 \( q(\lambda) = (1/T) \int_0^T (y_1 + y_2) \, dt \)

We approximate the solutions to \( T = 5 \) using a first order method with \( \Delta t = 0.1 \)
Cross-sections of $q(\lambda)$

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
Approximate cross-sections of generalized contours

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
Solving the stochastic inverse problem

Recall the goal is to compute probabilities of arbitrary events $A \subset \Lambda$

At this point, we can compute probabilities of events formed by the unions of generalized contours

Distribution on the contours

We can compute the probability of event $A$ but not event $B$

We need further structure
An event $\mathcal{A}$ is defined by the contours it intersect and the amount of each contour it contains.

**Underlying geometric measure on $\Lambda$**
We assume there exists parameter volume measure $\mu_{\Lambda}$ on $\Lambda$.

This measure determines the volume of sets in the parameter space.

It should be specified as part of the model.
The **deterministic** model determines a likelihood

\[ L(q \mid \lambda) = \delta(q - q(\lambda)) \]

\( \delta \) is the unit mass distribution at 0

The **Law of Total Probability** relates the two probability distributions as

\[ \rho_D(q) = \int_{\Lambda} L(q \mid \lambda) \sigma_{\Lambda}(\lambda) \, d\mu_{\Lambda}(\lambda) \]

Given \( \rho_D \), we invert the integral equation to compute \( \sigma_{\Lambda}(\lambda) \)

Many combinations of \( \sigma_{\Lambda}(\lambda) \) and \( \mu_{\Lambda} \) can give the same output distribution
Given $\sigma_\Lambda(\lambda)$, we can compute probabilities of events $A$

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

**Theorem**

Given $A \subset \Lambda$, $P(A)$ can be approximated using a simple function representation of $\sigma_\Lambda(\lambda)$ that only requires volume computations in $\Lambda$

A constructive proof outlines a computational measure theoretic algorithm for approximating the probability of $A \subset \Lambda$
Nonlinear system: 2 parameters

Statement of problem

Solve

\[ \begin{align*}
\lambda_1 x_1^2 + x_2^2 &= 1, \\
x_1^2 - \lambda_2 x_2^2 &= 1.
\end{align*} \]

The quantity of interest \( q(\lambda) \) is the first coordinate of the solution in the first quadrant

We impose a normal distribution on \( q(\lambda) \)

We use a product Lebesgue measure for the parameter distance measure

We let \( \Lambda = [.79, .99] \times [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}] \) and use 289 uniformly spaced points to construct an approximation
Nonlinear system: 2 parameters

Plot of the map from the parameter space to the output
We use 1156 small cells

On the left: we use the volume measure to determine how "much" of the contours are contained in the specified event. On the right: We use the probability of these contours being selected to determine the probability of the event.
We approximate the solutions to $T = 5$ using a first order method with $\Delta t = 0.1$

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
Example: Brusselator model

Plots of $P(b_i)$ for cells $b_i$

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 $P(b_i)$ for cells $b_i$ with $P(b_i) > 0.0001$
Conclusion
The adjoint problem describes the stability and provides an efficient way to compute sensitivity and derivatives of a quantity of interest.

Using this information, we can:

- Compute accurate error estimates for numerical solutions of complex differential equations
- Devise efficient error control strategies
- Conduct forward sensitivity analysis using relatively few model evaluations
- Carry out inverse sensitivity analysis without accept/reject sampling, MCMC
- Analyze error in computed distributions and design efficient computational methods