DISSERTATION

COMPUTATIONAL MEASURE THEORETIC APPROACH TO INVERSE SENSITIVITY ANALYSIS: METHODS AND ANALYSIS

Submitted by
Troy Daniel Butler
Department of Mathematics

In partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Summer 2009

UMI Number: 3385166

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI
Dissertation Publishing

UMI 3385166
Copyright 2009 by ProQuest LLC.
All rights reserved. This edition of the work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

July 10, 2009

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY TROY DANIEL BUTLER ENTITLED "COMPUTATIONAL MEASURE THEORETIC APPROACH TO INVERSE SENSITIVITY ANALYSIS: METHODS AND ANALYSIS" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Dr. Jean Opsomer

Simon J. Tavener

Dr. Simon Tavener

Dr. Jiangguo Liu

Adviser: Dr. Donald Estep

Simon J. Tavener

Department Head: Dr. Simon Tavener

ABSTRACT OF DISSERTATION

COMPUTATIONAL MEASURE THEORETIC APPROACH TO INVERSE SENSITIVITY ANALYSIS: METHODS AND ANALYSIS

We consider the inverse problem of quantifying the uncertainty of inputs to a finite dimensional map, e.g. determined implicitly by solution of a nonlinear system, given specified uncertainty in a linear functional of the output of the map. The uncertainty in the output functional might be suggested by experimental error or imposed as part of a sensitivity analysis. We describe this problem probabilistically, so that the uncertainty in the quantity of interest is represented by a random variable with a known distribution, and we assume that the map from the input space to the quantity of interest is smooth. We derive an efficient method for determining the unique solution to the problem of inverting through a many-to-one map by computing set-valued inverses of the input space which combines a forward sensitivity analysis with the Implicit Function Theorem. We then derive an efficient computational measure theoretic approach to further invert into the entire input space resulting in an approximate probability measure on the input space.

We provide detailed error analysis for inverse problems involving nonlinear ordinary differential equations and semilinear elliptic partial differential equations.

Troy Daniel Butler
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523
Summer 2009

ACKNOWLEDGEMENTS

I extend thanks to my advisor, Dr. Don Estep, for his mentorship, guidance, insight, and advice throughout this research project. From him I learned many valuable techniques of mathematics. He has shaped the lens through which I view the world of mathematics. This project would neither be worded nor explained as well as it is without his tremendous help. Any errors or opaque and confusing explanations only reflect my inabilities as an author.

A special thanks is extended to Dr. Ken Klopfenstein. If it were not for him, then I would never have begun this journey. He opened my eyes to the beauty of mathematics. Also to Dr. Eugene Allgower for his guidance and friendship throughout my MS studies.

Also to my family for pushing me to learn and do well in school at a very young age. I never truly appreciated all of the structure and discipline of my parents until I started graduate school. A special thanks to Aunt Pam for always encouraging me to follow my passion.

And to my wife, Lauren; I started down the road toward a PhD and marriage at the same time. Before getting myself into trouble, I define the end of the road for the latter as actually getting married and *not* the end of a marriage. Having completed the latter, it was only with her constant "encouragement" that I finish my PhD in a timely fashion so that I may "get a real job."

This work was supported by numerous grants and fellowships. Thanks especially to the CSU Mathematics Department, and the PRIMES program (NSF IGERT Grant DGE-0221595) for their fellowships and grants. Other grants include: Department of Energy (DE-FG02-05ER25699), National Science Foundation (MSPA-CSE-0434354).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

SENSITIVITY ANALYSIS

Consider the problem of predicting the global behavior of a complex system, e.g. climate change. A solution to such a problem often involves combining observations from different sources and different scales, e.g. ranging from local field observations to satellite images, to form input data and parameters. Knowledge of which parameters have the greatest effect on the output of the model is extremely useful, e.g.

- We may want to quantify the effects of uncertainty in data and parameters as well as error in evaluation on model predictions
- We may be able to predict which kinds of experimental data are most important for producing reliable predictions

At their heart, these are sensitivity analysis problems. We define sensitivity analysis as the study of a response of a system to variations in data and parameters. We use sensitivity analysis to determine which parameters have the greatest effect on information computed from the system, and to search for parameter values producing optimal values of the output.

There are two directions in sensitivity analysis: forward sensitivity analysis and inverse sensitivity analysis. In forward sensitivity analysis, we vary the input data and parameters and observe the effect on the output of the system. For inverse sensitivity analysis, we start with uncertainty in the output of the system and seek to determine the variations in the input data and parameters that produce this uncertainty in the output. The uncertainty in the output of the system is either observed, e.g. variations in empirical data, or imposed, e.g. modeling measurement error. Often times the parameter space has high dimension while the observation space has low dimension resulting in a "many-to-one" map. In that situation, the inverse sensitivity analysis problem is an ill-posed problem, i.e. there is typically more than one solution.

There are two predominant approaches to sensitivity analysis problems: deterministic and statistical. The deterministic approach uses derivatives to determine sensitivity at a point, and requires the map from input to output to be smooth. If no information is known about the map from input to output, then the statistical approach is often preferred. Statistical approaches consider variation in the input data and parameters to reflect uncertainty or be the consequence of error, and models the input data and parameters as random variables associated with some distribution. The model output becomes a new random variable with a new distribution. We often use sampling methods to determine the new distribution. In other words, this is a density estimation problem. Both of these approaches are useful, and they can be combined [20].

1.1 Forward Sensitivity Analysis

The motivation for solving complex systems, e.g. climate change, is often to obtain a low-dimensional quantity of interest, e.g. temperature, rather than the entire output of the system, e.g. the entire state of the climate, which is often too complex to analyze in any meaningful way. We study quantities of interest obtained from linear functionals of the solution to the system.

Considering functionals is typically physically motivated. For example, in experiments [18, 22], radioactively marked tumor cells were added to the blood stream of laboratory mice and levels of cancer cells were measured over time. The proposed model has two compartments in order to account for the fact that the rate of decay does not follow a simple exponential decay model. Let x_1 and x_2 denote the number of cancer cells in the capillaries and lung tissue, respectively. The loss of cancer cells from the capillaries by being dislodged and carried away by the blood and the transfer from the capillaries to lung tissue is modelled by linear functions $-\lambda_1 x_1$ and $-\lambda_2 x_1$, respectively. The loss of cancer cells in the lung tissue is modelled by $-\lambda_3 x_2$. This leads to the system of differential equations

$$\begin{cases} \dot{x}_1 = -(\lambda_1 + \lambda_2)x_1 \\ \dot{x}_2 = \lambda_2 x_1 - \lambda_3 x_2 \end{cases} \tag{1.1.1}$$

In experiments, state variables can *not* be measured individually, but the *total amount of radioactivity*, $x_1 + x_2$ is measurable. Thus, due to physical restrictions and not model complexity, we solve the system in order to obtain the quantity of interest, $q(\lambda) = q(x(\lambda))$, represented by the linear functional $q(\lambda) = x_1 + x_2$. We use the notation $q(\lambda)$ throughout to emphasize the dependence of the quantity of interest on the choice of input parameters. We analyze theoretical sensitivities of x_1 and x_2, but only the sensitivity of the quantity of interest is experimentally verifiable.

We present two different methodologies to solve the forward sensitivity analysis problem based on the motivation for solving such a problem. We use a forward derivative analysis to obtain sensitivities of the entire output of the system to the input data and parameters. We use an adjoint derivative analysis to obtain sensitivities of a quantity of interest to the input data and parameters.

1.1.1 Forward Derivative Analysis

Given a model, we compute the partial derivatives of the solution with respect to parameters along with the solution by solving an extended model. We illustrate this with a simple linear model. Suppose the model is defined by the linear system

$$Ax = b. \tag{1.1.2}$$

Here, $x \in \mathbb{R}^n$ is a vector of states, $A \in \mathbb{R}^{m \times n}$ describes the relations between states, and $b \in \mathbb{R}^m$ is the data. Suppose $b = b(\lambda)$ and $A = A(\lambda)$ depend on parameter $\lambda \in \mathbb{R}^p$, then $x = x(\lambda)$ is implicitly a function of λ. Differentiating (1.1.2) with respect to λ yields the linear system

$$A\frac{\partial x}{\partial \lambda} + \frac{\partial A}{\partial \lambda}x = \frac{\partial b}{\partial \lambda}, \tag{1.1.3}$$

which we solve for $\partial x/\partial \lambda$. Note that the system used to solve for the sensitivity $\partial x/\partial \lambda$ depends on the solution to (1.1.2). Thus, we obtain the sensitivity by solving the extended model defined by the linear system

$$\begin{cases} Ax = b \\ A\,\frac{\partial x}{\partial \lambda} + \frac{\partial A}{\partial \lambda}x = \frac{\partial b}{\partial \lambda}. \end{cases} \tag{1.1.4}$$

We use similar steps to form extended models for nonlinear systems and differential equations. Consider the following nonlinear differential equation

$$\begin{cases} \dot{x} = f(x; \lambda_1), & 0 < t, \\ x(0) = \lambda_0 \end{cases} \tag{1.1.5}$$

To compute the sensitivities, we solve the extended system

$$\begin{cases} \dot{x} = f(x; \lambda_1), & 0 < t, \\ x(0) = \lambda_0, & \\ \frac{\partial \dot{x}}{\partial \lambda} = f'(x; \lambda)\frac{\partial x}{\partial \lambda} + \frac{\partial f}{\partial \lambda}(x; \lambda), & 0 < t, \\ \frac{\partial x}{\partial \lambda}(0) = B, & \end{cases} \tag{1.1.6}$$

where f' is the Jacobian of f with respect to x, and B is a nonzero matrix if and only if at least one of the components of λ_0 is a parameter subject to variation.

Example 1.1.1. *Recall the model for malignant tumors is*

$$\begin{cases} \dot{x}_1 = -(\lambda_1 + \lambda_2)x_1, \\ \dot{x}_2 = \lambda_2 x_1 - \lambda_3 x_2. \end{cases} \tag{1.1.7}$$

We assume the initial conditions are given and static. We calculate the sensitivities by solving the extended system

$$\begin{cases} \dot{x} = \begin{pmatrix} -(\lambda_1 + \lambda_2) & 0 \\ \lambda_2 & -\lambda_3 \end{pmatrix} x, \\ \frac{\partial \dot{x}}{\partial \lambda} = \begin{pmatrix} -(\lambda_1 + \lambda_2) & 0 \\ \lambda_2 & -\lambda_3 \end{pmatrix} \frac{\partial \mathbf{x}}{\partial \lambda} + \begin{pmatrix} -x_1 & -x_1 & 0 \\ 0 & x_1 & -x_2 \end{pmatrix}, \end{cases} \tag{1.1.8}$$

where $\lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix}^{\mathsf{T}}$. Solving the extended system numerically, we obtain the sensitivities of x_1 and x_2 with respect to all the parameters for all time. We summarize these sensitivities in figures 1.1 and 1.2.

Example 1.1.2. *Consider the classical SIR-epidemic model represented by the following nonlinear system of differential equations*

$$\begin{cases} \dot{S} = \mu K - \beta SI - \mu S, \\ \dot{I} = \beta SI - \gamma I - \mu I - \alpha I, \\ \dot{R} = \gamma I - \mu R. \end{cases} \tag{1.1.9}$$

Here, S, I, and R are the state variables representing susceptible, infected, and recovered individuals, respectively, and $K = S + I + R$ denotes the total population. The parameters are μ, β, γ, and α, where

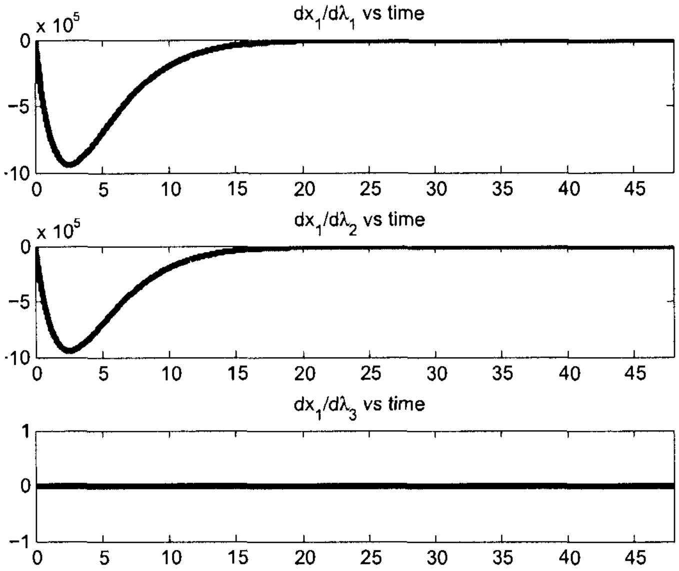

Figure 1.1: Forward derivative sensitivity analysis for x_1 in malignant tumor model. The top, middle, and buttom plots are the partial derivatives of x_1 with respect to λ_1, λ_2, and λ_3, respectively. The top two plots show that in the first couple hours, the number of cancer cells in the capillaries is most sensitive to changes in the rate of blood flow, λ_1, and the rate of transfer of the cells to the lungs, λ_2. The sensitivities decrease quickly after the first couple of hours, and by hour 15 the number of cancer cells in the capillaries are no longer significantly sensitive to changes in these parameters. The bottom plot shows that the number of cancer cells in the capillaries is never sensitive to the rate of loss of cancer cells in the lung tissue. Suppose there is a medication that promotes the rate of transfer of cells from the blood to the lungs or increases the rate of blood flow, and the medical goal is to decrease the number of cancer cells in the capillaries significantly, then the best time to adminster such a drug is approximately two and a half hours into the observation of the mouse, when the sensitivities are largest in magnitude

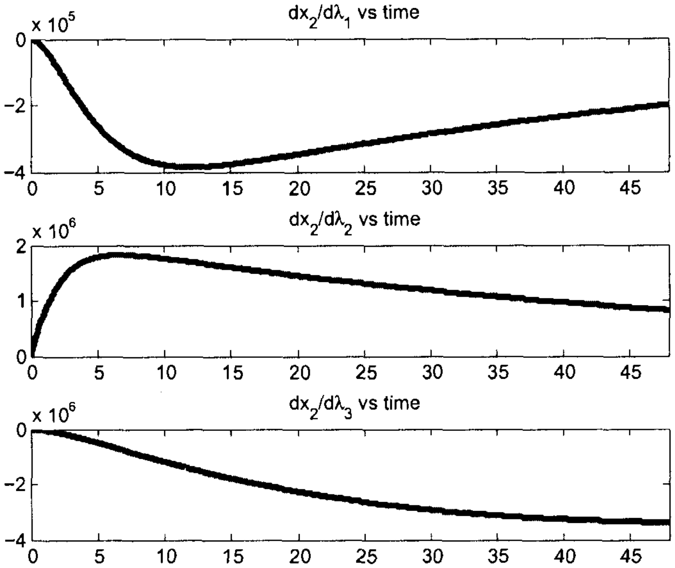

Figure 1.2: Forward derivative sensitivity analysis for x_2 in malignant tumor model. The top, middle, and buttom plots are the partial derivatives of x_2 with respect to λ_1, λ_2, and λ_3, respectively. Comparing the plots, we first note a difference of scale between the top plot and the middle and bottom plots. The sensitivites become an order of magnitude higher in the bottom two plots compared to the the sensitivity shown in the top plot. The sensitivities of the number of cancer cells in the lungs with respect to the rate of blood flow, λ_1, and the rate of transfer of the cells to the lungs, λ_2, are about half of their respective peaks by the end of two days. However, the sensitivity with respect to the rate of loss from the lungs, λ_3, continues to increase at the end of two days. As with the previous plots, this information can be used to determine appropriate times of adminstering different medical treatments or drugs in order to obtain the most extreme results in reducing the number of cancer cells.

- *μ is the birth and natural death rate, with all births placed in S class*
- *β is the infection rate*
- *γ is the recovery rate*
- *α is the fatality rate due to the disease*

We assume the initial conditions are given and not subject to variation. We calculate the sensitivities by solving the extended system

$$\begin{cases} \dot{S} = \mu K - \beta S I - \mu S, \\ \dot{I} = \beta S I - \gamma I - \mu I - \alpha I, \\ \dot{R} = \gamma I - \mu R, \\ \frac{\partial \dot{x}}{\partial \lambda} = \begin{pmatrix} -\beta I & -\beta S + \mu & \mu \\ \beta I & \beta S - \gamma - \mu - \alpha & 0 \\ 0 & \gamma & -\mu \end{pmatrix} \frac{\partial x}{\partial \lambda} + \begin{pmatrix} -SI & 0 & I + R & 0 \\ SI & -I & -I & -I \\ 0 & I & -R & 0 \end{pmatrix}. \end{cases} \tag{1.1.10}$$

Here $x = \begin{pmatrix} S & I & R \end{pmatrix}^{\mathsf{T}}$ and $\lambda = \begin{pmatrix} \mu & \beta & \gamma & \alpha \end{pmatrix}^{\mathsf{T}}$.

While the method presented above provides a complete sensitivity analysis, there are some drawbacks to consider. First, the dimension of the extended model is substantially larger than the initial model if the number of parameters is much greater than the number of state variables. Second, there is no obvious error control or analysis other than *a priori* error bounds when solving only the forward problem. There is also the issue of what we do with all the information obtained via the forward derivative analysis. If the motivation is to obtain a quantity of interest, then post-processing of the results remains, and some or most of the sensitivities might be discarded to obtain this quantity of interest.

1.1.2 Adjoint Derivative Analysis

Consider again the linear model

$$Ax = b. \tag{1.1.11}$$

Suppose the quantity of interest $q(\lambda)$ is given by the linear functional $q(\lambda) = (x, \psi)$ for fixed vector $\psi \in \mathbb{R}^n$. We define the adjoint problem to be

$$A^\top \phi = \psi, \tag{1.1.12}$$

where $A^\top$ is the transpose (or adjoint) of the matrix A, and ϕ is the Green's vector. Note that

$$(x, \psi) = (x, A^\top \phi) = (Ax, \phi) = (b, \phi). \tag{1.1.13}$$

Now suppose the data b is subject to variation and is treated as a random variable with some distribution from which independent identically distributed samples of b are generated, and we want to know the distribution of the output. If we only use the forward problem, this requires solving the system (1.1.11) for each sampled data vector. In the adjoint analysis, we solve *one* adjoint problem *once*, and use an inner product to cheaply compute the output for each sample.

We now assume $A = A(\lambda)$ and $b = b(\lambda)$ depend on parameter λ. Now to obtain the quantity of interest for distinct λ using (1.1.13), we either need to solve (1.1.11) or (1.1.12) for each distinct value of λ. The sensitivity of the quantity of interest is obtained using similar steps as above and (1.1.3), which yields

$$\frac{\partial q(\lambda)}{\partial \lambda} = \left(A \, \frac{\partial b}{\partial \lambda} - \frac{\partial A}{\partial \lambda} x, \phi \right). \tag{1.1.14}$$

The sensitivity of the quantity of interest depends on the solution $x = x(\lambda)$ to (1.1.11). Thus, to compute the sensitivity of the quantity of interest we observe that for each distinct λ we solve (1.1.11) to obtain x, and we solve the adjoint problem each time with the new $A^\top$ since $A = A(\lambda)$ implies $\phi = \phi(\lambda)$. We then use a cheap inner product to compute the sensitivity. Compare this to the forward derivative analysis that also required solving two linear systems for each distinct λ. There does not appear to be an immediate benefit to this method. We show later that by linearizing (1.1.11) about reference parameters and the corresponding reference solutions, we need only solve a small number of adjoint problems once, and then use a global piecewise-linear approximation to $q(\lambda)$ that uses cheap inner products.

Adjoint problems for nonlinear differential equations are slightly more complicated than the adjoint problems for linear systems. If the forward problem is defined by a nonlinear map between two Banach spaces, then we require the map to be at least Fréchet differentiable. The smoothness of the map makes it possible to linearize a perturbation of the map around a reference solution/parameter pair, which in turn makes it possible to define an adjoint problem. Suppose we solve the differential equation

$$\begin{cases} \dot{x} = f(x; \lambda_1), & 0 < t \leq T, \\ x(0) = \lambda_0, \end{cases} \tag{1.1.15}$$

where the initial condition is also considered a parameter subject to variation, and the function f is smooth in both variables. Linear functionals take the form

$$q(\lambda) = \int_0^T (x(s; \lambda), \psi(s))\, ds, \tag{1.1.16}$$

where ψ is a function of time. We describe the effect of varying the parameter $\lambda = \begin{pmatrix} \lambda_1^\top & \lambda_0^\top \end{pmatrix}^\top$ around a *reference parameter value* $\gamma = \begin{pmatrix} \gamma_1^\top & \gamma_0^\top \end{pmatrix}^\top$.

Let y solve (1.1.15) for this reference parameter. We call y the *reference solution.* Let $e = x - y$ where x solves (1.1.15) at parameter λ near γ. Then e solves

$$\begin{cases} \dot{e} = f(x;\lambda_1) - f(y;\gamma_1), & 0 < t \leq T, \\ e(0) = \lambda_0 - \gamma_0. \end{cases} \tag{1.1.17}$$

The smoothness of f gives $f(x;\lambda_1) - f(y;\gamma_1) \approx f'(y;\gamma_1)e + \partial_{\lambda_1} f(y;\gamma_1)(\gamma_1 - \lambda_1)$, where $f'(y(t;\gamma);\gamma_1)$ is the Jacobian of f, and $\partial_{\lambda_1} f(y;\gamma_1)(\gamma_1 - \lambda_1)$ is the derivative of f with respect to λ evaluated at the reference solution/parameter pair. Substitution of this linear approximation to the perturbation of f into (1.1.17) yields a linear differential equation for e. We use this linear differential equation to form an adjoint. When we refer to linearizing a nonlinear map around a reference solution/parameter pair, we follow this process of linearizing the perturbation.

The generalized Green's function $\phi(t)$ solves the adjoint problem

$$\begin{cases} -\dot{\phi} - A^{\top}\phi = \psi, \; T > t \geq 0, \\ \phi(T) = \psi(T), \end{cases} \tag{1.1.18}$$

where $A := f'(y(t;\gamma);\gamma_1)$. A standard variational argument analogous to (1.1.13) gives

$$\frac{\partial q(\lambda)}{\partial \lambda}(\lambda - \gamma) \approx (\lambda_0 - \gamma_0, \phi(0)) + \int_0^T (\partial_{\lambda_1} f(y;\gamma_1)(\lambda_1 - \gamma_1), \phi)\, ds. \tag{1.1.19}$$

The last term on the right describes the effect of variations in the model parameters, and the second to last term describes the effect of variations in the initial conditions. Similar results hold for iterated maps, e.g. discrete time population models, and partial differential equations.

Example 1.1.3. *Recall that the quantity of interest for the malignant tumor model is the* total amount of radioactivity $q(\lambda) = x_1 + x_2$. *Suppose the goal* *is to compute the total amount of radioactivity after a day and a half, so we set the* dual data $\psi(t) = \delta(t-36)\begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}}$.

The linearized adjoint problem is

$$\begin{cases} -\dot{\phi} - \begin{pmatrix} -\lambda_1 - \lambda_2 & \lambda_2 \\ 0 & -\lambda_3 \end{pmatrix} \phi = \psi, \; T > t \geq 0, \\ \phi(T) = 0, \end{cases} \tag{1.1.20}$$

and the quantity of interest sensitivity, denoted $\nabla q(\lambda)$, *is*

$$\nabla q(\lambda) = \int_0^T \begin{pmatrix} -y_1 & 0 \\ -y_1 & y_1 \\ 0 & -y_2 \end{pmatrix} \phi \, ds. \tag{1.1.21}$$

Numerically solving the adjoint on the same time mesh with the same method as the forward problem yields the sensitivities shown in Table 1.1, which are compared with results obtained from the forward derivative analysis. The values produced from the two analyses are identical. This is not surprising since the numerical method is identical for both analyses. There are some advantages of the adjoint analysis over the forward derivative analysis, such as

- *The quantity of interest and its sensitivity are calculated directly, i.e. no further processing of the solution is required*
- *Adjoint analysis naturally extends to the effects of numerical error in the evaluation of the model*

Example 1.1.4. *Recall the classical* SIR*-epidemic model described by the nonlinear system of differential equations*

$$\begin{cases} \dot{S} = \mu K - \beta S I - \mu S, \\ \dot{I} = \beta S I - \gamma I - \mu I - \alpha I, \\ \dot{R} = \gamma I - \mu R. \end{cases} \tag{1.1.22}$$

	Adjoint Analysis Results	Forward Analysis
$q(\lambda)$	$9.4204e+006$	$9.4204e+006$
$\nabla q(\lambda)$	$1.0e+008\begin{pmatrix} -0.2533 \\ 1.0551 \\ -3.1388 \end{pmatrix}$	$1.0e+008\begin{pmatrix} -0.2533 \\ 1.0551 \\ -3.1388 \end{pmatrix}$

Table 1.1: Table of sensitivities of total amount of radioactivity at 36 hours

We rewrite this as

$$\dot{x} = f(x;\lambda), \tag{1.1.23}$$

where $x = \begin{pmatrix} S & I & R \end{pmatrix}^{\mathsf{T}}$ *and* $\lambda = \begin{pmatrix} \beta & \gamma & \mu & \alpha \end{pmatrix}^{\mathsf{T}}$*. Suppose the quantity of interest is the total population at the end of one week. The dual data* $\psi(t) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^{\mathsf{T}} \delta(t-7)$*, and the linearized adjoint problem is*

$$\begin{cases} -\dot{\phi} - \begin{pmatrix} -\beta I & -\beta S + \mu & \mu \\ \beta I & \beta S - \gamma - \mu - \alpha & 0 \\ 0 & \gamma & -\mu \end{pmatrix}^{\mathsf{T}} \phi = \psi, \; T > t \geq 0, \\ \phi(T) = 0. \end{cases} \tag{1.1.24}$$

There are viruses that have many strains that can vary from year to year, e.g. the flu. Thus, this is a perfect example for examining the effect of variations in the model parameters on the quantity of interest. Using (1.1.19) we see that

$$q(\lambda) \approx \int_0^T (y,\psi)\, ds + \int_0^T \left(\begin{pmatrix} -SI & 0 & I+R & 0 \\ SI & -I & -I & -I \\ 0 & I & -R & 0 \end{pmatrix} (\lambda - \bar{\lambda}), \phi \right) ds, \tag{1.1.25}$$

where y is the reference solution solved about reference parameter $\bar{\lambda}$.

1.2 Inverse Sensitivity Analysis

We study the following inverse sensitivity analysis problem, given a specified uncertainty in the output of a map, e.g. a quantity of interest calculated from the solution of a differential equation, determine variations in the input data and parameters that produce this uncertainty in the output. This inverse sensitivity analysis problem is typically ill-posed, i.e. there is typically more than one solution, since the solution involves inverting through a many-to-one map. One approach to deal with nonuniqueness of solutions is a probabilistic description of the input data and parameters. We describe the inverse sensitivity analysis problem assuming an abstract probabilistic formulation of the forward problem. Given

- a joint density, $\rho_{\mathbf{\Lambda}}(\lambda)$, in parameter (input) space $\mathbf{\Lambda} \subset \mathbb{R}^d$,
- a model $M(Y, \lambda)$ with solution $Y = G(\lambda)$ that depends (implicitly) on parameters λ in a smooth way,
- a linear functional $q(\lambda) = q(Y(\lambda))$,

determine the density, $\rho_{\mathcal{D}}(q)$, of measurable output data $q(\lambda)$. Note that we assume the model, e.g. the physics of the problem at hand, is well known and described by a smooth system of equations so the map $q(\lambda)$ is implicitly a smooth and *deterministic* function of λ. Even though this is a probabilistic formulation of the forward problem, we may use derivative information typically used in deterministic forward sensitivity analysis to cheaply approximate the density of $q(\lambda)$ [20, 11, 12, 8, 9, 6, 10]. We now state the abstract version of the inverse problem. Given

- a model $M(Y, \lambda)$ with solution $Y = G(\lambda)$,

- a linear functional $q(\lambda) = q(Y(\lambda))$,
- an *observed* density, $\rho_{\mathcal{D}}(q(\lambda)) = \rho_{\mathcal{D}}(q(Y(\lambda)))$, of output values $q(\lambda)$,

determine a *posterior* density, $\sigma_{\mathbf{\Lambda}}(\lambda)$ of parameters that produces the observed density. This inverse problem has an abstract formulation, where we pose a density for the output in order to observe the effect. It also has an experimental formulation, where the model output is meant to match the observed values of an experiment and the imposed density is associated with the experimental data. This observed density may reflect the uncertainty in the data or be the consequence of error. It is affected by our ability to measure it. There is a very strong connection between the inverse sensitivity and forward sensitivity problems that we exploit in our approach.

Before generalizing to higher dimensions, we present some simple one-dimensional inverse problems and the issues that arise in such problems.

Example 1.2.1. *Consider the 1-1 map $y = e^x$ over the interval $[0,1]$. If x is assumed a random variable with uniform distribution on $[0,1]$, then y is a random variable with a distribution determined from a simple change of variables formula, see Figure 1.3. Now suppose we start with this distribution on y and wish to invert through the map $y = e^x$. We can use a simple change of variables as with the forward problem to determine the distribution of random variable x is uniform on $[0,1]$. Suppose instead we linearize the forward problem using a deterministic sensitivity analysis and invert through the linear map to obtain an estimate of the distribution of the random variable x. Figures 1.4-1.6 show the results obtained using various piecewise linear approximations to $y = e^x$. These figures suggest that as* *the linear approximations converge to the nonlinear map, the approximate density functions of the random variable x converge in some sense to the true density function. We note that some liberties are taken in the inversion through the piecewise linear approximations since these do not form a function of y in terms of x (there are some values of y that can be mapped to two values of x), but we address this problem more generally in the next example. To obtain the results of Figures 1.4-1.6, we simply invert through each piece of the piecewise linear approximation independently as a way to navigate around this technical difficulty.*

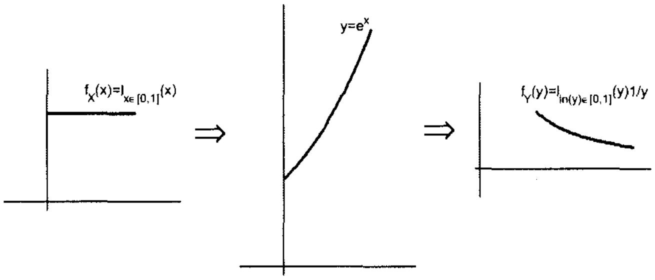

Figure 1.3: The forward problem of determining the density of random variable y (right) by passing the density of x (left) through the map $y = e^x$ (middle)

Example 1.2.2. *Consider the two-to-one map defined by $y = x(1 - x)$ on $[0, 1]$. Suppose y is a uniform random variable on $[0, 0.25]$, i.e. the range of $y = x(1 - x)$ on $[0, 1]$, and we want to determine the density of random variable x on $[0, 1]$. Except at $y = .25$, there are two possible x values for each value of y and a simple change of variables cannot be used to determine the density of x unless we restrict x to be in $[0, .5]$ or $[.5, 1]$ where the map is 1-1, see Figure 1.7. Thus, except at $y = .25$, we identify the set of two values of x found by inverting $y \in [0, .25)$ as the inverse.*

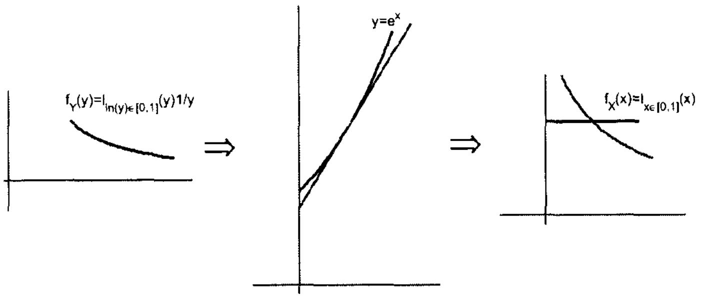

Figure 1.4: The inverse problem of approximating the density of random variable x (right plot, approximation is in red, exact is in blue) by passing the density of y (left) through a linear approximation of the map $y = e^x$ (middle plot, linear approximation is in red, exact is in blue)

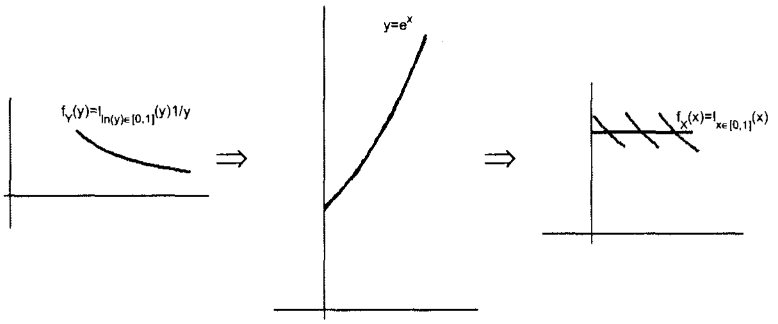

Figure 1.5: The inverse problem of approximating the density of random variable x (right plot, approximation is in red, exact is in blue) by passing the density of y (left) through a piecewise-linear approximation of the map $y = e^x$ (middle plot, linear approximation is in red, exact is in blue)

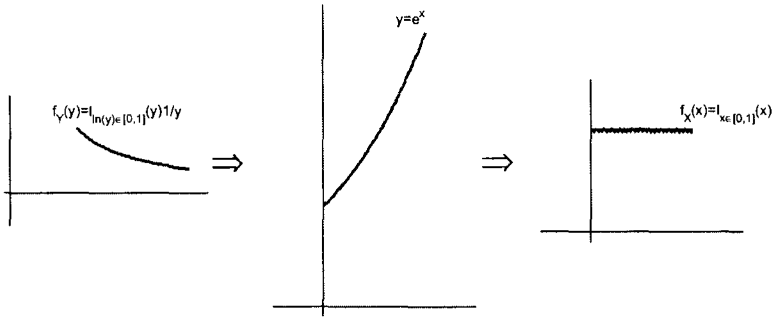

Figure 1.6: The inverse problem of approximating the density of random variable x (right plot, approximation is in red, truth is in blue) by passing the density of y (left) through a piecewise-linear approximation of the map $y = e^x$ (middle plot, linear approximation is in red, truth is in blue)

With this set-valued interpretation, all the inverses are indexed on the line-segment on the $x - axis$ between either $[0, .5]$ *or* $[.5, 1]$. *Given two distinct indexed points x_α and x_β in either line-segment of* $[0, .5]$ *or* $[.5, 1]$, *we might choose an indexing so that if $x_\alpha < x_\beta$, then the indices α and β satisfy $\alpha < \beta$. In this way, we define a "direction" to the inverse set so that the points in* $[0, .5]$ *are indexed in increasing order and the points in* $[.5, 1]$ *are indexed in decreasing order, and given a distribution on y, we can uniquely define the distribution of the inverse set on either* $[0, .5]$ *or* $[.5, 1]$ *by using consistent indexing. Thus, the distributions of the inverses is independent of the choice of inverse set! This approach is more complicated for a map that is not symmetric as seen in Example 1.2.3.*

Suppose we want to invert into the entire set of possible inputs $x \in$ $[0, 1]$ *that generate the output $y(x) \in$* $[0, 0.25]$. *We require some information on the probability of the various possible inputs such as a* prior *density function on the parameter space for a Bayesian approach. Given a value of y to invert through this two-to-one map, we find all possible values of x that map to this y value and use a given prior density function to determine the probability of each possible choice of x. For example, if the prior probability is uniform on* $[0, 1]$, *then each value of x found in Figure 1.7 for the given value of* $y = .15$ *is assigned probability* $.5$, *and more generally the posterior density is given by Figure 1.8. If the prior is defined as values in* $[0, 0.5]$ *being twice as likely as values in* $(.5, 1]$, *then the probability of the value of x to the left (right) of* $x = .5$ *in Figure 1.7 is* $2/3$ *(*$1/3$*) and the posterior density is given by Figure 1.8.*

Example 1.2.3. *Consider the map defined by*

$$y = \begin{cases} x(1-x), & 0 \leq x < .5, \\ -4(x-.5)^3 + .25, & .5 \leq x \leq 1. \end{cases} \tag{1.2.1}$$

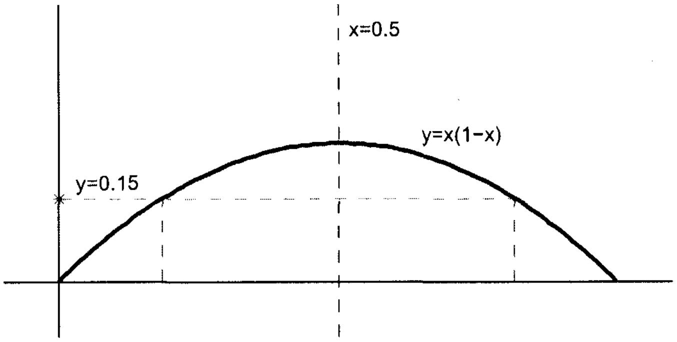

Figure 1.7: Plot of two-to-one map

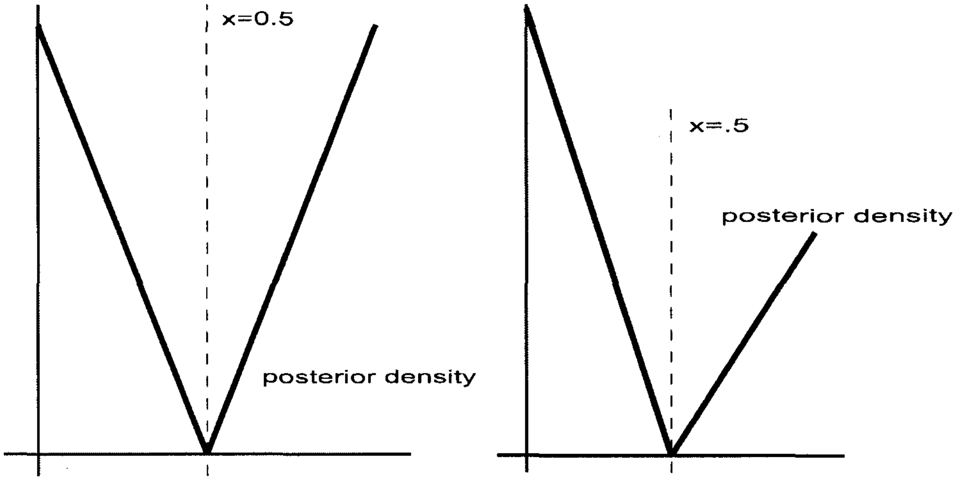

Figure 1.8: Left: A posterior density function if prior is uniform. Right: A posterior density function if prior is nonuniform

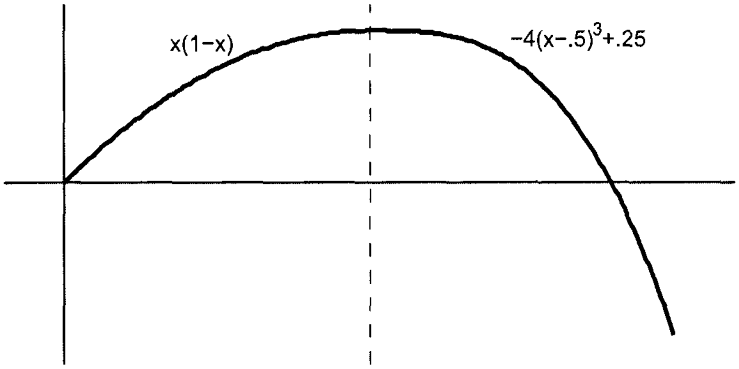

Figure 1.9: Non-symmetric map

We plot this map in Figure 1.9. As above, we assume y is a random variable with density function whose support is the range of the map, which in this case is $[-.25, .25]$. *We seek the density function of input random variable x given random variable y. Note that if we restrict x to be in* $[.5, 1]$, *then the map is a bijection so each value of y is uniquely associated with a unique value of x. If we restrict x to be in* $[0, .5]$, *then each value of x is associated with a unique value of y, but there are y values that are not mapped to from any value of x in this interval due to the lack of symmetry in this map. We can add a disjoint interval and restrict x to be in* $[0, .5] \cup ((1/16)^{1/3} + .5, 1]$, *which defines a bijection between x and y. We might define an indexing of* $[.5, 1]$ *as before, so the points are indexed in decreasing order. If we choose* $[0, .5] \cup ((1/16)^{1/3} + .5, 1]$, *then with the same indexing scheme, we begin indexing points in* $((1/16)^{1/3} + .5, 1]$ *in decreasing order, and then index points* $[0, .5]$ *in increasing order.*

Chapter 2

GENERALIZED CONTOURS: A COMPUTATIONAL APPROACH TO USING SET VALUED INVERSES

2.1 Basic theory

As foreshadowed in the 1-D examples of the previous chapter, we first define a unique solution to the inverse problem by indexing the "inverse sets." This does not require a probabilistic approach to inversion, but is rather a well posed "inverse" density estimation problem, in which we use a random variable on the output, and we compute the density of the random variable of our indexed inverse sets. In the next chapter, we use a probabilistic description, i.e. a measure-theoretic description, to further invert into the entire parameter space and make that problem mathematically rigorous. Our computational approach to inverting through a many-to-one map by using a set valued inverse uses the well known

Theorem 2.1.1 (Implicit Function Theorem). *Let $f(x,y) : \mathbb{R}^{n+m} \longrightarrow \mathbb{R}^m$ be continuously differentiable, where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. If for $(a,b) \in \mathbb{R}^{n+m}$, $f(a,b) = 0$ and $\left[\partial_{y_j} f_i(a,b)\right]$ is an invertible matrix, then there exists open U containing a, open V containing b and a differentiable function $g : U \longrightarrow V$ such that*

$$\{(x, g(x))\} = \{(x,y) | f(x,y) = 0\} \cap (U \times V). \qquad (2.1.1)$$

Under the assumption of a smooth map, if we are given a fixed output value $\bar{q}$, then the Implicit Function Theorem guarantees the existence of a $(d-1)$-dimensional manifold in $\mathbf{\Lambda}$ that is mapped to $\bar{q}$. Motivation and intuition comes from the two-dimensional case, $\lambda = (\lambda_1, \lambda_2)$, and in this dimension the manifolds are simply contours of the surface $q(\lambda_1, \lambda_2)$ (Fig. 2.1). Every point in $\mathbf{\Lambda}$ lies on a unique contour, so we first consider $\mathbf{\Lambda}$ as a set described by its contours.

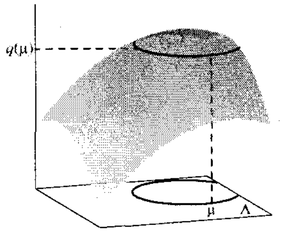

Figure 2.1: Each observation value corresponds to a unique contour curve

We need to find a way to describe the set of contours. In the set of these contours, there exists one-dimensional curves transverse to the contours that intersect each contour once and only once (lefthand illustration in Fig. 2.2). We can take one of these curves as the index for the set of contours. There is a bijection between the points on an index curve and the points in the range of the output $q(\Lambda)$. Therefore, any measure posed on the range of the output imposes a measure on the index curve. Thus, the intersections of the contours with the index curve is a random variable with a distribution uniquely defined by the distribution of the output $\rho_d(q(\lambda))$ (righthand illustration in Fig. 2.2). In other words, there exists a unique solution to the inverse sensitivity analysis problem in the set of contours.

However, determining the set of contours analytically is infeasible in practice. To construct a computable approximation, we use a piecewise-linear tangent plane approximation to the output surface $q(\lambda)$ to construct approximate contours.

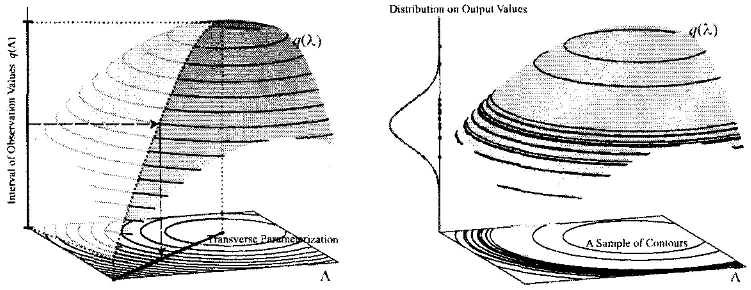
Interval of Observation Values $q(\lambda)$
$q(\lambda)$
Transverse Parameterization
Λ
Distribution on Output Values
$q(\lambda)$
A Sample of Contours
Λ

Figure 2.2: Left: On the horizontal plane, we show a transverse parameterization. Each point on the transverse parameterization corresponds to a unique contour curve, so the transverse parameterization acts as an index for the space of contour curves. There is a unique map from the points in the interval containing the observed output values to the points on the transverse parameterization. Right: We show a probability distribution imposed on the output values. A sample of output values drawn from this distribution corresponds to a unique sample of contour curves.

We study the inverse problem for a finite dimensional map q from the space of parameters to the output. Such a map can be defined implicitly, for example by the solution to a differential equation that depends on a finite number of parameters in the problem. With this in mind, we consider the finite dimensional nonlinear system of equations

$$f(x;\lambda) = b, \tag{2.1.2}$$

where $x \in \mathbb{R}^n$, parameter $\lambda \in \mathbf{\Lambda} \subset \mathbb{R}^d$ (we assume $\mathbf{\Lambda}$ is compact) is a random vector, and $f : \mathbb{R}^{n+d} \to \mathbb{R}^n$ is assumed smooth in both variables. The motivation for solving this system is to compute a quantity of interest, denoted $q(\lambda) = q(x(\lambda))$, which can be represented as a linear functional of the solution $x(\lambda)$. The Riesz representation theorem implies there exists

$\psi \in \mathbb{R}^n$ such that $q(\lambda) = \langle x, \psi \rangle$. Note that x depends implicitly on λ. Assuming that this dependence is smooth, then the dependence of q on λ is also smooth.

We define $\hat{q}(\lambda) := q(\lambda) - \bar{q}$, where $\bar{q} \in q(\mathbf{\Lambda})$ is fixed but arbitrary. By assumption, $\hat{q}(\lambda) : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and there exists $\bar{\lambda} \in \mathbf{\Lambda}$ such that $q(\bar{\lambda}) = \bar{q}$, which implies that $\hat{q}(\bar{\lambda}) = 0$. We are initially only interested in the cases where the quantity of interest varies as the parameters vary (and later where local extrema of the quantity of interest occur), so we assume that $\partial_{\lambda_d}\hat{q}(\bar{\lambda}) \neq 0$. Then, by the Implicit Function Theorem, there exists an open set $U_{\bar{\lambda}} \subset \mathbf{\Lambda}^{d-1}$, where $\mathbf{\Lambda}^{d-1} := \{\lambda^{d-1} := (\lambda_1, \ldots, \lambda_{d-1}) | \lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbf{\Lambda}\}$, containing $\bar{\lambda}^{d-1}$, an open set $V_{\bar{\lambda}} \subset \mathbf{\Lambda}_d$, where $\mathbf{\Lambda}_d := \{\lambda_d | \lambda \in \mathbf{\Lambda}\}$, and a differentiable function $g_{\bar{\lambda}} : U_{\bar{\lambda}} \to V_{\bar{\lambda}}$ such that

$$\{(\lambda^{d-1}, g_{\bar{\lambda}}(\lambda^{d-1}))\} = \{\lambda | q(\lambda) = \bar{q}\} \cap (U_{\bar{\lambda}} \times V_{\bar{\lambda}}). \tag{2.1.3}$$

Since the Implicit Function Theorem is a local result, there may be additional points in $\mathbf{\Lambda}$ that map to $\bar{q}$, but are not contained in the set defined by (2.1.3). Thus, given $\bar{q} \in q(\mathbf{\Lambda})$, we choose a collection of sets $\{U_{\bar{\lambda}} \times V_{\bar{\lambda}}\} = \bigcup_{\alpha \in A}\{U_{\bar{\lambda}_\alpha} \times V_{\bar{\lambda}_\alpha}\}$ where $\bigcup_{\alpha \in A}\{\bar{\lambda}_\alpha\}$ is the set of all $\lambda \in \mathbf{\Lambda}$ such that $q(\lambda) = \bar{q}$. Then using the same notation as in (2.1.3), the function $g_{\bar{\lambda}}(\lambda^{d-1})$ might be piecewise defined.

Definition 2.1.1. *The set in (2.1.3) is a $(d-1)$-dimensional manifold that is a natural inverse of $q(\lambda)$ given $\bar{q}$. We call this set the* ***generalized contour****.*

Theorem 2.1.2. *If we choose distinct $\bar{q}_1, \bar{q}_2 \in q(\mathbf{\Lambda})$, then the generalized contours for $\bar{q}_1$ and $\bar{q}_2$ are unique and do not intersect.*

Proof:

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

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

Remark 2.1.1. *We relax the restriction of $\partial_{\lambda_d}\hat{q}(\bar{\lambda}) \neq 0$ for a finite number of points in $\mathbf{\Lambda}$, where $q(\lambda)$ possibly attains a local extreme value. We ignore this set of points when considering the generalized contours.*

Our goal is to show that in the space of generalized contours, the inverse problem has a unique solution. We start by proving that there exists (possibly discontinuous) 1-dimensional curves that are transverse to the generalized contours. This allows us to *index* the family of generalized contours, and subsequently define the inverse into the space of generalized contours for a given distribution of $q(\lambda)$ uniquely. We give a constructive proof that is a useful algorithm. The algorithm produces discontinuous curves in $\mathbf{\Lambda}$ in general.

Definition 2.1.2. *We call any such curve that has the property that it intersects each generalized contour once and only once a* ***transverse parameterization*** *(TP).*

Theorem 2.1.3. *Suppose f is smooth in (2.1.2) and $q(\lambda)$ is a linear functional of the solution to (2.1.2). If $q(\lambda)$ is a random variable with distribution $F_q(q(\lambda))$, then there exists a TP in $\mathbf{\Lambda}$, the distribution of the intersections of the generalized contours on the TP, which is a random variable, is unique.*

We interpret this theorem to say that there is a unique solution to the inverse problem in the space of generalized contours.

Proof:

We claim that such a transverse curve can be constructed from a finite number of connected curves. To prove this, first fix $\epsilon > 0$ and $\epsilon > \delta > 0$, and take $I_{q,\epsilon} = [q_m + \epsilon, q_M - \epsilon]$. If $\mathbf{\Lambda}$ is compact, then the existence of transverse curves is guaranteed by the smoothness of $q(\lambda)$. To construct such a curve, we begin at a point $\gamma_M \in \mathbf{\Lambda}$ such that $q(\gamma_M) = q_M - \delta$, and follow the direction of the negative gradient until the curve either intersects the boundary or a minimum or saddle is reached, and denote this point γ_m. From smoothness, we have that exactly one contour for each value of $q(\lambda)$ between $(q(\gamma_m), q(\gamma_M))$ has been intersected by this curve. If $(q(\gamma_m), q(\gamma_M))$ does not completely cover $I_{q,\epsilon}$, then we select a point $\tau_m \in \mathbf{\Lambda}$ such that $q(\tau_m) = q_m + \delta$, and follow the direction of the gradient until the curve either intersects the boundary or a maximum or saddle is reached, and denote this point τ_M. We now check if $(q(\gamma_m), q(\gamma_M)) \cup (q(\tau_m), q(\tau_M))$

covers $I_{q,\epsilon}$. If so, then we eliminate any part of the second curve that gives an overlap with contours intersected by the first. Otherwise, we continue to create this possibly discontinuous curve with the same steps above trying to cover the output interval defined by $(q(\tau_M), q(\gamma_m))$. This process produces a countable number of connected curves whose union forms a (possibly discontinuous) transverse curve through the generalized contours that corresponds to a countable open cover of $I_{q,\epsilon}$, which is compact, so there is a finite subcover of $I_{q,\epsilon}$, which implies that only a finite number of steps is needed to construct such a (possibly discontinuous) transverse curve to the generalized contours of $I_{q,\epsilon}$. Thus, there exists a (possibly discontinuous) 1-dimensional curve in $\mathbf{\Lambda}$ that is orthogonal to the generalized contours it intersects, and intersects each generalized contour exactly once.

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

2.2 Approximation of the space of generalized contours

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

Definition 2.2.1. *The generalized contour above is a $(d-1)$-dimensional hyperplane and we refer to this as a* ***generalized linear contour****.*

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

2.3 Local linearization of the linear functional

The goal is to approximate the map $q(\lambda)$ with a piecewise-linear map $\tilde{q}(\lambda)$ since it is possible to calculate the generalized contours for this approximate map. We show how to linearize the map $q(\lambda)$ locally, and prove the approximate contours converge to the true contours locally as the number of linearization points increases.

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

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

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

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

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

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

$$A^\top \phi = \psi, \tag{2.3.1}$$

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

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

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

$$\left[g_{\bar{\lambda}}(\lambda^{d-1}) - \tilde{g}_{\bar{\lambda}}(\lambda^{d-1})\right] \left[\phi^{\top} \partial_{\lambda_d} f(y, \mu)\right] = - \langle \mathcal{R}, \phi\rangle . \tag{2.3.2}$$

By assumption,

$$\partial_{\lambda_d} q(\lambda) = \phi^{\top} \partial_{\lambda_d} f(y, \mu) \neq 0,$$

so we rewrite (2.3.2) as

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

where

$$C^{-1} = -\phi^{\top} \partial_{\lambda_d} f(y, \mu),$$

is a nonzero constant determined entirely by the reference point (y, μ). Thus, if we define

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

where $\|\|_2$ denotes the standard Euclidean norm, then as $\|U_{\bar{\lambda}}\| \to 0$, $\|\mathbf{R}\|_2 \to 0$, which implies that $\left|g_{\bar{\lambda}}(\lambda^{d-1}) - \tilde{g}_{\bar{\lambda}}(\lambda^{d-1})\right| \to 0$. We summarize this as the following

Theorem 2.3.1. *The generalized linear contours converge pointwise to the true contours locally in* $\mathbf{\Lambda}$.

Remark 2.3.1. *This also applies to differential equations that depend on a finite set of parameters. For ordinary differential equations, we require the same assumptions as the standard existence and uniqueness theorems to guarantee the smoothness of $q(\lambda)$. For partial differential equations, we have similar smoothness assumptions, and discuss these along with an in depth error analysis in more detail in later chapters.*

2.4 Global linearization of the linear functional

We extend the local linearization technique to obtain a global piecewise-linear approximation of the linear functional over all of $\mathbf{\Lambda}$. We first define a partition $\{B_i\}_{i=1}^{M}$ of $\mathbf{\Lambda}$. We refer to the B_i as *cells* even though we might use Voronoi tesselations or other more complicated geometric objects to partition $\mathbf{\Lambda}$. We apply the local linearization technique described above for each box, and defining

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

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

$$\tilde{q}(\lambda) := \sum_{i=1}^{M} \left(q(\mu_i) + \langle \nabla q(\mu_i), (\lambda - \mu_i) \rangle \right) \mathbf{1}_{B_i}(\lambda), \tag{2.4.1}$$

where μ_i is the reference parameter value chosen in cell B_i. For the finite system of nonlinear equations, we have

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

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

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

The generalized linear contour of $\tilde{q}$ given $\bar{q}$ is a collection of line segments in $\mathbf{\Lambda}$. Using the same notation as above,

$$\left| g_{\bar{\lambda}}(\lambda^{d-1}) - \tilde{g}_{\bar{\lambda}}(\lambda^{d-1}) \right| \leq C \sum_{i=1}^{M} \left| \langle \mathcal{R}_i, \phi_i \rangle \right|,$$

where

$$C^{-1} = \min_i \left\{ \left| \phi_i^\top \partial_{\lambda_d} f(y_i, \mu_i) \right| \right\}.$$

Thus, as $||B_i|| \to 0$ (or as $M \to \infty$ when the number of sample points are distributed uniformly), the generalized linear contour converges pointwise to the generalized contour. We demonstrate the convergence of generalized linear contours to true contours in the two examples below.

Example 2.4.1. *Suppose* $q(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 \exp\left[-(\lambda_1^2 + 1.25\lambda_2^2 - 1)\right]$ *over* $[0, 2] \times [0, 2]$. *We approximate* q *over some partition* $\{B_i\}$ *of* $[0, 2] \times [0, 2]$, *where each* B_i *is a square of the same size, and we linearize around the mid-point of each* B_i *to form* $\tilde{q}$ *as in (2.4.1). We plot various contour curves and two TP's on each plot. The black dotted line (with stars marking where it enters and exits each* B_i*) is a TP formed by starting at the maximum of* $\tilde{q}$ *and following the direction of the negative gradient. The red dotted line (with plus signs marking where it enters and exits each square) is a TP formed by starting at the minimum of* $\tilde{q}$ *and following the direction of the positive gradient. The results are summarized in Fig 2.3 below.*

Example 2.4.2. *Suppose* $q(\lambda_1, \lambda_2) = \exp\left[\cos(\lambda_1) + \sin(\lambda_2)\right]$ *on* $[-2\pi - 0.1, 2\pi + 0.1]^2$. *We proceed as above to obtain the numerical results summarized in Fig 2.4 below.*

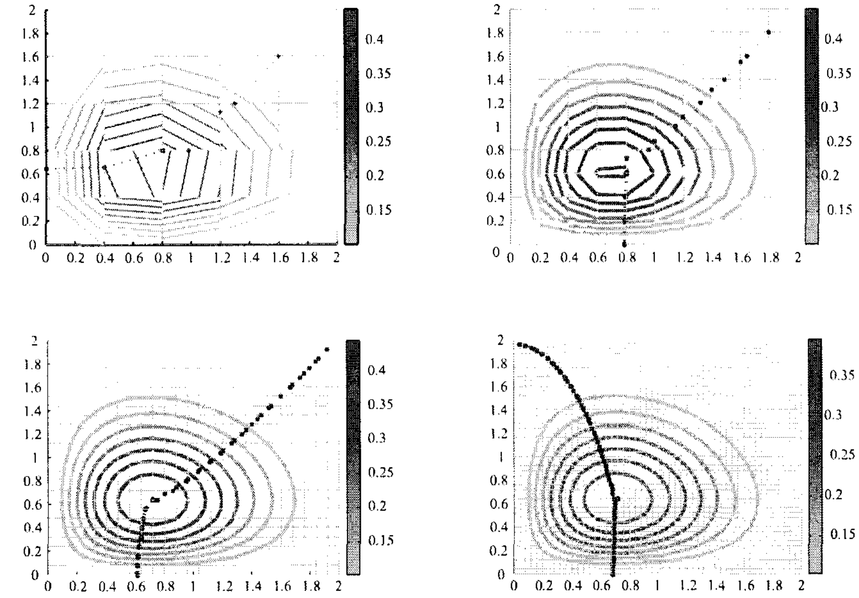

Figure 2.3: Contours of $\tilde{q}$ using 5 × 5 square cells (top left), 10 × 10 square cells (top right), 25 × 25 square cells (bottom left) and 50 × 50 square cells (bottom right). The TP is created using the same steps as in the proof of its existence and is denoted by the circle-dotted and plus-dotted lines. The circle-dotted line is constructed from the maximum of $q(\lambda)$ and follows the negative direction of the gradient of $q(\lambda)$, and the plus-dotted line is constructed from the minimum of $q(\lambda)$ and follows the direction of the gradient.

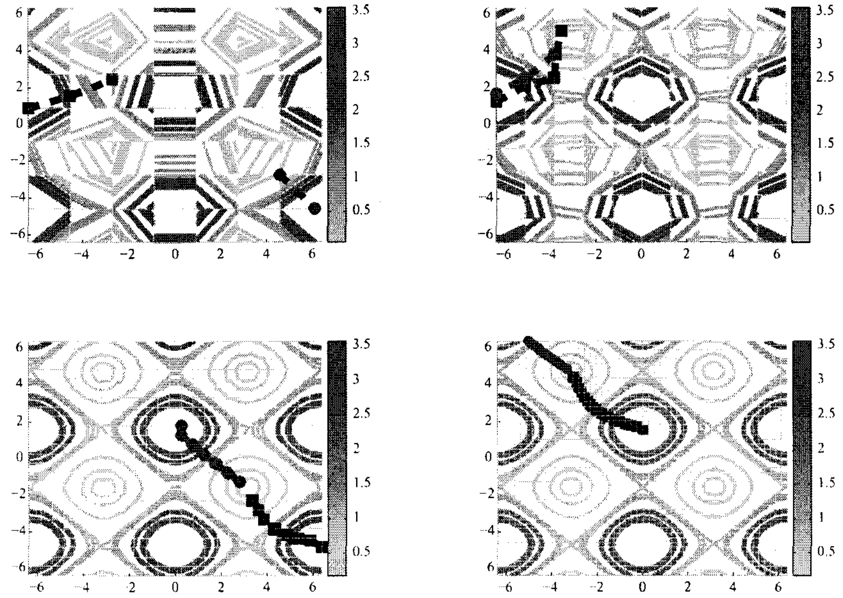

Figure 2.4: Contours of $\tilde{q}$ using 7 × 7 square cells (top left), 10 × 10 square cells (top right), 25 × 25 square cells (bottom left), and 50 × 50 square cells (bottom right). The TP is created using the same steps as in the proof of its existence and is denoted by the square-dotted and circle-dotted lines. The square-dotted line is constructed from the maximum of $q(\lambda)$ and follows the negative direction of the gradient of $q(\lambda)$, and the circle-dotted line is constructed from the minimum of $q(\lambda)$ and follows the direction of the gradient.

Chapter 3

PARAMETER ESTIMATION

3.1 Basic concept of parameter estimation

We define the solution of the inverse sensitivity problem as computing the probability of events (measurable sets) in parameter space. This requires further inversion from the computed density on the set of contours as described in the previous chapter. In order to assign a probability to a measurable set in $\mathbf{\Lambda}$, we recognize that such a set is defined by the contours it contains and the amount of each contour it contains, see Fig. 3.1. A measure specified on $\mathbf{\Lambda}$ quantifies the amount of each contour contained in

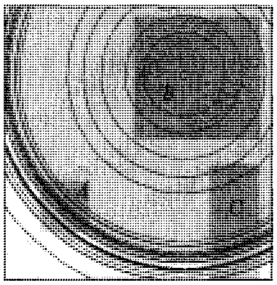

Figure 3.1: Plotted is a sample of contour lines in parameter space corresponding to a specified distribution on the output observation values along with three events. The Lebesgue uniform measure is specified as the measure on the parameter space. Event B has relatively low probability because while it has relatively large area, it contains relatively few contours. Event A has intermediate probability because while the area of event A is relatively small, A contains relatively many contours. The probability of event C is largest because it contains as many contours as A but has larger area.

any given set. Combining the results of the generalized contours with such a measure, and using the Monotone Convergence Theorem and additivity properties of measures, we develop an algorithm to estimate the probability of any measurable set in $\mathbf{\Lambda}$.

Alternately, results may be obtained by standard implementation of a Bayesian approach by sampling prior densities, solving the nonlinear model for each sample, accepting/rejecting these samples using the observed density, and finally by binning the accepted samples. Our methodology does not require any model solves other than the small number used to construct a piecewise-linear tangent plane approximation after which all calculations are done directly in parameter space. Furthermore, our method does not require sampling the prior density as it is used only as a normalized measure to determine the size of contours, so we never encounter convergence issues that often arise in Markov Chain Monte Carlo methods of sampling used to implement the Bayesian approach. Our method also gives the ability to directly test candidate prior densities without performing additional model solves.

3.2 Posterior densities

For smooth f in (2.1.2) and linear functional $q(\lambda)$, if $q(\lambda)$ is a random variable with distribution $F_q(q(\lambda))$, then for a fixed TP in $\mathbf{\Lambda}$, the distribution of the points of intersection of the generalized contours on the TP is unique.

We now discuss how to use the unique solution to the inverse problem to determine estimates of model parameters. We first observe if $I = [q_1, q_2]$ is an event with probability $P(I)$ (meaning the probability of the quantity of interest $q(\lambda)$ occuring in the interval I), then this corresponds to a measurable set in $\mathbf{\Lambda}$ that is defined as the set of all contours obtained by $q^{-1}(I)$. From the basic assumptions of smoothness and the nonintersecting property of the contours, the set of all contours is a set in $\mathbf{\Lambda}$ that is contained between the two contours defined by $q^{-1}(q_1)$ and $q^{-1}(q_2)$ (or possibly one of these contours and the boundary of $\mathbf{\Lambda}$). We assign this set the probability $P(I)$. In this way, we can assign a distribution to the space of contours that is described by the distribution of intersection points of the contours on the TP. In order to assign a probability to a measurable set in $\mathbf{\Lambda}$, we first recognize that such a set can be defined by the contours it contains and "how much" of each contour it contains. In order to assign a probability to such a set, it is clear that we need to quantify what is meant by "how much" of a contour is contained in the set, and to do this we need to have some concept of the "size" of a contour. Thus, we must identify a measure on $\mathbf{\Lambda}$.

Definition 3.2.1. *We normalize measures on* $\mathbf{\Lambda}$ *and call any such normalized measure a* ***joint measure*** *reflecting standard probabilistic terminology.*

Before proceeding further, we connect our notation with the standard nomenclature [25, 1]. Abstractly, we begin with a model $M(Y, \lambda)$ (e.g., finite system of nonlinear equations or differential equations) that depends on some **model parameters** λ. We refer to a particular choice of model parameters $\lambda \in \mathbf{\Lambda}$ as a **model selection** while the **model space** $\mathbf{\Lambda}$ is a manifold in which each point represents a possible model selection. We obtain information on model parameters via the **data** (or **observable parameters**) $q \in \mathcal{D}$ where the **data space** $\mathcal{D}$ is a manifold in which each point represents a possible measurement of the solution $Y = Y(\lambda)$ to the model $M(Y, \lambda)$. The data are determined by a linear functional of the solution to the model denoted $q(\lambda) = q(Y(\lambda))$. We define the **parameter manifold**, denoted $\mathcal{X}$, as the space whose points are defined by $x = (\lambda, q)$, where λ and q come from the model and data spaces, respectively. In other words, $\mathcal{X}$ is the joint manifold formed from the Cartesian product of the model and data manifolds, $\mathcal{X} := \mathbf{\Lambda} \times \mathcal{D}$.

We first assume that there exists measures on both the model and data space that are absolutely continuous with respect to the Lebesgue measure. This means we have a way of measuring *volume* in the model and data space, and we can measure volume in the joint manifold $\mathcal{X}$. We assume the manifolds have finite volume, and define the **homogeneous probability density** on $\mathcal{X}$ as

$$d\mu(x) = d\nu(x)/V,$$

where V is the total volume of $\mathcal{X}$ and $d\nu(x)$ is the **volume density** of the manifold $\mathcal{X}$. Thus, given a measurable set $A \subset \mathcal{X}$ we have a probability

$$\mu(A) = \int_A d\mu(x)$$

proportional to its volume $V(A)$. We assume that any probability measure P on $\mathcal{X}$ is absolutely continuous with respect to the homogeneous probability measure μ. By the Radon-Nikodym theorem, there exists positive function $f(x)$ such that

$$P(A) = \int_A f(x) d\mu(x) = \int_A \frac{dP(x)}{d\mu(x)} d\mu(x),$$

for any measurable $A \subset \mathcal{X}$. The function $f(x)$ is called a joint density function on $\mathcal{X}$.

Since $\mathcal{X}$ and its homogeneous probability density $d\mu$ come from a product space and product measures, we write

$$d\mu(x) = d\mu(\lambda, q) = d\mu_{\mathbf{\Lambda}}(\lambda) d\mu_{\mathcal{D}}(q),$$

where $d\mu_{\mathbf{\Lambda}}(\lambda)$ and $d\mu_{\mathcal{D}}(q)$ represent the homogeneous probability densities on model and data space, respectively.

Let $\Theta(\lambda, q)$ represent the joint (theoretical) probability density on $\mathcal{X}$ defined by the model, and use the fact that a joint probability density can be written as the product of a conditional and marginal probability density functions to obtain

$$\Theta(\lambda, q) = \theta(q \,|\, \lambda) d\mu_{\mathbf{\Lambda}}(\lambda), \tag{3.2.1}$$

where $\theta(q \,|\, \lambda)$ represents the probability density of an output given an input and we take the homogeneous probability density on model space as the marginal probability density function. Neglecting any uncertainties in the model, $\theta(q \,|\, \lambda) = \delta(q - q(\lambda))$. Equation (3.2.1) represents all the information on $\mathcal{X}$ we can obtain by solving the model.

Remark 3.2.1. *We ignore errors in model evaluation in this chapter. In a later chapter, we consider the effects of error in model evaluation and observed density.*

We use the probability density function $\rho_{\mathcal{D}}(q)$ to account for the uncertainty in the output, e.g. resulting from varying empirical observations or by modeling the error of the measurement equipment. We define the **prior information on $\mathbf{\Lambda}$** as the information obtained about the model parameters *independently* of the data, and we model this prior information with the probability density function $\rho_{\mathbf{\Lambda}}(\lambda)$. If we have no prior information, then we take $\rho_{\mathbf{\Lambda}}(\lambda) = d\mu_{\mathbf{\Lambda}}(\lambda)$. By definition of the prior information on $\mathbf{\Lambda}$, we write

$$\rho(\lambda, q) = \rho_{\mathbf{\Lambda}}(\lambda)\rho_{\mathcal{D}}(q), \tag{3.2.2}$$

to represent the joint (prior) probability density on $\mathcal{X}$ that takes into account all the prior information before statistical inversion and model parameter estimation.

We have defined two probability distributions on $\mathcal{X}$ that represent two different states of information. The combination of these two states of information defines the **posterior** state of information. It is widely accepted [25, 1, 19, 17] that the joint probability density on $\mathcal{X}$ that represents this posterior state of information is formed by the *conjunction* of the theoretical and prior information. Forming the conjunction leads to the posterior joint probability density function $\sigma(\lambda, q)$ defined by

$$\sigma(\lambda, q) = k\frac{\rho(\lambda, q)\Theta(\lambda, q)}{d\mu(\lambda, q)},$$

where k is a normalizing constant. Generic approaches to form a posterior joint probability density are widely understood and accepted in the Bayesian inference community developed with respect to applications to many fields of science [25, 1, 17, 19, 13]. We now investigate how to obtain model parameter estimates from this posterior joint density function.

3.3 New approach - computational measure theory

We propose a new method for finding the posterior density. By using the adjoint and generalized contours, this method uses weaker assumptions than those typically required to implement a Bayesian approach involving sampling from the posterior [13, 16, 17].

We do not require a prior density on model space. Rather, we assume only that the volume measure on model space is defined by the homogeneous density $d\mu_{\mathbf{\Lambda}}(\lambda)$. We later construct a volume measure using a product space structure and assuming prior densities on model space are given. With these assumptions,

$$\sigma(\lambda, q) = k\frac{d\mu_{\mathbf{\Lambda}}(\lambda)\rho_{\mathcal{D}}(q)\delta(q - q(\lambda))}{d\mu_{\mathcal{D}}(q(\lambda))}.$$

In the examples below, the data space is linear, making $d\mu_{\mathcal{D}}(q(\lambda))$ a constant, so

$$\sigma(\lambda, q) = \frac{1}{\nu}d\mu_{\mathbf{\Lambda}}(\lambda)\rho_{\mathcal{D}}(q)\delta(q - q(\lambda)). \tag{3.3.1}$$

where ν is the new normalization constant. We are interested in the posterior density on $\mathbf{\Lambda}$ calculated from (3.3.1) by integrating over $\mathcal{D}$, which yields

$$\sigma_{\mathbf{\Lambda}}(\lambda) = \frac{1}{\nu}d\mu_{\mathbf{\Lambda}}(\lambda)\rho_{\mathcal{D}}(q(\lambda)). \tag{3.3.2}$$

Our goal is to approximate the posterior density in (3.3.2) in a cost-effective way. Our approach is measure-theoretic in the sense that we approximate the posterior density using simple functions.

Theorem 3.3.1. *Given a measurable set $A \subset \mathbf{\Lambda}$, $P(A)$ can be approximated by a simple function approximation to (3.3.2), which only requires calculations of volumes in $\mathbf{\Lambda}$.*

Algorithm 3.3.1 (Approximate Posterior Probability Measure Method).

```
Fix simple function approximation, $\rho_{\mathcal{D}}^{(M)}(q)$, to $\rho_{\mathcal{D}}(q)$
$\rho_{\mathcal{D}}^{(M)}(q)$ induces partition $\cup_{i=1}^{N(M)}[q_{i-1}, q_i)$ of $\mathcal{D}$
For each $i = 1, \dots, N(M)$, $\rho_{\mathcal{D}}^{(M)}(q)$ is constant on $[q_{i-1}, q_i)$
$\cup_{i=1}^{N(M)}[q_{i-1}, q_i)$ induces partition of generalized contours $\{A_j\}_{j=1}^{N(M)}$ of $\mathbf{\Lambda}$
Let $P_j$ denote probability of $A_j$ given by $\int_{[q_{j-1}, q_j)} \rho_{\mathcal{D}}^{(M)}(q)\, d\mu_{\mathcal{D}}(q)$
```

```
Partition $\mathbf{\Lambda}$ with small cells $\{b_i\}_{i=1}^{M'}$
for $i = 1, \dots, M'$ do
    for $j = 1, \dots, N(M)$ do
        Calculate ratio of volumes of $b_i \cap A_j$ to $A_j$, store in matrix $V_{ij}$
    end for
    Set $P(b_i)$ equal to $\sum_{j=1}^{N(M)} V_{ij} P_j$
end for
Given event $A \subset \mathbf{\Lambda}$, estimate $P(A)$ using
    • inner sums, i.e. $\sum_i P(b_i)$ for $i \in I \subset \{1, \dots, M'\}$ with $b_i \subset A$,
    • outer sums, i.e. $\sum_i P(b_i)$ for $i \in I \subset \{1, \dots, M'\}$ with $b_i \cap A \neq \emptyset$,
    • average of inner and outer sums, or
    • $\int_A \sigma_{\mathbf{\Lambda},M'}(\lambda)\, d\lambda$
```

In the absence of a set A, we still carry out the first part of Alg. 3.3.1 to obtain a *discretized* approximation of the measure P on model space. Note that there are several discretizations that take place in Alg. 3.3.1. The first discretization is the simple function approximation of the output density $\rho_{\mathcal{D}}(q)$. This discretization induces the partition of $\mathbf{\Lambda}$ by generalized contours $\{A_j\}_{j=1}^{N(M)}$.

Remark 3.3.1. *For λ restricted between any two contours induced by a subinterval of a partition of $\mathcal{D}$ as above, $q(\lambda)$ is approximately a uniformly distributed random variable.*

This discretization allows the probability of any event $A \subset \mathbf{\Lambda}$ to be calculated using a ratio of volumes as seen in the proof below.

Remark 3.3.2. *The choice of $\{b_i\}_{i=1}^{M'}$ is arbitrary, and is not necessary to the approximation of $P(A)$. We choose $\{b_i\}_{i=1}^{M'}$ in order to approximate $P(A)$, for any event $A \subset \mathbf{\Lambda}$, without carrying out the calculations in the nested loops of Alg. 3.3.1 for each new event. If we only want to compute the probability of a particular event, $A \subset \mathbf{\Lambda}$, then we skip the step of partitionining $\mathbf{\Lambda}$ by $\{b_i\}_{i=1}^{M'}$ and replace the step in the nested loop by the following: Calculate ratio of volume of $A \cap A_j$ to volume of A_j, store in vector V_j. We may then approximate $P(A)$ by $\sum_{j=1}^{N(M)} V_j P_j$.*

Proof:

Suppose that $\{q_j\}_{j=0}^{N}$ is a partition of $\mathcal{D}$ such that $q_0 < q_1 < \cdots < q_N$, and if $E_j = [q_{j-1}, q_j]$, then $\mathcal{D} = \cup_j E_j$. Let $A_j = \{\lambda \,|\, q(\lambda) \in E_j\}$. We assume that the relationship between data and model parameters has been exploited so that $\mathbf{\Lambda} = \cup_j A_j$, which is to say that we are working with appropriate data and model spaces such that any choice of model parameters $\lambda \in \mathbf{\Lambda}$ corresponds to $q(\lambda) \in \mathcal{D}$ and any $q \in \mathcal{D}$ can be mapped to from some $\lambda \in \mathbf{\Lambda}$. We have that the probability of A_j from the posterior density is given by

$$P(A_j) = \int_{A_j} \int_{\mathcal{D}} \sigma(\lambda, q)\, dq\, d\lambda = \frac{1}{\nu} \int_{A_j} \rho_{\mathcal{D}}(q(\lambda))\, d\mu_{\mathbf{\Lambda}}(\lambda).$$

Given measurable set (or an **event**) $A \subset \mathbf{\Lambda}$, we use the law of total probability to write

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

Since the A_j are induced from the partition on data space, we rewrite $P(A \,|\, A_j) = P(\lambda \in A \,|\, q(\lambda) \in E_j)$ and $P(A_j) = P(q(\lambda) \in E_j)$. If $q(\lambda) \sim$

$\mathcal{U}(E_j)$, then $\rho_{\mathcal{D}}(q(\lambda))$ is constant for $\lambda \in A_j$, so by (3.3) we have

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

Hence, $P(\lambda \in A \mid q(\lambda) \in E_j) = P(A \mid A_j)$, and this probability can be calculated from the homogeneous density on model space since it only depends on measurable sets in $\mathbf{\Lambda}$ if we assume that $q(\lambda) \sim \mathcal{U}(E_j)$ for $\lambda \in A_j$, and its value is the ratio of volume of $A \cap A_j$ to the volume of A_j. Since the prior density on data space is a nonnegative integrable function, there exists a sequence of simple functions $\left\{\rho_{\mathcal{D}}^{(M)}(q)\right\}_{M=1}^{\infty}$ with

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

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

$$\begin{aligned} \int_E \rho_{\mathcal{D}}^{(M)}(q)\, dq &= \sum_{k=1}^{2^{2M}+1} \frac{k-1}{2^M} \mu_{\mathcal{D}}(E_{M,k} \cap E) \\ &\to P_{\mathcal{D}}(E) \text{ as } M \to \infty. \end{aligned}$$

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

We can estimate $P(A)$ using the inner and outer sums described by Alg. 3.3.1 since $P(A) = \sup\{P(K) : K \subset A,\ K \text{ compact}\}$ and $P(A) = \inf\{P(U) : A \subset U,\ \text{U open}\}$.

Remark 3.3.3. *In Alg. 3.3.1, the probabilities of cells b_i are computed exactly as the set A in the proof above, and the ratio of the volume of b_i to the sets A_j is determined using the generalized contours.*

Note that as we refine the partition $\{I_j\}$ on data space, which in turn refines the partition $\{A_j\}$ on model space, we need to also refine the mesh that defines the partition $\{b_i\}$ on model space. The reason is that we assign a probability $P(b_i)$ to each cell b_i that in essence approximates the posterior density on model space by the simple function

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

If the partition $\{b_i\}$ remains fixed as the approximation of $\rho_D(q)$ by simple functions is refined by the partition $\{I_j\}$, then the approximation of the posterior density on model space fails to improve even though we might obtain slightly better estimates for $P(b_i)$ on each b_i.

Note that the calculation of volumes is a computational geometry problem and the technical details are covered in Chapter 7.

3.3.1 Example

We present results for the finite-dimensional nonlinear system of equations given by

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

where λ_1 and λ_2 are the parameters. Geometrically, solutions $x = (x_1, x_2)^T$ to the system represent intersections of the hyperbola and ellipse. The quantity of interest is the second component of the solution in the first-quadrant, i.e., $q(\lambda) = q(x(\lambda)) = x_2 = \langle x, \psi \rangle$, where $\psi = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$. According to (2.3.1), the adjoint problem is

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

where $\mu = (\mu_1, \mu_2)^T$ and $\mathbf{y} = (y_1, y_2)^T$ are the reference parameter and reference solution for the forward problem. In order to capture interesting behavior of the solution with respect to the parameter values, we choose $\mathbf{\Lambda} = [.79, .99] \times [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}]$ based on analysis of the forward problem [20]. We use 6 uniformly spaced mesh points in both the λ_1 and λ_2 directions of $\mathbf{\Lambda}$ to create cells $\{B_i\}_{i=1}^{25}$ that partition Λ, and choose the centroid of each cell as the reference parameter $\mu_i = (\mu_{1,i}, \mu_{2,i})^T$ in that cell and solve the forward problem to obtain reference solutions $y_i = (y_{1,i}, y_{2,i})^T$ at these points, and then solve for the generalized Green's vector $\phi_i = (\phi_{1,i}, \phi_{2,i})^T$ at the reference point (μ_i, y_i). According to (2.4.1), we obtain a global piecewise-linear approximation $\tilde{q}$ to q defined as

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

We first assume that the data $\tilde{q}(\lambda)$ is a random variable with normal distribution on the data space defined by $\tilde{q}(\mathbf{\Lambda})$ (Fig 3.2). We assume that the model space $\mathbf{\Lambda}$ is linear and the homogeneous measure is the Lebesgue measure. When implementing a Monte Carlo sampling algorithm for the posterior density, this corresponds to assuming independent *prior* densities that are uniform for each parameter. We implement the algorithm from the previous section to calculate $P(b_i)$ for small cells for each fine partition of $\mathbf{\Lambda}$ and determine the probabilities of events $A \subset \mathbf{\Lambda}$ (Fig 3.3).

3.4 Classical approach - accept/reject sampling

As way of comparison, we describe the classical approach to studying the posterior density by generating independent identically distributed samples from the distribution defined by this density. We present some of the technical difficulties arising from this approach.

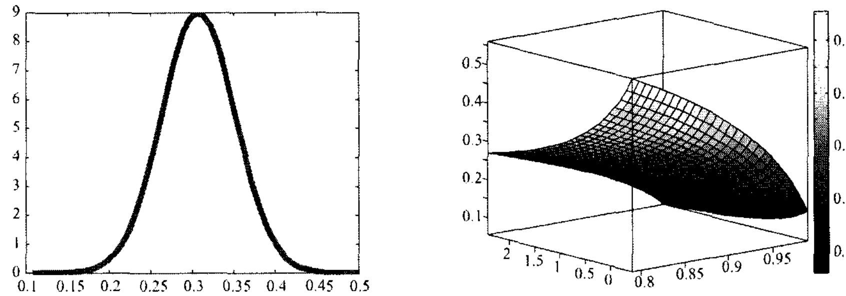

Figure 3.2: Left: Uncertainty of output is modeled as a random variable with a Normal distribution. Right: Map $q : \mathbf{\Lambda} \to \mathbb{R}$.

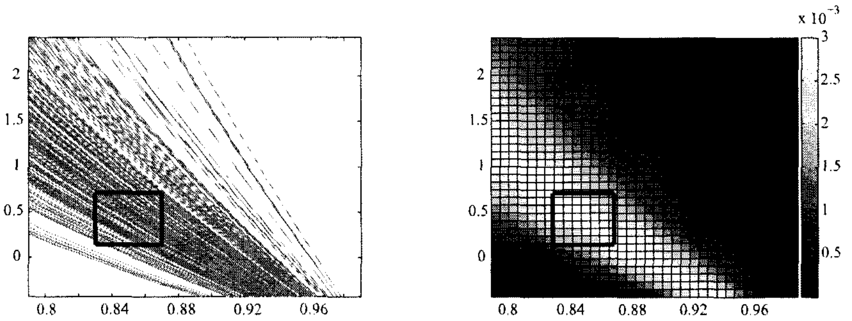

Figure 3.3: Left: We determine which contours are contained in event $A \subset \mathbf{\Lambda}$ and use a joint measure to determine how much of each contour is inside the cell, and then use this with the probability of these contours being selected to determine the probability of parameters being chosen from this cell. Right: Using a normalized Lebesgue measure as the joint measure, we estimate the probabilities of small cells and can use an inner and outer estimate to obtain an approximation of the probability of parameters being chosen inside the event A

We summarize the method for obtaining the posterior density function. There is an additional assumption in this approach of the prior density, which is needed for sampling. Note that in our approach, a joint measure is used to measure volumes in $\mathbf{\Lambda}$. Given

- the prior joint probability density, $\rho(\lambda, q)$, on the joint manifold $\mathcal{X}$
- the model $M(Y, \lambda)$ with solution $Y = Y(\lambda)$
- the linear functional $q(\theta) = q(Y(\theta))$
- the theoretical joint probability density, $\Theta(\lambda, q)$ on the joint manifold $\mathcal{X}$

determine the *posterior* joint density function, $\sigma(\lambda, q)$, determined from the conjunction of the prior and theoretical joint densities as

$$\sigma(\lambda, q) = k\frac{\rho(\lambda, q)\Theta(\lambda, q)}{d\mu(\lambda, q)},$$

where k is a normalizing constant. We define the **posterior density in model space** as the marginal density function,

$$\sigma_{\mathbf{\Lambda}}(\lambda) = \int_{\mathcal{D}} \sigma(\lambda, q)\, dq.$$

We say that the posterior density in model space is obtained by integrating out the information we have about the data. We rewrite the posterior density in model space using (3.2.1) and (3.2.2) as

$$\sigma_{\mathbf{\Lambda}}(\lambda) = k\rho_{\mathbf{\Lambda}}(\lambda) \int_{\mathcal{D}} \frac{\rho_{\mathcal{D}}(q)\theta(q \mid \lambda)}{d\mu_{\mathcal{D}}(q)}\, dq.$$

The posterior density in model space is paramount in describing parameter estimates, but the integral involved in this calculation is generally viewed as analytically intractable. There exists methods, e.g. maximum-likelihood, to generate estimates of certain characteristics of $\sigma_{\mathbf{\Lambda}}(\lambda)$, e.g. the mean, but these estimates generally do a poor job of describing the distribution of parameter values, e.g. consider a multimodal distribution. Thus, we desire a more detailed description of the distribution defined by $\sigma_{\mathbf{\Lambda}}(\lambda)$.

Monte Carlo sampling provides a way to generate samples, or **realizations** of the random vector, from the distribution defined by $\sigma_{\mathbf{\Lambda}}(\lambda)$ without having to calculate the integral. We let $\{\lambda_{(i)}\}_{i=1}^{N}$ represent the first N samples generated via Monte Carlo from the distribution defined by $\sigma_{\mathbf{\Lambda}}(\lambda)$. The distribution of the samples converges to the distribution defined by $\sigma_{\mathbf{\Lambda}}(\lambda)$ by the Central Limit Theorem at a rate proportional to $1/\sqrt{N}$. Thus, the first issue to consider when using Monte Carlo methods is the slow convergence of the method assuming samples are independent.

The Gibbs and Metropolis algorithms are the two main algorithms used in generating samples from probability distributions. We consider the Metropolis algorithm because it is widely used [17, 15, 25, 21].

3.4.1 Metropolis algorithm

The Metropolis algorithm is based upon the Fundamental Theorem of Simulation.

Theorem 3.4.1 (Fundamental Theorem of Simulation [21]). *Simulating*

$$X \sim f(x)$$

is equivalent to simulating

$$(X, U) \sim \mathcal{U}\left\{(x, u)\, :\, 0 < u < f(x)\right\}.$$

The Fundamental Theorem of Simulation is the basis of all accept-reject algorithms for generating random samples from a *target* density $f(x)$. It is often quite difficult to generate random numbers directly from a given arbitrary density $f(x)$, but it is relatively easy to generate samples in the joint density of X and U by generating samples in an even larger easier to generate set and accepting only those samples that satisfy the constraint $0 < u < f(x)$. This is in fact the classical accept-reject algorithm. More sophisticated accept-reject algorithms exist (e.g., the envelope accept-reject algorithm), but further exposition on this subject is unnecessary for our purposes.

The Metropolis algorithm generates samples from the target density $f(x)$ by generating proposed samples via random walks from a Markov chain represented by a conditional density $w(y|x)$, which is typically taken to be symmetric (i.e., $w(y|x) = w(x|y)$), and accepting a proposed sample based on criteria from the Fundamental Theorem of Simulation. The algorithm produces a Markov Chain Monte Carlo (MCMC) set of samples from a given density.

Algorithm 3.4.1 (Metropolis Sampling Method). *Given*

- *target density* $f(x)$
- *conditional density* $w(y|x)$ *producing a Markov chain*
- *initial sample* X_0

```
for $j = 1, \ldots, N$ do
    Generate proposed sample $Y \sim w(y|X_{j-1})$
    Generate $U \sim \mathcal{U}(0,1)$
    if $0 < U < \frac{f(Y)w(X_{j-1}|Y)}{f(X_{j-1})w(Y|X_{j-1})}$ then $X_j = Y$
    *else*$X_j = X_{j-1}$
    *end if*
*end for*
```

It is left to the user to determine appropriate conditional densities $w(\cdot|x)$. The density $w(\cdot|x)$ should be easy to simulate from so that Alg. 3.4.1 can be implemented as efficiently as possible. Often $w(\cdot|x)$ is chosen to be symmetric, i.e. $w(y|x) = w(x|y)$, and Alg. 3.4.1 is interpreted as generating samples of $f(x)$ via a random walk. If we choose the conditional density w to be symmetric, then $w(X_i|Y) = w(Y|X_i)$. In this case, the algorithm automatically accepts any "move to higher probability" and with probability $f(Y_i)/f(X_i)$ accepts a move to lower probability. Under some natural conditions on $w(\cdot|x)$, Alg 3.4.1 produces a set of samples $\{X_j\}_{j=0}^N$ with the property that as $N \to \infty$, the sample distribution converges to the stationary distribution that is equal to the distribution defined by $f(x)$. A minimal necessary condition [21] that the stationary distribution of the Markov chain is given by $f(x)$ is that

$$\text{supp } f \subset \bigcup_{x \in \text{supp } f} \text{supp } w(\cdot|x). \tag{3.4.1}$$

This condition ensures that the chain has the necessary properties (irreducibility, positive recurrence, aperiodicity, and ergodicity) so that the stationary distribution is $f(x)$.

Remark 3.4.1. *Suppose $\{X_j\}_{j=0}^N$ is a set of random samples generated according to Alg. 3.4.1. Care must be taken in using these samples to calculate statistics of the random process defined by $f(x)$. This should be clear since the set of samples is not necessarily a set of independent identically* *distributed (iid) samples. For instance, we might have repeated occurences of the same value due to repeated rejections. The lack of independence implies that much of the theory devoted to sequences of independent (and iid) samples doesn't directly apply. For instance, none of the various laws of large numbers, i.e. the weak law, Kolmogorov's strong law, and Khinchine's strong law, can be used to discuss the convergence of the sample mean to the true mean. Similarly, the hypothesis of the Central Limit Theorem is not satisifed.*

Using $\{X_j\}_{j=0}^{N}$ to approximate a statistic such as the mean of $f(x)$, a weighted mean is used where each sample is associated with a weight chosen as a function of the number of successive rejections [21]. For more information about the convergence of a Markov chain to its stationary distribution, we refer the interested reader to [21, 15, 16]

We turn our attention to the problem of sampling the posterior density defined by the inverse problem. The posterior density can be written as [25]

$$\sigma_{\mathbf{\Lambda}}(\lambda) = k\rho_{\mathbf{\Lambda}}(\lambda)\int_{\mathcal{D}} \frac{\rho_{\mathcal{D}}(q)\theta(q\,|\,\lambda)}{d\mu_{\mathcal{D}}(q)}\,dq = k\rho_{\mathbf{\Lambda}}(\lambda)L(\lambda),$$

where

$$L(\lambda) := \int_{\mathcal{D}} \frac{\rho_{\mathcal{D}}(q)\theta(q\,|\,\lambda)}{d\mu_{\mathcal{D}}(q)}\,dq$$

is defined to be the likelihood function, and $\rho_{\mathbf{\Lambda}}(\lambda)$ is the prior density on $\mathbf{\Lambda}$. There is a functional relation $q = q(\lambda)$ between output data (the quantity of interest) and parameters λ, so $\theta(q\,|\,\lambda) = \delta(q - q(\lambda))$ and

$$L(\lambda) = \rho_{\mathcal{D}}(q(\lambda)).$$

We use Alg. 3.4.1 to generate samples from target density $\sigma_{\mathbf{\Lambda}}(\lambda)$. Let $w(\cdot|\lambda)$ be given with property (3.4.1). Observe that

$$\frac{\sigma_{\mathbf{\Lambda}}(Y)w(X_{j-1}|Y)}{\sigma_{\mathbf{\Lambda}}(X_{j-1})w(Y|X_{j-1})} = \frac{L(Y)}{L(X_{j-1})}\left\{\frac{\rho_{\mathbf{\Lambda}}(Y)w(X_{j-1}|Y)}{\rho_{\mathbf{\Lambda}}(X_{j-1})w(Y|X_{j-1})}\right\}. \tag{3.4.2}$$

Remark 3.4.2. *Suppose (3.4.1) holds with $f = \rho_{\mathbf{\Lambda}}$. We write*

$$\frac{\rho_{\mathbf{\Lambda}}(Y)w(X_{j-1}|Y)}{\rho_{\mathbf{\Lambda}}(X_{j-1})w(Y|X_{j-1})} = \frac{\tilde{w}(X_{j-1}|Y)}{\tilde{w}(Y|X_{j-1})}. \tag{3.4.3}$$

Using (3.4.3) in (3.4.2), we see that generating samples of $\sigma_{\mathbf{\Lambda}}(\lambda)$ is equivalent to using $L(\lambda)$ as the target density and generating proposed samples according to $\tilde{w}(\cdot|\lambda)$ in Alg. 3.4.1. Moreover, generating samples from $\tilde{w}(\cdot|\lambda)$ is equivalent to using $\rho_{\mathcal{D}}(\lambda)$ as the target density and generating proposed samples according to $w(\cdot|\lambda)$ in Alg. 3.4.1. Therefore, generating samples of $\sigma_{\mathbf{\Lambda}}(\lambda)$ is equivalent to using $L(\lambda)$ as the target density and generating proposed samples according to $\rho_{\mathcal{D}}(\lambda)$ in Alg. 3.4.1.

We assume that we are able to obtain as many samples of the prior density $\rho_{\mathcal{D}}(\lambda)$ as desired. Often the structure of $\rho_{\mathcal{D}}(\lambda)$ is simple enough that samples can either be simulated directly from $\rho_{\mathcal{D}}(\lambda)$ or by using simple methods, e.g. basic accept/reject algorithms [21, 25]. Thus, not only is it often easy to obtain samples of the prior density $\rho_{\mathcal{D}}(\lambda)$, these samples can generally be assumed to be iid. For example, common choices of prior densities represent the components of random vector λ as either uniform or normal random variables for which many pre-packaged random number generators exist.

With the assumption that we have easy access to iid samples of the prior density, we can alter Alg. 3.4.1 to iterate only upon acceptance of a sample. Thus, we produce a collection of iid samples of the posterior density. It is with such a set of iid samples generated as described above that we numerically compare the methodologies in a future paper [5].

Remark 3.4.3. *We are interested in answering the following question. Given event $A \subset \mathbf{\Lambda}$, what is $P(A)$? By only iterating in Alg. 3.4.1 upon* *acceptance, we produce a collection of iid samples of the posterior density and binning leads to an approximation of $P(A)$. Suppose we use Alg. 3.4.1 as stated, i.e. iterate upon rejection, then as discussed above, weights must be assigned to samples as a function of successive rejections [21], which is equivalent to using less samples, so in a sense, there is no cost savings by iterating upon rejection.*

Remark 3.4.4. *We assume the dimension of $\mathbf{\Lambda} \subset \mathbb{R}^d$ is small enough so that the above remark is valid. For large d, the "emptiness" of high-dimensional spaces can lead to unacceptably large rejection rates. Often [21, 25, 15, 16] the solution to this problem is not to use iid samples of $\rho_{\mathbf{\Lambda}}(\lambda)$ in Alg. 3.4.1, but use a random walk to explore $\mathbf{\Lambda}$ in small enough increments so that changes in the likelihood function are small, which leads to an acceptable rejection rate. However, the random walk must also have a step size large enough so that $\mathbf{\Lambda}$ is searched rapidly enough for Alg. 3.4.1 to be considered efficient. In high-dimensional spaces, there can be disconnected sets of high probability, i.e. near zero values of $\sigma_{\mathbf{\Lambda}}(\lambda)$ for λ not in these sets, in which case these sets must be located and Markov chains initiated in each set separately. This is a very hard problem and is problem specific. Furthermore, the samples generated are clearly dependent and there is typically too much spatial correlation between samples. The solution to this problem is typically to accept only a subset of accepted samples by perhaps fixing some positive integer m, and considering the subset $\{X_{n_k}\}$ of samples from Alg. 3.4.1 where $n_k = km$. This reduces the correlation between samples. How to go about choosing such a subset is problem specific.*

From the above remarks, we note some fundamental issues involved in MCMC sampling that effect the efficiency of Alg. 3.4.1 and should be considered.

Remark 3.4.5. *We now reconcile the assumption of prior densities in the sampling methodology to the assumption of an underlying measure in our new methodology and specifically to Alg. 3.3.1. In applying the computational measure-theoretic algorithm, we use the underlying volume measure on* $\mathbf{\Lambda}$*, and it is this measure that we use to calculate the ratios of volumes described in Alg. 3.3.1. There is no notion of prior density required in this approach. We can incorporate a prior density into Alg. 3.3.1 by interpreting a given prior density as defining a measure absolutely continuous with respect to the underlying measure on* $\mathbf{\Lambda}$*. Thus, a prior density is interpreted as assigning a different geometry to the contours used in the calculations of Alg. 3.3.1. This interpretation implies that a prior probability density is used as a different measure of the volumes described in Alg. 3.3.1. Similarly, if we only want to assume a notion of measure on* $\mathbf{\Lambda}$*, then we assign the prior density to be the normalized measure on* $\mathbf{\Lambda}$*, which is often done in standard Bayesian implementation.*

Remark 3.4.6. *The assumption of a prior density, considered as uninformative or not, is a constant source of debate among the statistical community. Our assumption of a measure on parameter space is dictated by experimental observation. In determining the measure, we seek to answer the question, "How do we measure distance in parameter space?" Often times this question has an obvious answer of Lebesgue measure. This is consistent with standard Bayesian implementation in that there exists an underlying volume measure typically assumed to be Lebesgue. There must be an underlying volume measure in order to specify a prior density. Specifically, consider the fact that probability theory and density functions are built upon the foundations of measure theory [2, 14]. A prior density is necessary* *for standard Bayesian implementation because sampling methods require a probability density. Since we perform no sampling, we do not require a prior density.*

3.4.2 Sampling the posterior probability distribution - an example

Recall that the posterior density in model space is

$$\sigma_{\mathbf{\Lambda}}(\lambda) = k\rho_{\mathbf{\Lambda}}(\lambda) \int_{\mathcal{D}} \frac{\rho_{\mathcal{D}}(q)\theta(q \mid \lambda)}{d\mu_{\mathcal{D}}(q)}\, dq.$$

It is typically easy to generate samples from the prior density in model space $\rho_{\mathbf{\Lambda}}(\lambda)$. We use this as the conditional density in the Metropolis algorithm. We define the **likelihood** function as

$$L(\lambda) = \int_{\mathcal{D}} \frac{\rho_{\mathcal{D}}(q)\theta(q \mid \lambda)}{d\mu_{\mathcal{D}}(q)}\, dq.$$

According to the Metropolis algorithm, given sample $\lambda_{(i)}$, we accept a proposed sample γ generated from $\rho_{\mathbf{\Lambda}}(\lambda)$ as sample $\lambda_{(i+1)}$ if $U \sim \mathcal{U}(0,1)$ and $0 < U < L(\gamma)/L(\lambda_{(i)})$. Continuing in this way by generating proposed samples from the prior density in model space and accepting or rejecting these samples using the likelihood function, we sample from the posterior density in model space. There is an expense in generating these N samples. It is not uncommon for rejection ratios to be higher than 70% even for "good" choices of prior densities in model space. The likelihood function is a measure of how well the parameters "fit" the solution of the model to the data q. Usually much trial and error goes into the choice of a prior distribution that generates parameter samples that "fit" the solution so that the rejection ratio is acceptable. As an example, consider the forward problem where

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

where λ_1, λ_2 are independent identically distributed $N(0, 1/25)$ random variables. In this case, $q(\lambda)$ is a random variable with $N(0, 2/25)$ distribution. For the inverse problem, we begin with the measurements on $q(\lambda)$ that define a $N(0, 2/25)$ distribution and seek to determine the prior distribution on $\mathbf{\Lambda}$ and use this with the model to obtain the posterior density on $\mathbf{\Lambda}$. If we find a distribution of samples on $\mathbf{\Lambda}$ that generates $q(\lambda)$ according to a $N(0, 2/25)$ distribution, then we accept this as a solution to the inverse problem and call the density of these samples a posterior density. Figures 3.4 - 3.10 show samples and the approximate distributions generating these samples obtained by kernel density estimation, and Figure 3.11 shows the approximate density of $q(\lambda)$ for all of these different model parameter distributions. Figures 3.11 shows that all these vastly different collections of samples of inputs generate the same output. These figures show that there are many distinct posterior densities. This problem arises for several reasons. We often do not have adequate information to select a "good" prior density on model space [13]. A poor choice of prior density on model space can lead to a nonconvergent Markov Chain of samples. In order for our method to produce these results, we require different volume measures, which is exactly how we interpret the use of prior densities as discussed above.

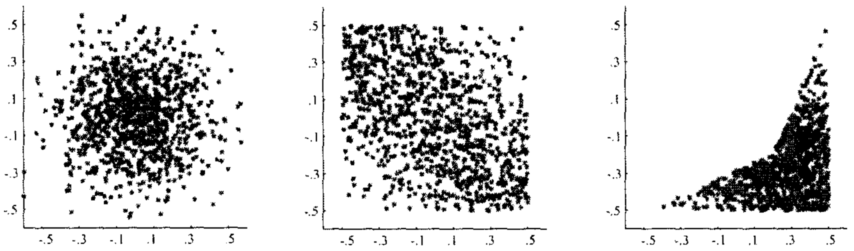

Figure 3.4: Left: 1000 random samples of parameters chosen as i.i.d. random variables with $N(0, 1/25)$ distributions. Middle: 1000 random samples of parameters chosen with respect to the unnamed density denoted $\rho_{\Lambda,1}(\lambda)$. Right: 1000 random samples of parameters chosen with respect to the unnamed density denoted $\rho_{\Lambda,2}(\lambda)$

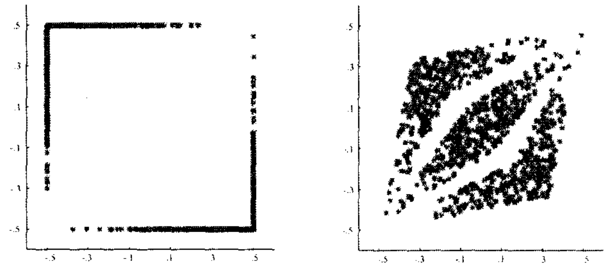

Figure 3.5: Left: 1000 random samples of parameters chosen with respect to the unnamed density denoted $\rho_{\Lambda,3}(\lambda)$. Right: 1000 Random samples of parameters chosen with respect to the unnamed density denoted $\rho_{\Lambda,4}(\lambda)$

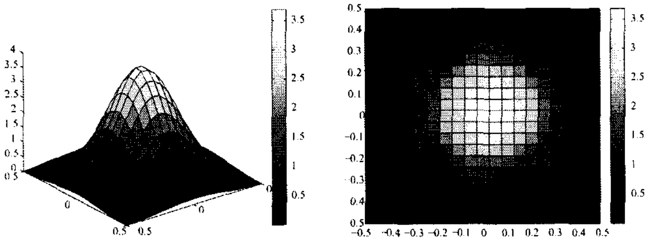

Figure 3.6: Kernel density estimate of the joint distribution of parameters sampled as i.i.d random variables with $N(0, 1/25)$ distributions

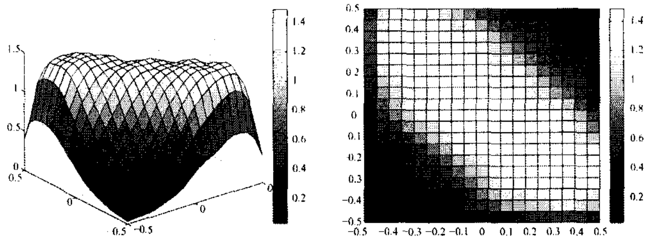

Figure 3.7: Kernel density estimate of the joint distribution of parameters sampled with respect to the density $\rho_{\mathbf{\Lambda},1}(\lambda)$

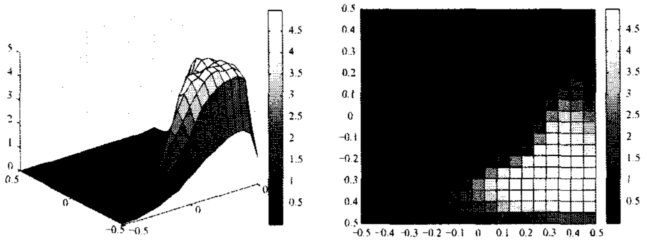

Figure 3.8: Kernel density estimate of the joint distribution of parameters sampled with respect to the density $\rho_{\mathbf{\Lambda},2}(\lambda)$

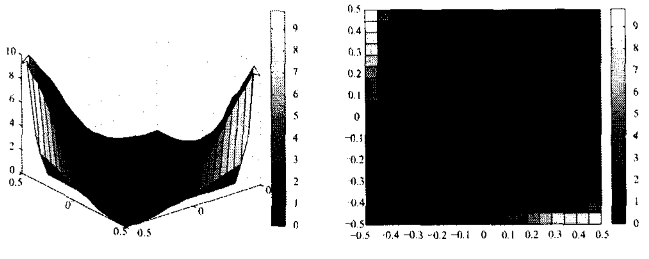

Figure 3.9: Kernel density estimate of the joint distribution of parameters sampled with respect to the density $\rho_{\mathbf{\Lambda},3}(\lambda)$

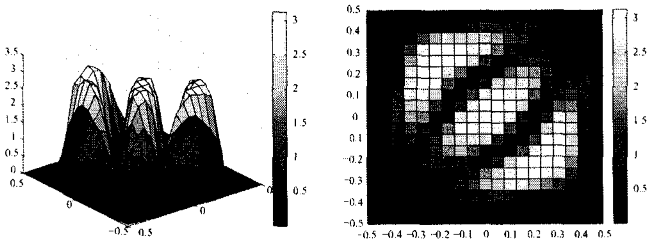

Figure 3.10: Kernel density estimate of the joint distribution of parameters sampled with respect to the density $\rho_{\mathbf{\Lambda},4}(\lambda)$

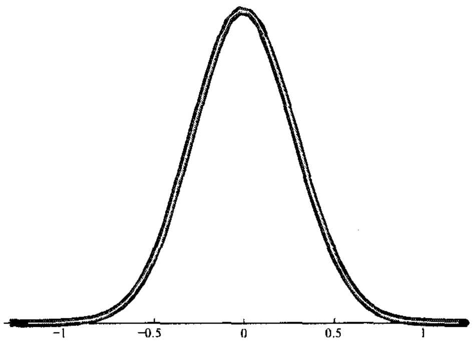

Figure 3.11: Kernel density estimate of $q(\lambda)$ with respect to Normally distributed parameters and parameters from $\rho_{\mathbf{\Lambda},1}(\lambda)$, $\rho_{\mathbf{\Lambda},2}(\lambda)$, $\rho_{\mathbf{\Lambda},3}(\lambda)$, and $\rho_{\mathbf{\Lambda},4}(\lambda)$. The kernel density estimates for $q(\lambda)$ for the different input distributions are identical.

Chapter 4

APPLICATIONS

4.1 Regions of high probability and testing priors

If we ignore model uncertainties, then $\Theta(\lambda, q) = \delta(q - q(\lambda))d\mu_{\mathbf{\Lambda}}(\lambda)$, and we have

$$\sigma(\lambda, q) = k\frac{\rho(\lambda, q)\delta(q - q(\lambda))}{d\mu_{\mathcal{D}}(q(\lambda))}.$$

Suppose there was an interval $I = [q_1, q_2] \subset \mathcal{D}$ of high probability, and consider the set in model space $A = \{\lambda \,|\, q(\lambda) \in I\}$. If A is not in the support of the prior density on model space, then the Monte Carlo method does not generate parameter samples that give model solutions in I. If A is in the support of the prior density on model space, but $\int_A \rho_{\mathbf{\Lambda}}(\lambda)\, d\lambda$ is small, then the Monte Carlo method either does not generate enough proposed samples from A, or it simply fails to converge. This is *not* a problem of Bayesian inference, but rather with the standard implementation of sampling from a prior density. Consider the fact that as the dimension of model space increases, the space itself becomes more "empty." This is illustrated by the example of inscribing a hypersphere in a hypercube [25]. The ratio of the volume of the hypersphere to the hypercube is given by

$$\frac{\pi^{n/2}}{2^{n-1}n\Gamma(n/2)},$$

which rapidly decreases to zero as the dimension n increases. Thus, given uniform priors on the hypercube, the probability of hitting points inside of the hypersphere goes to zero as the dimension increases. There are additional problems encountered with sampling methods. The problem of finding and sampling from the regions of highest problem is generally viewed as the most difficult problem and it is said that the particular *geometry* of the problem may be the key to solving this problem [25].

The generalized contours prove very useful in finding regions of high probability. We view any prior knowledge as simply imposing a certain geometry to the contours. Consider again $q(\lambda) = \lambda_1 + \lambda_2$, where $\mathbf{\Lambda} = [0, 1] \times [0, 1]$. Fig 4.1 shows the generalized contours for 500 samples of $q(\lambda)$ taken from a $N(0, 2/25)$ distribution along with the TP and the intersections of contours on the TP. Where the contours intersect the TP most densely corresponds to a region of high probability in the space of contours. This information might prove useful when selecting prior densities since it is determined entirely from the known output and model. Consequently, iid samples generated from a given prior density are on contours, and we can use the derivatives obtained from the adjoint sensitivity analysis to trace these values along the contours and find the intersection points on the TP. Using a 1-D statistical test, such as the Kolmogorov-Smirnov test, we can immediately determine if the prior density "samples the contours" correctly, i.e. if passing the prior density through the model generates the observed density on the quantity of interest. This is done without a single evaluation of the model. In fact, the densities of Fig 3.7-3.10 were created to sample the contours identically as the density shown in Fig 3.4.

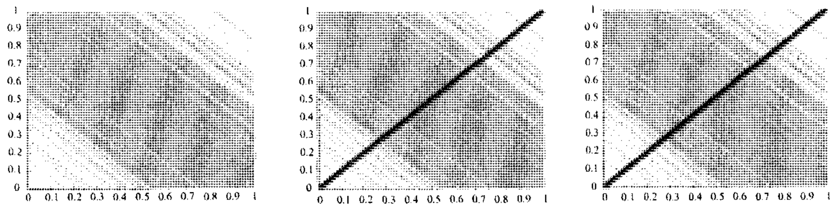

Figure 4.1: Left: Generalized contours from 500 samples of $q(\lambda) = \lambda_1 + \lambda_2$ generated from a $N(0, 2/25)$ distribution. Middle: The TP intersects each contour once and goes from the minimum of $q(\lambda)$ in the lower left corner to the maximum of $q(\lambda)$ in the upper right corner of the plot. Right: Intersections of contours on the TP are marked with a star and can be used to index the inverses and determine a unique distribution of the contours on the TP using any consistent indexing scheme

4.2 Examples

We first consider constrained geometric optimization problems with three and four parameters as the initial examples. We conclude the examples with a nonlinear ordinary differential equation with two parameters, and a nonlinear elliptic partial differential equation with two parameters.

4.2.1 Three-Parameter Constrained Optimization

We minimize the distance to the point $(1, -1, 1)$ subject to the constraint that the point lies on the surface $g = 4$, where

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

Geometrically, the parameters determine the shape of the ellipsoid that defines the constraint. Using the method of Lagrange multipliers we set up a system of nonlinear equations with four state variables and three parameters. We take the quantity of interest as the first state variable, which geometrically is interpreted as the first spatial coordinate in the solution to the constrained minimization problem. We set $\mathbf{\Lambda} = [.35, .65] \times [.28, .52] \times [.42, .78]$ and construct a piecewise-linear approximation using 125 points in $\mathbf{\Lambda}$. Assuming a normal distribution on $q(\lambda)$ and taking the joint measure on $\mathbf{\Lambda}$ to be a normalized Lebesgue measure, we use 3375 small cells $\{B_i\}$ in the algorithm to discretize the posterior probability measure on $\mathbf{\Lambda}$ and plot the probabilities at the mid-point of each cell with the color of the point determined by the probability of the small cell (Fig 4.2-4.3).

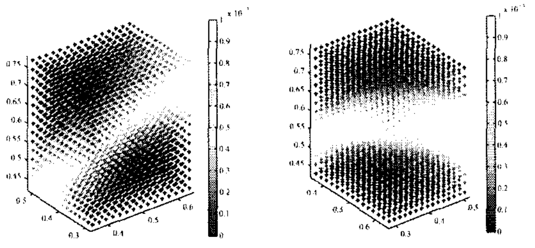

Figure 4.2: Using $15 \times 15 \times 15$ small cells, we estimate the probability of parameters selected from the cells using a normalized Lebesgue joint measure. Left: Standard 3-D view. Right: Standard 3-D view rotated 90 degrees clockwise.

4.2.2 Four-Parameter Constrained Optimization

We minimize the distance to the point $(5, 5, 5)$ subject to the constraints that the point lies on the intersection of the surfaces $g = 1$ and $h = 0$, where

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

Geometrically, $g = 1$ defines a hyperboloid of one sheet and $h = 0$ defines a plane through the origin, and the intersection of the two constraints is a

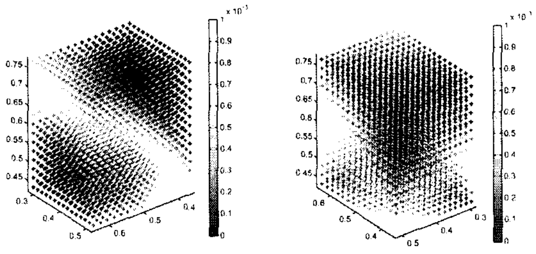

Figure 4.3: Using 15 × 15 × 15 small cells, we estimate the probability of parameters selected from the cells using a normalized Lebesgue joint measure. Left: Standard 3-D view rotated 180 degrees clockwise. Right: Standard 3-D view rotated 270 degrees clockwise.

closed curve. Using the method of Lagrange multipliers we set up a system of nonlinear equations with five state variables and four parameters. We take the quantity of interest as the first state variable, which geometrically is interpreted as the first spatial coordinate in the solution to the constrained minimization problem. We set $\mathbf{\Lambda} = [1.4, 2.6] \times [.7, 1.3] \times [1.4, 2.6] \times [.35, .65]$ and construct a piecewise-linear approximation using 750 points in $\mathbf{\Lambda}$. Assuming a normal distribution on $q(\lambda)$ and taking the joint measure on $\mathbf{\Lambda}$ to be a normalized Lebesgue measure, we use 60750 small cells $\{b_i\}$ in the algorithm to discretize the posterior probability measure on $\mathbf{\Lambda}$. By treating the fourth parameter as time, we can take "snapshots" for fixed λ_4 values of the approximated posterior probability density function in a similar way as for the three-parameter example above (Fig 4.4). The algorithm for approximating the posterior probability density allows us to locate regions of high probability by sorting through the probability of the fine cells $\{b_i\}$. We can rank order these cells and determine any cells of high probability within close proximity to other cells of high probability (Table 4.1).

Often sampling methods are used to search for the maximum-likelihood point. It is our opinion that this should be avoided since the MCMC methods that produce this point simply search for the maximum of the posterior density, and the probability of a single point with a continuous density function is zero. We approximate probabilities of events. If the goal is to produce a maximum-likelihood estimate, then our method can be used to search for regions of highest probability where it is perhaps more likely that a maximum of the posterior density occurred, replacing the simulated annealing typically used in a maximum-likelihood search algorithm. If we let the events $\{b_i\}$ become small, under a smoothness assumption, the probabilities of these events are related to the maximum-likelihood estimate.

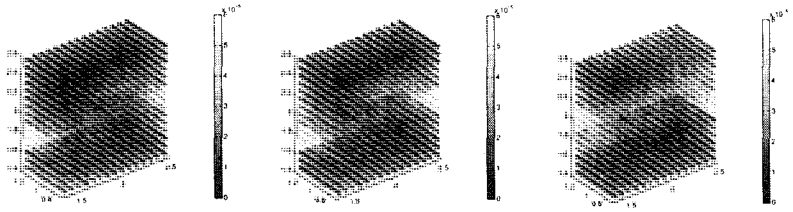

Figure 4.4: Using 15 × 15 × 15 × 18 small cells, we estimate the probability of parameters selected from the cells using a normalized Lebesgue joint measure. Left: The fourth parameter is set at its minimum value. Middle: The fourth parameter is set at its mid-point value. Right: The fourth parameter is set at its maximum value. Notice how the probabilities vary in space, i.e. with respect to the first three parameters, as we vary time, i.e. the fourth parameter

4.2.3 Two-Parameter ODE

We now study the nonlinear ordinary differential equation

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

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

Table 4.1: The ten small cells with highest probability are listed in ascending order in the first column. The second column gives the dimensions and location of these cells. We can use this information to determine where the largest regions of highest probability are located in a high-dimensional parameter space

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

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

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

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

where $A(t) := f'(y(t; \mu))$ is the Jacobian of $f = \lambda_1 \sin(\lambda_2 x)$ evaluated at $y(t; \mu)$, μ is a reference parameter, and $y(t; \mu)$ is the solution to (4.2.3) for this reference parameter. Compare this to (2.3.1). Using substitution, integration by parts, and Taylor's theorem, we arrive at a linear approximation to $q(\lambda)$ for parameters near μ, and analagous to the finite dimensional case, we obtain a global piecewise-linear approximation to $q(\lambda)$ over $\mathbf{\Lambda} = [.8, 1.2] \times [.1, \pi - .1]$ (Fig 4.5).

Remark 4.2.1. *There can be substantial error in the reference solutions and gradients used to linearize the problem, and the effect of this error on the contours is studied in later chapters.*

Given a distribution on $q(\lambda)$ and a joint measure on $\mathbf{\Lambda}$, we can estimate the posterior density as in the above examples (Fig 4.5).

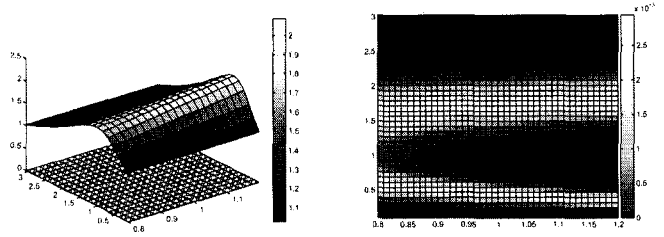

Figure 4.5: Left: Global piecewise-linear approximation to $q(\lambda)$. The cells in $\mathbf{\Lambda}$ illustrate the coarse discretization of this space for the forward problem of obtaining a piecewise-linear approximation and the circles in each cell indicate the reference parameter used to linearize $q(\lambda)$ in that cell. The piecewise-linear contours to this surface are used for the inverse problem exactly as with the nonlinear systems of equations examples given above. Right: Assuming a normal distribution of $q(\lambda)$, we use a grid of 40 × 40 small cells to estimate the posterior density function

4.2.4 Two-Parameter PDE

We now study a nonlinear partial differential equation

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

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

$$q(\lambda) = \langle u, \psi \rangle = \int_\Omega u(x,y)\psi(x,y)\, dA,$$

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

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

where $A := f'(w(x,y;\mu);\mu)$ is the Jacobian of $f = \lambda_1 \exp(\lambda_2 u)$ evaluated at $w(x,y;\mu)$, μ is a reference parameter, and $w(x,y;\mu)$ is the solution to (4.2.4) for this reference parameter. Using substitution, the weak form of (4.2.4), and Taylor's theorem, we arrive at a linear approximation to $q(\lambda)$ for parameters near μ, and just as with the previous examples, we obtain a global piecewise-linear approximation to $q(\lambda)$ over $\mathbf{\Lambda} = [.95, 1.05] \times [-.1, .1]$ (Fig 4.6). Given a distribution on $q(\lambda)$ and a joint measure on $\mathbf{\Lambda}$, we proceed exactly as with the other examples to produce an estimate of the posterior density (Fig 4.6). Similar to the ordinary differential equation above, the error in the reference solutions and gradients used to linearize the problem might prove significant in biasing the contours, and the effect of this error on the contours is studied in later chapters.

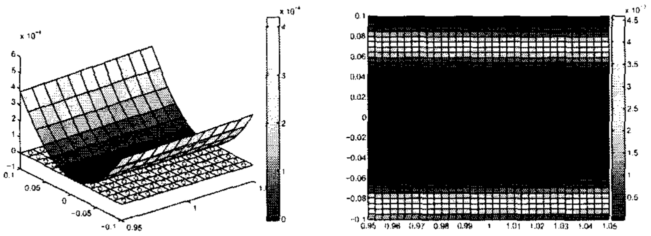

Figure 4.6: Left: Global piecewise-linear approximation to $q(\lambda)$. We used a 11×13 grid of coarse cells to discretize $\mathbf{\Lambda}$ and used the mid-point of each cells as the reference parameter in that cell. We can determine posterior density estimates as with all the other examples using the same algorithm now that we have reference solutions and gradient information. Right: Assuming a normal distribution of $q(\lambda)$, we use a 33×39 grid of small cells to estimate the posterior density function

Chapter 5

ERROR ANALYSIS FOR INVERSE SENSITIVITY PROBLEMS WITH ORDINARY DIFFERENTIAL EQUATIONS

We study three sources of computational error effecting the posterior density.

- There is a linearization error in the representation of the surface $q(\lambda)$, which we bound by an *a priori* expression.
- There is a numerical error in the solution of $q(\lambda)$ and its derivative at reference parameter values used to form the global piecewise-linear representation to this surface. This error is deterministic and we obtain an *a posteriori* estimate for this error.
- There is an error in using a finite collection of samples of a distribution on the output to "pass" this distribution through the response surface $q(\lambda)$ in order to determine the posterior density. This error is statistical and we obtain an *a posteriori* estimate for this error.

5.1 The effect of linearization on the inverse problem

We study the inverse problem for a finite dimensional map q from the space of parameters to the output defined implicitly by the solution to a differential equation that depends on a finite number of parameters in the model. We consider the initial value problem

$$\begin{cases} \dot{y} = f(y; \lambda_1), \ t > 0, \\ y(0) = \lambda_0, \end{cases} \tag{5.1.1}$$

where $y \in \mathbb{R}^n$, $f : \mathbb{R}^{n+p} \to \mathbb{R}^n$ is smooth, and $\lambda = (\lambda_1^\top, \lambda_0^\top)^\top \in \mathbf{\Lambda} \subset \mathbb{R}^d$ $(d = p + n)$ are the parameters. We solve (5.1.1) to calculate a linear functional of the solution, or a quantity of interest,

$$q(y) = \int_0^T \langle y, \psi \rangle \ dt. \tag{5.1.2}$$

We assume that the solution y of (5.1.1) depends (implicitly) on parameters λ in a smooth way and denote solutions of (5.1.1) as y_λ and the quantity of interest as $q(\lambda)$ to emphasize the implicit dependence of the quantity of interest on the parameters. The smooth dependence of solutions to (5.1.1) on parameters λ implies the dependence of the quantity of interest on λ is also smooth.

5.1.1 Local linearization of a quantity of interest

We seek to determine the effects of variations in parameters on the quantity of interest. We first consider the initial value problem at a reference parameter value $\mu = (\mu_1^\top, \mu_0^\top)^\top$,

$$\begin{cases} \dot{y}_\mu = f(y_\mu; \mu_1), \ t > 0, \\ y(0) = \mu_0. \end{cases} \tag{5.1.3}$$

We define (y_μ, μ) as the reference point. We define the **exact adjoint** based on the reference point as

$$\begin{cases} -\dot{\phi} - D_y f(y_\mu; \mu_1)^\top \phi = \psi, \ T > t \geq 0, \\ \phi(T) = 0. \end{cases} \tag{5.1.4}$$

We now consider (5.1.1) for λ that are near the reference parameter μ, and seek to determine $q(\lambda)$ given $q(\mu)$ and the generalized Green's function that solves (5.1.4).

Theorem 5.1.1. *[20] If $f(y;\lambda)$ is twice continuously differentiable with respect to both y and λ and Lipschitz continuous in both y and λ, then the quantity of interest is Fréchet differentiable at (y_μ, μ) with derivative $\nabla q(\mu) : \mathbb{R}^d \to \mathbb{R}$ given by*

$$\nabla q(\mu)[\lambda] = \langle (\lambda_0 - \mu_0), \phi(0) \rangle + \int_0^T \langle D_{\lambda_1} f(y_\mu; \mu)(\lambda_1 - \mu_1), \phi \rangle \, dt. \tag{5.1.5}$$

Additionally,

$$q(\lambda) \approx q(\mu) + \nabla q(\mu)[\lambda]. \tag{5.1.6}$$

We can prove (5.1.6) without proving that (5.1.5) is the Fréchet derivative of the quantity of interest, but proving (5.1.5) implies that the representation given in (5.1.6) is in fact a *linear* approximation to the quantity of interest. We refer to the linearized approximation of the quantity of interest in (5.1.6) as a HOPS (Higher Order Parameter Sample) representation [20]. We present a variation of the proof found in [20] that contains useful techniques for the *a posteriori* analysis.

Proof:

We first prove that (5.1.6) holds. We let $e(t) = y_\lambda - y_\mu$ and use a standard variational argument to obtain

$$\begin{aligned}
\int_0^T \langle e, \psi \rangle \, dt &= \int_0^T \left\langle e, -\dot{\phi} - D_y f(y_\mu; \mu_1)^\top \phi \right\rangle \, dt \\
&= -\int_0^T \left\langle e, \dot{\phi} \right\rangle \, dt - \int_0^T \langle e, D_y f(y_\mu; \mu_1)^\top \phi \rangle \, dt \\
&= -\left[\langle e(T), \phi(T) \rangle - \langle e(0), \phi(0) \rangle - \int_0^T \langle \dot{e}, \phi \rangle \, dt \right] - \\
&\quad \int_0^T \langle D_y f(y_\mu; \mu_1) e, \phi \rangle \, dt \\
&= \langle (\lambda_0 - \mu_0), \phi(0) \rangle + \int_0^T \langle \dot{e} - D_y f(y_\mu; \mu_1) e, \phi \rangle \, dt \\
&= \langle (\lambda_0 - \mu_0), \phi(0) \rangle \\
&\quad + \int_0^T \langle (\dot{y}_\lambda - \dot{y}_\mu) - D_y f(y_\mu; \mu_1)(y_\lambda - y_\mu), \phi \rangle \, dt \qquad (5.1.7)
\end{aligned}$$

Since f is assumed twice continuously differentiable, we apply Taylor's theorem to obtain

$$\begin{aligned}
f(y_\lambda; \lambda_1) &= f(y_\mu; \mu_1) + D_y f(y_\mu; \mu_1)(y_\lambda - y_\mu) + D_{\lambda_1} f(y_\mu; \mu_1)(\lambda_1 - \mu_1) \\
&\quad + R_2(y_\lambda, y_\mu; \lambda_1, \mu_1),
\end{aligned}$$

where $R_2(y_\lambda, y_\mu; \lambda_1, \mu_1) \sim \mathcal{O}(\|y_\lambda - y_\mu\|^2 + \|\lambda_1 - \mu_1\|^2)$. We solve the above equation for $-D_y f(y_\mu; \mu_1)(y_\lambda - y_\mu)$ to obtain

$$\begin{aligned}
-D_y f(y_\mu; \mu_1)(y_\mu - y_\lambda) &= f(y_\mu; \mu_1) - f(y_\lambda; \lambda_1) + D_{\lambda_1} f(y_\mu; \mu_1)(\lambda_1 - \mu_1) \\
&\quad + R_2(y_\lambda, y_\mu; \lambda_1, \mu_1).
\end{aligned} \qquad (5.1.8)$$

Substitution of (5.1.8) into (5.1.7) and using the fact that $\dot{y}_\mu - f(y_\mu; \mu_1) = 0$ and $\dot{y}_\lambda - f(y_\lambda; \lambda_1) = 0$ yields

$$\begin{aligned}
\int_0^T \langle e, \psi \rangle \, dt &= \langle (\lambda_0 - \mu_0, \phi(0) \rangle + \int_0^T \langle D_{\lambda_1} f(y_\mu; \mu_1)(\lambda_1 - \mu_1), \phi \rangle \, dt \\
&\quad + \int_0^T \langle R_2(y_\lambda, y_\mu; \lambda_1, \mu_1), \phi \rangle \, dt.
\end{aligned} \qquad (5.1.9)$$

Substituting $e = y_\lambda - y_\mu$, re-arranging the terms in (5.1.9), using the notation of both (5.1.2) and (5.1.5), and neglecting the higher-order term we have

$$q(\lambda) \approx q(\mu) + \nabla q(\mu)[\lambda].$$

We now prove that the Fréchet derivative of the quantity of interest is given by (5.1.5).

Since λ is "nearby" to μ, we set $\lambda = \mu + h$. We prove that

$$\lim_{h \to 0} \frac{|q(\lambda) - q(\mu) - \nabla q(\mu)[\lambda]|}{\|h\|} = 0. \tag{5.1.10}$$

Let $\nu(h)$ denote the numerator of (5.1.10). Then according to (5.1.9)

$$\begin{aligned} \nu(h) &= \left| \int_0^T \langle R_2(y_{\mu+h}, y_\mu; \mu_1 + h_1, \mu_1), \phi \rangle \, dt \right| \\ &\leq \int_0^T \|R_2(y_{\mu+h}, y_\mu; \mu_1 + h_1, \mu_1)\| \, \|\phi\| \, dt. \end{aligned}$$

We have that $\|R_2(y_{\mu+h}, y_\mu; \mu_1 + h_1, \mu_1)\| \leq C(\|y_{\mu+h} - y_\mu\|^2 + \|h_1\|^2)$. If $\|y_{\mu+h} - y_\mu\| \sim \mathcal{O}(\|h\|)$, then $R_2(y_{\mu+h}, y_\mu; \mu_1 + h_1, \mu_1) \sim \mathcal{O}(\|h\|^2)$, and the proof is complete. In the rest of the proof, we make use of the fact that $\|h_0\| \leq \|h\|$ and $\|h_1\| \leq \|h\|$. We use the integral form of (5.1.1),

$$y_\lambda(t) = \lambda_0 + \int_0^t f(y_\lambda(s); \lambda_1) \, ds,$$

to obtain

$$\begin{aligned} \|y_{\mu+h}(t) - y_\mu(t)\| &\leq \|h_0\| + \int_0^t \|f(y_{\mu+h}(s); \mu_1 + h_1) - f(y_\mu(s); \mu_1)\| \, ds \\ &\leq (LT + 1) \, \|h\| + L \int_0^t \|y_{\mu+h}(s) - y_\mu(s)\| \, ds, \end{aligned}$$

where the last inequality follows from the assumption of Lipschitz continuity of $f(y; \lambda)$. Gronwall's inequality gives

$$\|y_{\mu+h}(t) - y_\mu(t)\| \leq (LT + 1) \, \|h\| \, e^{Lt}.$$

This proves that $\|y_{\mu+h} - y_\mu\| \sim \mathcal{O}(\|h\|)$, and the proof is complete. □

To summarize the theorem above, *in the absence of numerical error*,

$$q(\lambda) \approx \int_0^T \langle y_\mu, \psi \rangle \, dt + \langle (\lambda_0 - \mu_0), \phi(0) \rangle + \int_0^T \langle D_{\lambda_1} f(y_\mu; \mu_1)(\lambda_1 - \mu_1), \phi \rangle \, dt \tag{5.1.11}$$

for λ close to μ.

We extend the local linearization technique to obtain a global piecewise-linear approximation of the linear functional over all of $\mathbf{\Lambda}$. This follows the presentation in 2.4. We first define a partition $\{B_i\}_{i=1}^M$ of $\mathbf{\Lambda}$. We apply the local linearization technique described above for each B_i, and define

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

We obtain a global piecewise-linear approximation $\hat{q}(\lambda)$ to $q(\lambda)$.

$$\hat{q}(\lambda) := \sum_{i=1}^{M} \left(q(\mu_i) + \langle \nabla q(\mu_i), (\lambda - \mu_i) \rangle \right) \mathbf{1}_{B_i}(\lambda), \tag{5.1.12}$$

where μ_i is the reference parameter value chosen in B_i.

5.1.2 Effect of using generalized linear contours on the inverse problem

We use a piecewise-linear tangent plane approximation to the surface $q(\lambda)$, which we denoted $\hat{q}(\lambda)$. The generalized contours of $q(\lambda)$ are approximated by the generalized contours of $\hat{q}(\lambda)$, and we refer to these approximate contours as *generalized linear contours.* Similar to above, the inverse problem has a unique solution in $\mathbf{\Lambda}$ considered as a set defined by generalized linear contours. The generalized linear contours converge pointwise to the generalized contours as the "size" of the largest cell, B_i, decreases to zero.

Let B_i and B_j denote cells used in (5.1.12) such that $i \neq j$ and the cells share a boundary. Consider a generalized contour, denoted $q^{-1}(\bar{q})$, that is connected across the boundary of these cells, and let $\hat{q}^{-1}(\bar{q})$ denote the associated generalized linear contour in both B_i and B_j. The generalized linear contour, $\hat{q}^{-1}(\bar{q})$, is typically discontinuous across such a boundary since

$$[q(\mu_i) + \langle \nabla q(\mu_i), (\lambda - \mu_i)\rangle] - [q(\mu_j) + \langle \nabla q(\mu_j), (\lambda - \mu_j)\rangle] \qquad (5.1.13)$$

is typically nonzero for almost every $\lambda \in \partial B_i \cap \partial B_j$. The generalized linear contours are simply the level sets of $\hat{q}(\lambda)$ and (5.1.13) gives the discontinuity of $\hat{q}(\lambda)$ at each $\lambda \in \partial B_i \cap \partial B_j$, so the maximum of (5.1.13) over all cells B_i and B_j with a shared boundary is a measure of the smoothness of using a piecewise-linear approximation to $q(\lambda)$ to solve the inverse problem. We let D denote this measure of smoothness of $\hat{q}(\lambda)$, so

$$D = \max\{|q(\mu_i) - q(\mu_j) + \langle \nabla q(\mu_i), (\lambda - \mu_i)\rangle - \langle \nabla q(\mu_j), (\lambda - \mu_j)\rangle| : \\ \lambda \in \partial B_i \cap \partial B_j \neq 0\}$$

Since the generalized linear contours converge pointwise to the generalized contours, we can make D small and neglect the effect of using a piecewise-linear approximation to $q(\lambda)$ to solve the inverse problem. It is important to note that D is computable and cheap to obtain. This is easy to see since (5.1.13) is a linear function, and if the B_i are polygonal geometric objects, then only a finite number of points need be calculated in order to determine the extreme values of (5.1.13) for $\lambda \in \partial B_i \cap \partial B_j$. Furthermore, we claim that D is a measure of the error in using generalized linear contours to approximate the true generalized contours for sufficiently small cells $\{B_i\}$. As above, let $q^{-1}(\bar{q})$ and $\hat{q}^{-1}(\bar{q})$ denote a generalized contour and its approximation by a generalized linear contour, respectively, in cells B_i and B_j, where $i \neq j$ and the cells share a boundary. Using an analogous argument as in the case of a finite dimensional nonlinear system, the error in the approximation of a generalized contour by a generalized linear contour in cell B_i is given by

$$\left[q^{-1}(\bar{q}) - \hat{q}^{-1}(\bar{q})\right] = C \int_0^T \langle R_2(y_\lambda, y_{\mu_i}; \lambda_1, \mu_{i,1}), \phi \rangle \, dt.$$

Here, C is a constant depending on $\nabla q(\mu_i)$. Replacing i with j in the above equation gives an expression for the error in cell B_j. Therefore, the error in the approximation of the generalized contour by a generalized linear contour in B_i and B_j is $\mathcal{O}(\|R_2(y_\lambda, y_{\mu_i}; \lambda_1, \mu_{i,1})\| + \|R_2(y_\lambda, y_{\mu_j}; \lambda_1, \mu_{j,1})\|)$. This is also true of the term D, so for small D, we consider it a measure of the error in the approximation of generalized contours.

The calculation of D is an *a priori* measure, computed before solving the inverse problem. This measure allows us to determine if (5.1.12) is sufficient for use in solving the inverse problem. Henceforth, we assume that (5.1.12) is a sufficient representation of $q(\lambda)$, and neglect any error in the solution of the inverse problem arising from using generalized linear contours.

5.2 A posteriori analysis

We have shown that in the absence of numerical error, we can use a piecewise-linear approximation to the surface $q(\lambda)$ with negligible error occuring from this representation. Typically, we only have a numerical approximation of the reference solution y_μ of (5.1.1). Additionally, error is introduced in the solution of the inverse problem, defined as a posterior density on $\mathbf{\Lambda}$ [3, 25, 17, 19, 13, 1], by finite sampling of distributions.

We perform an *a posteriori* error analysis for both the deterministic error arising from using a numerical solution for (5.1.1) and also the statistical error arising from finite sampling.

5.2.1 Review of standard *a posteriori* error analysis

We will solve ordinary differential equations assuming a discontinuous Galerkin method. Let Y denote the numerical solution to

$$\begin{cases} \dot{y} = f(y,t), \ 0 < t \leq T, \\ y(0) = y_0. \end{cases} \tag{5.2.1}$$

Let $e = y - Y$, where y solves (5.2.1) exactly. We linearize around Y in the sense described in Chapter 1 to arrive at the adjoint problem

$$\begin{cases} -\dot{\phi} = f'(Y,t)^{\top}\phi + \psi_1(t), \ T > t \geq 0, \\ \phi(T) = \psi_2. \end{cases} \tag{5.2.2}$$

If $\psi_1(t) \equiv 0$, then the quantity of interest is $(e(T), \psi_2)$. If $\psi_2 = 0$, then the quantity of interest is $\int_0^T (e(t), \psi_1(t))\, dt$. The following discussion on estimating the error in a quantity of interest is a summary of [23]. For convenience, we let $f' = f'(Y,t)$ in the remainder of this discussion.

Estimating the error in a quantity of interest valued at the endpoint

Assume $\psi_1(t) \equiv 0$ in (5.2.2). Take the inner product of the adjoint problem with e and integrate from 0 to T to obtain

$$-\int_0^T (\dot{\phi}, e)\, dt - \int_0^T ((f')^{\top}, e)\, dt = 0. \tag{5.2.3}$$

The problem is solved numerically by dividing $[0,T]$ into N subintervals, $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$. The error computation is also computed interval by interval, so we break up (5.2.3) into a sum of integral equations over each interval, integrate by parts over each interval, and use properties of inner products to get

$$-\sum_{n=1}^{N} (e,\phi)|_{t_{n-1}}^{t_n} + \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (\phi, \dot{e})\, dt - \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (\phi, f'e)\, dt = 0. \tag{5.2.4}$$

Since $e = y - Y$ might be discontinuous at the boundaries of each interval, we expand the first term on the right hand side of (5.2.4) to

$$-\sum_{n=1}^{N}(e,\phi)|_{t_{n-1}}^{t_n} = (e(0),\phi(0)) + \sum_{n=2}^{N}([Y]_{n-1},\phi_{n-1}) - (e(T),\phi(T)). \quad (5.2.5)$$

Here, $[Y]_{n-1}$ denotes the jump discontinuity of Y at t_{n-1} calculated as the difference between the right and left sided limits of Y at t_{n-1}, respectively. We use ϕ_{n-1} as a shorthand for $\phi(t_{n-1})$. Substitution of (5.2.5) into (5.2.4) and re-arranging the terms yields

$$\begin{aligned}(e(T),\phi(T)) &= (e(0),\phi(0)) + \sum_{n=2}^{N}([Y]_{n-1},\phi_{n-1}) \\ &\quad + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(\phi,\dot{e})\,dt - \int_{t_{n-1}}^{t_n}(\phi,f'e)\,dt. \qquad (5.2.6)\end{aligned}$$

We substitute $e = y - Y$ into (5.2.6), and use the first-order approximation that $f'(y - Y) \approx f(y) - f(Y)$ so that

$$\begin{aligned}(e(T),\phi(T)) &= (e(0),\phi(0)) + \sum_{n=2}^{N}([Y]_{n-1},\phi_{n-1}) \\ &\quad + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(\phi,\dot{y})\,dt - \int_{t_{n-1}}^{t_n}(\phi,f(y))\,dt \qquad (5.2.7) \\ &\quad + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(\phi,\dot{Y})\,dt - \int_{t_{n-1}}^{t_n}(\phi,f(Y))\,dt.\end{aligned}$$

Since $\dot{y} - f(y) = 0$, (5.2.7) is simplified to

$$(e(T),\phi(T)) = (e(0),\phi(0)) + \sum_{n=2}^{N}([Y]_{n-1},\phi_{n-1}) - \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(\dot{Y} - f(Y),\phi)\,dt. \quad (5.2.8)$$

Note that (5.2.8) is a computable estimate of the error. We can use Galerkin orthogonality to introduce terms such as the projection of ϕ into the space of test functions used in solving (5.2.1) that allow us to rewrite (5.2.8) so that the error generated by discretization of the domain is estimated. Additional terms can be added to take into account the effect of quadrature. We direct the interested reader to [23].

Estimating the error in a quantity of interest with nonzero $\psi_1(t)$

Assume $\psi_2 = 0$ and $\psi_1(t)$ is nonzero for some $t \in (0, T)$. We use inner products and integrate as above to get

$$\int_0^T (e, \psi_1)\, dt = -\int_0^T (e, \dot{\phi})\, dt - \int_0^T (e, f'\phi)\, dt.$$

Following the analysis above, we integrate by parts over each interval to get

$$\begin{aligned}\int_0^T (e, \psi_1)\, dt &= (e(0), \phi(0)) + \sum_{n=2}^{N} ([Y]_{n-1}, \phi_{n-1}) \\ &\quad + \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (\dot{e}, \phi)\, dt - \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (f'e, \phi)\, dt. \qquad (5.2.9)\end{aligned}$$

Note that $\phi(T) = 0$ since $\psi_2 = 0$, so the term involving $e(T)$ does not appear in (5.2.9). Using the same arguments as above we have

$$\int_0^T (e, \psi_1)\, dt = (e(0), \phi(0)) + \sum_{n=2}^{N} ([Y]_{n-1}, \phi_{n-1}) - \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (\dot{Y} - f(Y), \phi)\, dt. \qquad (5.2.10)$$

Note that the right hand sides of (5.2.10) and (5.2.8) are identical.

5.2.2 The effects of deterministic error in the evaluation of the approximate map

We let Y_μ denote the numerical solution to (5.1.1) at the reference parameter μ. This is the numerical approximation around which we linearize the forward problem in order to construct an adjoint. We define the **approximate adjoint** using (5.1.4) with "perturbed" operator $D_y f(Y_\mu; \mu_1)$, and let Φ denote the solution to this approximate adjoint, which is computed using a numerical scheme for

$$\begin{cases} -\dot{\Phi} - D_y f(Y_\mu; \mu_1)^\top \Phi = \psi, \ T > t \geq 0, \\ \Phi(T) = 0. \end{cases} \tag{5.2.11}$$

We assume the numerical error of this solution is negligible. This assumption is reasonable since (5.2.11) is a linear differential equation typically solved using a higher-order method than used to solve (5.1.1), e.g. when we use a piecewise-linear discontinous Galerkin method for the forward problem, we use a quadratic continuous Galerkin method for the adjoint problem. We can alter the analysis to include the error in Φ, but it makes the presentation tedious.

Remark 5.2.1. *In Theorem 5.1.1, we assume $f(y; \lambda)$ is twice continuously differentiable with respect to both y and λ. This assumption of smoothness implies the Lipschitz continuity of $f(y; \lambda)$.*

Therefore, for Y_μ sufficiently close to y_μ over short time,

$$\|D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1)\|_V \leq K \|y_\mu - Y_\mu\|_U , \tag{5.2.12}$$

where $\|\cdot\|_V$ and $\|\cdot\|_U$ are the $L^2([0, T])$ norm of some appropriate matrix and vector norms of the arguments, respectively.

We now seek to quantify the effect of the numerical error of using Y_μ in the construction of the adjoint problem for the representation formula given by (5.1.6).

Let $\check{q}(\lambda)$ denote the approximate quantity of interest calculated using (5.1.6) with Y_μ and Φ in place of y_μ and ϕ, which is to say that

$$\check{q}(\lambda) \approx \int_0^T \langle Y_\mu, \psi \rangle \, dt + \langle (\lambda_0 - \mu_0), \Phi(0) \rangle + \int_0^T \langle D_{\lambda_1} f(Y_\mu; \mu_1)(\lambda_1 - \mu_1), \Phi \rangle \, dt. \tag{5.2.13}$$

The error of using the approximate Y_μ on the representation formula for $q(\lambda)$ is given by $q(\lambda) - \check{q}(\lambda)$. Taking the difference of (5.1.11) and (5.2.13) gives

$$\begin{aligned} q(\lambda) - \check{q}(\lambda) \approx & \int_0^T \langle (y_\mu - Y_\mu), \psi \rangle \, dt \\ & + \underbrace{\langle (\lambda_0 - \mu_0), (\phi(0) - \Phi(0)) \rangle}_{\mathbf{I}} \\ & + \underbrace{\int_0^T (\langle D_y f(y_\mu; \mu_1)(\lambda_1 - \mu_1), \phi \rangle - \langle D_y f(Y_\mu; \mu_1)(\lambda_1 - \mu_1), \Phi \rangle) \, dt}_{\mathbf{II}} . \end{aligned} \tag{5.2.14}$$

The first term on the right-hand side of (5.2.14) is a linear functional of the error $y_\mu - Y_\mu$ and it can be estimated by standard *a posteriori* techniques described above. The goal is to determine estimates of terms **I** and **II**. Terms **I** and **II** measure the effect that the numerical solutions Y_μ and Φ have in the estimate for $q(\lambda)$. Specifically, **I** measures the effect of using an approximate adjoint on the sensitivity of $q(\lambda)$ to changes in the initial conditions of (5.1.1). Term **II** measures the effect of using Y_μ and Φ on the sensitivity of $q(\lambda)$ to changes in model parameters of (5.1.1).

Remark 5.2.2. *The terms* **I** *and* **II** *are functions of the vector* $\lambda - \mu$. *The dependence is linear, and the analysis below produces estimates that also depend on this vector linearly so that the error estimates for these terms are also linear functions of this vector. Thus, following the analysis described below for* p *linearly independent vectors* $\lambda - \mu$, *we obtain a set of error estimates such that the error defined by* **I** *and* **II** *for any vector* $\lambda - \mu$ *can be written as a linear combination from this set of error estimates.*

Estimating term I

We first observe that this term is a linear functional of the error arising from solving the exact adjoint with an approximate adjoint. Standard *a* *posteriori* techniques exist for estimating a linear functional of the error of the *forward* solutions $y_\mu - Y_\mu$ using an adjoint analysis. We use a similar technique here. We define the **adjoint to the *approximate* adjoint** as

$$\begin{cases} \dot{w} = D_y f(Y_\mu; \mu_1) w, \ 0 < t \le T, \\ w(0) = (\lambda_0 - \mu_0) \end{cases}$$

Since $\dot{w} - D_y f(Y_\mu; \mu) = 0$, we have

$$\begin{aligned} 0 &= \int_0^T \langle \dot{w} - D_y f(Y_\mu; \mu_1) w, (\phi - \Phi) \rangle \, dt \\ &= \int_0^T \langle \dot{w}, (\phi - \Phi) \rangle \, dt - \int_0^T \langle D_y f(Y_\mu; \mu_1) w, (\phi - \Phi) \rangle \, dt \\ &= \langle w(T), (\phi(T) - \Phi(T)) \rangle - \langle (w(0), (\phi(0) - \Phi(0)) \rangle \\ &\quad - \int_0^T \left\langle w, (\dot{\phi} - \dot{\Phi}) \right\rangle dt - \int_0^T \langle w, D_y f(Y_\mu; \mu_1)^\top (\phi - \Phi) \rangle \, dt \\ &= - \langle (\lambda_0 - \mu_0), (\phi(0) - \Phi(0)) \rangle \\ &\quad + \int_0^T \left\langle w, \left[-\dot{\phi} - D_y f(Y_\mu; \mu_1)^\top \phi + \dot{\Phi} + D_y f(Y_\mu; \mu_1)^\top \Phi \right] \right\rangle dt. \end{aligned}$$

This gives

$$\mathbf{I} = \int_0^T \left\langle w, \left[-\dot{\phi} - D_y f(Y_\mu; \mu_1)^\top \phi + \dot{\Phi} + D_y f(Y_\mu; \mu_1)^\top \Phi \right] \right\rangle dt. \quad (5.2.15)$$

By adding and subtracting $D_y f(Y_\mu; \mu_1)^\top \phi$ to the differential equation in (5.1.4) for the exact adjoint, we have

$$-\dot{\phi} - D_y f(Y_\mu; \mu_1)^\top \phi = \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1) \right]^\top \phi + \psi. \quad (5.2.16)$$

Substituting (5.2.16) into (5.2.15) and using (5.2.11), we have

$$\begin{aligned} \mathbf{I} &= \int_0^T \left\langle w, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1) \right]^\top \phi \right\rangle dt \\ &= \int_0^T \left\langle w, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1) \right]^\top \Phi \right\rangle dt & (5.2.17) \\ &\quad + \int_0^T \left\langle w, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1) \right]^\top (\phi - \Phi) \right\rangle dt. & (5.2.18) \end{aligned}$$

We claim the second term on the right-hand side of the last equation is higher-order and can be neglected. We prove this claim later. For now, we assume it is true, and go about estimating the first term on the right-hand side. If $f(y;\lambda)$ is three-times continuously differentiable, then we use Taylor's theorem and ignore the higher-order term to get

$$\begin{aligned}
&\int_0^T \left\langle w, \left[D_y f(y_\mu;\mu_1) - D_y f(Y_\mu;\mu_1)\right]^\top \Phi \right\rangle dt \\
&\approx \int_0^T \left\langle w, \underbrace{\left[\Phi^\top \otimes \mathbf{J}\right]}_{n\times n^2} \underbrace{\left[D_y\left(\mathrm{vec}(D_y f(Y_\mu;\mu_1)^\top)\right)\right]}_{n^2\times n}(y_\mu - Y_\mu)\right\rangle dt \\
&= \int_0^T \left\langle \left[D_y\left(\mathrm{vec}(D_y f(Y_\mu;\mu_1)^\top)\right)\right]^\top \left[\Phi^\top \otimes \mathbf{J}\right]^\top w, (y_\mu - Y_\mu)\right\rangle dt.
\end{aligned}$$

Here, $\mathbf{J}$ denotes the $n \times n$ identity matrix and the vector operator denoted vec is a map from $\mathbb{R}^{l\times m} \to \mathbb{R}^{lm}$ defined by stacking the columns (in order) of a matrix to form a column vector. We let

$$\psi_I = \left[D_y\left(\mathrm{vec}(D_y f(Y_\mu;\mu_1)^\top)\right)\right]^\top \left[\Phi^\top \otimes \mathbf{J}\right]^\top w.$$

We now have that the first term on the right-hand side is a linear functional of the error $y_\mu - Y_\mu$. This term is estimable by standard *a posteriori* techniques as described above.

We now prove the claim that the second term is of higher-order and can be neglected. Let $\eta = \phi - \Phi$, then

$$\begin{cases} -\dot{\eta} = D_y f(y_\mu;\mu_1)^\top \phi - D_y f(Y_\mu;\mu_1)^\top \Phi, \ T > t \geq 0 \\ \eta(T) = 0. \end{cases} \tag{5.2.19}$$

From Remark 5.2.1, we have that the first derivatives $f(y, \lambda_1)$ are Lipschitz continuous, so if Y_μ is sufficiently close to y_μ over $[0, T]$, then

$$D_y f(Y_\mu;\mu_1)^\top = D_y f(y_\mu;\mu_1)^\top + \epsilon(t), \ t \in [0, T], \tag{5.2.20}$$

where $\epsilon(t)$ is a perturbation matrix satisfying $\|\epsilon(t)\| \leq C \|y_\mu - Y_\mu\|$, for some $C > 0$ and all $t \in [0, T]$. Substituting (5.2.20) into (5.2.19) gives,

$$\begin{cases} -\dot{\eta} = D_y f(y_\mu; \mu_1)^\top \eta + \epsilon(t)\Phi(t), \ T > t \geq 0 \\ \eta(T) = 0. \end{cases} \tag{5.2.21}$$

Let $\Sigma(t)$ denote the fundamental matrix of (5.2.22), then

$$\eta(t) = -\Sigma(t) \int_T^t [\Sigma(s)]^{-1} \epsilon(s)\Phi(s)\, ds.$$

This implies that

$$\|\eta(t)\| \leq \|\Sigma(t)\| \int_0^T \left\|\Sigma(s)^{-1}\right\| \|\epsilon(s)\| \|\Phi(s)\| \, ds \leq C \|y_\mu - Y_\mu\|_U . \tag{5.2.22}$$

Here, $\|\cdot\|_U$ is interpreted as before to mean the $L^2([0, T])$ norm of a given vector norm of the argument, and $C > 0$ is some constant that bounds the product of $\sup_{t\in[0,T]} \|\Sigma(t)\|$, $\sup_{t\in[0,T]} \|\Sigma(t)^{-1}\|$, and $\sup_{t\in[0,T]} \|\Phi(t)\|$. Thus, by Lipschitz continuity of the first derivatives of $f(y; \lambda)$ and (5.2.22),

$$\left| \int_0^T \left\langle w, [D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1)]^\top (\phi - \Phi) \right\rangle dt \right| \quad \leq \quad C \|y_\mu - Y_\mu\|_U^2 .$$

Thus, this is higher-order, and we have established an estimate of term **I**.

Estimating term II

By adding and subtracting $\langle D_\lambda f(Y_\mu; \mu_1)(\lambda_1 - \mu_1), \phi\rangle$ to the integrand in **II**, we rewrite **II** so that **II** = **IIa** + **IIb**, where

$$\begin{aligned} \mathbf{IIa} &= \int_0^T \langle (D_\lambda f(y_\mu; \mu_1) - D_\lambda f(Y_\mu; \mu_1))(\lambda_1 - \mu_1), \phi\rangle \, dt \\ \mathbf{IIb} &= \int_0^T \langle D_\lambda f(Y_\mu; \mu_1)(\lambda_1 - \mu_1), (\phi - \Phi)\rangle \, dt. \end{aligned}$$

We now estimate **IIa** and **IIb**.

Estimating term IIa

By adding and subtracting $\langle (D_\lambda f(y_\mu;\mu_1) - D_\lambda\ f(Y_\mu;\mu_1))(\lambda_1 - \mu_1), \Phi\rangle$ to the integrand in **IIa**, we rewrite **IIa** so that **IIa** = **IIaa** + **IIab**, where

$$\begin{aligned}\mathbf{IIaa} &= \int_0^T \langle (D_\lambda f(y_\mu;\mu_1) - D_\lambda f(Y_\mu;\mu_1))(\lambda_1 - \mu_1), (\phi - \Phi)\rangle\ dt \\ \mathbf{IIab} &= \int_0^T \langle (D_\lambda f(y_\mu;\mu_1) - D_\lambda f(Y_\mu;\mu_1))(\lambda_1 - \mu_1), \Phi\rangle\ dt.\end{aligned}$$

We show that **IIaa** is of higher-order, and can be neglected. We have shown above that

$$\|\phi - \Phi\| \leq C\, \|y_\mu - Y_\mu\|_U\,,$$

for some constant $C > 0$, and from Remark 5.2.1 we have that the first-derivatives of $f(y;\lambda)$ are Lipschitz continuous, so

$$|\mathbf{IIaa}| \leq C\, \|y_\mu - Y_\mu\|_U^2\,,$$

for some constant $C > 0$. Thus, **IIaa** $\sim \mathcal{O}(\|y_\mu - Y_\mu\|^2)$, and is neglected in the estimate.

Again assuming that $f(y;\lambda)$ is three-times continuously differentiable, then

$$\begin{aligned}(D_\lambda f(y_\mu;\mu_1) - D_\lambda f(Y_\mu;\mu_1))(\lambda_1 - \mu_1) \ &\approx \\ &\left[(\lambda_1 - \mu_1)^\top \otimes \mathbf{J}\right] \left[D_y\left(\mathrm{vec}(D_\lambda f(Y_\mu;\mu_1))\right)\right](y_\mu - Y_\mu)\,.\end{aligned}$$

We substitute this estimate into **IIab** so that

$$\begin{aligned}\mathbf{IIab} \ &\approx \int_0^T \left\langle \left[(\lambda_1 - \mu_1)^\top \otimes \mathbf{J}\right] \left[D_y\left(\mathrm{vec}(D_\lambda f(Y_\mu;\mu_1))\right)\right](y_\mu - Y_\mu)\,, \Phi\right\rangle\ dt \\ &= \int_0^T \left\langle (y_\mu - Y_\mu), \left[D_y\left(\mathrm{vec}(D_\lambda f(Y_\mu;\mu_1))\right)\right]^\top \left[(\lambda_1 - \mu_1)^\top \otimes \mathbf{J}\right]^\top \Phi\right\rangle\ dt.\end{aligned}$$

We let

$$\psi_{IIab} = \left[D_y\left(\mathrm{vec}(D_\lambda f(Y_\mu;\mu_1))\right)\right]^\top \left[(\lambda_1 - \mu_1)^\top \otimes \mathbf{J}\right]^\top \Phi.$$

Thus, we have represented **IIab** as a linear functional of the error in $y_\mu - Y_\mu$, which can be estimated by standard *a posteriori* techniques as described above.

Estimating term IIb

We let

$$\psi_{IIb} = D_\lambda f(Y_\mu, \mu_1)(\lambda_1 - \mu_1),$$

so that

$$\mathbf{IIb} = \int_0^T \langle \psi_{IIb}, (\phi - \Phi) \rangle \; dt.$$

Thus, **IIb** is a linear functional of the error in the adjoint solutions $\phi - \Phi$. We apply standard *a posteriori* techniques used to estimate linear functionals of the error in the forward solutions to estimate the error in the adjoint solutions as was done above. We again define an adjoint to the *approximate* adjoint as

$$\begin{cases} \dot{z} - D_y f(Y_\mu; \mu_1) z = \psi_{IIb}, \;\; 0 < t \leq T, \\ z(0) = 0. \end{cases}$$

We perform a standard variational argument to obtain

$$\begin{aligned}
\mathbf{IIb} &= \int_0^T \langle (\dot{z} - D_y f(Y_\mu; \mu_1) z, (\phi - \Phi) \rangle \; dt \\
&= \int_0^T \langle \dot{z}, (\phi - \Phi) \rangle \; dt - \int_0^T \langle D_y f(Y_\mu; \mu_1) z, (\phi - \Phi) \rangle \; dt \\
&= \underbrace{\langle z(T), (\phi(T) - \Phi(T)) \rangle}_{\phi(T) - \Phi(T) = 0} - \underbrace{\langle z(0), (\phi(0) - \Phi(0)) \rangle}_{z(0) = 0} \\
&\quad - \int_0^T \left\langle z, (\dot{\phi} - \dot{\Phi}) \right\rangle \; dt - \int_0^T \langle z, D_y f(Y_\mu; \mu_1)^\top (\phi - \Phi) \rangle \; dt \\
&= \int_0^T \left\langle z, -\dot{\phi} - D_y f(Y_\mu; \mu_1) \phi + \dot{\Phi} + D_y f(Y_\mu; \mu_1) \Phi \right\rangle \; dt.
\end{aligned}$$

Using (5.2.19)-(5.2.21) in the right-hand side above, we have

$$\begin{aligned}\mathbf{IIb} &= \int_0^T \left\langle z, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1)\right]^\top \phi \right\rangle dt \\ &= \int_0^T \left\langle z, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1)\right]^\top \Phi \right\rangle dt \\ &\quad - \int_0^T \left\langle z, \left[D_y f(y_\mu; \mu_1) - D_y f(Y_\mu; \mu_1)\right]^\top (\phi - \Phi) \right\rangle dt\end{aligned}$$

The two terms on the right-hand side are analagous to (5.2.17) and (5.2.18). The second term on the right-hand side has already been proven to be higher-order. Therefore, the second term is neglected in the estimate. The first term is estimated similarly to how (5.2.17) was estimated. We define

$$\tilde{\psi}_{IIb} = \left[D_y \left(\text{vec}(D_y f(Y_\mu; \mu_1)^\top)\right)\right]^\top \left[\Phi^\top \otimes \mathbf{J}\right]^\top z,$$

and the first term is approximated by

$$\int_0^T \left\langle \tilde{\psi}_{IIb}, (y_\mu - Y_\mu) \right\rangle dt,$$

which is a linear functional of the error of $y_\mu - Y_\mu$, and is estimable by standard *a posteriori* techniques as described above.

5.2.3 Summarizing the effect of deterministic error

We consider λ that are near the reference parameter μ, and seek to determine $q(\lambda)$ given $q(\mu)$ and the generalized Green's function that solves the adjoint problem. In the absence of numerical error of the solution to the initial value problem, we saw that

$$q(\lambda) \approx q(\mu) + \langle (\lambda_0 - \mu_0), \phi(0) \rangle + \int_0^T \langle D_{\lambda_1} f(y_\mu; \mu_1)(\lambda_1 - \mu_1), \phi \rangle \, dt,$$

which we refer to as the HOPS approximation for ordinary differential equations.

We numerically solve the initial value problem (5.1.3) and the numerical error should not be neglected. We determined estimates of the effect of this error on the HOPS approximation. We observed that such estimates can be made by extending standard *a posteriori* techniques and applying smoothness assumptions to $f(y;\lambda)$. By solving several additional adjoint problems, and establishing *a priori* bounds on some terms, proving these are of higher-order and can be neglected, an error estimate on the effect of using numerical solutions and perturbed adjoint problems was obtained.

Theorem 5.2.1. *Let Y_μ and Φ denote the numerical solutions to the initial value problem (5.1.3) and the approximate adjoint problem (5.2.11), respectively.*

Use standard a posteriori *techniques to estimate the error $y_\mu - Y_\mu$ defined as $e_0 := \int_0^T \langle (y_\mu - Y_\mu), \psi \rangle \, dt$.*

Let p_m be the number of model parameters and p_i the number of initial conditions ($p_m + p_i = p$)

for $i = 1, \dots, p$ ***do***

  if $i \leq p_m$ ***then***

    Let z denote the solutions to the adjoint to the approximate adjoint problem

$$\begin{cases} \dot{x} - D_y f(Y_\mu;\mu_1)x = D_\lambda f(Y_\mu,\mu_1)\delta_i, \ 0 < t \leq T, \\ x(0) = 0, \end{cases}$$

where δ_i denotes the i^{th} standard basis vector in $\mathbb{R}^{p_m}$

    Set

$$\psi_{IIab} = \left[D_y\left(vec(D_\lambda f(Y_\mu;\mu_1))\right)\right]^\top \left[\delta_i^\top \otimes \mathbf{J}\right]^\top \Phi,$$

$$\tilde{\psi}_{IIb} = \left[D_y\left(vec(D_y f(Y_\mu;\mu_1)^\top)\right)\right]^\top \left[\Phi^\top \otimes \mathbf{J}\right]^\top z.$$

```
        Solve (5.2.11) with data given by the above vectors and calculate the standard error representations given by
            $e_1^i := \int_0^T \langle (y_\mu - Y_\mu), \psi_{IIab} \rangle \, dt, \; e_2^i := \int_0^T \left\langle \tilde{\psi}_{IIb}, (y_\mu - Y_\mu) \right\rangle \, dt.$
    else
        Let $w$ denote the solutions to the adjoint to the approximate adjoint problem
            $\begin{cases} \dot{x} - D_y f(Y_\mu; \mu_1) x = 0, \; 0 < t \leq T, \\ x(0) = \delta_i, \end{cases}$
        where $\delta_i$ denotes the $i^{th}$ standard basis vector in $\mathbb{R}^{p_i}$
        Set
            $\psi_I = \left[ D_y \left( vec(D_y f(Y_\mu; \mu_1)^\top) \right) \right]^\top \left[ \Phi^\top \otimes \mathbf{J} \right]^\top w$
        Solve (5.2.11) with data given by the above vectors and calculate the standard error representations given by
            $e_3^i := \int_0^T \langle \psi_I, (y_\mu - Y_\mu) \rangle \, dt$
    end if
end for
The estimate of the effect of error on the HOPS representation in direction u given by
    $u = \begin{pmatrix} \nu_1 & \cdots & \nu_{p_m} & \rho_1 & \cdots & \rho_{p_i} \end{pmatrix}^\top$
is denoted $e_u$, and given by
    $e_u = e_0 + \sum_{i=1}^{p_m} \nu_i \sum_{j=1}^{2} e_j^i + \sum_{i=1}^{p_i} \rho_i e_3^i.$
```

Time Step	True Error	Term **IIab**	Term **IIb**	Error Estimate	Ratio
0.2	-2.18×10^{-4}	-1.93×10^{-4}	1.63×10^{-5}	-1.77×10^{-4}	0.8114
0.1	-2.92×10^{-5}	-2.72×10^{-5}	-2.90×10^{-8}	-2.72×10^{-5}	0.9317
0.05	-3.77×10^{-6}	-3.59×10^{-6}	-1.57×10^{-7}	-3.75×10^{-6}	0.9952
0.025	-4.79×10^{-7}	-4.62×10^{-7}	-2.98×10^{-8}	-4.91×10^{-7}	1.0262
0.0125	-5.95×10^{-8}	-5.85×10^{-8}	-4.36×10^{-9}	-6.28×10^{-8}	1.0557

Table 5.1: Results for $T = 1$

Example 5.2.1. *Consider the nonlinear problem with changing stability given by*

$$\begin{cases} \dot{x} + (0.25 + \sin(\lambda t))x^2 = 0, \\ x(0) = 1, \end{cases}$$

with exact solution

$$x = \frac{\lambda}{(\lambda + 1) + .25\lambda t - \cos(\lambda t)}.$$

We take the quantity of interest to be the value of $x(T)$ at various T. The exact adjoint solution is

$$\phi = C[(\lambda + 1) + .25\lambda t - \cos(\lambda t)]^2,$$

where $C = [(\lambda + 1) + .25\lambda T - \cos(\lambda T)]^{-2}$. We study the effect of error around the reference parameter $\mu = \pi$ and denote solution at this reference parameter as y. Recall from the above theorem that we present error gradients resulting from using unit perturbations to the affine map $q(\lambda)$. The results are summarized in Tables 5.1-5.3.

Time Step	True Error	Term **IIab**	Term **IIb**	Error Estimate	Ratio
0.2	-6.13×10^{-4}	-3.78×10^{-4}	-2.71×10^{-4}	-6.5×10^{-4}	1.0599
0.1	-6.76×10^{-5}	-4.41×10^{-5}	-3.73×10^{-5}	-8.14×10^{-5}	1.2044
0.05	-7.85×10^{-6}	-5.28×10^{-6}	-4.75×10^{-6}	-1.00×10^{-5}	1.2767
0.025	-9.46×10^{-7}	-6.44×10^{-7}	-5.96×10^{-7}	-1.24×10^{-7}	1.3106
0.0125	-1.21×10^{-7}	-7.94×10^{-8}	-7.45×10^{-8}	-1.54×10^{-7}	1.2767

Table 5.2: Results for $T = 4$

Time Step	True Error	Term **IIab**	Term **IIb**	Error Estimate	Ratio
0.2	-8.21×10^{-4}	-3.46×10^{-4}	-3.73×10^{-4}	-7.19×10^{-4}	0.8763
0.1	-8.66×10^{-5}	-3.28×10^{-5}	-5.03×10^{-5}	-8.31×10^{-5}	0.9601
0.05	-9.78×10^{-6}	-3.43×10^{-6}	-6.34×10^{-6}	-9.77×10^{-6}	0.9994
0.025	-1.17×10^{-6}	-3.86×10^{-7}	-7.91×10^{-7}	-1.18×10^{-6}	1.0057
0.0125	-1.46×10^{-7}	-4.55×10^{-8}	-9.85×10^{-8}	-1.44×10^{-7}	0.9847

Table 5.3: Results for $T = 10$

5.2.4 Statistical error

We examine the effect of finite sampling on the posterior density. We present a brief review of the salient results and methodology for obtaining the posterior density, and then follow a similar analysis to [8] to obtain an *a posteriori* error analysis for the approximate distribution resulting from finite sampling.

The posterior density

The goal in solving the inverse problem is to determine probabilities of events $A \subset \mathbf{\Lambda}$. We procede by constructing a posterior density that defines a measure, or in probabilistic language a distribution, on $\mathbf{\Lambda}$ in order to determine the probabilities of these events. Recall the following

Theorem 5.2.2. *Given a measurable set $A \subset \mathbf{\Lambda}$, we can approximate $P(A)$ using a simple function approximation to (3.3.1), which only requires calculations of volumes in $\mathbf{\Lambda}$.*

The proof of the above theorem lead to Alg. 3.3.1. The algorithm assigns a probability $P(b_i)$ to each cell b_i, where $\{b_i\}_{i=1}^{M'}$ partitions $\mathbf{\Lambda}$, resulting in an approximate posterior density on model space by the simple function

$$\sigma_{\mathbf{\Lambda}}(\lambda) \approx \sigma_{\mathbf{\Lambda},M'}(\lambda) = \sum_{k=1}^{M'} P(b_i)\mathbf{1}_{b_i}(\lambda). \tag{5.2.23}$$

We focus on the nested loops of Alg. 3.3.1 applied to the cells $\{b_i\}$ with a fixed simple function approximation to $\rho_{\mathcal{D}}(q)$.

Remark 5.2.3. *We assume smoothness of $\rho_{\mathcal{D}}(q)$ so that standard calculus results hold, i.e. so that instead of inducing a partition on $\mathcal{D}$ from the simple function approximation as in the proof of the above theorem, we can* *instead create a fine partition on $\mathcal{D}$ then use this partition to approximate Riemann integrals of $\rho_{\mathcal{D}}(q)$. This is done in practice, and when $\rho_{\mathcal{D}}(q)$ is available, we might use the true probability of subintervals partitioning $\mathcal{D}$ to fix the simple function approximation in Alg. 3.3.1, which is analogous to using an integral mean value theorem for each subinterval. Under these assumptions that the simple function approximation calculates the true probability for each subinterval, we seek to determine an estimate for the error resulting from the case where the simple function approximation is instead approximated by binning a finite collection of samples from $\rho_{\mathcal{D}}(q)$.*

With this in mind, we let $\rho_{\mathcal{D}}$ denote the simple function approximation in the discussion below. The error in using finite sampling effects the calculations of $P(b_i)$. We emphasize the calculation of $P(b_i)$ in Alg. 5.2.1 below.

Algorithm 5.2.1 (Approximate Probabilities of Cells Partitioning $\mathbf{\Lambda}$).

```
Partition $\mathbf{\Lambda}$ with small cells $\{b_i\}_{i=1}^{M'}$
for $i = 1, \ldots, M'$ (number of small cells) do
    Use $\rho_{\mathbf{\Lambda}}$ to calculate volume of $b_i$
    Calculate $q_{i,m} = \min\{q(\lambda) \mid \lambda \in b_i\}$
    Calculate $q_{i,M} = \max\{q(\lambda) \mid \lambda \in b_i\}$
    Use $\rho_{\mathcal{D}}$ to calculate probability of event $[q_{i,m}, q_{i,M}]$
    Use generalized contours to determine set $A_i := q^{-1}([q_{i,m}, q_{i,M}]) \subset \mathbf{\Lambda}$
    Use $\rho_{\mathbf{\Lambda}}$ to calculate volume of $A_i$
    Set $P(b_i)$ to be the product of the probability of $[q_{i,m}, q_{i,M}]$ and the ratio of the volume of $b_i$ to the volume of $A_i$
end for
```

We seek to determine the effect of the error in the distribution defined by (5.2.23) on $\mathbf{\Lambda}$ when the probability of $[q_{i,m}, q_{i,M}]$ in Alg. 5.2.1 is determined from a finite collection of samples of the distribution of $q(\lambda)$.

A posteriori error analysis for (5.2.23)

Let $F_q(t)$ denote the probability distribution function of $q(\lambda)$. If we can evaluate $\rho_{\mathcal{D}}(q)$ or $F_q(t)$ directly, then no sampling is necessary, and from the above algorithm we have for $1 \le i \le M'$

$$P(b_i) = F_q(q(b_i)) \frac{\int_{b_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}, \tag{5.2.24}$$

where $q(b_i)$ indicates the interval in $\mathcal{D}$ that is mapped to by $q(\lambda)$ for $\lambda \in b_i$.

We let $F(t)$ denote the probability distribution on $\mathbf{\Lambda}$ defined by (5.2.23), where $t \in \mathbb{R}^d$, and

$$F(t) = P(\{\lambda \mid \lambda \le t\}) = P(\lambda \le t). \tag{5.2.25}$$

Here the inequality, $\lambda \le t$, is considered component-wise. Using (5.2.24) in (5.2.25) gives

$$F(t) = \sum_{i=1}^{M'} F_q(q(b_i)) \frac{\int_{b_i \cap \{\lambda \le t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}.$$

We use a sample distribution function for $F_q(t)$ computed from a finite collection of *error free* sample values $\{Q_1, \ldots, Q_N\}$, which we denote $F_{q,N}(t)$ so that

$$F_{q,N}(t) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}(Q_n \le t).$$

This leads to an approximation of $F(t)$ by the sample distribution function $F_N(t)$ defined as

$$F_N(t) = \sum_{i=1}^{M'} F_{q,N}(q(b_i)) \frac{\int_{b_i \cap \{\lambda \le t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}.$$

The deterministic error estimated above leads to an error in the calculation of $q(b_i)$. We let $\tilde{q}(b_i)$ denote this calculation with error, and define the approximate sample distribution function $\tilde{F}_N(t)$ as

$$\tilde{F}_N(t) = \sum_{i=1}^{M'} F_{q,N}(\tilde{q}(b_i)) \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}. \qquad (5.2.26)$$

We calculate probabilities using (5.2.26) and seek to determine the error $F(t) - \tilde{F}_N(t)$. We decompose the error as in [8] to get

$$\left|F(t) - \tilde{F}_N(t)\right| \leq \underbrace{|F(t) - F_N(t)|}_{\mathbf{I}} + \underbrace{\left|F_N(t) - \tilde{F}_N(t)\right|}_{\mathbf{II}}. \qquad (5.2.27)$$

We first consider **I**:

$$\mathbf{I} \leq \sum_{i=1}^{M'} |F_q(q(b_i)) - F_{q,N}(q(b_i))| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}. \qquad (5.2.28)$$

From [8] and [24], we have that

$$P\left(\sup_{t \in \mathbb{R}} |F_q(t) - F_{q,N}(t)| \geq \epsilon\right) \leq Ce^{-2\epsilon^2 N}, \text{ for all } \epsilon > 0,$$

where $C > 0$ is some constant not depending on F_q, or in other words, for any $\epsilon > 0$,

$$\sup_{t \in \mathbb{R}} |F_q(t) - F_{q,N}(t)| \leq C \left(\frac{\log(\epsilon^{-1})}{2N}\right)^{1/2} \qquad (5.2.29)$$

with probability greater than $1 - \epsilon$. Using these results in (5.2.28) gives

$$\mathbf{I} \leq C \left(\frac{\log(\epsilon^{-1})}{2N}\right)^{1/2} \sum_{i=1}^{M'} \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)} \leq C \left(\frac{\log(\epsilon^{-1})}{2N}\right)^{1/2}$$

with probability greater than $1 - \epsilon$.

Next, we consider **II** of (5.2.27).

From the *a posteriori* deterministic error analysis, we have an approximation for the error in $\tilde{q}(b_i)$ for each small cell b_i. For convenience, we choose the fine partition $\{b_i\}$ so that for each $1 \leq i \leq M'$, $b_i \subset B_j$ for some $1 \leq j \leq M$. Thus, for all small cells $b_i \subset B_j$ for a fixed j, there is the same deterministic error term associated with $\tilde{q}(b_i)$ since we neglect the effects of linearization on the inverse problem as described above. Let E_j, $1 \leq j \leq M$ denote the deterministic error associated with each $\tilde{q}(b_i)$ for all $b_i \subset B_j$. Set $E = \max_j |E_j|$. Let $M_i = \max q(b_i)$ and $m_i = \min q(b_i)$. Using an analogous argument as in [8],

$$\begin{aligned}
\mathbf{II} &\leq \sum_{i=1}^{M'} |F_{q,N}(M_i + E) - F_{q,N}(m_i - E)| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)} \\
&\leq \sum_{i=1}^{M'} |F_q(M_i + E) - F_{q,N}(m_i + E)| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)} \\
&\quad + \sum_{i=1}^{M'} |F_q(m_i - E) - F_{q,N}(m_i - E)| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)} \\
&\quad + \sum_{i=1}^{M'} |F_q(M_i + E) - F_q(m_i - E)| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}.
\end{aligned}$$

Using (5.2.29) for the first two terms on the right-hand side of the inequality we have that for any $\epsilon > 0$

$$\begin{aligned}
\mathbf{II} &\leq 2C \left(\frac{\log(\epsilon^{-1})}{2N} \right)^{1/2} \\
&\quad + \sum_{i=1}^{M'} |F_q(M_i + E) - F_q(m_i - E)| \frac{\int_{b_i \cap \{\lambda \leq t\}} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}{\int_{A_i} \rho_{\mathbf{\Lambda}}(\lambda)\, d\mu_{\mathbf{\Lambda}}(\lambda)}
\end{aligned}$$

with probability greater than $1 - \epsilon$. Assuming Lipschitz continuity of the distribution F_q and letting L denote the Lipschitz constant, then

$$\mathbf{II} \leq 2C \left(\frac{\log(\epsilon^{-1})}{2N} \right)^{1/2} + LE \max_{1 \leq i \leq M'} (M_i - m_i)$$

with probability greater than $1 - \epsilon$.

We summarize these results as a theorem analogous to the one presented in [8].

Theorem 5.2.3. *There exists constant C not depending on any distribution such that for all $\epsilon > 0$*

$$\left|F(t) - \tilde{F}_N(t)\right| \leq 3C \left(\frac{\log(\epsilon^{-1})}{2N}\right)^{1/2} + LE \max_{1 \leq i \leq M'} (\max q(b_i) - \min q(b_i))$$

with probability greater than $1 - \epsilon$.

Remark 5.2.4. *Much of the above analysis is directly applicable to the forward problem of passing a distribution on $\mathbf{\Lambda}$ through the response surface $q(\lambda)$ to determine a distribution $F_q(t)$ of the output. The deterministic error analysis is identical, and only slight changes need be made to the statistical error analysis since the sample distribution $F_{q,N}(t)$ is no longer computed from error free samples as above. In the above analysis, we assume that even if we do not have access to $F_q(t)$ directly, we are at least able to generate (or are provided with) independent identically distributed samples according to this distribution to compute $F_{q,N}(t)$. In the forward problem, we do not know* a priori *what $F_q(t)$ is, and estimate it entirely from $F_{q,N}(t)$ except that now the samples are generated by passing samples of a distribution on $\mathbf{\Lambda}$ through the surface $q(\lambda)$ so that now the samples are no longer error free. Having computed the deterministic errors associated with the output samples, E_j, as above, we follow the* a posteriori *statistical analysis of [8] verbatim.*

Remark 5.2.5. *In the case where we can evaluate $\rho_{\mathcal{D}}(q)$ or $F_q(t)$ directly, so no sampling is necessary, the analysis simplifies greatly and we simply examine the effect of the deterministic error. We can take the bound in the theorem above and by "sending N to infinity" arrive at the correct bound*

$$\left|F(t) - \tilde{F}(t)\right| \leq LE \max_{1 \leq i \leq M'} (\max q(b_i) - \min q(b_i)).$$

Here $\tilde{F}(t)$ is the distribution calculated using exact values of $\rho_{\mathcal{D}}(q)$, but with deterministic errors effecting the calculation of $P(b_i)$ in Alg. 3.3.1.

Example 5.2.2. *Consider the same problem as in Example 5.2.1. We restrict λ to be in the interval $[\pi - 0.1, \pi + 0.1]$ and use the linear approximation for the quantity of interest constructed in the calculations of the previous example for $T = 1$ with a step size of* 0.05. *This implies that $E = 3.75 \times 10^{-6}$ and using just one HOPS point we have that $\max q(b_i) - \min q(b_i)$ is approximately* 0.01138. *Assume the output is uniformly distributed so that $L = 87.8231$, which is found by considering the output distribution on $[\min q(b_i), \max q(b_i)]$. This gives the bound on $\left|F(t) - \tilde{F}(t)\right|$ as 3.75×10^{-6}. This is* exactly *the numerical error. This should be the case for this particular example and checks the consistency of the above analysis. A uniformly distributed output has a uniformly distributed input since this is a 1-1 linear map. The only error comes from the numerical error in this case, and is reflected in the result.*

Chapter 6

ERROR ANALYSIS FOR INVERSE SENSITIVITY PROBLEMS WITH SEMILINEAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

Compared to the case of ordinary differential equations already presented, there are the same various sources of computational error effecting the posterior density defined as the solution to the inverse problem. We present both the analysis for the effect of the linearization and the numerical error in the solution of $q(\lambda)$ and its derivative at reference parameter values used to form the global piecewise-linear representation to this surface. This error is deterministic and we obtain an *a posteriori* estimate for this error. The statistical sources of error have the same analysis as already presented for ordinary differential equations.

6.1 The effect of linearization on the inverse problem

We study the inverse problem for a finite dimensional map q from the space of parameters to the output defined implicitly by the solution to a semilinear elliptic partial differential equation that depends on a finite number of parameters in the model. We consider the boundary value problem

$$\begin{cases} -\nabla \cdot (a\nabla u) = f(u;\lambda), \text{ in } \Omega, \\ u = 0, \text{ on } \partial\Omega, \end{cases} \tag{6.1.1}$$

where $u = u(x) = (u_1(x), \ldots, u_m(x))^\top$ is the solution, $f : \mathbb{R}^{m+p} \to \mathbb{R}^m$ is the forcing term depending on parameters and solution, and $\lambda = (\lambda_1, \ldots, \lambda_p)^\top$ are the parameters. Let $V = (H_0^1(\Omega))^m$ We seek a solution $u \in V$ to the variational (weak) formulation of (6.1.1) satisfying

$$B[u, v] = (f(u; \lambda), v)_{L^2}, \ \forall v \in V. \tag{6.1.2}$$

Here $B[\cdot, \cdot]$ is the bilinear form defined by $B[u, v] = \sum_{i=1}^m (a\nabla u_i, \nabla v_i)_{L^2}$. We use the notation $(\cdot, \cdot)_{L^2}$ to denote the $(L^2)^m$ inner product.

We require f to be Lipschitz continuous in some invariant region for the solutions u to guarantee the existence of such solutions as is usually required in the proofs using fixed point theorems.

We solve (6.1.2) to calculate a linear functional of the solution, or a quantity of interest $q(u) = (u, \psi)_V$ for some $\psi \in V$. As before, we write $q(\lambda)$ to emphasize the dependence of solutions to (6.1.2) on parameters λ.

6.1.1 Local linearization of a quantity of interest

We seek to determine the effects of variations in parameters on the quantity of interest. We first solve (6.1.2) at reference parameter μ to obtain reference solution $w \in V$. As before, we linearize around this solution and solve an adjoint problem for $\phi \in V$ satisfying

$$B[v, \phi] = (v, D_u f(w; \mu)^\top \phi)_{L^2} + (v, \psi)_V, \ \forall v \in V. \tag{6.1.3}$$

We now consider (6.1.2) for λ that are near the reference parameter μ, and seek to determine $q(\lambda)$ given $q(\mu)$ and the generalized Green's function ϕ that solves (6.1.3). Let u denote the solution associated with λ and set

$e = u - w \in V$. We have that

$$\begin{aligned}(e,\psi)_V &= B[e,\phi] - (e, D_u f(w;\mu)^\top \phi)_{L^2}\\ &= B[e,\phi] - (D_u f(w;\mu)e, \phi)_{L^2}\\ &= (f(u,\lambda) - f(w,\mu) - D_u f(w;\mu)(u-w), \phi)_{L^2}\\ &= (D_\lambda f(w;\mu)(\lambda-\mu), \phi)_{L^2} + (\mathcal{R}, \phi)_{L^2},\end{aligned}$$

where $\mathcal{R}$ is a remainder term from the first-order Taylor expansion of f and is of higher order. Neglecting this term, we have the HOPS representation formula [20]

$$q(\lambda) \approx q(\mu) + (D_\lambda f(w;\mu)(\lambda-\mu), \phi)_{L^2}.$$

We rewrite this as

$$q(\lambda) \approx q(\mu) + \nabla q(\mu)[\lambda - \mu], \tag{6.1.4}$$

where $\nabla q(\mu)[\cdot] = (\cdot, D_\lambda f(w;\mu)^\top \phi)_{L^2}$ denotes the Fréchet derivative of $q(\lambda)$ at $\lambda = \mu$ [20].

As before, we extend the local linearization technique to obtain a global piecewise-linear approximation of the linear functional over all of $\mathbf{\Lambda}$ to obtain (5.1.12). The inverse problem statement is now identical to that presented before. The analysis of the effect of using generalized linear contours on the inverse problem is identical to the presentation on ordinary differential equations.

6.1.2 Review of standard *a posteriori* error analysis

Let U denote the numerical solution to

$$\begin{cases}-\nabla \cdot (a\nabla u) = f(u), \text{ in } \Omega,\\ u = 0, \text{ on } \partial\Omega,\end{cases} \tag{6.1.5}$$

Let $e = u - U$, where u solves (6.1.5) exactly. We linearize around U in the sense described in Chapter 1 to arrive at the formal adjoint problem

$$\begin{cases} -\nabla \cdot (a\nabla\phi) = f'(U)^{\top}\phi + \psi, \text{ in } \Omega, \\ \phi = 0, \text{ on } \partial\Omega, \end{cases} \tag{6.1.6}$$

For convenience, we let $f' = f'(U)$ in the remainder of this discussion.

We work with the weak forms using e and ϕ to get

$$B[e, \phi] = (e, (f')^{\top}\phi)_{L^2} + (e, \psi)_V. \tag{6.1.7}$$

Substituting $e = u - U$ and using the same first order approximation as in the analysis for ordinary differential equations, we have

$$(e, \psi)_V = B[U, \phi] - (f(U), \phi)_{L^2}. \tag{6.1.8}$$

Similary to the analysis for ordinary differential equations, we solve (6.1.5) by triangulating Ω into K elements, and we can similarly break up both (6.1.7) and (6.1.8) as a summation over each element contribution to the error.

6.1.3 The effects of deterministic error in the evaluation of the approximate map

We let W denote the numerical solution to (6.1.2) at the reference parameter μ. This is the numerical approximation around which we linearize the forward problem in order to construct an adjoint. We define the approximate adjoint using (6.1.3) with "perturbed" operator $D_u f(W; \mu)$, and let Φ denote the solution to this approximate adjoint, which is computed using a numerical scheme, e.g. second order continuous Galerkin finite element method, for

$$B[v, \Phi] = (v, D_u f(W; \mu)^{\top}\Phi)_{L^2} + (v, \psi)_V, \ \forall v \in V. \tag{6.1.9}$$

As with the analysis for ordinary differential equations, we assume the numerical error of this solution is negligible, and we have the same assumption that f is three times continuously differentiable. Therefore, for W sufficiently close to w over short time,

$$\|D_u f(w;\mu) - D_w f(W;\mu)\|_V \leq K \|w - W\|_V .$$

We now seek to quantify the effect of the numerical error of using W in the construction of the adjoint problem for the representation formula given by (6.1.4).

Let $\check{q}(\lambda)$ denote the approximate quantity of interest calculated using (6.1.4) with W and Φ in place of w and ϕ, which is to say that

$$\check{q}(\lambda) \approx (W, \psi)_V + (D_\lambda f(W;\mu)(\lambda - \mu), \Phi)_{L^2}. \tag{6.1.10}$$

The effect of the numerical error in W on the representation formula for $q(\lambda)$ is

$$\begin{aligned} q(\lambda) - \check{q}(\lambda) \;\approx\; & \underbrace{(D_u f(w;\mu)(\lambda - \mu), \phi)_{L^2} - (D_u f(W;\mu)(\lambda - \mu), \Phi)_{L^2}}_{\textbf{II}} \\ & + \underbrace{(w - W, \psi)_V}_{\textbf{I}} . \end{aligned} \tag{6.1.11}$$

Term **I** is a linear functional of the error $w - W$ and it can be estimated by standard *a posteriori* techniques as described above. The goal is to determine an estimate of term **II** that is a measure of the effect that the numerical solutions W and Φ have in the estimate for $q(\lambda)$. Specifically, **II** measures the effect of using W and Φ on the sensitivity of $q(\lambda)$ to changes in model parameters of (6.1.2). The dependence on $\lambda - \mu$ is linear, and the analysis below produces estimates that also depend on this vector linearly so that the error estimates for these terms are also linear functions of this vector. Thus, following the analysis described below for p linearly independent vectors $\lambda - \mu$, we obtain a set of error estimates such that the error for any vector $\lambda - \mu$ can be written as a linear combination from this set of error estimates, which completes the proof of Theorem 6.1.1.

Estimating term II

By adding and subtracting $(D_\lambda f(W;\mu)(\lambda-\mu),\phi)_{L^2}$ to the integrand in **II**, we rewrite **II** so that **II** = **IIa** + **IIb**, where

$$\begin{aligned} \mathbf{IIa} &= ((D_\lambda f(w;\mu) - D_\lambda f(W;\mu))(\lambda-\mu), \phi)_{L^2} \\ \mathbf{IIb} &= (D_\lambda f(W;\mu)(\lambda-\mu), (\phi-\Phi))_{L^2} \end{aligned}$$

We now estimate **IIa** and **IIb**.

Estimating term IIa

By adding and subtracting $((D_\lambda f(w;\mu){-}D_\lambda f(W;\mu))(\lambda-\mu),\Phi)_{L^2}$ to the integrand in **IIa**, we rewrite **IIa** so that **IIa** = **IIaa** + **IIab**, where

$$\begin{aligned} \mathbf{IIaa} &= ((D_\lambda f(w;\mu) - D_\lambda f(W;\mu))(\lambda-\mu), \phi-\Phi)_{L^2} \\ \mathbf{IIab} &= ((D_\lambda f(w;\mu) - D_\lambda f(W;\mu))(\lambda-\mu), \Phi)_{L^2}. \end{aligned}$$

We show that **IIaa** is of higher-order, and can be neglected. We claim that

$$\|\phi - \Phi\|_V \le C\,\|w - W\|_V^{1/2},$$

for some constant $C > 0$, and we have that the first-derivatives of $f(u;\lambda)$ are Lipschitz continuous, so

$$|\mathbf{IIaa}| \le C\,\|w - W\|_V^{3/2},$$

for some constant $C > 0$. Thus, **IIaa** is neglected in the estimate. To prove this claim, we first observe that

$$B[v, \phi - \Phi] = (v, D_u f(w; \mu)^\top \phi - D_u f(W; \mu)^\top \Phi)_{L^2}, \ \forall v \in V.$$

If W is sufficiently close to w, then $D_u f(W; \mu)^\top = D_u f(w; \mu)^\top + \epsilon$, where ϵ is a function satisfying $\|\epsilon\|_V \leq C \|w - W\|_V$. Thus,

$$B[v, \phi - \Phi] = (v, D_u f(w; \mu)^\top (\phi - \Phi))_{L^2} + (v, \epsilon\Phi)_{L^2}, \ \forall v \in V. \qquad (6.1.12)$$

Let $\eta = \phi - \Phi \in V$ and assume that $D_u f(w; \mu) \in \otimes_{i=1}^m (L^\infty(\Omega))^m$. Then η solves the weak form of the elliptic problem $L\eta = g$ in Ω and $\eta = 0$ on $\partial\Omega$, where $L\eta = -\nabla \cdot (a\nabla\eta) - D_u f(w; \mu)^\top \eta$ and $g = \epsilon\Phi$. With unique adjoint solutions, $g \in (L^2(\Omega))^m \cap (L^\infty(\Omega))^m$, we have from standard regularity results that

$$\|\eta\|_V \leq \|\eta\|_{(H^2(\Omega))^m} \leq C \|g\|_{L^2}.$$

Observe that if Ω has finite measure, then $(L^2(\Omega))^m \subset (L^1(\Omega))^m$ and $\|g\|_{L^2} \leq \|g\|_{L^1}^{1/2} \|g\|_{L^\infty}^{1/2}$ [14], and by Hölder's inequality

$$\|g\|_{L^1}^{1/2} \leq \|\epsilon\|_{L^2}^{1/2} \|\Phi\|_{L^2}^{1/2}.$$

Thus, we have

$$\|\phi - \Phi\|_V \leq C \|w - W\|_V^{1/2},$$

for some constant C. This proves the claim.

Remark 6.1.1. *The constant C above depends on $\|\epsilon\|_{L^\infty}^{1/2}$. If the Lipschitz inequality giving $\|\epsilon\|_{L^2} \leq K \|w - W\|_V^{1/2}$ can be extended to the $(L^\infty(\Omega))^m$ norm on ϵ, then the above term that is neglected can in some sense be considered $\mathcal{O}(2)$ instead of $\mathcal{O}(3/2)$, which makes it more like the case of ordinary differential equations.*

Again assuming that $f(u;\lambda)$ is three-times continuously differentiable, then

$$(D_\lambda f(w;\mu) - D_\lambda f(W;\mu))(\lambda-\mu) \approx \\ \left[(\lambda-\mu)^\top \otimes \mathbf{J}\right] \left[D_u\left(\text{vec}(D_\lambda f(W;\mu))\right)\right](w-W)\,.$$

We substitute this estimate into **IIab** so that

$$\begin{aligned} \mathbf{IIab} &\approx \left(\left[(\lambda-\mu)^\top \otimes \mathbf{J}\right] \left[D_u\left(\text{vec}(D_\lambda f(W;\mu))\right)\right](w-W)\,,\Phi\right)_{L^2} \\ &= \left((w-W), \left[D_u\left(\text{vec}(D_\lambda f(W;\mu))\right)\right]^\top \left[(\lambda-\mu)^\top \otimes \mathbf{J}\right]^\top \Phi\right)_{L^2}. \end{aligned}$$

We let

$$\psi_{IIab} = \left[D_u\left(\text{vec}(D_\lambda f(W;\mu))\right)\right]^\top \left[(\lambda-\mu)^\top \otimes \mathbf{J}\right]^\top \Phi.$$

Thus, we have represented **IIab** as a linear functional of the error in $w-W$, which can be estimated by standard *a posteriori* techniques as described above.

Estimating term IIb

We let

$$\psi_{IIb} = D_\lambda f(W,\mu)(\lambda-\mu),$$

so that

$$\mathbf{IIb} = (\psi_{IIb}, (\phi-\Phi))_{L^2}.$$

Thus, **IIb** is a linear functional of the error in the adjoint solutions $\phi-\Phi$. We apply standard *a posteriori* techniques used to estimate linear functionals of the error in the forward solutions to estimate the error in the adjoint solutions as was done above. We again define an adjoint to the *approximate* adjoint as satisfying the weak equation

$$B[z,v] - (D_u f(W;\lambda)z, v)_{L^2} = \psi_{IIb},\ \forall v \in V.$$

We perform a standard variational argument to obtain

$$\begin{aligned}\mathbf{IIb} &= B[z,\phi-\Phi] - (D_u f(W;\mu)z, \phi-\Phi)_{L^2} \\ &= B[z,\phi-\Phi] - (z, D_u f(W;\mu)^{\top}(\phi-\Phi))_{L^2}\end{aligned}$$

Using (6.1.12) in the right-hand side above, we have

$$\begin{aligned}\mathbf{IIb} &= (z, [D_u f(w;\mu) - D_u f(W;\mu)]^{\top}\phi)_{L^2} \\ &= (z, [D_u f(w;\mu) - D_u f(W;\mu)]^{\top}\Phi)_{L^2} \\ &\quad +(z, [D_u f(w;\mu) - D_u f(W;\mu)]^{\top}(\phi-\Phi))_{L^2}.\end{aligned}$$

The second term on the right-hand side has already been proven to be higher-order. Thus, it is neglected in the estimate. The first term is estimated by using second derivatives of f in the first argument to approximate the difference $D_u f(w;\mu) - D_u f(W;\mu)$. This is identical to the ODE case. We define

$$\tilde{\psi}_{IIb} = \left[D_u\left(\mathrm{vec}(D_u f(W;\mu)^{\top})\right)\right]^{\top}\left[\Phi^{\top}\otimes\mathbf{J}\right]^{\top} z,$$

and the first term is approximated by

$$(\tilde{\psi}_{IIb}, (w-W))_{L^2}$$

which is a linear functional of the error of $w - W$, and is estimable by standard *a posteriori* techniques as described above.

6.1.4 Summarizing the effect of deterministic error

There are many notational and a few subtle changes in the analysis from the ODE case presented in the previous chapter, yet the bulk of the analysis is identical in content. We summarize the results in the following

Theorem 6.1.1. *Let W and Φ denote the numerical solutions to (6.1.2) and the (6.1.9), respectively.*

Use standard a posteriori *techniques to estimate the error $w - W$ defined as $e_0 := (w - W, \psi)_{L^2}$*

Let p denote the number of model parameters.

for** $i = 1, \dots, p$ **do

Let z denote the solutions to the adjoint to the approximate adjoint problem

$$B[z, v] - (D_u f(W; \lambda) z, v)_{L^2} = D_\lambda f(W, \mu)\delta_i, \ \forall v \in V.$$

where δ_i denotes the i^{th} standard basis vector in $\mathbb{R}^p$. Set

$$\psi_{IIab} = \left[D_u \left(vec(D_\lambda f(W; \mu))\right)\right]^\top \left[\delta_i^\top \otimes \mathbf{J}\right]^\top \Phi,$$
$$\tilde{\psi}_{IIb} = \left[D_u \left(vec(D_u f(W; \mu)^\top)\right)\right]^\top \left[\Phi^\top \otimes \mathbf{J}\right]^\top z.$$

Solve (6.1.9) with data given by the above vectors and calculate the standard error representations given by

$$e_1^i := (w - W, \psi_{IIab})_{L^2}, \ e_2^i := (\tilde{\psi}_{IIb}, w - W)_{L^2}.$$

end for

The estimate of the effect of error on the HOPS representation in direction ν given by

$$\nu = \begin{pmatrix} \nu_1 & \cdots & \nu_p \end{pmatrix}^\top$$

is denoted e_ν, and given by

$$e_\nu = e_0 + \sum_{i=1}^{p} \nu_i \sum_{j=1}^{2} e_j^i.$$

Chapter 7

MULTIPLE QUANTITIES OF INTEREST

As seen in Alg. 3.3.1, the calculation of $P(b_i)$ is a computational geometry problem where the volumes of $b_i \cap A_j$ and A_j need be calculated for each j.

We first recall the case where there is one quantity of interest. Recall that $\mathbf{\Lambda} \subset \mathbb{R}^d$. We assume that none of the cells b_i intersect more than one coarse cell B_k. For simplicity, we assume the cells b_i are boxes with each edge parallel to one of the coordinate axes. Over any B_k we use the linear approximation

$$\tilde{q}(\lambda)|_{\lambda \in B_k} = q(\mu) + \langle \nabla q(\mu), (\lambda - \mu) \rangle, \tag{7.0.1}$$

where we use μ in place of μ_k to avoid any confusion with components of vectors in the discussion below.

If $b_i \cap A_j = \emptyset$, set $V_{ij} = 0$ in Alg. 3.3.1. Otherwise, $b_i \cap A_j$ defines a closed convex polytope. There exists a half-space representation $C\lambda \leq p$ for $b_i \cap A_j$, where C is a $(2 + 2d) \times d$ matrix, p is a $(2 + 2d) \times 1$ vector, and the $\lambda \in \mathbb{R}^d$ that satisfy $C\lambda \leq p$ define the closed convex polytope $b_i \cap A_j$. To see this, let $[q_{j-1}, q_j)$ denote the interval of $\mathcal{D}$ associated with the induced region A_j. Since the boundary of A_j is a set of measure zero, we take A_j to be induced from $[q_{j-1}, q_j]$ in the following derivations with no effect on the calculation of $P(b_i)$. Since $b_i \subset B_k$ for some k, we are only interested in the part of A_j contained in B_k. We use (7.0.1) for calculations in B_k, so in B_k the region A_j is defined by the inequalities

$$\begin{aligned} q(\mu) + \langle \nabla q(\mu), (\lambda - \mu) \rangle &\leq q_j, \\ q(\mu) + \langle \nabla q(\mu), (\lambda - \mu) \rangle &\geq q_{j-1}. \end{aligned}$$

We rewrite these inequalites as

$$\begin{aligned} \langle \nabla q(\mu), \lambda \rangle &\leq q_j - q(\mu) + \langle \nabla q(\mu), \mu \rangle, \\ \langle -\nabla q(\mu), \lambda \rangle &\leq q(\mu) - q_{j-1} - \langle \nabla q(\mu), \mu \rangle. \end{aligned}$$

Set the first two rows of C to be $\nabla q(\mu)^\top$ and $-\nabla q(\mu)^\top$, respectively. Set the first two entries of p to be $q_j - q(\mu) + \langle \nabla q(\mu), \mu \rangle$ and $q(\mu) - q_{j-1} - \langle \nabla q(\mu), \mu \rangle$, respectively. The remaining rows of C and p are determined by b_i. Let $\bigotimes_{n=1}^{d} [\lambda_{n,i,\min}, \lambda_{n,i,\max}]$ denote b_i, then b_i is defined by the following inequalities

$$\begin{aligned} \lambda_1 &\leq \lambda_{1,i,\max}, \\ -\lambda_1 &\leq -\lambda_{1,i,\min}, \\ &\vdots \\ \lambda_n &\leq \lambda_{n,i,\max}, \\ -\lambda_n &\leq -\lambda_{n,i,\min}, \\ &\vdots \\ \lambda_d &\leq \lambda_{d,i,\max}, \\ -\lambda_d &\leq -\lambda_{d,i,\min}. \end{aligned}$$

Construct the remaining rows of C and p as follows. For $n = 1, \ldots, d$, set rows $2 + (2n - 1)$ and $2 + 2n$ of C (p) to be $e_n^\top$ and $-e_n^\top$ ($\lambda_{n,i,\max}$ and

$-\lambda_{n,i,\min})$, respectively, where e_n denotes the n^{th} standard basis vector in $\mathbb{R}^d$. We now have a half-space representation for the closed convex polytope defined by $b_i \cap A_j$. We use a code that utilizes the Quickhull algorithm for computing the volume of $b_i \cap A_j$. It is necessary to find the points defining the convex hull of $b_i \cap A_j$. In $\mathbb{R}^d$, a point can be defined by the intersection of d manifolds of dimension $(d-1)$. We search for points defining the convex hull of $b_i \cap A_j$ by solving $\tilde{C}\lambda = \tilde{p}$, where $\tilde{C}$ is any one of $d2^d$ invertible matrices created by choosing a specific set of d rows of C, and $\tilde{p}$ is a $d \times 1$ vector formed by choosing the same set of rows of p. For any point to be on the convex hull it must satisfy the inequality $C\lambda \leq p$. If a point solves $\tilde{C}\lambda = \tilde{p}$ and satisfies the inequality $C\lambda \leq p$, then that point is stored. After all $d2^d$ points are found and possibly stored, the code returns the volume of the convex polytope defined by $b_i \cap A_j$.

Remark 7.0.2. *We see that C has $d2^d$ invertible submatrices of size $d \times d$ by first rewriting C in the block matrix form*

$$C = \begin{pmatrix} D \\ F_1 \\ \dots \\ F_d \end{pmatrix},$$

where $D, F_1, \ldots, F_d$ are all $2 \times d$ matrices where the two rows are linearly dependent. The linearly dependent rows correspond to parallel hyperplanes in $\mathbb{R}^d$, and there are $d+1$ pairs of such hyperplanes. By construction, a row from F_n is linearly independent of any row from F_m where $n \neq m$. By assumption, a row from D is linearly independent from the the rows of F_n for each $n = 1, \ldots, d$. Geometrically, the points on the convex hull of $b_i \cap A_j$ are determined from intersection points of d hyperplanes that satisfy the inequality $C\lambda \leq p$. We choose d pairs from the $d+1$ pairs hyperplanes, *and there are d ways to do this. For each choice of d pairs, we choose one hyperplane from each pair to obtain d hyerplanes that intersect at a point. This results in 2^d choices of hyperplanes for each choice of d pairs of hyperplanes. Thus, there are $d2^d$ invertible submatrices of size $d \times d$.*

The volume returned is stored as $v(i,j)$. The volume of A_j can be determined from the sum

$$\sum_{i=1}^{M'} v(i,j).$$

The ratio of volumes of $b_i \cap A_j$ to A_j can be determined from the calculation

$$\frac{v(i,j)}{\sum_{i=1}^{M'} v(i,j)}$$

The value of $P(b_i)$ is equal to $\sum_{j=1}^{N(M)} V_{ij} P_j$ according to Alg. 3.3.1, where V_{ij} is the matrix with (ij)-entry the ratio of volumes of $b_i \cap A_j$ to A_j, so $P(b_i)$ can be determined from the calculation

$$\sum_{j=1}^{N(M)} \left[\frac{v(i,j)}{\sum_{i=1}^{M'} v(i,j)} \right] P_j.$$

Now consider the problem of multiple quantities of interest. The level of computational difficulty is not increased, but there is a subtle change to initial calculations in Alg. 3.3.1. The output space $\mathcal{D}$ is now a subset of $\mathbb{R}^m$ where m is the number of quantities of interest. The map $q(\lambda)$ is a vector-valued function, where each component function $q_l(\lambda)$ for $l = 1, \ldots, m$ is approximated by a piecewise-linear function obtained from an adjoint analysis. For simplicity, we assume that $\rho_{\mathcal{D}}(q)$ can be approximated by a simple function approximation

$$\rho_{\mathcal{D}}^{(M)}(q) = \sum_{j=1}^{M_{\mathcal{D}}} P(b_{\mathcal{D},j}) \mathbf{1}_{b_{\mathcal{D},j}}(q),$$

where $\{b_{\mathcal{D},j}\}_{j=1}^{M_q}$ partitions $\mathcal{D}$ into $M_{\mathcal{D}}$ fine boxes. Let $\bigotimes_{l=1}^{m}[q_{l,j,\min}, q_{l,j,\max}]$ denote box $b_{\mathcal{D},i}$. For any fixed $l \in \{1, \dots, m\}$, the interval $[q_{l,j,\min}, q_{l,j,\max}]$ induces a region of contours $A_{l,j}$ in $\mathbf{\Lambda}$. Since we consider multiple quantities of interest, set $A_j = \bigcap_{l=1}^{m} A_{l,j}$. Analogous to the problem with a single quantity of interest, for any $\lambda \in A_j$, $q(\lambda)$ is approximately a uniform random variable with probability $P(b_{\mathcal{D},j})$. The remainder of Alg. 3.3.1 is unchanged with only the details of calculating the volumes of $b_i \cap A_j$ in $\mathbf{\Lambda}$ slightly different from above.

The matrix C is now a $(2m + 2d) \times d$ matrix, p is now a $(2m + 2d) \times 1$ vector, and the $\lambda \in \mathbb{R}^d$ that satisfy $C\lambda \leq p$ still define the closed convex polytope $b_i \cap A_j$. For $l = 1, \dots, m$ set the $2m - 1$ and $2m$ rows of C (p) to be $\nabla q_l(\mu)^\top$ and $-\nabla q_l(\mu)^\top$ $(q_{l,j,\max} - q_l(\mu) + \langle \nabla q_l(\mu), \mu \rangle$ and $q_l(\mu) - q_{l,j,\min} - \langle \nabla q_l(\mu), \mu \rangle)$, respectively. The remaining rows of C and p are constructed exactly as before. There are now

$$\gamma := \frac{(2m + 2d)!}{d!(2m + 2)!}$$

choices of d pairs of hyperplanes, and 2^d choices for each selection of d pairs of hyperplanes to determine the possible points on the convex hull defined by $b_i \cap A_j$. Thus, there are $\gamma 2^d$ linear problems to be solved.

Example 7.0.1. *Consider again*

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

The solution of which represents the intersection of an ellipse and hyperbola. We initially considered the quantity of interest to be the first coordinate of the solution in the first quadrant. We now consider two quantities of interest taken to be the first and second coordinate of the solution in the first quadrant, respectively.

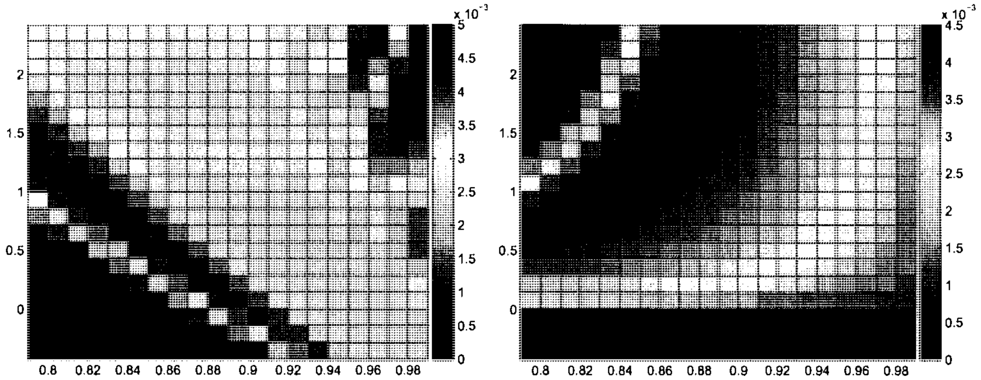

Figure 7.1: Left Plot: Posterior resulting from first quantity of interest. Right Plot: Posterior resulting from second quantity of interest

Distributions of these quantities of interest are determined by solving the forward problem with $\mathbf{\Lambda} = [.79, .99] \times [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}]$ *and uniform distribution on* λ_1 *and normal distribution on* λ_2*. The results are summarized in Fig 7.1-7.2*

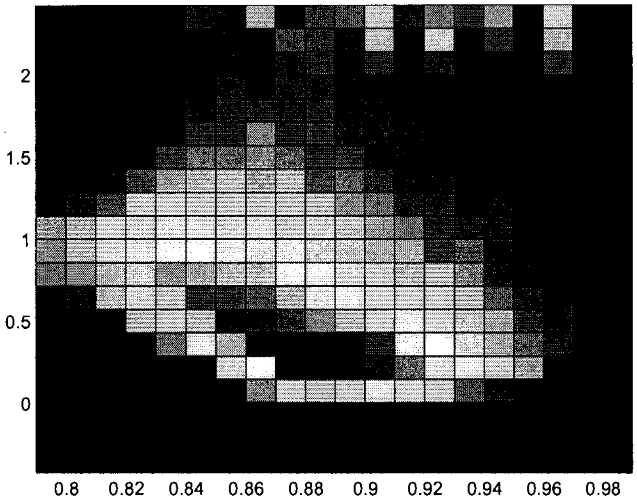

Figure 7.2: The result of using both quantity of interest maps and output distributions to obtain a posterior. Elements of both distributions in this "combined" distribution are clearly visible

Bibliography

[1] J.M. Bernardo. Reference posterior distributions for Bayesian infercncc. *J.R. Statist. Soc.*, 41:113–147, 1979.

[2] P. Billingsley. *Probability and Measure.* John Wiley & Sons, Inc., 1995.

[3] T. Butler and D. Estep. A computational measure theoretic approach to inverse sensitivity problems I: Basic method and analysis. *submitted to SIMA, October 2008.*

[4] T. Butler and D. Estep. A computational measure theoretic approach to inverse scnsitivity problcms III: Applications. *In preparation*, 2009.

[5] T. Butler and D. Estep. A computational measure theoretic approach to inverse sensitivity problems IV: Numerical studies. *In preparation*, 2009.

[6] D. Estep, M. J. Holst, and A. Målqvist. Nonparametric density estimation for randomly perturbed elliptic problems III: Convergence and a priori analysis. In prcparation, 2008.

[7] D. Estep, M. G. Larson, and R. D. Williams. Estimating the error of numerical solutions of systems of reaction-diffusion equations. *Mem. Amer. Math. Soc.*, 146(696):viii+109, 2000.

[8] D. Estep, Målqvist A.and, and S. Tavener. Nonparametric density estimation for randomly perturbed elliptic problems I: Computational methods, a postcriori analysis, and adaptivc crror control. In preparation, 2008.

[9] D. Estep, A. Målqvist, and S. Tavener. Nonparametric density estimation for randomly perturbed clliptic problems II: Applications and adaptive modeling. In preparation, 2008.

[10] D. Estep, B. Mckeown, D. Neckels, and J. Sandelin. GAASP: Globally Accurate Adaptive Sensitivity Package, 2006. write to estep@math.colostate.edu for information.

[11] D. Estep and D. Neckels. Fast and reliable methods for determining the evolution of uncertain parameters in differential equations. *J. Comput. Physics*, 213:530–556, 2006.

[12] D. Estep and D. Neckels. Fast methods for determining the evolution of uncertain parameters in reaction-diffusion equations. *Computer Methods in Applied Mechanics and Engineering*, 196:3967–3979, 2007.

[13] J.P. Huelsenbeck et al. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.*, 51:673–688, 2002.

[14] G. Folland. *Real Analysis.* John Wiley & Sons, Inc., 1999.

[15] J.E. Gentle. *Random Number Generation and Monte Carlo Methods.* Springer, 2003.

[16] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* CRC Press, 1995.

[17] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems.* Springer, 2005.

[18] D. Kaplan and L. Glass. *Understanding Nonlinear Dynamics.* Springer, 1995.

[19] D.C Knill and W. Richards. *Perception as Bayesian Inference.* Cambridge University Press, 1996.

[20] D. Neckels. *Variational methods for Uncertainty Quantification.* PhD thesis, Department of Mathematics, Colorado State University, Fort Collins, CO 80523, 2005.

[21] C.P. Robert and George Casella. *Monte Carlo Statistical Methods.* Springer, 2004.

[22] R. Clark Robinson. *An Introduction to Dynamical Systems Continuous and Discrete.* Prentice Hall, 2004.

[23] J. Sandelin. *Global Estimate and Control of Model, Numerical, and Parameter Error.* PhD thesis, Department of Mathematics, Colorado State University, Fort Collins, CO 80523, 2006.

[24] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc., 1980.

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