DISSERTATION

INFINITE DIMENSIONAL STOCHASTIC INVERSE PROBLEMS

Submitted by

Lei Yang

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Advisor: Donald Estep

F. Jay Breidt
Simon Tavener
Wen Zhou

ABSTRACT

INFINITE DIMENSIONAL STOCHASTIC INVERSE PROBLEMS

In many disciplines, mathematical models such as differential equations, are used to characterize physical systems. The model induces a complex nonlinear measurable map from the domain of physical parameters to the range of observable Quantities of Interest (QoI) computed by applying a set of functionals to the solution of the model. Often the parameters can not be directly measured, and people are confronted with the task of inferring information about values of the parameters given the measured or imposed information about the values of the QoI. In such applications, there is generally significant uncertainty in the measured values of the QoI. Uncertainty is often modeled using probability distributions. For example, a probability structure imposed on the domain of the parameters induces a corresponding probability structure on the range of the QoI. This is the well known Stochastic Forward Problem that is typically solved using a variation of the Monte Carlo method. This dissertation is concerned with the Stochastic Inverse Problems (SIP) where the probability distributions are imposed on the range of the QoI, and problem is to compute the induced distributions on the domain of the parameters. In our formulation of the SIP and its generalization for the case where the physical parameters are functions, main topics including the existence, continuity and numerical approximations of the solutions are investigated.

Chapter 1 introduces the background and previous research on the SIP. It also gives useful theorems, results and notation used later. Chapter 2 begins by establishing a relationship between Lebesgue measures on the domain and the range, and then studies the form of solution of the SIP and its continuity properties. Chapter 3 proposes an algorithm for computing the solution of the SIP, and discusses the convergence of the algorithm to the true solution. Chapter 4 exploits the fact that a function can be represented by its coefficients with respect to a basis, and extends the SIP framework to allow for cases where the domain representing the basis coefficients is a count-

able cube with decaying edges, referred to as the infinite dimensional SIP. We then discusses how its solution can be approximated by the SIP for which the domain is the finite dimensional cube obtained by taking a finite dimensional projection of the countable cube. Chapter 5 begins with an algorithm for approximating the solution of the infinite dimensional SIP, and then proves the algorithm converges to the true solution. Chapter 6 gives a numerical example showing the effects of different decay rates and the relation to truncation to finite dimensions. Chapter 7 reviews popular probabilistic inverse problem methods and proposes a combination of the SIP and statistical models to address problems encountered in practice.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my advisor Prof. Donald Estep. His vision on this subject made my dissertation possible. Without his patience, knowledge and support, I would never have finished my dissertation. I could not imagine having a better advisor and mentor for my Ph.D studies.

Besides my advisor, I would like to thank the rest of my dissertation committee: Prof. F. Jay Breidt, Prof. Simon Tavener, and Prof. Wen Zhou: for their insightful comments and advice. Prof. F. Jay Breidt has also given me a lot of help in my other research projects.

My sincere thanks also goes to Prof. Troy D. Butler and Dr. Nishant Panda for their valuable suggestions and insights on my research.

I would like to thank many people in the Statistics Department at Colorado State University. I am thankful to my fellow graduate students for the stimulating discussions and all the fun we had together. I thank all the faculty for their help in answering my questions and giving me advice. I thank the staff for their prompt help in every problem I had.

I would like to thank my family for their unconditional love and support throughout the years. Without encouragement and help from my parents, I could not have made it through this challenging journey. I thank my husband, Fei Jiang, for accompanying me through the good times and bad. I thank my daughter Nicole, for making my life more colourful and worthwhile.

The research in this dissertation was supported in part by the US Department of Energy (DE-SC0009279).

DEDICATION

To my grandma Zhongxiu Wu, who took care of me, taught me to be strong, and passed away on Dec. 17th, 2016 without me being by her side.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Deterministic Inverse Problem

Mathematical models such as differential equations, are used to characterize physical systems of many scientific and engineering disciplines, such as physics, mechanical engineering and meteorology. Solutions of these mathematical model are usually determined by a few driving parameters. In many cases, the solution itself is not of interest. Instead, Quantities of Interest (QoI) computed as functionals of solution of the mathematical model are the focus of study.

Let $Q_1 : \mathbf{\Lambda} \to S$, a map from the parameter space $\mathbf{\Lambda}$ to solution space S, be induced by the mathematical model. Let $Q_2 : S \to \mathcal{D}$ map the mathematical model solution to QoI. Then the composite operator $Q = Q_1 \circ Q_2$, mapping the parameter to QoI, is the full model under consideration. To simplify later discussion, we assume $\mathcal{D} = Q(\mathbf{\Lambda})$.

While Q is defined theoretically, the exact functional form of Q is usually unavailable. In practice, Q is often solved by numerical methods and subject to errors, which potentially have strong impact on the output and pose important research questions. To simplify discussion, we shall not address errors arising in this process and treat Q as being computed exactly. But we acknowledge the fact that evaluation of Q may be expensive and slow.

We refer to the process of computing $Q(\lambda)$ for a given λ as the deterministic forward problem, which, despite its computational cost, is usually straightforward. On the contrary, when given one $y \in \mathcal{D}$, The task of determining one $\lambda \in \mathbf{\Lambda}$ such that $Q(\lambda) = y$ for a given $y \in \mathcal{D}$ is not so straightforward as the there might be multiple solutions or the solution does not depend continuously on the data.

If we define the inverse image of any set $A \subset \mathcal{D}$ as

$$Q^{-1}(A) = \{\lambda : Q(\lambda) \in A, \lambda \in \mathbf{\Lambda}\},$$

then $Q^{-1}(\{y\})$ is a uniquely defined set and is the set-valued solution of deterministic inverse problem for $y \in \mathcal{D}$. We simply write $Q^{-1}(\{y\})$ as $Q^{-1}(y)$. This definition of set-valued solutions gets past the issue of non-uniqueness in the usual deterministic inverse problem, but poses new questions as to how continuity can be defined in set setting.

1.2 Stochastic Inverse Problem

We look at a motivating example from Breidt et al. (2011), which we paraphrase as follows. A batch of metal plates are made to withstand high temperature. While purity of the alloy and thickness are part of the manufacturing specification, they are inevitably subject to variations, and affect the heat distribution when a heat source is placed on the plate. The heat distribution can modelled by the heat equation under a given load for a given conductivity which is further determined by the alloy composition and the thickness of the plates. At the same time, temperature measurements on certain locations can be collected on a sample of plates after they are heated for a specific period of time.

If we treat the purity of alloy and the thickness as a random vector Λ taking values in $\mathbf{\Lambda}$, and temperature measurements as realizations from distribution of a random vector Y taking values in $\mathcal{D}$, and assume other parameters for the heat equation such as the initial condition and the boundary condition for the heat equation as fixed and known, then the heat equation induces a deterministic map Q from parameter space $\mathbf{\Lambda}$ to the space of temperature measurements $\mathcal{D}$. The probability distribution of Λ determines the probability distribution of Y. Specifically, to understand how stochasticity is propagated by Q, we need to further assume that Q is a measurable map from measurable space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ to measurable space $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$. Then any $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ induces a $P_{\mathcal{D}}$ defined in the following way: for all $A \in \mathcal{B}_{\mathcal{D}}$, $P_{\mathcal{D}}(A) = P_{\mathbf{\Lambda}}(Q^{-1}(A))$. One common approach of getting independent random samples distributed as $P_{\mathcal{D}}$ is to generate independent random samples distributed as $P_{\mathbf{\Lambda}}$ and apply Q to each of the sample points. The problem of defining, computing and sampling from $P_{\mathcal{D}}$ for a given $P_{\mathbf{\Lambda}}$ is known as the Stochastic Forward Problem (SFP).

The Stochastic Inverse Problem (SIP), as its name suggests, is the direct inverse of the SFP. Specifically, for $Q : \mathbf{\Lambda} \to \mathcal{D}$, SIP seeks to identify and compute a probability measure $P_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ that induces the given probability measure $P_{\mathcal{D}}$ on $\mathcal{D}$. Just as with the deterministic inverse problem, there might be multiple solutions to one given $P_{\mathcal{D}}$. Further restrictions may yield a unique solution.

In recent years, a theoretical framework of the SIP have been established through a series of papers (Breidt et al., 2011; Butler et al., 2012, 2014). They also propose computational algorithms for the SIP which are implemented in the BET package (Butler, Estep, Tavener Method - A python based package for measure-theoretic stochastic inverse and forward problems. McDougall 2016). Among the papers, Butler et al. (2014) presents the latest and most comprehensive work, but also leaves some problems which inpire the reformulation of the SIP constituting a major part of our work. Next, we present the formulation of the SIP in Butler et al. (2014).

1.3 A Formulation of the SIP

The formulation of Butler et al. (2014) is based on the setup that there is a deterministic map $Q : \mathbf{\Lambda} \to \mathcal{D}$ and $\mathbf{\Lambda}$ is compact subset in $\mathbb{R}^n$, $\mathcal{D} \subset \mathbb{R}^m$, and $n \geq m$. They also implicitly assume $\mathcal{D} = Q(\mathbf{\Lambda})$, so that $Q^{-1}(y)$ is well-defined for every $y \in \mathcal{D}$. They assume Q is (piecewise) continuously differentiable, and the component maps of Q are geometrically distinct (GD), characterized as:

Definition 1.3.1 (Geometrically Distinct (GD))**.** *Component maps of m-dimensional piecewise-smooth vector valued map $Q(\lambda)$ are geometrically distinct (GD) if the Jacobian of Q has full rank.*

By the Implicit Function Theorem 1.5.1 introduced later, the inverse set $Q^{-1}(y)$ is a piecewise $n - m$ dimensional manifold in $\mathbf{\Lambda}$ called generalized contour for each $y \in \mathcal{D}$. In the measure theory formulation, Butler et al. (2014) use the solution of deterministic inverse problem of point y consisting of $Q^{-1}(y)$, which is a set-valued solution that can be interpreted as an equivalence class. Given a probability measure space $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, P_{\mathcal{D}})$, Q induces a σ-algebra $\mathcal{C}_{\mathbf{\Lambda}}$, the point of which is the generalized contours $Q^{-1}(y)$, and probability $P_{\mathcal{C}_{\mathbf{\Lambda}}}$ on $\mathbf{\Lambda}$. Therefore, as a counterpart to deterministic inverse problem, $(\mathbf{\Lambda}, \mathcal{C}_{\mathbf{\Lambda}}, P_{\mathcal{C}_{\mathbf{\Lambda}}})$ is the direct solution of the SIP for $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, P_{\mathcal{D}})$.

While $P_{\mathcal{C}_{\mathbf{\Lambda}}}$ is a solution to the SIP, they determine a probability measure $P_{\mathbf{\Lambda}}$ defined with respect to $\mathcal{B}_{\mathbf{\Lambda}}$ instead of $\mathcal{C}_{\mathbf{\Lambda}}$, such that $P_{\mathcal{D}}(A) = P_{\mathbf{\Lambda}}(Q^{-1}(A))$ for $A \in \mathcal{B}_{\mathcal{D}}$. $P_{\mathcal{D}}$ is called the induced measure of $\mu_{\mathbf{\Lambda}}$ via Q.

Defining an equivalence relation on $\mathbf{\Lambda}$ as "being mapped to the same point in $\mathcal{D}$", the resulting partion is essentially a decomposion of $\mathbf{\Lambda}$ as union of manifolds $\{Q^{-1}(y)\}$ indexed by $y \in \mathcal{D}$. This space of equivalent class, $\{Q^{-1}(y), y \in \mathcal{D}\}$, is denoted as $\mathcal{L}$. Then Q as a map from $\mathcal{L}$ to $\mathcal{D}$ is one-to-one and onto, thus is an isomorphism between $\mathcal{L}$ and $\mathcal{D}$, and Q induces $(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, P_{\mathcal{L}})$ by the probability space $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, P_{\mathcal{D}})$.

1.3.1 Disintegration Theorem

In practice, people want to know the probabilities of sets in $\mathcal{B}_{\mathbf{\Lambda}}$ instead of $\mathcal{B}_{\mathcal{D}}$, so it is desirable to compute a solution of the SIP with respect to $\mathcal{B}_{\mathbf{\Lambda}}$. For $A \in \mathcal{B}_{\mathcal{D}}$, define $\mathcal{E}_A \in \mathcal{B}_{\mathcal{L}}$ such that $Q(\mathcal{E}_A) = Q(A)$, then define the equivalence map $\pi_{\mathcal{L}} : \mathbf{\Lambda} \to \mathcal{L}$ by $\pi_{\mathcal{L}}(A) = \mathcal{E}_A$. To analyze the relation between solutions regarding the two different σ-algebras, Butler et al. (2014) adapt the disintegration theorem from Chang and Pollard (1997) as follows.

Theorem 1.3.2 (Disintegration Theorem)**.** *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps, and assume that Ψ is a measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ and $\mu_{\mathcal{L}}$ is its induced measure on $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$ via $\pi_{\mathcal{L}}$. There is a family of measures $\{\Psi_\ell\}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ defined for almost every $\ell \in \mathcal{L}$ such that*

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

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

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

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

As already discussed, $P_{\mathcal{D}}$ induces $P_{\mathcal{L}}$ as the solution of the SIP through the isomorphism between $\mathcal{D}$ and $\mathcal{L}$. If there is a $P_{\mathbf{\Lambda}}$, defined with respect to $\mathcal{B}_{\mathbf{\Lambda}}$, as a solution of the SIP for probability measure $P_{\mathcal{D}}$, the Disintegration Theorem yields

$$P_{\mathbf{\Lambda}}(A) = \int_{\mathcal{E}_A} P_N(A;\ell)dP_{\mathcal{L}}(\ell) = \int_{\mathcal{E}_A} \left(\int_{\pi_{\mathcal{L}}^{-1}(\ell)\cap A} P_N(\lambda;\ell) \right) dP_{\mathcal{L}}(\ell),$$

where $\{P_N(\lambda;\ell)\}$ is a family of measures concentrated on $\pi_{\mathcal{L}}^{-1}(\ell)$ for each $\ell \in \mathcal{L}$.

1.3.2 Uniform Ansatz

While the marginal distribution on $\mathcal{L}$ is entirely determined by the SIP, the family of conditional distributions $\{P_N(\cdot;\ell)\}$ are not determined by the inverse of Q. Butler et al. (2012) specify an ansatz assuming the values of $\{P_N(\cdot;\ell)\}$. Defining $\mu_{\mathcal{L}} = \pi_{\mathcal{L}}\mu_{\mathbf{\Lambda}}$, disintegrating the volumn (Lebesgue) measure $\mu_{\mathbf{\Lambda}}$ with respect to $\mu_{\mathcal{L}}$ determines a family of conditional distributions $\{\mu_N(\cdot;\ell)\}$ by

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

In the absence of knowledge regarding the conditional probability distributions $\{P_N(\lambda;\ell)\}$, Butler et al. (2014) use an ansatz that assigns probability mass evenly on the generalized contours, i.e. the “Uniform Ansatz" defined as

$$dP_N(\lambda;\ell) = \rho_N(\lambda;\ell)d\mu_N(\lambda;\ell) \quad \text{with} \quad \rho_N(\cdot;\ell) = \frac{1}{\mu_N(\pi_{\mathcal{L}}^{-1};\ell)(\ell)} \quad \ell \in \mathcal{L}.$$

When one has no preference over different regions on the same generalized contour, letting $\{P_N(\cdot;\ell)\}$ be proportional to $\{\mu_N(\cdot;\ell)\}$ is the most reasonable choice. The Uniform Ansatz determines the probability measure as the solution to the SIP in an almost sure sense. Numerical methods to the SIP under the Uniform Ansatz given by Butler et al. (2014) is discussed in next section.

1.3.3 Numerical Method

The goal of solving the SIP numerically is to approximate $P_{\mathbf{\Lambda}}(A)$ for $A \in \mathcal{B}_{\mathbf{\Lambda}}$. According to Butler et al. (2014), this involves three approximation issues:

1. approximation of events in $\mathcal{B}_{\mathcal{D}}$
2. approximation of events in $\mathcal{B}_{\mathbf{\Lambda}}$
3. approximation of events in $\mathcal{C}_{\mathbf{\Lambda}}$

They let $\{I_i\}$ and $\{b_j\}$ be generating sequences for σ-algebras $\mathcal{B}_{\mathcal{D}}$ and $\mathcal{B}_{\mathbf{\Lambda}}$, and $\{I_i\}_{i=1}^{M}$ and $\{b_j\}_{j=1}^{N}$ be a finite partition of $\mathcal{D}$ and $\mathbf{\Lambda}$ with subsets taken from these sequences. Any event in $\mathcal{B}_{\mathcal{D}}$, $\mathcal{B}_{\mathbf{\Lambda}}$, can be approximated by such partitions, and since $\mathcal{C}_{\mathbf{\Lambda}} \subset \mathcal{B}_{\mathbf{\Lambda}}$, events in $\mathcal{B}_{\mathbf{\Lambda}}$ can also be approximated by partition $\{b_j\}_{j=1}^{N}$. They refer to both the rectangles from a regular grid and cells from a Voronoi tesselation generated from random sampling as proper generating sequence.

They aim to compute the probability of each b_j as an approximation of the solution of the SIP by the following algorithm, which have variations in different scenarios.

Algorithm 1: Approximation to inverse probability.

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

In Theorem 4.6, where they prove the convergence for computed probabilty of a given set, they assume $P_{\mathcal{D}}$ has a density with respect to $\mu_{\mathcal{D}}$ that is piecewise continuous. In the proof, they show the above algorithm amounts to approximating the solution probability density on $\mathbf{\Lambda}$ by a simple function taking values on the partition sets. They show that the simple function converges to the true probability density a.e. and by using the dominated convergence theorem, they prove the convergence of the computed probabilities.

1.4 Disintegration theorem

Butler et al. (2014) apply the disintegration theorem in $\mathcal{L}$, the space of equivalence classes represented as manifolds in $\mathbf{\Lambda}$ and analyze the computed approximations using measure theory in $\mathbf{\Lambda}$, we use the disintegration theorem in $\mathcal{D}$. The original disintegration theorem (Chang and Pollard, 1997, Theorem 1,2) is stated in the range of a map, the data that is used to create a distribution on $\mathcal{D}$ is represented in $\mathcal{D}$, and the algorithms in Butler et al. (2014) involve computations in $\mathcal{D}$. So here we consider the disintegration theorem in $\mathcal{D}$.

Theorem 1.4.1 (Disintegration Theorem 2)**.** *Let $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a measurable space and $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable map with GD component maps, and assume that Ψ is a measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ and $\Psi_{\mathcal{D}}$ is its induced measure on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ via Q. There is a family of measures $\{\Psi_y\}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ uniquely defined for $\Psi_{\mathcal{D}}$-almost every $y \in \mathcal{D}$ such that*

$$\Psi_y(Q^{-1}(y)) = 1, \Psi_y(\mathbf{\Lambda} \backslash Q^{-1}(y)) = 0,$$

which gives the following disintegration of Ψ:

$$\Psi(A) = \int_{Q(A)} \Psi_y(A) d\Psi_{\mathcal{D}}(y) = \int_{Q(A)} \left(\int_{Q^{-1}(y)\cap A} d\Psi_y(\lambda) \right) d\Psi_{\mathcal{D}}(y)$$

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

1.5 The implicit function theorem

Generally speaking, the implict function theorem states how an equation for a function of several variables determines the relation among variables such that some variables can be expressed as a function of other variables.

We use two versions of the implicit function theorem. The first one from P224, Rudin et al. (1964) is a standard result in multivariate calculus. This one is applied for many results in the formulation of SIP, for example, it is key to prove that $Q^{-1}(y)$ is a manifold.

Theorem 1.5.1. *Let f be a C^1-mapping of an open set $X \times Y \subset \mathbb{R}^{n+m}$, where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are both open sets, into $\mathbb{R}^n$, such that $f(\bar{x}, \bar{y}) = 0$ for some point $(\bar{x}, \bar{y}) \in E$. Put $A_x = J_{f,x}(\bar{x}, \bar{y})$ and assume that A_x is invertible. Also put $A_y = J_{f,y}(\bar{x}, \bar{y})$. Then there exist open sets $U \subset X$ and $W \subset Y$, with $(\bar{x}, \bar{y}) \in U \times W$, having the following property:*

To every $y \in W$ corresponds a unique x such that

$$x \in U \quad \textit{and} \quad f(x, y) = 0.$$

If this x is defined to be $g(y)$, then g is a C^1-mapping of W into U, $g(b) = a$,

$$f(g(y), y) = 0 \quad (y \in W),$$

and

$$g'(b) = -(A_x)^{-1} A_y.$$

The scope of above theorem is limited to a Euclidean space, so it is inadequate for cases where infinite dimensional spaces are involved. For the infinite dimensional case, we use Theorem 3 of Border (2013).

Theorem 1.5.2. *Let X be an open subset in $\mathbb{R}^n$, let Y be a metric space, and let $f : X \times Y \to \mathbb{R}^n$ be continuous. Suppose the derivative $J_{f,x}$ of f with respect to x exists at each point (x, y) and is* *continuous on $X \times Y$. Assume that $J_{f,x}(\bar{x}, \bar{y})$ is invertible. Let*

$$f(\bar{x}, \bar{y}) = 0.$$

Then there are neighborhoods $U \subset X$ and $W \subset Y$ of $\bar{x}$ and $\bar{y}$, and a function $\xi : W \to U$ such that:

1. *$f(\xi(y); y) = 0$ for $y \in W$.*
2. *For each $y \in W$, $\xi(y)$ is the unique solution to $f(x, y) = 0$ lying in U. In particular, then $\xi(\bar{y}) = \bar{x}$.*
3. *ξ is continuous on W.*

1.6 Background in differential geometry

A (topological) manifold is a topological space that is locally Euclidean, i.e. for every point on it, there is a neighborhood that is isomorphic to an open ball in $\mathbb{R}^n$. We deal with a very specific manifold prevalent in the SIP formulation: a k-dimensional smooth manifold in $\mathbb{R}^n$. Even though we only assume Q is C^1 and the differential structure on the manifold is C^1, Theorem 2.10 in Chapter 2 of Hirsch (1976) asserts that for $r \geq 1$, every C^r manifold is C^r diffeomorphic to a C^∞ manifold, so consideration of a C^∞ manifold, or smooth manifold, is sufficient for most purposes. We use manifold and smooth manifold interchangeably. We apply definitions from Schlichtkrull (2013) to k-dimensional manifold in $\mathbb{R}^n$ unless otherwise specified.

Definition 1.6.1 (Smooth function)**.** *If f has continuous partial derivatives up to order r, then f is called a C^r−function. A function which is C^r for all r is called C^∞ or smooth.*

Definition 1.6.2. *A parametrized manifold in $\mathbb{R}^n$ is a smooth map $\sigma : U \to \mathbb{R}^n$, where $U \subset \mathbb{R}^k$ is a non-empty open set. It is called regular if the $n \times k$ Jacobian matrix $J_\sigma(x)$ has rank k at all $x \in U$, that is, σ has geometrically distinct component maps. An k-dimensional parametrized manifold is a parametrized manifold $\sigma : U \to \mathbb{R}^n$ with $U \subset \mathbb{R}^k$, which is regular.*

Definition 1.6.3. *Let $A \subset \mathbb{R}^k$ and $B \subset \mathbb{R}^n$. A map $f : A \to B$ which is continuous, bijective and has a continuous inverse is called a homeomorphism.*

Definition 1.6.4. *A regular parametrized manifold $\sigma : U \to \mathbb{R}^n$ which is a homeomorphism $U \to \sigma(U)$, is called an embedded parametrized manifold.*

Definition 1.6.5. *Let $A \subset \mathbb{R}^n$. A subset $B \subset A$ is said to be relatively open if it has the form $B = A \cap W$ for some open set $W \subset \mathbb{R}^n$.*

Definition 1.6.6. *An k-dimensional manifold in $\mathbb{R}^n$ is a non-empty set $S \subset \mathbb{R}^n$ satisfying the following property for each point $p \in S$. There exists an open neighborhood $N(p)$ of p and an k-dimensional embedded parametrized manifold $\sigma : U \to \mathbb{R}^n$ with image $\sigma(U) = S \cap N(p)$. The surrounding space $\mathbb{R}^n$ is said to be the ambient space of the manifold.*

A k-dimensional manifold is locally a k-dimensional embedded parametrized manifold, which further admits a specific parameterization as the following Lemma.

Lemma 1.6.7. *Let $S \subset \mathbb{R}^n$ be non-empty. Then S is an k-dimensional manifold if and only if it satisfies the following condition for each $p \in S$: There exists an open neighborhood $N(p) \subset \mathbb{R}^n$ of p, such that $S \cap N(p)$ is the graph of a smooth function h, where $n - k$ of the variables $x_1, \cdots, x_n$ are considered as functions of the remaining k variables.*

By the Implicit Function Theorem 1.5.1, we can show that $Q^{-1}(y), y \in \mathcal{D}$ is a $n - m$ dimensional manifold in $\mathbb{R}^n$, just as the next theorem.

Theorem 1.6.8. *Let $f : \mathbf{\Omega} \to \mathbb{R}^m$ be a C^1 function, where $m \leq n$ and where $\mathbf{\Omega} \subset \mathbb{R}^n$ is open, and let $c \in \mathbb{R}^m$. If it is not empty, the set*

$$\mathcal{S} = \{p \in \mathbf{\Omega} | f(p) = c, rank(J_f(p)) = m\}$$

is an $n - m$-dimensional smooth manifold in $\mathbb{R}^n$.

For a C^1 map $f : U \to V$, where U, V are open subsets of $\mathbb{R}^n$, $\mathbb{R}^m$, we know the linear approximation of f is given by the Jacobian matrix J_f. Next, we define differential as the linear approximation of a map between manifolds $f : \mathcal{M} \to \mathcal{N}$. The next definition is from Wilson (2012, Definition 9).

Definition 1.6.9 (Tangent Space $T_x\mathcal{M}$, Differential)**.** *Suppose that $\mathcal{M} \subset \mathbb{R}^n$ is a k-dimensional manifold. Let $\phi : U \to \mathcal{M}$, where U is an open set in $\mathbb{R}^k$, be a local parameterization around some point $x \in \mathcal{M}$ with $\phi(0) = x$. We define the tangent space $T_x\mathcal{M}$ to be the image of the map $J_\phi(0) : \mathbb{R}^k \to \mathbb{R}^n$ at 0. Note that $T_x\mathcal{M}$ is an k-dimensional subspace of $\mathbb{R}^n$, and its translation $x + T_x\mathcal{M}$ is the linear approximation to ϕ at x.*

Given a $C^r, r \geq 1$ map of manifolds $f : \mathcal{M} \to \mathcal{N}$, and local parameterizations $\phi : U \to \mathcal{M}, \phi(0) = x \in \mathcal{M}$ and $\psi : V \to N, \psi(0) = f(x) \in \mathcal{N}$. Let h be the map $h = \psi^{-1} \circ f \circ \phi : U \to V$. We can define the differential of f at x by

$$df_x : T_x\mathcal{M} \to T_{f(x)}\mathcal{N} \quad df_x = J_\psi(0) \circ J_h(0) \circ J_\phi(0)^{-1}.$$

The spaces $T_x\mathcal{M}$, $T_{f(x)}\mathcal{N}$, and the differential df_x are independent of choice of ϕ and ψ.

Next definition is from Wilson (2012, Definition 12).

Definition 1.6.10 (Submersion;Immersion)**.** *A $C^r, r \geq 1$ map of smooth manifolds $f : \mathcal{M} \to \mathcal{N}$ is a submersion if its differential is surjective at every point. The map is an immersion if the differential is injective at every point.*

A submersion is an extension of a projection map, and it is also an open map. The GD condition insures the map Q is a submersion when $n > m$.

Proposition 1.6.11 (Properties of Submersions)**.** *Let $\mathcal{M}$ and $\mathcal{N}$ be smooth manifolds, and suppose $f : \mathcal{M} \to \mathcal{N}$ is a submersion. Then f is an open map.*

On the contrary, when the differential df_x is not surjective, x is called critical point, and its value is called critical value, fomally defined as follows.

Definition 1.6.12 (Critical values)**.** *Given a $C^r, r \geq 1$ map of manifolds $f : \mathcal{M} \to \mathcal{N}$, the set of critical values of f is defined as*

$$\{f(p) \in \mathcal{N} | \operatorname{rank}(df_p) < \dim \mathcal{N}\}.$$

Next classical result from Sard (1965) shows that when f is smooth enough, the set of critical values is negligible.

Theorem 1.6.13 (Morse-Sard)**.** *If $\mathcal{N}$ is an n-dimensional manifold and $\mathcal{M}$ is an m-dimensional manifold, $f : \mathcal{N} \to \mathcal{M}$ is a C^k map for $k > \max\{n - m, 0\}$, then the set of critical values of f has measure zero.*

1.7 Some useful results

In this section we introduce some useful results.

1.7.1 Smith-Volterra-Cantor set

One famous set often used in construction of counter examples is the Smith-Volterra-Cantor set, or fat Cantor set, denoted as $C^{\frac{1}{4}}$. It is constructed by recursively removing the central $1/2^{2n}$ from each bar beginning with $[0, 1]$. This set is nowhere dense and consequently has no interior, so it is the boundary of itself. Yet it has a positive measure of $1/2$. The next figure taken from Wikipedia (2017) shows the first five steps towards construction of the fat Cantor set.

Figure 1.1: The first five steps towards construction of the Smith-Volterra-Cantor set.

1.7.2 Lebesgue Differentiation theorem

For a Lebesgue integrable real valued function f on $\mathbb{R}^n$, the integral of f over any measurable set A is $\int_A f d\mu$. The derivative of integral at x is defined as

$$\lim_{r \to 0} \frac{\int_{N(x,r)} f d\mu}{\int_{N(x,r)} d\mu}.$$

The Lebesgue Differentiation Theorem (Stein and Shakarchi, 2005, Chapter 3, Theorem 1.3), states the derivative of integral exists and is equal to f at almost every point $x \in \mathbb{R}^n$.

Theorem 1.7.1. *If f is integrable on $\mathbb{R}^n$, then*

$$\lim_{r \to 0} \frac{\int_{N(x,r)} f d\mu}{\int_{N(x,r)} d\mu} = f(x) \quad \textit{for almost every } x.$$

1.7.3 Uniform integrability

The well known dominated convergence theorem shows that if a sequence of functions $\{f_i\}$ on $\mathbf{\Omega}$ converge to f a.e. and are dominated by an integrable function g, then f is integrable and $\{f_i\}$ converges to f in L^1, that is $\int_{\mathbf{\Omega}} |f_i - f| d\mu_{\mathbf{\Omega}} \to 0$

Sometimes a dominating function is hard to find, and uniform integrability of $\{f_n\}$ yields the same convergence result. First, we shall give the definition of uniform integrability.

Definition 1.7.2 (Uniform integrability)**.** *Let $(\mathbf{\Omega}, \mathcal{B}_{\mathbf{\Omega}}, \mu_{\mathbf{\Omega}})$ be a measure space. A set $\Phi \subset L^1(\mathbf{\Omega})$ is called uniformly integrable if to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that*

$$\int_E |f| d\mu_{\mathbf{\Omega}} < \epsilon$$

whenever $f \in \Phi$ and $\mu_{\mathbf{\Omega}}(E) < \delta$.

Next theorem from Rudin (1987, Chapter 6, Problem 10) states the desired convergence is implied from uniform integrability.

Theorem 1.7.3 (Vitali convergence theorem). *If $\mu_{\mathbf{\Omega}}(\mathbf{\Omega}) < \infty$, and $\{f_i\}$ be a sequence that is uniformly integrable and $f_i \to f$ a.e., and $f < \infty$ a.e., then $f \in L^1(\mathbf{\Omega})$ and $\int_{\mathbf{\Omega}} |f_i - f| d\mu_{\mathbf{\Omega}} \to 0$.*

1.8 Some notation

In this section we define some notation used throughout this dissertation.

Let X be a metric space with metric d. For a point $x_0 \in X$, we denote an open ball centered at x_0 as $N(x_0, r) = \{x : d(x, x_0) < r, x \in X\}$. When the radius is not important, we may just use $N(x_0)$ to refer to an open ball centered at x_0.

For $A \subset X$, we define the interior of A to be the set

$$\text{int}(A) = \{x \in A | \text{exists } W(x, r_x) \subset A, r_x > 0\}$$

We define the closure of A to be the set

$$\overline{A} = \{x \in X | x = \lim_{i \to \infty} a_i, \text{with } \{a_i\}_{i=1}^{\infty} \subset A\}.$$

It can be proved that int(A) is open and $\overline{A}$ is closed, and int$(A) \subset A \subset \overline{A}$. The boundary of A is usually defined as $\partial A = \overline{A} \setminus \text{int}(A)$.

The symmetric difference between two sets A and B is denoted by $\triangle$,

$$A \triangle B = (A \setminus B) \bigcup (B \setminus A).$$

1.9 Goal of this dissertation

While the previous formulation of the SIP is directly based on the space of equivalent class $\mathcal{L}$, formulated and analyzed as computations in $\mathbf{\Lambda}$, there is an isomorphism between $\mathcal{L}$ and $\mathcal{D}$, and the Borel $\sigma-$algebra and probability on $\mathcal{L}$ are all induced by their counterparts on $\mathcal{D}$, we may restate the results in Butler et al. (2014) in terms of $\mathcal{D}$. Furthermore, the available data and sample points with respect to which computations are carried out are all on $\mathcal{D}$ rather than on $\mathcal{L}$. So it is entirely possible and natural to reformulate the SIP with respect to $\mathcal{D}$.

The formulation with respect to $\mathcal{D}$ enables us to address many questions include the relation between $\mu_{\mathcal{L}}$ and $P_{\mathcal{L}}$; and the specification of density $\rho_N(\cdot;\ell)$ on each generalized contour under the Uniform Ansatz.

Our work in this dissertation begins with a significant reformulation of the SIP in terms of $\mathcal{D}$. Although our formulation does not directly involve $\mathcal{L}$, we address similar questions in the disintegration with respect to $\mathcal{D}$, and use them as a starting point to give a precise form of the solution to the SIP. We also propose a modified version of algorithm and show its convergence under some mild conditions.

We refer to the SIP with respect to $Q : \mathbf{\Lambda} \to \mathcal{D}$ where $\mathbf{\Lambda} \subset \mathbb{R}^n$ as "SIP with finite dimensional domain", or "finite dimensional SIP". The second part of our work concerns the extension to the SIP for which the domain $\mathbf{\Lambda}$ is $[0,1]^\infty$ endowed with metric d. Motivation for this extension comes from applications where functions rather than finite dimensional vectors are potential inputs to the mathematical model Q. Accordingly, this extension is called "SIP with infinite dimensional domain", or "infinite dimensional SIP". We discuss the existence of solution, and its theoretical and numerical approximations. Lastly, we present a numerical example for the infinite dimensional SIP.

The last part is a review of other popular approaches of inverse problems where the probability distributions of physical parameters are of interest. While the SIP starts with a given probability distribution, it needs to be coupled with statistical methods to solve any practical problems where only data are available. The introduction of SIP coupled with statistical models shows new perspective on the SIP and suggests new research directions.

Chapter 2

Reformulation of the SIP in terms of computations on the range

In this chapter, we present a comprehensive reformulation of the SIP in terms of the range. The problem setup resembles the problem setup in Butler et al. (2014). We let $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a C^1 function, where $\mathbf{\Lambda} \subset \mathbb{R}^n$ is compact and $\mathcal{D} \subset \mathbb{R}^m$ and $m \leq n$. We assume $\mathcal{D} = Q(\mathbf{\Lambda})$, however we sometimes expand $\mathcal{D}$ to a slightly bigger open set. It is easy to infer which case $\mathcal{D}$ belongs to from the context. We assume J_Q is full rank on $\mathbf{\Lambda}$. The results can be generalized to the case that J_Q is full rank except for a finite union of $n-1$-dimensional manifolds in $\mathbf{\Lambda}$, by union or summation. This condition is called geometrically distinct of component maps in Butler et al. (2014). When $n > m$, Q : $\text{int}\mathbf{\Lambda} \to Q(\text{int}\mathbf{\Lambda})$ is a submersion and when $n = m$, it is a diffeomorphism.

To impose probability structures on $\mathbf{\Lambda}$ and $\mathcal{D}$, we let $\mathbf{\Lambda}$ and $\mathcal{D}$ inherit the Borel $\sigma-$algebra and Lebesgue measures from measure spaces $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu_{\mathbb{R}^n})$ and $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m}, \mu_{\mathbb{R}^m})$ respectively, and denote the measure spaces as $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ and $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$. When we say a measure is defined on a space without specifying its σ-algebra, we implicitly take the σ-algebra to be the default Borel σ-algebra.

For $Q : \mathbf{\Lambda} \to \mathcal{D}$ where $\mathcal{D} = Q(\mathbf{\Lambda})$, we define the set function Q^{-1} as

$$Q^{-1}(B) = \{\lambda : \lambda \in \mathbf{\Lambda}, Q(\lambda) \in B\},$$

for $B \subset \mathcal{D}$. When $\mathcal{D}$ is a larger set, we generalize Q^{-1} by letting $Q^{-1}(B) = \varnothing$, for $B \subset \mathcal{D}, B \cap Q(\mathbf{\Lambda}) = \varnothing$, then $Q^{-1}(B) = Q^{-1}(B \cap Q(\mathbf{\Lambda}))$ for $B \subset \mathcal{D}$ in general. We write $Q^{-1}(y)$ for $Q^{-1}(\{y\})$ for simplicity.

In later sections, Q is assumed to be defined on a larger set $U_{\mathbf{\Lambda}}$ containing $\mathbf{\Lambda}$, making it a function $Q : U_{\mathbf{\Lambda}} \to \mathcal{D}$, where $Q(U_{\mathbf{\Lambda}}) \subset \mathcal{D}$. Then Q^{-1} should be defined in terms of $U_{\mathbf{\Lambda}}$ accordingly, that is

$$Q^{-1}(B) = \{\lambda : \lambda \in U_{\mathbf{\Lambda}}, Q(\lambda) \in B\},$$

for $B \subset \mathcal{D}$.

For $Q : U_{\mathbf{\Lambda}} \to \mathcal{D}$, we use $Q^{-1}(B) \cap \mathbf{\Lambda}$ to denote the preimage restricted to $\mathbf{\Lambda}$.

We define the induced measure $\tilde{\mu}_{\mathcal{D}}$ on $\mathcal{D}$ by $\tilde{\mu}_{\mathcal{D}}(A) = \mu_{\mathbf{\Lambda}}(Q^{-1}(A))$ for $A \in \mathcal{B}_{\mathcal{D}}$, and denote it as $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$. This notation also applies to other induced measures.

We begin the formulation in Section 2.1 by showing that $\tilde{\mu}_{\mathcal{D}}$ is dominated by $\mu_{\mathcal{D}}$ with density of a given expression. In Section 2.2, we apply the Disintegration Theorem to $\mu_{\mathbf{\Lambda}}$ and $P_{\mathbf{\Lambda}}$, and by utilizing the relation between $\tilde{\mu}_{\mathcal{D}}$ and $\mu_{\mathcal{D}}$ and a more precise definition of the Uniform Ansatz, we define a unique density as solution of the SIP. In Section 2.3, we present an simple example. In Section 2.4, we discuss conditions that guarantee the continuity of the solution density of the SIP.

2.1 Relationship between the Lebesgue measures on the domain and the range of Q

The goal of the SIP for Q is to define and calculate a probability measure $P_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ for a given probability measure $P_{\mathcal{D}}$ on $\mathcal{D}$, such that $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$. It is usually assumed that $P_{\mathcal{D}}$ is dominated by $\mu_{\mathcal{D}}$. Under this assumption, whether the solution probability measure $P_{\mathbf{\Lambda}}$ is dominated by $\mu_{\mathbf{\Lambda}}$ is closely related to the relationship between $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$ and Lebesgue measure $\mu_{\mathcal{D}}$.

In next theorem, we prove $\tilde{\mu}_{\mathcal{D}}$ is absolutely continuous with respect to $\mu_{\mathcal{D}}$, with the density being an integration over each manifold $Q^{-1}(y)$. This result is the cornerstone of the theory of the SIP.

Theorem 2.1.1. *Let $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a C^1 function, where $\mathbf{\Lambda} \subset \mathbb{R}^n$ is compact with codomain $\mathcal{D} \subset \mathbb{R}^m$ where $m \leq n$, and the Jacobian J_Q is full rank. Let $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ be the Lebesgue measures on $\mathbf{\Lambda}$ and $\mathcal{D}$. Let $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$, then $d\tilde{\mu}_{\mathcal{D}} = \tilde{\rho}_{\mathcal{D}}(y)d\mu_{\mathcal{D}}$, where*

$$\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds.$$

Note this formula also applies for $y \in \mathcal{D} \setminus Q(\mathbf{\Lambda})$, for which $\tilde{\rho}_{\mathcal{D}}(y) = 0$ and $Q^{-1}(y) = \varnothing$.

Proof. We prove that for any generalized rectangle $[a, b] \subset \mathbb{R}^m$, $b > a$,

$$\tilde{\mu}_{\mathcal{D}}([a, b]) = \int_{[a,b]} \int_{Q^{-1}(y)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds d\mu_{\mathcal{D}}(y).$$

The general result follows from the Caratheodory Extension Theorem and the fact that family of sets $\{[a, b]\}$ is a semi-algebra.

Since J_Q is full rank for every $\lambda \in Q^{-1}([a, b])$, there are m indices $i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, m\}$ such that $J_Q^{(m)} = \left(J_Q^{(i_1)}, J_Q^{(i_2)}, \ldots, J_Q^{(i_m)}\right)$ is invertible at λ, where $J_Q^i = i^{th}$ column of J_Q. Since J_Q is continuous, there is an open ball $N(\lambda, r)$ of radius r centered at λ, such that J_Q^m is invertible in $N(\lambda, r) \cap Q^{-1}([a, b])$.

$Q^{-1}([a, b])$ is a closed subset of $\mathbf{\Lambda}$ so it is also compact. From $\{N(\lambda, r)\}$, we choose a collection U_j with center points $\lambda_1, \ldots, \lambda_J$, such that $Q^{-1}([a, b]) \subset \bigcup_{j=1}^{J} U_j$. Further we let $U_1 \leftarrow U_1$, $U_j \leftarrow U_j \setminus \overline{(\bigcup_{k=1}^{j-1} U_k)}$ for $j > 1$ so that U_j are disjoint open sets whose union is $\mathbf{\Lambda}$ except for a negligible set.

In any $U_j \cap Q^{-1}([a, b])$, we have $(\lambda) \sim (\lambda^{(m_j)}; \hat{\lambda}^{(m_j)})$ of which $\lambda^{(m_j)}$ are m entries of λ for which the Jacobian $J_Q^{(m_j)}$ is invertible, and $\hat{\lambda}^{(m_j)}$ the remaining $n - m$ entries with Jacobian $\hat{J}_Q^{(m_j)}$. Define $\pi_j : \mathbb{R}^n \to \mathbb{R}^{n-m}$ as $\pi_j(\lambda) = \hat{\lambda}^{(m_j)}$. Then define $V_j = \{(Q(\lambda), \pi_j(\lambda)); \lambda \in U_j \cap Q^{-1}([a, b])\}$, and for any $y \in [a, b]$, define $V_{jy} = \{\pi_j(\lambda); \lambda \in Q^{-1}(y) \cap U_j\}$. For now, we fix j and drop superscript. When we consider all $U_j \cap Q^{-1}([a, b])$ we rewrite superscript j.

V_j is diffeomorphic to $U_j \cap Q^{-1}([a, b])$ with:

$$d\lambda^{(m)} d\hat{\lambda}^{(m)} = \left|\det(J_Q^{(m)})^{-1}\right| dy d\hat{\lambda}^{(m)}.$$

Fixing y, in each U_j, by the implicit function theorem, $\lambda^{(m)} = f(\hat{\lambda}^{(m)})$, with the $J_f = (J_Q^{(m)})^{-1} \hat{J}_Q^{(m)}$, so the volume form of manifold $Q^{-1}(y) \cap U_j$ has the local representation as

$$
\begin{aligned}
ds &= \sqrt{\det((J_Q^{(m)})^{-1}\hat{J}_Q^{(m)}, I)((J_Q^{(m)})^{-1}\hat{J}_Q^{(m)}, I)^T} d\hat{\lambda}^{(m)} \\
&= \sqrt{\det((J_Q^{(m)})^{-1}\hat{J}_Q^{(m)}\hat{J}_Q^{(m)T}(J_Q^{(m)})^{-T} + I)} d\hat{\lambda}^{(m)} \\
&= k(\lambda) d\hat{\lambda}^{(m)},
\end{aligned}
$$

where

$$
\begin{aligned}
k(\lambda) &= \sqrt{\det((J_Q^{(m)})^{-1}(\hat{J}_Q^{(m)}\hat{J}_Q^{(m)T} + J_Q^{(m)}J_Q^{(m)T})(J_Q^{(m)})^{-T})} \\
&= \sqrt{\det(J_Q J_Q^T)} \left|\det(J_Q^{(m)})^{-1}\right|.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\tilde{\mu}_{\mathcal{D}}([a,b]) = \mu_{\mathbf{\Lambda}}(Q^{-1}([a,b])) &= \sum_{j=1}^{J} \mu_{\mathbf{\Lambda}}(U_j \cap Q^{-1}([a,b])) \\
&= \sum_{j=1}^{J} \int_{U_j \cap Q^{-1}([a,b])} d\lambda_i^{(m_j)} d\hat{\lambda}_i^{(m_j)} \\
&= \sum_{j=1}^{J} \int_{V_j} \left|\det(J_Q^{(m_j)})^{-1}\right| dy d\hat{\lambda}_i^{(m_j)} \\
&= \int_{[a,b]} dy \sum_{j=1}^{J} \int_{V_{jy}} \left|\det(J_Q^{(m_j)})^{-1}\right| d\hat{\lambda}_i^{(m)} \\
&= \int_{[a,b]} dy \sum_{j=1}^{J} \int_{V_{jy}} \frac{\left|\det(J_Q^{(m_j)})^{-1}\right|}{k_j(\lambda)} k_j(\lambda) d\hat{\lambda}_i^{(m)} \\
&= \int_{[a,b]} dy \sum_{j=1}^{J} \int_{V_{jy}} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds \\
&= \int_{[a,b]} \int_{Q^{-1}(y)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds d\mu_{\mathcal{D}}(y).
\end{aligned}
$$

$\square$

The expression for $\tilde{\rho}_{\mathcal{D}}(y)$ shows it does not depend on the parameterization of the generalized contour.

We know that $\tilde{\rho}_{\mathcal{D}}(y) = 0$ when $y \notin Q(\mathbf{\Lambda})$. We want to find conditions that guarantee $\tilde{\rho}_{\mathcal{D}}(y) > 0$.

Theorem 2.1.2. *For $y \in \mathcal{D}$, such that $Q^{-1}(y) \cap int\mathbf{\Lambda} \neq \varnothing$, $\tilde{\rho}_{\mathcal{D}}(y) > 0$.*

Proof. If $y \in \mathcal{D}$, and $Q^{-1}(y) \cap \text{int}\mathbf{\Lambda} \neq \varnothing$, there exists $\lambda \in \text{int}\mathbf{\Lambda}$ such that $Q(\lambda) = y$. So there exists a neighborhood of λ, $N(\lambda)$, with $N(\lambda) \subset \text{int}\mathbf{\Lambda}$. $Q^{-1}(y) \cap N(\lambda)$ is diffeomorphic to an open set in $\mathbb{R}^{n-m}$. As a result,

$$\tilde{\rho}_{\mathcal{D}}(y) > \int_{Q^{-1}(y)\cap N(\lambda)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds > 0.$$

□

Theorem 2.1.3. *If $\mu_{\mathcal{D}} \left(Q(\mathbf{\Lambda}) \setminus Q(int\mathbf{\Lambda}) \right) = 0$, $\{y : y \in Q(\mathbf{\Lambda}), \tilde{\rho}_{\mathcal{D}}(y) = 0\}$ is a $\mu_{\mathcal{D}}$-null set.*

Proof. The results follows from the previous theorem that $\{y : y \in Q(\mathbf{\Lambda}), \tilde{\rho}_{\mathcal{D}}(y) = 0\} \subset Q(\mathbf{\Lambda}) \setminus Q(\text{int}\mathbf{\Lambda})$. □

Theorem 2.1.3 shows that when $\mu_{\mathcal{D}} \left(Q(\mathbf{\Lambda}) \setminus Q(\text{int}\mathbf{\Lambda}) \right) = 0$, $\tilde{\mu}_{\mathcal{D}}$ is dominated by $\mu_{\mathcal{D}}$ on $Q(\mathbf{\Lambda})$, hence

$$\frac{1}{\tilde{\rho}_{\mathcal{D}}(y)} d\tilde{\mu}_{\mathcal{D}} = d\mu_{\mathcal{D}}.$$

In this case, we say $\tilde{\rho}_{\mathcal{D}}$ and $\rho_{\mathcal{D}}$ are equivalent on $Q(\mathbf{\Lambda})$.

2.2 The Solution of the SIP

We assume that $P_{\mathcal{D}}$ is dominated by $\mu_{\mathcal{D}}$, i.e. $dP_{\mathcal{D}} = \rho_{\mathcal{D}} d\mu_{\mathcal{D}}$. This assumption allows any theoretic result regarding $\mu_{\mathcal{D}}$ to be carried over to $P_{\mathcal{D}}$.

In Butler et al. (2014), a connection between $P_{\mathcal{L}}$ and $P_{\mathbf{\Lambda}}$ is established by identifying the isomorphism between $(\mathcal{L}, \mathcal{B}_{\mathcal{L}}, P_{\mathcal{L}})$ and $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, P_{\mathcal{D}})$ and using Disintegration Theorem 1.3.2. Alternatively, we use Theorem 1.4.1 to perform disintegration for the measure on $\mathbf{\Lambda}$ and its induced measure on $\mathcal{D}$, including $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$. For $A \in \mathcal{B}_{\mathbf{\Lambda}}$,

$$P_{\mathbf{\Lambda}}(A) = \int_{Q(A)} P_N(A; y) dP_{\mathcal{D}}(y),$$

where $\{P_N(\cdot; y)\}$ is a family of probability measures on $\mathbf{\Lambda}$ concentrated on $Q^{-1}(y)$, that is unique for $P_{\mathcal{D}}$-almost all y. Consider the disintegration of Lebesgue measure

$$\mu_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \mu_N(A; y) d\tilde{\mu}_{\mathcal{D}}(y),$$

where $\{\mu_N(\cdot; y)\}$ is a probability measure on $\mathbf{\Lambda}$ concentrated on $Q^{-1}(y)$, that is unique for $\tilde{\mu}_{\mathcal{D}}$-almost all y.

In light of the two disintegrations, the relation between the unknown $P_{\mathbf{\Lambda}}$ and known $\mu_{\mathbf{\Lambda}}$ depends on the relation between $\{P_N(\cdot; y)\}$ and $\{\mu_N(\cdot; y)\}$. Such relation can not be inferred from the problem itself and must be specified beforehand, and any choice of such relation is called an ansatz with respect to $\mu_{\mathbf{\Lambda}}$. As in Butler et al. (2014), in the absense of other knowledge, we assume $P_N(\cdot; y) \equiv \mu_N(\cdot; y)$ for $\tilde{\mu}_{\mathcal{D}}$-almost all y. The above identity holds because $\{P_N(\cdot; y)\}$ and $\{\mu_N(\cdot; y)\}$ are all probability measures by Theorem 1.4.1.

Definition 2.2.1 (Uniform Ansatz)**.** *Disintegration of unknown $P_{\mathbf{\Lambda}}$ incurs a family of unknown conditional probability distributions $\{P_N(\cdot; y)\}$ that is unique for $P_{\mathcal{D}}$-almost all y, and the disintegration of the Lebesgue measure $\mu_{\mathbf{\Lambda}}$ generates a family of $\{\mu_N(\cdot; y)\}$ that is unique for $\tilde{\mu}_{\mathcal{D}}$-almost all y. We refer to the assumption $P_N(\cdot; y) \equiv \mu_N(\cdot; y)$ as the uniform ansatz of $P_{\mathbf{\Lambda}}$ with respect to $\mu_{\mathbf{\Lambda}}$.*

Theorem 2.2.2. *Let $Q : \mathbf{\Lambda} \to \mathcal{D}$ be a measurable function from the measure space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ to $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$, and $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$. Assuming $d\tilde{\mu}_{\mathcal{D}} = \tilde{\rho}_{\mathcal{D}}(y)d\mu_{\mathcal{D}}$ and $\tilde{\rho}_{\mathcal{D}}(y) > 0$, for $\mu_{\mathcal{D}}$-almost all y. Given a probability measure $P_{\mathcal{D}}$ on $\mathcal{D}$ as $dP_{\mathcal{D}} = \rho_{\mathcal{D}}(y)d\mu_{\mathcal{D}}$ for $\mu_{\mathcal{D}}$-almost all y. There is a probability measure $P_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ satisfying;*

1. $QP_{\mathbf{\Lambda}} = P_{\mathcal{D}}$;

2. *the Uniform Ansatz with respect to Lebesgue measure* $\mu_{\mathbf{\Lambda}}$.

$P_{\mathbf{\Lambda}}$ *is uniquely determined as* $dP_{\mathbf{\Lambda}} = \rho_{\mathbf{\Lambda}}(\lambda)d\mu_{\mathbf{\Lambda}}$ *up to an* $\mu_{\mathbf{\Lambda}}-$*almost a.e. sense, with*

$$\rho_{\mathbf{\Lambda}}(\lambda) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}.$$

Proof. By the Disintegration Theorem 1.4.1, for $A \in \mathcal{B}_{\mathbf{\Lambda}}$, if a solution $P_{\mathbf{\Lambda}}$ exists, it satisfies

$$P_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(y)\cap A} dP_N(\lambda; y) dP_{\mathcal{D}}(y).$$

where $\{P_N(\cdot; y)\}$ is uniquely determined up to $P_{\mathcal{D}}$-almost all y. By assumption, $P_{\mathcal{D}}$ has a density $\rho_{\mathcal{D}}$ with respect to $\mu_{\mathcal{D}}$. For Lebesgue measure $\mu_{\mathbf{\Lambda}}$, we have

$$\mu_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(y)\cap A} d\mu_N(\lambda; y) d\tilde{\mu}_{\mathcal{D}}(y),$$

where $\{\mu_N(\cdot; y)\}$ is uniquely determined up to $\tilde{\mu}_{\mathcal{D}}$-almost all y. The Uniform Ansatz implies $dP_N(\lambda; y) = d\mu_N(\lambda; y)$, and Theorem 2.1.1 and Corollary 2.1.2 shows $\frac{1}{\tilde{\rho}_{\mathcal{D}}(y)} d\tilde{\mu}_{\mathcal{D}}(y) = d\mu_{\mathcal{D}}(y)$, both in a $\tilde{\mu}_{\mathcal{D}}$-a.e. sense (since $\tilde{\mu}_{\mathcal{D}}$ and $\mu_{\mathcal{D}}$ dominate each other, $\tilde{\mu}_{\mathcal{D}}$-a.e. is equivalent to $\mu_{\mathcal{D}}$-a.e.).

Combining the disintegration of $\mu_{\mathbf{\Lambda}}$, the Uniform Ansatz and the expression for $P_{\mathcal{D}}$, we have

$$\begin{aligned}
P_{\mathbf{\Lambda}}(A) &= \int_{Q(A)} \int_{Q^{-1}(y)\cap A} dP_N(\lambda; y) dP_{\mathcal{D}}(y) \\
&= \int_{Q(A)} \int_{Q^{-1}(y)\cap A} dP_N(\lambda; y) \rho_{\mathcal{D}}(y) d\mu_{\mathcal{D}}(y) \\
&= \int_{Q(A)} \int_{Q^{-1}(y)\cap A} d\mu_N(\lambda; y) \rho_{\mathcal{D}}(y) \frac{1}{\tilde{\rho}_{\mathcal{D}}(y)} d\tilde{\mu}_{\mathcal{D}}(y) \\
&= \int_{Q(A)} \int_{Q^{-1}(y)\cap A} \rho_{\mathcal{D}}(Q(\lambda)) \frac{1}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))} d\mu_N(\lambda; y) d\tilde{\mu}_{\mathcal{D}}(y) \\
&= \int_A \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))} d\mu_{\mathbf{\Lambda}}.
\end{aligned}$$

The density $\frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}$ is unique in a $\mu_{\mathbf{\Lambda}}-$a.e. sense, though it may take different values on a subset of $Q^{-1}(A)$ where $\mu_{\mathbf{\Lambda}}(Q^{-1}(A)) = \tilde{\mu}_{\mathcal{D}}(A) = 0$. $\square$

One implication of the Uniform Ansatz is that $\rho_{\mathbf{\Lambda}}(\lambda)$ depends on $Q(\lambda)$, or $\rho_{\mathbf{\Lambda}}$ is same on $Q^{-1}(y)$. In fact, this gives another characterization of the Uniform Ansatz without directly invoking the Disintegration Theorem.

Definition 2.2.3 (Uniform Ansatz 2)**.** *Suppose the conditions in Theorem 2.2.2 hold. For the unknown $P_{\mathbf{\Lambda}}$ such that $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$, requiring $dP_{\mathbf{\Lambda}} = \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}}$ and $\rho_{\mathbf{\Lambda}}(\lambda_1) = \rho_{\mathbf{\Lambda}}(\lambda_2)$ for all $\lambda_1, \lambda_2 \in \mathbf{\Lambda}$ such that $Q(\lambda_1) = Q(\lambda_2)$ uniquely determines $\rho_{\mathbf{\Lambda}}$ up to a $\mu_{\mathbf{\Lambda}}$-a.e. sense. Such requirement is referred to as the Uniform Ansatz of $P_{\mathbf{\Lambda}}$ with respect to $\mu_{\mathbf{\Lambda}}$.*

Proof. Suppose $\rho_{\mathbf{\Lambda}}$ exists so that $dP_{\mathbf{\Lambda}} = \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}}$ and $\rho_{\mathbf{\Lambda}}(\lambda_1) = \rho_{\mathbf{\Lambda}}(\lambda_2)$ for all $\lambda_1, \lambda_2 \in \mathbf{\Lambda}$ such that $Q(\lambda_1) = Q(\lambda_2)$. Define $\hat{\rho}_{\mathcal{D}}$ so $\hat{\rho}_{\mathcal{D}}(y) = \rho_{\mathbf{\Lambda}}(\lambda)$ for any $\lambda \in Q^{-1}(y)$. Conversely, $\rho_{\mathbf{\Lambda}}(\lambda) = \hat{\rho}_{\mathcal{D}}(Q(\lambda))$. Then for $A \in \mathcal{B}_{\mathcal{D}}$,

$$P_{\mathcal{D}}(A) = P_{\mathbf{\Lambda}}(Q^{-1}(A)) = \int_{Q^{-1}(A)} \rho_{\mathbf{\Lambda}}(\lambda) d\mu_{\mathbf{\Lambda}} = \int_A \hat{\rho}_{\mathcal{D}}(y) d\tilde{\mu}_{\mathcal{D}} = \int_A \hat{\rho}_{\mathcal{D}}(y)\tilde{\rho}_{\mathcal{D}}(y) d\mu_{\mathcal{D}}$$

At the same time,

$$P_{\mathcal{D}}(A) = \int_A \rho_{\mathcal{D}}(y) d\mu_{\mathcal{D}}.$$

As a result, $\hat{\rho}_{\mathcal{D}}(y)\tilde{\rho}_{\mathcal{D}}(y) = \rho_{\mathcal{D}}(y)$, and we get

$$\hat{\rho}_{\mathcal{D}}(y) = \frac{\rho_{\mathcal{D}}(y)}{\tilde{\rho}_{\mathcal{D}}(y)}.$$

So

$$\rho_{\mathbf{\Lambda}}(\lambda) = \hat{\rho}_{\mathcal{D}}(Q(\lambda)) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}.$$

The density $\rho_{\mathbf{\Lambda}}$ is uniquely defined up to an $\mu_{\mathbf{\Lambda}}$-a.e. sense because it is allowed to take different values on a subset of $Q^{-1}(A)$ where $\mu_{\Lambda}(Q^{-1}(A)) = \tilde{\mu}_{\mathcal{D}}(A) = 0$. □

The contour $Q^{-1}(y)$ may shrink to a point when y goes to the boundary of $\mathcal{D}$. Since $\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds$, $\tilde{\rho}_{\mathcal{D}}$ converges to 0 when $Q^{-1}(y)$ shrinks to a point. In such case, if $\rho_{\mathcal{D}}$ does not goes to 0 at a same rate as $\tilde{\rho}_{\mathcal{D}}$ near the boundary of $\mathcal{D}$, $\rho_{\mathbf{\Lambda}}$ may go to infinity at the corresponding generalized contours $Q^{-1}(y)$. But such phenomenon is not a cause of concern, as $\rho_{\mathbf{\Lambda}}$ still integrates to 1, and it occurs because some arbitrary $\rho_{\mathcal{D}}$ is imposed on $\mathcal{D}$. Should $P_{\mathcal{D}}$ be generated by some $P'_{\mathbf{\Lambda}}$ of which the density is finite, the solution density $\rho_{\mathbf{\Lambda}}$ of the solution $P_{\mathbf{\Lambda}}$ under the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$ does not go to infinity at the boundary.

Although the results we have shown are expressed in terms of $\mu_{\mathbf{\Lambda}}$, especially the Uniform Ansatz, which is specifically stated with respect to $\mu_{\mathbf{\Lambda}}$, one can easily replace $\mu_{\mathbf{\Lambda}}$ with another measure $\mu'_{\mathbf{\Lambda}}$ that are equivalent to $\mu_{\mathbf{\Lambda}}$ and adapt the key steps in the above formulation to construct a solution with respect to $\mu'_{\mathbf{\Lambda}}$.

2.3 Example

We consider a simple example, with $\mathbf{\Lambda} = \{\lambda = (\lambda_1, \lambda_2) : \lambda_1^2 + \lambda_2^2 \leq 1\}$ and map $Q : \mathbf{\Lambda} \to \mathcal{D}, (\lambda_1, \lambda_2) = \lambda_1 + 2\lambda_2$.

Using Lagrange multipliers, we find that $Q(\lambda_1, \lambda_2)$ achieves minimum and maximum at points $(-\frac{\sqrt{5}}{5}, -\frac{2\sqrt{5}}{5})$ and $(\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5})$ respectively. Since $\mathbf{\Lambda}$ is connected, the range $\mathcal{D} = Q(\mathbf{\Lambda})$ is also connected, and $\mathcal{D} = [-\sqrt{5}, \sqrt{5}]$.

We want to check $d\tilde{\mu}_{\mathcal{D}}(y) = \tilde{\rho}_{\mathcal{D}}(y) d\mu_{\mathcal{D}}(y)$, where

$$\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)} \frac{1}{\sqrt{(J_Q J_Q^T)}} ds.$$

Let $y_1 = y = \lambda_1 + 2\lambda_2, y_2 = \lambda_1$, and let $\lambda = (\lambda^{(1)}, \hat{\lambda}^{(1)})$, where $\lambda^{(1)} = \lambda_2$, $\hat{\lambda}^{(1)} = \lambda_1$. Then, $J_Q = (J_Q^{(1)}, \hat{J}_Q^{(1)})$, where $J_Q^{(1)} = \dfrac{\partial Q}{\partial \lambda_2}$, $\hat{J}_Q^{(1)} = \dfrac{\partial Q}{\partial \lambda_1}$, so

$$\begin{aligned}
\left|\frac{\partial(\lambda_1, \lambda_2)}{\partial(y_1, y_2)}\right| &= \left|\det \begin{pmatrix} \frac{\partial \lambda_1}{\partial y_1} & \frac{\partial \lambda_1}{\partial y_2} \\ \frac{\partial \lambda_2}{\partial y_1} & \frac{\partial \lambda_2}{\partial y_2} \end{pmatrix}\right| = \left|\det \begin{pmatrix} \frac{\partial y_1}{\partial \lambda_1} & \frac{\partial y_1}{\partial \lambda_2} \\ \frac{\partial y_2}{\partial \lambda_1} & \frac{\partial y_2}{\partial \lambda_2} \end{pmatrix}\right|^{-1} \\
&= \left|\det \begin{pmatrix} \frac{\partial Q}{\partial \lambda_1} & \frac{\partial Q}{\partial \lambda_2} \\ 1 & 0 \end{pmatrix}\right|^{-1} = \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|} = \frac{1}{\left|J_Q^{(1)}\right|},
\end{aligned}$$

thus $d\lambda_1 d\lambda_2 = \dfrac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|} dy d\lambda_1$. We define $\pi_1 : \mathbb{R}^2 \to \mathbb{R}$ as $\pi_1((\lambda_1, \lambda_2)) = \lambda_1$, then for $[a,b] \subset \mathcal{D}$. We define $V = \{(Q(\lambda), \pi_1(\lambda)) : \lambda \in Q^{-1}([a,b])\}$, and $V_y = \{\pi_1(\lambda) : \lambda \in Q^{-1}(y)\}$. This gives

$$
\begin{aligned}
\tilde{\mu}_{\mathcal{D}}([a,b]) = \mu_{\mathbf{\Lambda}}(Q^{-1}([a,b])) &= \int_{Q^{-1}([a,b])} d\lambda_1 d\lambda_2 \\
&= \int_V \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|}\Bigg| \left(\lambda_1, \frac{y-\lambda_1}{2}\right) dy d\lambda_1 \\
&= \int_{[a,b]} dy \int_{V_y} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|}\Bigg| \left(\lambda_1, \frac{y-\lambda_1}{2}\right) d\lambda_1.
\end{aligned}
$$

For a given y, the Implicit Function Theorem implies that $\dfrac{d\lambda_2}{d\lambda_1} = -\dfrac{\frac{\partial Q}{\partial \lambda_1}}{\frac{\partial y}{\partial \lambda_2}}$, thus $ds = \sqrt{d\lambda_1^2 + d\lambda_2^2} = \sqrt{1 + \dfrac{\frac{\partial Q}{\partial \lambda_1}^2}{\frac{\partial Q}{\partial \lambda_2}^2}}\Bigg|(\lambda_1, \frac{y-\lambda_1}{2}) d\lambda_1$. Then,

$$
\begin{aligned}
\int_{[a,b]} dy \int_{V_y} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|}\Bigg|(\lambda_1, \frac{y-\lambda_1}{2}) d\lambda_1 &= \int_{[a,b]} dy \int_{Q^{-1}(y)} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|} \frac{1}{\sqrt{1 + \frac{\frac{\partial Q}{\partial \lambda_1}^2}{\frac{\partial Q}{\partial \lambda_2}^2}}} ds \\
&= \int_{[a,b]} dy \int_{Q^{-1}(y)} \frac{1}{\sqrt{\frac{\partial Q}{\partial \lambda_2}^2 + \frac{\partial Q}{\partial \lambda_1}^2}} ds.
\end{aligned}
$$

Since $\sqrt{(J_Q J_Q^T)} = \frac{\partial Q}{\partial \lambda_2}^2 + \frac{\partial Q}{\partial \lambda_1}^2$, Theorem 2.1.1 holds.

For this problem, we compute $\tilde{\rho}_{\mathcal{D}}(y)$ in practice, using a parameterization of $Q^{-1}(y)$ with respect to λ_1 or λ_2. This amounts to computing $\int_{\pi_1 Q^{-1}(y)} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|} d\lambda_1$ or $\int_{\pi_2 Q^{-1}(y)} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_1}\right|} d\lambda_2$ instead of $\int_{Q^{-1}(y)} \frac{1}{\sqrt{\frac{\partial Q}{\partial \lambda_2}^2 + \frac{\partial Q}{\partial \lambda_1}^2}} ds$.

We choose to compute $\tilde{\rho}_{\mathcal{D}}(y)$ for a given y using parameterization with respect to λ_1. First, we find boundaries of $\pi_1\left(Q^{-1}(y)\right)$ by

$$
\begin{cases}
\lambda_1^2 + \lambda_2^2 = 1, \\
\lambda_1 + 2\lambda_2 = y.
\end{cases}
$$

The solution is $\lambda_1 = \dfrac{y \pm 2\sqrt{5-y^2}}{5}$, so $\pi_1\left(Q^{-1}(y)\right) = \left[\dfrac{y-2\sqrt{5-y^2}}{5}, \dfrac{y+2\sqrt{5-y^2}}{5}\right]$. Therefore,

$$\begin{aligned}\tilde{\rho}_{\mathcal{D}}(y) &= \int_{\pi_1 Q^{-1}(y)} \frac{1}{\left|\frac{\partial Q}{\partial \lambda_2}\right|}\Big|(\lambda_1, \frac{y-\lambda_1}{2}) d\lambda_1 \\ &= \int_{\left[\frac{y-2\sqrt{5-y^2}}{5}, \frac{y+2\sqrt{5-y^2}}{5}\right]} \frac{1}{2} d\lambda_1 = \frac{2\sqrt{5-y^2}}{5}.\end{aligned}$$

As shown in Figure 2.1, here the $\tilde{\rho}_{\mathcal{D}}$ has a simple interpretation: the area under the graph of $\tilde{\rho}_{\mathcal{D}}$ on the right side and the area of the region between two corresponding generalized contours on the circle on the left side are the same.

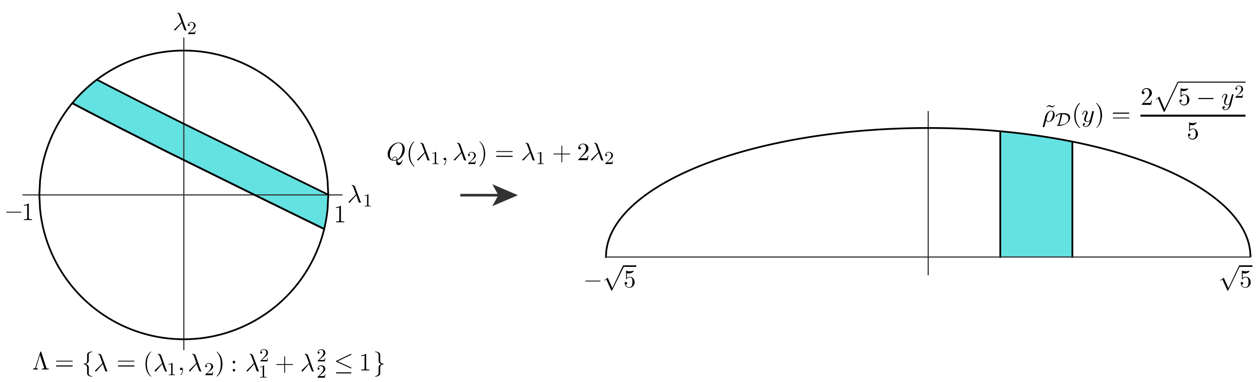


Figure 2.1: A simple example for computing $\tilde{\rho}_{\mathcal{D}}$.

If $P_{\mathcal{D}}$ is a probability measure on $\mathcal{D}$, with $dP_{\mathcal{D}}(y) = \rho_{\mathcal{D}}(y) d\mu_{\mathcal{D}}$, and we assume there is a probability distribution $P_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ with 2 conditions:

1. $QP_{\mathbf{\Lambda}} = P_{\mathcal{D}}$,

2. the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$,

then $P_{\mathbf{\Lambda}}$ is has the form

$$dP_{\mathbf{\Lambda}}(\lambda_1, \lambda_2) = \frac{\rho_{\mathcal{D}}(Q(\lambda_1, \lambda_2))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda_1, \lambda_2))} d\mu_{\mathbf{\Lambda}}.$$

Here, we define $\rho_{\mathbf{\Lambda}}(\lambda_1, \lambda_2) = \dfrac{\rho_{\mathcal{D}}(Q(\lambda_1, \lambda_2))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda_1, \lambda_2))}$.

Suppose we assume a uniform distribution on $\mathcal{D}$, then $dP_{\mathcal{D}}(y) = \frac{1}{2\sqrt{5}} d\mu_{\mathcal{D}}(y)$. We have,

$$\rho_{\mathbf{\Lambda}}(\lambda_1, \lambda_2) = \frac{\dfrac{1}{2\sqrt{5}}}{\dfrac{2\sqrt{5-(\lambda_1+2\lambda_2)^2}}{5}} = \frac{\sqrt{5}}{4\sqrt{5-(\lambda_1+2\lambda_2)^2}}.$$

Note that when $(\lambda_1, \lambda_2) \to (-\frac{\sqrt{5}}{5}, -\frac{2\sqrt{5}}{5}), (\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5})$, $\rho_{\mathbf{\Lambda}}(\lambda_1, \lambda_2) \to \infty$, since $Q^{-1}(\sqrt{-5})$ (or $Q^{-1}(\sqrt{5})$) degenerates to a point.

We check $\rho_{\mathbf{\Lambda}}(\lambda_1, \lambda_2)$ is indeed a probability distribution by

$$\begin{aligned}
\int_{\Lambda} \frac{\sqrt{5}}{4\sqrt{5-(\lambda_1+2\lambda_2)^2}} d\lambda_1 d\lambda_2 &= \int_{(-\sqrt{5},\sqrt{5})} \int_{\left[\frac{y-2\sqrt{5-y^2}}{5}, \frac{y+2\sqrt{5-y^2}}{5}\right]} \frac{\sqrt{5}}{4\sqrt{5-y^2}} \frac{1}{2} d\lambda_1 dy \\
&= \int_{(-\sqrt{5},\sqrt{5})} \frac{\sqrt{5}}{4\sqrt{5-y^2}} \frac{4}{5} \sqrt{5-y^2} \frac{1}{2} dy \\
&= \int_{(-\sqrt{5},\sqrt{5})} \frac{1}{2\sqrt{5}} dy = 1.
\end{aligned}$$

2.4 Continuity of the density

In practice, we often assume $\rho_{\mathcal{D}}$ is continuous a.e.. It is also desirable for the solution density of the SIP to be continuous a.e. as well. By the formula

$$\rho_{\mathbf{\Lambda}}(\lambda) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))},$$

this holds if $\tilde{\rho}_{\mathcal{D}}$ is also continuous a.e..

In this section, we explore sufficient conditions that guarantee that $\tilde{\rho}_{\mathcal{D}}$ is continuous. This investigation also gives insights into the understanding of the geometry in the SIP, and provides a sufficient condition for convergence of a numerical algorithm approximating the solution of the SIP.

We proved $\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds$ in Theorem 2.1.1. Since Q is C^1, J_Q is continuous and consequently $\frac{1}{\sqrt{\det(J_Q J_Q^T)}}$ is continuous. A formal definition of "continuity" of the set $Q^{-1}(y)$ is difficult. At least it should satisfy

$$\sup_{\lambda' \in Q^{-1}(y')} d(\lambda', Q^{-1}(y)) \to 0 \text{ as } y' \to y. \tag{2.1}$$

Whether (2.1) holds depends on how the boundary of $Q^{-1}(y)$ changes as y changes, which in turn depends on how $Q^{-1}(y)$ interacts with $\partial\mathbf{\Lambda}$.

To characterize the intersection of $Q^{-1}(y)$ and $\partial\mathbf{\Lambda}$, we assume that Q is defined on an open set $U_{\mathbf{\Lambda}}$ containing $\mathbf{\Lambda}$ and $\partial\mathbf{\Lambda}$ is locally determined by $B(\lambda) = 0$, where B is a C^1 function defined on $U_{\mathbf{\Lambda}}$. We exclude the case that the boundary has redundant part by assuming $\mathbf{\Lambda} = \overline{\text{int}\mathbf{\Lambda}}$. Then for each piece of $\partial\mathbf{\Lambda}$, int$\mathbf{\Lambda}$ lies on one side of $\partial\mathbf{\Lambda}$, so int$\mathbf{\Lambda}$ is locally determined by $B(\lambda) < 0$ or $B(\lambda) > 0$. For the generalized contour $Q^{-1}(y)$ in $U_{\mathbf{\Lambda}}$, the boundary of $Q^{-1}(y)$ in $\mathbf{\Lambda}$ is $Q^{-1}(y) \cap \partial\mathbf{\Lambda}$, which is determined by the augmented system:

$$\begin{cases} Q(\lambda) - y = 0, \\ B(\lambda) = 0. \end{cases}$$

While the domain of Q is enlarged to $U_{\mathbf{\Lambda}}$, we still define $\tilde{\mu}_{\mathcal{D}}$ with respect to $\mathbf{\Lambda}$, that is

$$\tilde{\mu}_{\mathcal{D}}(B) = \mu_{\mathbf{\Lambda}}(Q^{-1}(B) \cap \mathbf{\Lambda}),$$

for $B \subset \mathcal{D}$. Abusing the notation, we still denote it as $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$. The next theorem gives a sufficient condition for $\tilde{\rho}_{\mathcal{D}}(y)$ to be continuous at point y, namely the intersection of $\partial\mathbf{\Lambda}$ and $Q^{-1}(y)$ is nondegenerate, i.e. the Jacobian of the augmented system $\begin{pmatrix} J_Q \\ J_B \end{pmatrix}$ is full rank.

Theorem 2.4.1. *Let $Q : U_{\mathbf{\Lambda}} \to \mathbb{R}^m$ be a C^1 map, where $U_{\mathbf{\Lambda}}$ is an open set in $\mathbb{R}^n$. Define $\tilde{\rho}_{\mathcal{D}}$ with respect to the compact set $\mathbf{\Lambda} \subset U_{\mathbf{\Lambda}}$, so*

$$\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)\cap\mathbf{\Lambda}} \frac{1}{\sqrt{(J_Q J_Q^T)}} ds. \tag{2.2}$$

Suppose $\partial\mathbf{\Lambda}$ is locally determined by $B(\lambda) = 0$, where $B : U_{\mathbf{\Lambda}} \to \mathbb{R}^m$ is a piecewise C^1 map. Given y_0, if

1. *There exists $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ such that $J_Q^{(m)} = \left(J_Q^{(i_1)}, J_Q^{(i_2)}, \dots, J_Q^{(i_m)}\right)$ is full rank on $Q^{-1}(y_0)$,*
2. *$\begin{pmatrix} J_Q \\ J_B \end{pmatrix}$ is full rank on $\partial\mathbf{\Lambda} \cap Q^{-1}(y_0)$.*

Then $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous in a neighborhood of y_0.

Proof. Without loss of generality, we assume $i_1, \dots, i_m = 1, \dots, m$. Let $\lambda^{(m)} = (\lambda_1, \dots, \lambda_m)$, $\hat{\lambda}^{(m)} = (\lambda_{m+1}, \dots, \lambda_n)$, then λ can be formally written as $\lambda \sim (\lambda^{(m)}, \hat{\lambda}^{(m)})$. Define $\pi_{\hat{m}} : \mathbb{R}^n \to \mathbb{R}^{n-m}$ as $\pi_{\hat{m}}(\lambda) = \hat{\lambda}^{(m)}$.

For a $\lambda_0 \in \mathbf{\Lambda} \cap Q^{-1}(y_0)$, since $J_Q^{(m)}(\lambda_0)$ is full rank, the Implicit Function Theorem 1.5.1 implies that $Q(\lambda) - y = 0$ determines a unique continuous function $h : N(\hat{\lambda}_0^{(m)}, r) \times N(y_0, r) \to N(\lambda_0^{(m)}, r)$. Without loss of generality we assume h is defined on $\overline{N(\hat{\lambda}_0^{(m)}, r)} \times \overline{N(y_0, r)}$. Since the collection $\{N(\hat{\lambda}^{(m)}, r) \times N(\lambda^{(m)}, r)\}$ covers the compact set $\mathbf{\Lambda} \cap Q^{-1}(y_0)$, there is a finite subcover $\{N(\hat{\lambda}_k^{(m)}, r_k) \times N(\lambda_k^{(m)}, r_k)\}, k = 1, \dots, K$. Let $r_0 = \min_{k=1}^K r_k$. There is a unique continuous function $h : \left(\bigcup_{k=1}^K \overline{N(\hat{\lambda}_k^{(m)}, r_k)}\right) \times \overline{N(y_0, r_0)} \to \bigcup_{k=1}^K N(\lambda_k^{(m)}, r_k)$. We have

$$\tilde{\rho}_{\mathcal{D}}(y_0) = \int_{\pi_{\hat{m}}(\mathbf{\Lambda}\cap Q^{-1}(y_0))} \frac{1}{\left|J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y_0), \hat{\lambda}^{(m)})\right|} d\hat{\lambda}^{(m)}.$$

Since $J_Q^{(m)}(\lambda)$ is a continuous function,

$$\frac{1}{\left|J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)})\right|}$$

is continuous on a compact domain $\left(\bigcup_{k=1}^{K} \overline{N(\hat{\lambda}_k^{(m)}, r_k)}\right) \times \overline{N(y_0, r_0)}$, which means it is uniformly continuous and bounded. Let $S > 0$ be bound, thus

$$\frac{1}{\left|J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)})\right|} < S.$$

Since $\partial\mathbf{\Lambda}$ is piecewise smooth, for almost all $\lambda_0 \in \partial\mathbf{\Lambda}$, there exists r such that $B : N(\lambda_0, r) \to \mathbb{R}$ is a C^1 map and determines $\partial\mathbf{\Lambda} \cap N(\lambda_0, r)$ by $B(\lambda) = 0$.

Take one such point $\lambda_0 \in \partial\mathbf{\Lambda} \cap Q^{-1}(y_0)$, then there exists $N(\lambda_0, r)$ such that $\partial\mathbf{\Lambda} \cap Q^{-1}(y_0)$ is determined by $Q^*(\lambda) = \begin{pmatrix} Q(\lambda) - y_0 \\ B(\lambda) \end{pmatrix} = 0$ with

$$J_{Q^*} = \begin{pmatrix} J_Q \\ J_B \end{pmatrix} = \begin{pmatrix} J_Q^{(m)} & \hat{J}_Q^{(m)} \\ J_B^{(m)} & \hat{J}_B^{(m)} \end{pmatrix}$$

having full rank because of Condition 2. Since $J_{Q^*}(\lambda_0)$ and $J_Q^{(m)}(\lambda_0)$ are full rank, there exists $j \in m+1, \dots, n$ such that

$$\begin{pmatrix} J_Q^{(m)} & (\hat{J}_Q^{(m)})_j \\ J_B^{(m)} & (\hat{J}_B^{(m)})_j \end{pmatrix} (\lambda_0)$$

is invertible.

Define $\lambda^{(m+1)} = (\lambda^{(m)}, \lambda_j), \hat{\lambda}^{(m+1)} = \hat{\lambda}^{(m)} \setminus \lambda_j$, where we write $\lambda \sim (\lambda^{(m+1)}, \hat{\lambda}^{(m+1)})$. By the Implicit Function Theorem 1.5.1, there exists a unique continuous function,

$$g : N(\hat{\lambda}_0^{(m+1)}, r') \times N(y_0, r') \to N(\lambda_0^{(m)}, r') \times N(\lambda_{0j}, r'),$$

which satisfies

$$Q^*(g(\hat{\lambda}^{(m+1)}, y), \lambda^{(m)}, \lambda_j) = 0.$$

Without loss of generality, we assume g is defined on $\overline{N(\hat{\lambda}_0^{(m+1)}, r')} \times \overline{N(y_0, r')}$. By compactness of the domain, g is uniformly continuous. The set $\{N(\hat{\lambda}^{(m+1)}, r') \times N(\lambda^{(m)}, r') \times N(\lambda_j, r')\}$ covers $\partial\mathbf{\Lambda} \cap Q^{-1}(y)$ for $y \in N(y_0, r')$, and we assume r' is sufficiently small so that

$$N(\hat{\lambda}^{(m+1)}, r') \times N(\lambda^{(m)}, r') \times N(\lambda_j, r') \subset \bigcup_{k=1}^{K} \left(N(\hat{\lambda}_k^{(m)}, r_k) \times N(\lambda_k^{(m)}, r_k) \right).$$

Since $\partial\mathbf{\Lambda} \cap Q^{-1}(y_0)$ is compact, a finite subcover $\{N(\hat{\lambda}_\ell^{(m+1)}, r'_\ell) \times N(\lambda_\ell^{(m)}, r'_\ell) \times N(\lambda_{\ell j}, r'_\ell)\}, \ell = 1, \ldots, L$ exists, and let $r'_0 = \min\{r'_\ell, \ell = 1, \cdots, L; r_0\}$ then for any $y \in N(y_0, r'_0)$, the sets also cover $\partial\mathbf{\Lambda} \cap Q^{-1}(y)$. So Condition 2 holds for all $y \in N(y_0, r'_0)$. Thus

$$\bigcup_{\ell=1}^{L} \left(N(\hat{\lambda}_\ell^{(m+1)}, r'_\ell) \times N(\lambda_\ell^{(m)}, r'_\ell) \times N(\lambda_{\ell j}, r'_\ell) \times N(y_0, r'_0)\} \right)$$

is covered by $\{N(\hat{\lambda}_k^{(m)}, r_k) \times N(\lambda_k^{(m)}, r_k) \times N(y_0, r_0)\}$.

As a result, $\{N(\hat{\lambda}_k^{(m)}, r_k) \times N(\lambda_k^{(m)}, r_k)\}$ covers $\mathbf{\Lambda} \cap Q^{-1}(y)$ for all $y \in N(y_0, r'_0)$, thus we have

$$\tilde{\rho}_{\mathcal{D}}(y) = \int_{\pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y))} \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)}) \right|} d\hat{\lambda}^{(m)}.$$

For $y \in N(y_0, r'_0)$, let $C = \pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y_0)) \triangle \pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y))$. Thus,

$$\begin{aligned} |\tilde{\rho}_{\mathcal{D}}(y_0) - \tilde{\rho}_{\mathcal{D}}(y)| \leq & \int_{\pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y_0))} \left| \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y_0), \hat{\lambda}^{(m)}) \right|} - \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)}) \right|} \right| d\hat{\lambda}^{(m)} \\ & + \int_C S d\hat{\lambda}^{(m)} \end{aligned}$$

For the first term, uniform continuity implies that given ϵ, there exists $r_\epsilon < r_0'$ such that for $y \in N(y_0, r_\epsilon)$, and $\hat{\lambda}^{(m)} \in \left(\bigcup_{k=1}^K N(\hat{\lambda}_k^{(m)}, r_k)\right)$,

$$\left| \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y_0), \hat{\lambda}^{(m)}) \right|} - \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)}) \right|} \right| < \epsilon.$$

Thus,

$$\int_{\pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y_0))} \left| \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y_0), \hat{\lambda}^{(m)}) \right|} - \frac{1}{\left| J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)}) \right|} \right| d\hat{\lambda}^{(m)} \leq \hat{\mu}^{(m)}(\pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y_0)))\epsilon$$

For the second term, we need $\hat{\mu}^{(m)}(B)$ to be small when y is close to y_0.

Since $\{\pi_{\hat{m}} \left(N(\hat{\lambda}_\ell^{(m+1)}, r_\ell') \times N(\lambda_\ell^{(m)}, r_\ell') \times N(\lambda_{\ell j}, r_\ell') \right) = N(\hat{\lambda}_\ell^{(m+1)}, r_\ell') \times N(\lambda_{\ell j}, r_\ell')\}$ covers $\pi_{\hat{m}}(\partial \mathbf{\Lambda} \cap Q^{-1}(y))$ when $y \in N(y_0, r_0')$, then

$$\begin{aligned} \hat{\mu}^{(m)}(B) &= \hat{\mu}^{(m)}(\pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y_0)) \triangle \pi_{\hat{m}}(\mathbf{\Lambda} \cap Q^{-1}(y))) \\ &\leq \sum_{\ell=1}^L \hat{\mu}^{(m)} \left(\left(\pi_{\hat{m}} \left(\mathbf{\Lambda} \cap Q^{-1}(y_0) \right) \triangle \pi_{\hat{m}} \left(\mathbf{\Lambda} \cap Q^{-1}(y) \right) \right) \cap \left(N(\hat{\lambda}_\ell^{(m+1)}, r_\ell') \times N(\lambda_{\ell j}, r_\ell') \right) \right), \end{aligned}$$

and

$$\hat{\mu}^{(m)} \left(\left(\pi_{\hat{m}} \left(\mathbf{\Lambda} \cap Q^{-1}(y_0) \right) \triangle \pi_{\hat{m}} \left(\mathbf{\Lambda} \cap Q^{-1}(y) \right) \right) \cap \left(N(\hat{\lambda}_\ell^{(m+1)}, r_\ell') \times N(\lambda_{\ell j}, r_\ell') \right) \right) \leq \int_{N(\hat{\lambda}_\ell^{(m+1)}, r_\ell')} \left| g(\hat{\lambda}_\ell^{(m+1)}, y_0) - g(\hat{\lambda}_\ell^{(m+1)}, y) \right| d\hat{\lambda}^{(m+1)}.$$

By the uniform continuity of g, there exists an $r_\epsilon' \leq r_0'$ such that for all $y \in N(y_0, r_\epsilon')$, $\hat{\mu}^{(m)}(B) < \epsilon$. Finally, we have for all $y \in N(y_0, r)$, $r = \min(r_\epsilon, r_\epsilon')$,

$$
\begin{aligned}
|\tilde{\rho}_{\mathcal{D}}(y_0) - \tilde{\rho}_{\mathcal{D}}(y)| \leq & \int_{\pi_{\hat{m}}(\mathbf{\Lambda}\cap Q^{-1}(y_0))} \left| \frac{1}{\left|J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y_0), \hat{\lambda}^{(m)})\right|} - \frac{1}{\left|J_Q^{(m)}(h(\hat{\lambda}^{(m)}, y), \hat{\lambda}^{(m)})\right|} \right| d\hat{\lambda}^{(m)} \\
& + \int_B S d\hat{\lambda}^{(m)} \\
\leq & \hat{\mu}^{(m)}(\pi_{\hat{m}}(\mathbf{\Lambda}\cap Q^{-1}(y_0)))\epsilon + S\epsilon.
\end{aligned}
$$

So $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous at y_0. Furthermore, the two conditions hold for all y in $N(y_0, r_0')$, so $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous in $N(y_0, r)$ as well. $\square$

In next theorem, we remove the requirement that $Q^{-1}(y_0)$ is uniformly parameterized.

Theorem 2.4.2 (Continuity of $\tilde{\rho}_{\mathcal{D}}$)**.** *Let $Q : U_{\mathbf{\Lambda}} \to \mathcal{D}$ be a C^1 map, where $U_{\mathbf{\Lambda}}$ is an open set in $\mathbb{R}^n$, and $\mathcal{D}$ is a subset of $\mathbb{R}^m$. Suppose $\mathbf{\Lambda}$ is a compact subset of $U_{\mathbf{\Lambda}}$ and J_Q is full rank on $\mathbf{\Lambda}$. Define $\tilde{\rho}_{\mathcal{D}}$ with respect to $\mathbf{\Lambda}$, thus*

$$
\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)\cap\mathbf{\Lambda}} \frac{1}{\sqrt{(J_Q J_Q^T)}} ds. \tag{2.3}
$$

Suppose $\partial\mathbf{\Lambda}$ is locally determined by $B(\lambda) = 0$, where $B : U_{\mathbf{\Lambda}} \to \mathbb{R}^m$ is a piecewise C^1 map. Given y_0, if $\begin{pmatrix} J_Q \\ J_B \end{pmatrix}$ is full rank on $\partial\mathbf{\Lambda}\cap Q^{-1}(y_0)$ or $\partial\mathbf{\Lambda}\cap Q^{-1}(y_0) = \varnothing$, $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous in a neighborhood of y_0.

Proof. For all $\lambda \in Q^{-1}(y_0)$, there exists $N(\lambda, r)$ such that there exists $\{i_1, \ldots, i_m\} \subset \{1, \ldots, n\}$ with $J_Q^{(m)} = \left(J_Q^{(i_1)}, J_Q^{(i_2)}, \ldots, J_Q^{(i_m)}\right)$ being full rank on $Q^{-1}(y_0)$ in $N(\lambda, r)$. Obviously, the boundary of $N(\lambda, r)\cap\mathbf{\Lambda}$, which is the union of part of $\partial\mathbf{\Lambda}$ and part of $\partial N(\lambda, r)$ is piecewise smooth. Suppose $\partial\left(N(\lambda, r)\cap\mathbf{\Lambda}\right)$ is locally determined by $B'(\lambda) = 0$. When r is sufficiently small, $\begin{pmatrix} J_Q \\ J_{B'} \end{pmatrix}$ is full rank on $\partial\left(N(\lambda, r)\cap\mathbf{\Lambda}\right)\cap Q^{-1}(y_0)$ by continuity of J_Q. The collection $\{N(\lambda, r)\}$ admits a finite sub cover $\{N(\lambda_k, r_k)\}, k = 1, \ldots, K$ of $\mathbf{\Lambda}\cap Q^{-1}(y_0)$. Taking r sufficiently small, the collection covers $\mathbf{\Lambda}\cap Q^{-1}(y)$, for y in $N(y_0, r)$. Define $N_k = N(\lambda_k, r_k) \setminus \left(\bigcup_{j=1}^{k-1} N(\lambda_j, r_j)\right)$, the boundary of $N_k \cap \mathbf{\Lambda}$ is piecewise smooth. Suppose $\partial\left(N_k \cap \mathbf{\Lambda}\right)$ is locally determined by $B_k(\lambda) = 0$, then, $\begin{pmatrix} J_Q \\ J_{B_k} \end{pmatrix}$ is full rank on $\partial\left(N_k \cap \mathbf{\Lambda}\right)$, and

$$\tilde{\rho}_{\mathcal{D}}(y) = \int_{Q^{-1}(y)\cap\mathbf{\Lambda}} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds = \sum_{k=1}^{K} \int_{Q^{-1}(y)\cap(N_k\cap\mathbf{\Lambda})} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds.$$

So for all $y \in N(y_0, r)$,

$$|\tilde{\rho}_{\mathcal{D}}(y) - \tilde{\rho}(y_0)| \leq \sum_{k=1}^{K} \left| \int_{Q^{-1}(y)\cap(N_k\cap\mathbf{\Lambda})} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds - \int_{Q^{-1}(y_0)\cap(N_k\cap\mathbf{\Lambda})} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds \right|$$

Theorem 2.4.1 implies that $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous at y_0. Since $\int_{Q^{-1}(y)\cap(N_k\cap\mathbf{\Lambda})} \frac{1}{\sqrt{\det(J_Q J_Q^T)}} ds$ is continuous in a neighborhood of y_0 as Theorem 2.4.1 suggests, with finite K, $\tilde{\rho}_{\mathcal{D}}(y)$ is continuous in a neighborhood of y_0. □

Figure 2.2 shows when the generalized contour $Q^{-1}(y_0)$ intersects with the boundary of $\mathbf{\Lambda}$, or it does not intersect with the boundary of $\mathbf{\Lambda}$, y_0 is a continuity points of $\tilde{\rho}_{\mathcal{D}}$.

Moreover, for $y \in Q(\mathbf{\Lambda})$ satisfying conditions of this theorem, $\tilde{\rho}_{\mathcal{D}}(y) > 0$. This is trivially true when $\partial\mathbf{\Lambda} \cap Q^{-1}(y) = \varnothing$, since then $\text{int}\mathbf{\Lambda} \cap Q^{-1}(y) \neq \varnothing$. When $\partial\mathbf{\Lambda} \cap Q^{-1}(y) \neq \varnothing$, at the intersection of $Q(\lambda) - y = 0$ and $\partial\mathbf{\Lambda}$, the tangent space of $Q(\lambda) - y = 0$ intersects with both sides of $\partial\mathbf{\Lambda}$, so it must intersect with $\text{int}\mathbf{\Lambda}$.

Now that we have the local results, we want to study the global properties of $\tilde{\rho}_{\mathcal{D}}$. Theorem 2.4.2 indicates that if $\tilde{\rho}_{\mathcal{D}}$ is discontinuous at point y_0 and $\partial\mathbf{\Lambda} \cap Q^{-1}(y) \neq \varnothing$, then the Jacobian $\begin{pmatrix} J_Q \\ J_B \end{pmatrix}$ is rank deficient on a subset of $\partial\mathbf{\Lambda} \cap Q^{-1}(y)$. Treating Q as a map from the manifold $\partial\mathbf{\Lambda}$ to $\mathcal{D}$, a point on $\partial\mathbf{\Lambda}$ such that $\begin{pmatrix} J_Q \\ J_B \end{pmatrix}$ is rank deficient is a critical point for the map. The image of

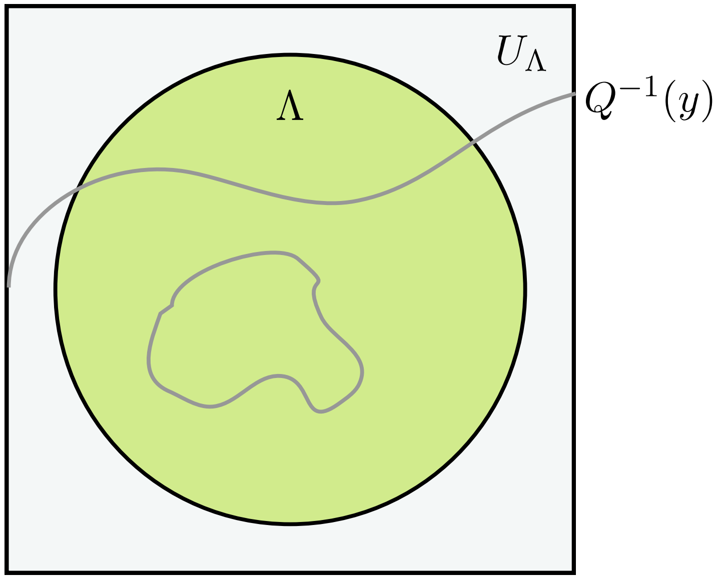


Figure 2.2: Examples of continuity points of $\tilde{\rho}_{\mathcal{D}}$.

critical point is called a critical value. Sard's Theorem 1.6.13 states that the set of critical values has measure zero when Q is sufficiently smooth.

Theorem 2.4.3. *If Q is C^r map with $r > n - m - 1$, then the set of discontinuity points of $\tilde{\rho}_{\mathcal{D}}$ has measure zero.*

Proof. The set of discontinuity points of $\tilde{\rho}_{\mathcal{D}}$ is a subset of set of critical values of $Q : \partial\mathbf{\Lambda} \to \mathcal{D}$. Since $\partial\mathbf{\Lambda}$ is $n-1$ dimensional piecewise smooth manifold, and $\mathcal{D}$ is an m-dimensional smooth manifold. By Sard's theorem 1.6.13, when $r > (n-1) - m = n - m - 1$, the set of critical values has measure zero. So the set of discontinuity points has measure zero. $\square$

Moreover, as conditions of Theorem 2.4.2 guarantee $\tilde{\rho}_{\mathcal{D}}(y) > 0$ for $y \in Q(\mathbf{\Lambda})$, Theorem 2.4.3 guarantees $\tilde{\rho}_{\mathcal{D}}$ is positive on $Q(\mathbf{\Lambda})$ a.e. when Q is sufficiently smooth.

Chapter 3

Numerical analysis for solution of the SIP

In this chapter, we discuss different aspects of numerical methods to solve the SIP. Our ultimate goal is to show that the approximation computed by an algorithm modified from Algorithm 1 converge to the probability measure $P_{\mathbf{\Lambda}}$ that solves the SIP under the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$.

This chapter is divided into five sections. In Section 3.1, we discuss the approximation of sets by discretization of the sample space. In Section 3.2, we point out that any Lebesgue measurable function is approximated by the ratio of the integral to the volumn of a set containing the point. In Section 3.3, we give a modified version of Algorithm 1, namely Algorithm 2, and prove the function generated by this algorithm converges to the solution of the SIP. In Section 3.4, we give an example where Algorithm 2 fails, due to the failure of the boundary condition. Lastly, an error analysis is conducted in Section 3.5.

For the first two sections, we let $\mathbf{\Omega}$ be a compact subset of $\mathbb{R}^n$, and $\mathcal{B}_{\mathbf{\Omega}}$ be the Borel σ-algebra and Lebesgue measure $\mu_{\mathbf{\Omega}}$ inherited from $\mathbb{R}^n$. Notationally, we avoid using $\mathbf{\Lambda}$ or $\mathcal{D}$ because we want to emphasize that the approximations in these two sections work for a general measure space, not necessarily a space related to the SIP.

3.1 Approximation of sets

It is well known that for any Lebesgue measurable set $A \subset \mathbb{R}^n$ for which $\mu(A) < \infty$, given $\epsilon > 0$, A can be approximated by a finite collection of disjoint rectangles $\{R_j\}_1^N$ in the sense that $\mu(A \triangle \bigcup_1^N R_j) < \epsilon$ (Folland, 2013).

Here we restrict attention to the task of approximating a set with negligible boundary. Our goal of set approximation can be summarized as this: We want to define a process for generating partitions on $\mathbf{\Omega} \subset \mathbb{R}^n$, such that any set in $A \in \mathcal{B}_{\mathbf{\Omega}}$ with $\mu_{\mathbf{\Omega}}(\partial A) = 0$ can be approximated by element sets from the partition.

Before we tackle this problem, we show the Voronoi tessellation generated from a Poisson point process on $\mathbf{\Omega}$ is one such partition generating process. The Voronoi tessellation $\{\mathcal{V}_j\}_{j=1}^N$ of $\mathbf{\Omega}$ corresponding to a sample $\left\{\omega^{(j)}\right\}_{j=1}^N$ is the partition with Voronoi cells defined to be

$$\mathcal{V}_j = \left\{\omega \in \mathbf{\Omega} \,:\, d_v(\omega^{(j)}, \omega) \leq d_v(\omega^{(i)}, \omega), \text{for all } i = 1, \ldots, N\right\},$$

where $d_v(\cdot, \cdot)$ is a metric on $\mathbf{\Omega}$.

The Voronoi coverage A_N of $A \in \mathcal{B}_{\mathbf{\Omega}}$ is defined as

$$A_N = \bigcup_{\omega^{(j)} \in A, 1 \leq j \leq N} \mathcal{V}_j.$$

In other words, we use the Voronoi cell center to decide which cells to include in the approximation. We prove A_N converges to A under certain conditions by a proof adapted from Khmaladze and Toronjadze (2001, Lemma 2.1).

Lemma 3.1.1. *Given a probability density $f(\omega) > 0$ a.e. on $\mathbf{\Omega}$ with respect to the Lebesgue measure, if $A \in \mathcal{B}_{\mathbf{\Omega}}$ such that $\mu_{\mathbf{\Omega}}(\partial A) = 0$, then*

$$\mu_{\mathbf{\Omega}}(A_N \triangle A) \to 0, \text{ a.s. } N \to \infty.$$

Proof. Suppose $A \in \mathcal{B}_{\mathbf{\Omega}}$, $\mu_{\mathbf{\Omega}}(A) > 0$ and $\mu_{\mathbf{\Omega}}(\partial A) = 0$. Let $A^\delta := (A + \delta B(0,1)) \cap \mathbf{\Omega}$ denote the Minkowski sum of A and $\delta B(0,1)$ restricted to domain $\mathbf{\Omega}$, where $B(0,1)$ is the unit ball in $\mathbb{R}^n$. Let $A_\delta = ((A^c)^\delta)^c$. Since $\mu_{\mathbf{\Omega}}(\partial A) = 0$, for all $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $\partial A \subset A^{\delta(\epsilon)} \backslash A_{\delta(\epsilon)}$ and $\mu_{\mathbf{\Omega}}(A^{\delta(\epsilon)} \backslash A_{\delta(\epsilon)}) < \epsilon$.

Let $r(\omega^{(j)}) := \max_{\omega \in \mathcal{V}(\omega^{(j)})} d_v(\omega, \omega^{(j)})$ and

$$r_N := \max_{1 \leq j \leq N} r(\omega^{(j)}) = \max_{\omega \in \mathbf{\Omega}} \min_{1 \leq j \leq N} d_v(\omega, \omega^{(j)}).$$

We prove that $r_N \to 0$ a.s.. For any $h \in \mathbb{N}$, let $\mathcal{H}_h$ denote the finite set of hypercubes in $\mathbb{R}^n$ with edge-length $1/h$ partitioning compact $\mathbf{\Omega}$. By strong law of large numbers,

$$\max_{H\in\mathcal{H}_h}\left|\frac{\sum_{j=1}^{N}\mathbb{1}_{\{\omega^{(j)}\in H\}}}{N}-\int_H f(\omega)\,d\mu_{\mathbf{\Omega}}(\omega)\right|\to 0 \text{ a.s. } N\to\infty$$

and

$$\min_{H\in\mathcal{H}_h}\int_H f(\omega)\,d\mu_{\mathbf{\Omega}}(\omega)>0 \text{ for all } h.$$

This implies that there must be at least one sample point $\omega^{(j)}$ in each $\mathcal{H}_h$ a.s. when N is sufficiently large. Since $\omega \in \mathbf{\Omega}$ must fall into one of $\mathcal{H}_h$, $r_N \leq \sqrt{n}/h$. So for all ϵ, there exists N such that $A_{\delta(\epsilon)} \subset A_N \subset A^{\delta(\epsilon)}$. □

The key for the approximation is that r_N, the upper bound of radius of the Voronoi tessellation goes to 0 a.s.. Intuitively, this means the partition sets become small uniformly when $N \to \infty$.

Definition 3.1.2. *A rule for defining sequences of samples* $\left\{\omega^{(j)}\right\}_{j=1}^{N} \subset \mathbf{\Omega}$ *is* $\mathcal{B}_{\mathbf{\Omega}}$***-consistent*** *if*

$$r_N := \max_{1\leq j\leq N} r(\omega^{(j)}) = \max_{1\leq j\leq N}\max_{\omega\in\mathcal{V}(\omega^{(j)})} d_v(\omega,\omega^{(j)}) \to 0 \textit{ a.s., as } N\to\infty.$$

Drawing i.i.d samples from probability distribution with $f(\omega) > 0$ a.e. is a $\mathcal{B}_{\mathbf{\Omega}}$-consistent rule.

For any partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$ of $\mathbf{\Omega}$, there does not necessarily exist a point ω_i associated with each set I_i as is true for the Voronoi tessellation based on a point process. In order to measure the size of each element of the partition, we simply use

$$\operatorname{diam} I_i = \sup_{\omega_1,\omega_2\in I_i} d_v(\omega_1,\omega_2).$$

In Lemma 3.1.1, the properties of approximation rely on properties of the point process by which the Voronoi tessellation is generated. In general, we consider a sequence of partition $\{\mathcal{I}_m\}_{m=1}^{\infty} = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^{\infty}$ that is not associated with point process. The sequence of partitions does not need to be nested. The only condition the sequence needs to satisfy is

$$\lim_{m\to\infty}\max_{1\le i\le M_m}\operatorname{diam} I_i = 0.$$

The sequence of Voronoi tessellations generated by a sequence of samples from $\mathcal{B}_{\mathbf{\Omega}}$-consistent rule satisfies the above condition a.s.. For a general partition $\mathcal{I} = \{I_i\}_{i=1}^M$, one can associate each I_i with an arbitrary point $\omega_i \in I_i$, and define a set approximation of A as

$$A_M = \bigcup_{\omega_i\in A, 1\le i\le M} I_i.$$

We can prove the following theorem.

Theorem 3.1.3. *Given a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^\infty = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^\infty$, with*

$$\lim_{m\to\infty}\max_{1\le i\le M_m}\operatorname{diam} I_i = 0,$$

if $A \in \mathcal{B}_{\mathbf{\Omega}}$ satisfies $\mu_{\mathbf{\Omega}}(\partial A) = 0$, then

$$\mu_{\mathbf{\Omega}}(A_{M_m}\triangle A) \to 0, \text{ as } m\to\infty.$$

3.2 Approximation of density

For an integrable non-negative function $\rho_{\mathbf{\Omega}}$ defined on $(\mathbf{\Omega}, \mathcal{B}_{\mathbf{\Omega}}, \mu_{\mathbf{\Omega}})$, given a partition $\mathcal{I} = \{I_i\}_{i=1}^M$, we are interested in whether $\rho_{\mathbf{\Omega}}$ can be approximated by

$$\rho_{\mathbf{\Omega},M}(\omega) = \sum_{i=1}^M p_i \mathbb{1}_{I_i}(\omega), \quad p_i = \frac{\int_{I_i}\rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}}}{\int_{I_i} d\mu_{\mathbf{\Omega}}}. \tag{3.1}$$

Consider a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^\infty = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^\infty$, we want $\rho_{\mathbf{\Omega},M_m}(\omega) \to \rho_{\mathbf{\Omega}}(\omega)$ if certain conditions hold for $\{\mathcal{I}_m\}_{m=1}^\infty$. We define

$$\mathcal{I}_m(\omega) = \bigcup_{\omega \in I_i, 1 \le i \le M_m} I_i.$$

Theorem 3.2.1. *If ω is an interior point of one and only one I_i from all $\mathcal{I}_m$, and $\mathrm{diam}(\mathcal{I}_m(\omega)) \to 0$, then $\rho_{\mathbf{\Omega},M_m}(\omega) \to \rho_{\mathbf{\Omega}}(\omega)$.*

Proof. This follows from the Lebesgue Differentiation Theorem 1.7.1. □

For a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^{\infty}$, if $\lim_{m\to\infty} \max_{1\le i\le M_m} diam(I_i) = 0$, for any ω except points on boundaries of partition sets $\bigcup_{m=1}^{\infty}\bigcup_{i=1}^{M_m} \partial I_i$, $diam(\mathcal{I}_m(\omega)) \to 0$.

Note that $\mu_{\mathbf{\Omega}}(\bigcup_{i=1}^{M_m} \partial I_i) = 0$ for any $\mathcal{I}_m$, $\mu_{\mathbf{\Omega}}(\bigcup_{m=1}^{\infty}\bigcup_{i=1}^{M_m} \partial I_i) = 0$, which means the exception set of Theorem 3.2.1 is negligible.

Combing all the above results yields the following theorem.

Theorem 3.2.2. *For a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^{\infty} = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^{\infty}$ satisfying*

$$\lim_{m\to\infty} \max_{1\le i\le M_m} diam(I_i) = 0,$$

$\rho_{\mathbf{\Omega},M_m} \to \rho_{\mathbf{\Omega}}$ a.e. as $m \to \infty$.

When the density $\rho_{\mathbf{\Omega}}$ is unbounded as is $\rho_{\mathbf{\Lambda}}$ in Section 2.3, it is hard to find a function dominating $\{\rho_{\mathbf{\Omega},M}\}$. To facilitate discussions concerning convergence of measures, we show that $\{\rho_{\mathbf{\Omega},M}\}$ are uniform integrable.

Theorem 3.2.3. *$\{\rho_{\mathbf{\Omega},M}\}$ is a uniformly integrable collection.*

Proof. For any $\epsilon > 0$, there exists a t such that $\int_{\{\rho_{\mathbf{\Omega}}>t\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} < \epsilon$. Let $\delta := \int_{\{\rho_{\mathbf{\Omega}}>t\}} d\mu_{\mathbf{\Omega}} = \mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega}} > t\})$.

Since $\rho_{\mathbf{\Omega},M}$ is a simple function, there is a t' such that $\mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega},M} \ge t'\}) \ge \delta$ and $\mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega},M} > t'\}) < \delta$. Define $E^M := \{\rho_{\mathbf{\Omega},M} > t'\} \cup F$, where F is any set satisfying $F \subset \{\rho_{\mathbf{\Omega},M} = t'\}$ and $\mu_{\mathbf{\Omega}}(F) = \delta - \mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega},M} > t'\})$. Then for all $E \in \mathcal{B}_{\mathbf{\Omega}}$ such that $\mu_{\mathbf{\Omega}}(E) \le \delta$,

$$\int_E \rho_{\mathbf{\Omega},M} d\mu_{\mathbf{\Omega}} \le \int_{E^M} \rho_{\mathbf{\Omega},M} d\mu_{\mathbf{\Omega}} = \int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega},M} d\mu_{\mathbf{\Omega}} + \mu_{\mathbf{\Omega}}(F)\cdot t'. \tag{3.2}$$

By definition of $\rho_{\mathbf{\Omega},M}$, we have

$$\int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega},M} d\mu_{\mathbf{\Omega}} = \int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}},$$

$$\int_{\{\rho_{\mathbf{\Omega},M}=t'\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} = \mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega},M} = t'\}) \cdot t'.$$

So, there is an $F' \subset \{\rho_{\mathbf{\Omega},M} = t'\}$, such that

$$\int_{F'} d\mu_{\mathbf{\Omega}} = \mu_{\mathbf{\Omega}}(F), \int_{F'} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} \geq \mu_{\mathbf{\Omega}}(F) \cdot t'.$$

So (3.2) implies

$$\int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega},M} d\mu_{\mathbf{\Omega}} + \mu_{\mathbf{\Omega}}(F)t' \leq \int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} + \int_{F'} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}}.$$

Since $\mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega},M} > t'\} \cup F') = \delta = \mu_{\mathbf{\Omega}}(\{\rho_{\mathbf{\Omega}} > t\})$, we have

$$\int_{\{\rho_{\mathbf{\Omega},M}>t'\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} + \int_{F'} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} \leq \int_{\{\rho_{\mathbf{\Omega}}>t\}} \rho_{\mathbf{\Omega}} d\mu_{\mathbf{\Omega}} < \epsilon$$

□

3.3 Approximation to solution of the SIP

In this section, we prove the main theorem regarding approximation of $P_{\mathbf{\Lambda}}(A)$ for $A \in \mathcal{B}_{\mathbf{\Lambda}}$ with $P_{\mathbf{\Lambda}}(\partial A) = 0$.

On $\mathcal{D}$ and $\mathbf{\Lambda}$, we generate partitions $\{I_i\}_{i=1}^{M}$ and $\{b_j\}_{j=1}^{N}$ respectively. Let each b_j be associated with a point $\lambda_j \in b_j$. Then $\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i) \cap b_j)$ in Algorithm 1 is further approximated as $\mu_{\mathbf{\Lambda}}(b_j)$ or 0 depending on whether λ_j is inside $Q^{-1}(I_i)$ or not. Adapted from Algorithm 1, the algorithm we consider is

Algorithm 2: Approximation to inverse probability.

Generate approximating sets $\{I_i\}_{i=1}^{M}$ of $\mathcal{D}$

Generate approximating sets $\{b_j\}_{j=1}^{N}$ of $\mathbf{\Lambda}$, associate every b_j with a point $\lambda_j \in b_j$

Fix and normalize the simple function approximation $\rho_{\mathcal{D},M} = \sum_{i=1}^{M} p_i \mathbb{1}_{I_i}(q)$

Compute $M \times N$ matrix V this way:

if $Q(\lambda_j) \in I_i$ **then**

  $V_{ij} = \mu_{\mathbf{\Lambda}}(b_j)$

else

  $V_{ij} = 0$

end

foreach $j = 1, \cdots, N$ **do**

  Set $P(b_j)$ to $\sum_{i=1}^{M} p_i \left(V_{ij} / \sum_{j=1}^{N} V_{ij} \right)$

end

To analyze the convergence properties of this algorithm, we denote the computed $P(b_j)$ as $P_{\mathbf{\Lambda},M,N}(b_j)$, and use $P_{\mathbf{\Lambda}}$ to denote the true probability measure. For $A \in \mathcal{B}_{\mathbf{\Lambda}}$ with $P_{\mathbf{\Lambda}}(\partial A) = 0$, define A_N as

$$A_N = \bigcup_{\lambda_j \in A, 1 \leq j \leq N} b_j,$$

then use

$$P_{\mathbf{\Lambda},M,N}(A_N) = \sum_{\lambda_j \in A, 1 \leq j \leq N} P_{\mathbf{\Lambda},M,N}(b_j)$$

as an approximation of $P_{\mathbf{\Lambda}}(A)$. The next theorem shows that when $\{I_i\}_{i=1}^{M}$ and $\{b_j\}_{j=1}^{N}$ both become fine, $P_{\mathbf{\Lambda},M,N}(A_N)$ converges to $P_{\mathbf{\Lambda}}(A)$.

Theorem 3.3.1. *Let $Q : \mathbf{\Lambda} \to \mathcal{D}$ be the map satisfying the conditions in Theorem 2.1.1, with the additional requirement $\mu_{\mathbf{\Lambda}}(\partial \mathbf{\Lambda}) = 0$. Given a probability measure $P_{\mathcal{D}}$ on $\mathcal{D}$ that is dominated by $\mu_{\mathcal{D}}$. Let $P_{\mathbf{\Lambda}}$ be the solution of the SIP under the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$. For event $A \in \mathcal{B}_{\mathbf{\Lambda}}$ with $\mu_{\mathbf{\Lambda}}(\partial A) = 0$, there exists a sequence of approximations $P_{\mathbf{\Lambda},M_m,N_n}(A_{N_n})$ using* *simple function approximations and requiring only calculations of volumes in* $\mathbf{\Lambda}$ *based on partition sequence* $\{\mathcal{I}_m\}_{m=1}^{\infty} = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^{\infty}$ *on* $\mathcal{D}$ *and* $\{\mathcal{T}_n\}_{n=1}^{\infty} = \{\{b_j\}_{j=1}^{N_n}\}_{n=1}^{\infty}$ *on* $\mathbf{\Lambda}$ *that satisfies*

$$\lim_{m\to\infty}\lim_{n\to\infty} P_{\mathbf{\Lambda},M_m,N_n}(A_{N_n}) = P_{\mathbf{\Lambda}}(A),$$

if $\lim_{m\to\infty}\max_{1\leq i\leq M_m} diam(I_i) = 0$ ***and*** $\lim_{n\to\infty}\max_{1\leq i\leq N_n} diam(b_j) = 0$.

Proof. Since $\mu_{\mathcal{D}}$ and $\tilde{\mu}_{\mathcal{D}}$ are equivalent with $d\tilde{\mu}_{\mathcal{D}} = \tilde{\rho}_{\mathcal{D}} d\mu_{\mathcal{D}}$, $P_{\mathcal{D}}$ has a density $\rho'_{\mathcal{D}} = \dfrac{\rho_{\mathcal{D}}}{\tilde{\rho}_{\mathcal{D}}}$ with respect to $\tilde{\mu}_{\mathcal{D}}$. Given any partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$ on $\mathcal{D}$, we define the simple function

$$\rho'_{\mathcal{D},M} = \sum_{i=1}^{M} p'_i \mathbb{1}_{(I_i)}(y), \quad p'_i = \frac{P_{\mathcal{D}}(I_i)}{\tilde{\mu}_{\mathcal{D}}(I_i)} = \frac{\int_{I_i} \frac{\rho_{\mathcal{D}}}{\tilde{\rho}_{\mathcal{D}}} d\tilde{\mu}_{\mathcal{D}}}{\int_{I_i} d\tilde{\mu}_{\mathcal{D}}}, \tag{3.3}$$

as an approximation of $\tilde{\rho}_{\mathcal{D}}$. We also define

$$\rho_{\mathbf{\Lambda},M} = \sum_{i=1}^{M} p'_i \mathbb{1}_{Q^{-1}(I_i)}(\lambda), \quad p'_i = \frac{P_{\mathcal{D}}(I_i)}{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i))} = \frac{P_{\mathcal{D}}(I_i)}{\tilde{\mu}_{\mathcal{D}}(I_i)} = \frac{\int_{I_i} \frac{\rho_{\mathcal{D}}}{\tilde{\rho}_{\mathcal{D}}} d\tilde{\mu}_{\mathcal{D}}}{\int_{I_i} d\tilde{\mu}_{\mathcal{D}}}, \tag{3.4}$$

as an approximation of $\rho_{\mathbf{\Lambda}}$. By definition, $\rho_{\mathbf{\Lambda},M}(\lambda) = \rho'_{\mathcal{D},M}(Q(\lambda))$. For a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^{\infty} = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^{\infty}$, we write corresponding sequence of densities as $\{\rho'_{\mathcal{D},M_m}\}_{m=1}^{\infty}$ and $\{\rho_{\mathbf{\Lambda},M_m}\}_{m=1}^{\infty}$. By Theorem 3.2.2, since $\lim_{m\to\infty}\max_{1\leq i\leq M_m} diam(I_i) = 0$, $\rho'_{\mathcal{D},M_m}$ converges to $\rho'_{\mathcal{D}} = \dfrac{\rho_{\mathcal{D}}}{\tilde{\rho}_{\mathcal{D}}}$, $\tilde{\mu}_{\mathcal{D}}$-a.e., so $\rho_{\mathbf{\Lambda},M_m} \to \rho_{\mathbf{\Lambda}}$, $\mu_{\mathbf{\Lambda}}$-a.e., where $\rho_{\mathbf{\Lambda}}(\lambda) = \dfrac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}$ is the solution to the SIP under the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$.

Given a partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$ on $\mathcal{D}$ with its associated $\rho_{\mathbf{\Lambda},M}$, partition sequence $\{\mathcal{T}_n\}_{n=1}^{\infty} = \{\{b_j\}_{j=1}^{N_n}\}_{n=1}^{\infty}$ on $\mathbf{\Lambda}$, we can associate each b_j with a point $\lambda_j \in b_j$, and construct

$$\rho_{\mathbf{\Lambda},M,N_n} = \sum_{j=1}^{N_n} p_j \mathbb{1}_{b_j}(\lambda), \quad p_j = \frac{P_{\mathcal{D}}(I_i)}{\sum_{\lambda_k \in Q^{-1}(I_i)} \mu_{\mathbf{\Lambda}}(b_k)} \text{ if } \lambda_j \in Q^{-1}(I_i), \tag{3.5}$$

as an approximation of $\rho_{\mathbf{\Lambda},M}$. We can prove $\rho_{\mathbf{\Lambda},M,N_n} \to \rho_{\mathbf{\Lambda},M}$ a.e. if

$$\lim_{n\to\infty}\max_{1\leq j\leq N_n} diam(b_j) = 0.$$

Since $\mu_{\mathcal{D}}(\partial I_i) = 0$ and $\mu_{\mathbf{\Lambda}}(\partial \mathbf{\Lambda}) = 0$, $\mu_{\mathbf{\Lambda}}(\partial(Q^{-1}(I_i))) = 0$, by Theorem 3.1.3,

$$\sum_{\lambda_k \in Q^{-1}(I_i)} \mu_{\mathbf{\Lambda}}(b_k) \to \mu_{\mathbf{\Lambda}}(Q^{-1}(I_i)) \text{ if } \lim_{n\to\infty} \max_{1\leq j\leq N_n} diam(b_j) = 0.$$

Furthermore, for almost all $\lambda \in \text{int}Q^{-1}(I_i)$, $\lambda \in \text{int}b_j$ for one b_j in $\mathcal{T}_n = \{b_j\}_{j=1}^{N_n}$ for all $n, 1 \leq n < \infty$. As a result, $\rho_{\mathbf{\Lambda},M,N_n}(\lambda) \to \rho_{\mathbf{\Lambda},M}(\lambda)$ for almost all $\lambda \in \text{int}Q^{-1}(I_i)$ for an $I_i \in \mathcal{I}$. Since

$$\mu_{\mathbf{\Lambda}}\left(\bigcup_{i=1}^{M} \partial(Q^{-1}(I_i))\right) \leq \sum_{i=1}^{M} \mu_{\mathbf{\Lambda}}(\partial(Q^{-1}(I_i))) = 0,$$

$\rho_{\mathbf{\Lambda},M,N_n} \to \rho_{\mathbf{\Lambda},M}$ a.e. as $n \to \infty$.

Actually, the convergence of $\rho_{\mathbf{\Lambda},M,N_n}$ to $\rho_{\mathbf{\Lambda},M}$ is stronger than the pointwise sense. For every I_i in partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$, when $\mathcal{T}_n = \{b_j\}_{j=1}^{N_n}$ is sufficiently fine, i.e $\max_{1\leq j\leq N_n} diam(b_j)$ is sufficiently small, we can ensure that $|p_j - p'_i| < \epsilon$ for any j such that b_j has its associated $\lambda_j \in Q^{-1}(I_i)$, and

$$\mu_{\mathbf{\Lambda}}\left(\left(\bigcup_{\lambda_j \in Q^{-1}(I_i)} b_j\right) \triangle Q^{-1}(I_i)\right) < \epsilon.$$

Since M is finite, we can combine the above results for each $Q^{-1}(I_i)$ to guarantee that

$$|\rho_{\mathbf{\Lambda},M,N_n}(\lambda) - \rho_{\mathbf{\Lambda},M}(\lambda)| < \epsilon, \text{ for } \lambda \in \mathbf{\Lambda}' \subset \mathbf{\Lambda} \text{ with } \mu_{\mathbf{\Lambda}}(\mathbf{\Lambda} \setminus \mathbf{\Lambda}') < \epsilon.$$

Consequently, $|\max \rho_{\mathbf{\Lambda},M,N_n} - \max \rho_{\mathbf{\Lambda},M}| < \epsilon$.

Next, we want to use the Vitali Convergence Theorem 1.7.3 to conclude convergence of the integrals of functions that converge a.e., provided that the functions are uniformly integrable. Before that, we need to show $\{\rho_{\mathbf{\Lambda},M_m}\}_{m=1}^{\infty}$, and $\{\rho_{\mathbf{\Lambda},M,N_n}\}_{n=1}^{\infty}$ for a fixed M, are uniformly integrable.

For a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^{\infty} = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^{\infty}$, the sequence of densities $\{\rho_{\mathbf{\Lambda},M_m}\}_{m=1}^{\infty}$ is uniformly integrable by Theorem 3.2.3.

Fix the partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$, we show $\{\rho_{\mathbf{\Lambda},M,N_n}\}_{n=1}^{\infty}$ constructed from partition sequence $\{\mathcal{T}_n\}_{n=1}^{\infty} = \{\{b_j\}_{j=1}^{N_n}\}_{n=1}^{\infty}$ is uniformly integrable if $\lim_{n\to\infty} \max_{1\leq i\leq N_n} diam(b_j) = 0$.

Given $\epsilon > 0$, there exists $\delta > 0$ such that $\int_E \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}} < \epsilon$ for all $E \in \mathcal{B}_{\mathbf{\Lambda}}$ satisfying $\mu_{\mathbf{\Lambda}}(E) \leq \delta$. When the partition $\mathcal{T}_n = \{b_j\}_{j=1}^{N_n}$ on $\mathbf{\Lambda}$ is sufficiently fine, we have $|\rho_{\mathbf{\Lambda},M,N_n}(\lambda) - \rho_{\mathbf{\Lambda},M}(\lambda)| < \epsilon/\delta$ for $\lambda \in \mathbf{\Lambda}' \subset \mathbf{\Lambda}$ where $\mu_{\mathbf{\Lambda}}(\mathbf{\Lambda} \setminus \mathbf{\Lambda}') < \epsilon/(\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta)$, and $|\max \rho_{\mathbf{\Lambda},M,N_n} - \max \rho_{\mathbf{\Lambda},M}| < \epsilon/\delta$, so

$$\begin{aligned}
\int_E \rho_{\mathbf{\Lambda},M,N_n} d\mu_{\mathbf{\Lambda}} &= \int_{E \cap \mathbf{\Lambda}'} \rho_{\mathbf{\Lambda},M,N_n} d\mu_{\mathbf{\Lambda}} + \int_{E \setminus \mathbf{\Lambda}'} \rho_{\mathbf{\Lambda},M,N_n} d\mu_{\mathbf{\Lambda}} \\
&\leq \int_{E \cap \mathbf{\Lambda}'} (\rho_{\mathbf{\Lambda},M} + \epsilon/\delta) d\mu_{\mathbf{\Lambda}} + \frac{\epsilon}{\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta} (\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta) \\
&\leq \int_E \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}} + \epsilon + \epsilon < 3\epsilon.
\end{aligned}$$

for all $E \in \mathcal{B}_{\mathbf{\Lambda}}$ satisfying $\mu_{\mathbf{\Lambda}}(E) \leq \delta$. So the sequence $\{\rho_{\mathbf{\Lambda},M,N_n}\}$ is uniformly integrable.

If $A \in \mathcal{B}_{\mathbf{\Lambda}}$ satisfies $\mu_{\mathbf{\Lambda}}(\partial A) = 0$, A is approximated by

$$A_{N_n} = \bigcup_{\lambda_j \in A} b_j,$$

and $P_{\mathbf{\Lambda}}(A)$ is approximated by

$$P_{\mathbf{\Lambda},M_m,N_n}(A_{N_n}) = \sum_{\lambda_j \in A} P_{\mathbf{\Lambda},M_m,N_n}(b_j) = \int_{A_{N_n}} \rho_{\mathbf{\Lambda},M_m,N_n} d\mu_{\mathbf{\Lambda}}.$$

We prove $P_{\mathbf{\Lambda},M_m,N_n}(A_{N_n})$ converges to $P_{\mathbf{\Lambda}}(A)$ in two steps.

First, we show for a given partition $\mathcal{I} = \{I_i\}_{i=1}^{M}$ on $\mathcal{D}$,

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m,N_n} \mathbb{1}_{A_{N_n}} - \rho_{\mathbf{\Lambda},M_m} \mathbb{1}_A| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } n \to \infty. \tag{3.6}$$

Since the Vitali Convergence Theorem implies

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m,N_n} \mathbb{1}_A - \rho_{\mathbf{\Lambda},M_m} \mathbb{1}_A| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } n \to \infty, \tag{3.7}$$

and uniform integrability of $\rho_{\mathbf{\Lambda},M_m,N_n}$ implies

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m,N_n}\mathbb{1}_{A_{N_n}} - \rho_{\mathbf{\Lambda},M_m,N_n}\mathbb{1}_A| d\mu_{\mathbf{\Lambda}} = \int_{A_{N_n}\triangle A} \rho_{\mathbf{\Lambda},M_m,N_n} d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } n \to \infty, \tag{3.8}$$

combining (3.7) and (3.8) yields (3.6).

Secondly,

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m}\mathbb{1}_A - \rho_{\mathbf{\Lambda}}\mathbb{1}_A| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } m \to \infty, \tag{3.9}$$

is implied by the Vitali Convergence Theorem.

Combing the (3.6) and (3.9), we have

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m,N_n}\mathbb{1}_{A_{N_n}} - \rho_{\mathbf{\Lambda}}\mathbb{1}_A| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } n \to \infty \text{ and then } m \to \infty,$$

hence

$$\lim_{m\to\infty}\lim_{n\to\infty}\int_{A_{N_n}} \rho_{\mathbf{\Lambda},M_m,N_n} d\mu_{\mathbf{\Lambda}} = \int_A \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}} = P_{\mathbf{\Lambda}}(A).$$

□

The order of first letting $n \to \infty$ and then letting $m \to \infty$ is important. For example, one can show that for a fixed n, $\rho_{\mathbf{\Lambda},M_m,N_n} \to 0$ a.e. as $m \to \infty$, thus

$$\lim_{n\to\infty}\lim_{m\to\infty} \rho_{\mathbf{\Lambda},M_m,N_n} = 0 \text{ a.e.},$$

while the proof of Theorem 3.3.1 shows

$$\lim_{m\to\infty}\lim_{n\to\infty} \rho_{\mathbf{\Lambda},M_m,N_n} = \rho_{\mathbf{\Lambda}} \text{ a.e.}.$$

Theorem 3.3.1 is not the whole story in terms of approximation. In practice, it might be the case that $\mu_{\mathbf{\Lambda}}(b_j)$ are not known, so $P_{\mathbf{\Lambda},M,N}(b_j) = \dfrac{\mu_{\mathbf{\Lambda}}(b_j)}{\sum_{\lambda_k\in Q^{-1}(I_i)}\mu_{\mathbf{\Lambda}}(b_k)} P_{\mathbf{\Lambda}}(Q^{-1}(I_i))$ cannot be evaluated directly. Instead we further use an approximation $\hat{\mu}_{\mathbf{\Lambda}}(b_k)$ of $\mu_{\mathbf{\Lambda}}(b_k)$, to approximate $P_{\mathbf{\Lambda},M,N}(b_j)$ by $\hat{P}_{\mathbf{\Lambda},M,N}(b_j) = \dfrac{\hat{\mu}_{\mathbf{\Lambda}}(b_j)}{\sum_{\lambda_k\in Q^{-1}(I_i)}\hat{\mu}_{\mathbf{\Lambda}}(b_k)} P_{\mathbf{\Lambda}}(Q^{-1}(I_i))$, where $\hat{\mu}_{\mathbf{\Lambda}}(b_k)$ is an approximation of $\mu_{\mathbf{\Lambda}}(b_k)$.

However, when $\mathcal{T}_n = \{b_j\}_{j=1}^{N_n}$ has certain properties, an accurate approximation of $\mu_{\mathbf{\Lambda}}(b_j)$ may not be necessary. For example, when $\mu_{\mathbf{\Lambda}}(b_j)$ approximately follows a distribution with mean $1/N$, which is the case when $\mathcal{T}_n$ is Voronoi tessellation generated by a homogeneous Poisson process. Using $1/N$ as approximation of $\mu_{\mathbf{\Lambda}}(b_j)$ yields

$$\sum_{\lambda_j \in A} \hat{P}_{\mathbf{\Lambda},M,N_n}(b_j) \to \sum_{\lambda_j \in A} P_{\mathbf{\Lambda},M,N_n}(b_j) \text{ as } n \to \infty.$$

Here, $\hat{P}_{\mathbf{\Lambda},M,N_n}(b_j) = \dfrac{P_{\mathbf{\Lambda}}(Q^{-1}(I_i))}{\#k : \lambda_k \in Q^{-1}(I_i)}$ for $\lambda_j \in Q^{-1}(I_i)$.

That is because

$$\sum_{\lambda_j \in A} P_{\mathbf{\Lambda},M,N_n}(b_j) \to \int_A \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}} \text{ as } n \to \infty,$$

and for each i, by the Law of Large Numbers,

$$\frac{\#j : \lambda_j \in Q^{-1}(I_i) \cap A}{\#k : \lambda_k \in Q^{-1}(I_i)} \to \frac{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i) \cap A)}{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i))} = \frac{P_{\mathbf{\Lambda},M}(Q^{-1}(I_i) \cap A)}{P_{\mathbf{\Lambda},M}(Q^{-1}(I_i))} \text{ as } n \to \infty,$$

hence

$$\sum_{\lambda_j \in A} \hat{P}_{\mathbf{\Lambda},M,N_n}(b_j) \to \sum_{i=1}^{M} P_{\mathbf{\Lambda},M}(Q^{-1}(I_i) \cap A) = \int_A \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}}.$$

3.4 Counter Example

We construct a counter example using fat Cantor set $C^{\frac{1}{4}}$ to show when the approximation might fail. We need not worry about this counter example because in practice it is highly unlikely that someone use a Fat Cantor set or other set with non-negligible boundary as the domain.

Let

$$\mathbf{\Lambda} = ([0,1] \times [0,1]) \bigcup \left(C^{\frac{1}{4}} \times [1,2] \right),$$

and $Q(\lambda_1, \lambda_2) = \lambda_1$, so $\mathcal{D} = [0,1]$.

For a uniform probability measure on $\mathcal{D}$, the solution of SIP with the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$ is

$$\rho_{\mathbf{\Lambda}}(\lambda) = \rho_{\mathbf{\Lambda}}((\lambda_1, \lambda_2)) = \begin{cases} 1/4, & \lambda_1 \in C^{\frac{1}{4}}, \\ 1/2, & \lambda_1 \in [0,1] \setminus C^{\frac{1}{4}}. \end{cases}$$

It is continuous at $\lambda = (\lambda_1, \lambda_2)$ for which $\lambda_1 \in [0,1] \setminus C^{\frac{1}{4}}$, and discontinuous at λ for which $\lambda_1 \in C^{\frac{1}{4}}$. We review the key steps in the proof of Theorem 3.3.1 for the SIP for this problem to explain how the approximation algorithm works (and fails).

First, we impose a partition $\mathcal{I} = \{I_i\}_{i=1}^M$ on $\mathcal{D} = [0,1]$, and use

$$\rho_{\mathbf{\Lambda},M} = \sum_{i=1}^{M} p_i' \mathbf{1}_{Q^{-1}(I_i)}(\lambda), \quad p_i' = \frac{P_{\mathcal{D}}(I_i)}{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i))} = \frac{P_{\mathcal{D}}(I_i)}{\tilde{\mu}_{\mathcal{D}}(I_i)} = \frac{\int_{I_i} \frac{\rho_{\mathcal{D}}}{\tilde{\rho}_{\mathcal{D}}} d\tilde{\mu}_{\mathcal{D}}}{\int_{I_i} d\tilde{\mu}_{\mathcal{D}}}, \tag{3.10}$$

as an approximation of $\rho_{\mathbf{\Lambda}}$. This is equivalent to assuming the distribution is uniform on each $Q^{-1}(I_i)$, and using $\rho_{\mathbf{\Lambda},M}$ to approximate $\rho_{\mathbf{\Lambda}}$ on every $Q^{-1}(y)$ for $y \in I_i$. The Lebesgue Differentiation Theorem 1.7.1 guarantees $\rho_{\mathbf{\Lambda},M} \to \rho_{\mathbf{\Lambda}}$ a.e.. So this step works even for such a bizarre domain, provided that the $\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i))$ can be evaluated exactly.

Secondly, we impose a partition $\mathcal{T} = \{b_j\}_{j=1}^N$ on $\mathbf{\Lambda}$, and use the partition sets to approximate each $Q^{-1}(I_i)$. This approximation is difficult because $Q^{-1}(I_i)$ has non-negligible boundary. The theory of approximating a set with non-negligible boundary by Voronoi tesselation generated by random sample has not been established. The second step fails if the approximation algorithm is wrong. For example, if we use set inclusion as the standard, we would falsely identify $[0,1] \times [0,1]$ as $\mathbf{\Lambda}$, and end up approximating a uniform density on $[0,1] \times [0,1]$ as the solution.

3.5 Sources of error in the computation

There are two types of error in the approximation of the probability measure $P_{\mathbf{\Lambda}}$ using the numerical algorithm. The first type of error come from the partitions $\mathcal{T} = \{b_j\}_{j=1}^N$ and $\mathcal{I} = \{I_i\}_{i=1}^M$, and lack of information of $\mu_{\mathbf{\Lambda}}(b_j)$. The second type of error arises from the numerical evaluation of Q. We only analyze the first type. The error of approximation of $P_{\mathbf{\Lambda}}(A)$ for $A \in \mathcal{B}_{\mathbf{\Lambda}}$ can be decomposed as:

$$
\begin{aligned}
\sum_{\lambda_j \in A} \hat{P}_{\mathbf{\Lambda},M,N}(b_j) - P_{\mathbf{\Lambda}}(A) = & \int_A \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}} - \int_A \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}} && \mathbf{I} \\
& + \int_{A_N} \rho_{\mathbf{\Lambda},M,N} d\mu_{\mathbf{\Lambda}} - \int_A \rho_{\mathbf{\Lambda},M} d\mu_{\mathbf{\Lambda}} && \mathbf{II} \\
& + \sum_{\lambda_j \in A} \hat{P}_{\mathbf{\Lambda},M,N}(b_j) - \sum_{\lambda_j \in A} P_{\mathbf{\Lambda},M,N}(b_j) && \mathbf{III}
\end{aligned}
$$

Term **I** is associated with the partition $\mathcal{I} = \{I_i\}_{i=1}^M$ on $\mathcal{D}$.

Given $\mathcal{I}$, term **II** is associated with the partition $\mathcal{T} = \{b_j\}_{j=1}^N$ on $\mathbf{\Lambda}$. A finer $\mathcal{I}$ requires a finer $\mathcal{T}$ to keep term **II** small.

Term **III** arises due to lack of information of $\mu_{\mathbf{\Lambda}}(b_j)$. When $\mathcal{T}$ is regular grid where every $\mu_{\mathbf{\Lambda}}(b_j)$ is the same and known exactly, term **III** vanishes. When $\mathcal{T}$ is Voronoi Tesselation generated by random sample, we may just use $\hat{\mu}_{\mathbf{\Lambda}}(b_j) = 1/N$, then term **III** is inevitable. In other practical cases, using a precise approximation of $\mu_{\mathbf{\Lambda}}(b_j)$ yields a smaller term **III**.

Chapter 4

Formulation and Solution of the SIP for an Infinite Dimensional Domain

4.1 Introduction

In many applications, the inputs of the mathematical models are functions rather than finite dimensional vectors. To motivate such an extension, recall the metal plates example. Previously, we treat the purity of alloy and the thickness as two parameters taking values in a two dimensional parameter space on which the probability measure can be solved by the SIP. This choice of parameters relies on the assumption that the plates are all homogeneous in their physical properties, i.e. the purity of alloy and the thickness are the same at any location on a plate. It is physically natural to consider the situation in which the plates are inhomogeneous in their physical properties and treat the purity of alloy and the thickness as two functions of the location on the plate. The previous SIP framework can be extended to allow for the description and computation of probability distributions over input functions. To that end, we need to extend the domain of map Q.

We assume that the input function λ is square integrable on a measurable space $(\mathbf{\Omega}, \mathcal{B}_{\mathbf{\Omega}}, \mu_{\mathbf{\Omega}})$, thus $\lambda \in L^2(\mathbf{\Omega})$. However, $L^2(\mathbf{\Omega})$ contains functions not of interest, for example, very rough and highly oscillatory functions that are neither physical nor allowable as inputs into the differential equation. Like the SIP with finite dimensional domain, we consider a compact subset of $L^2(\mathbf{\Omega})$ as the domain.

The Riesz-Fischer Theorem implies $L^2(\mathbf{\Omega})$ is a complete space. Moreover, when $\mathbf{\Omega}$ is a subset of $\mathbb{R}^n$ equipped with the inherited Borel σ-algebra and measure, $L^2(\mathbf{\Omega})$ is separable (Bruckner et al., 1997) and has a countable orthonormal basis. For any given basis, there is an isometric isomorphism between the function space $L^2(\mathbf{\Omega})$ and the space of the coefficients, which is a space of square summable sequences ℓ^2, and the isometric isomorphism preserves the inner product

(Conway, 1994, chap. 1). Therefore, a measure structure on ℓ^2 or on a subset of ℓ^2 is equivalent to the measure structure on $L^2(\mathbf{\Omega})$ or on a subset of $L^2(\mathbf{\Omega})$. This enables us to consider compact subset of ℓ^2 as the domain $\mathbf{\Lambda}$, thereby allowing more straightforward generalization from $\mathbb{R}^n$ to ℓ^2.

Common orthonormal basis functions include the standard Fourier basis for $L^2([0,1])$, which consist of orthonormal trigonometric functions:

$$\{\frac{1}{2}, \cos(2\pi jt), \sin(2\pi jt), j = 1, 2, \cdots\}.$$

Thus

$$\lambda(t) = (1/2)a_0 + \sum_{j=1}^{\infty} a_j \cos(2\pi jt) + \sum_{j=1}^{\infty} b_j \sin(2\pi jt),$$

where the Fourier coefficients are given by

$$a_j = 2\int_0^1 \lambda(t)\cos ntdt, \quad j = 0, 1, 2, \cdots$$
$$b_j = 2\int_0^1 \lambda(t)\sin ntdt, \quad j = 1, 2, 3, \cdots$$

for all $\lambda(t) \in L^2([0,1])$. Of course, there are other orthonormal bases, for example, the Haar wavelets form a countable orthonormal basis on $L^2([0,1])$ (Daubechies, 1992) as well. An advantage of certain bases like the standard Fourier basis or Haar wavelets is that the decay rate of the coefficients with respect to index characterize the smoothness of functions. For example, a continuous periodic function whose first k derivatives are continuous but whose $k+1$ derivative is discontinuous has Fourier coefficients that decay at a rate of $1/n^{k+2}$ after some finite index (Grafakos, 2008). There are wavelet bases that have similar properties (Daubechies, 1992, Section 2.9). It is possible to characterize Sobolev spaces in terms of decay of coefficients in such bases.

This a good point to recall the importance of the Hilbert cube, which is the Cartesian product of the sequence of intervals $\{[0, \frac{1}{n}]\}_{n=1}^{\infty}$, in the study of measure structures in infinite dimensional spaces. Any product of countably infinite intervals of decaying lengths is homeomorphic to the hilbert cube, and is generally called a "countable dimensional cube". Thus, we can directly relate a subspace of $L^2(\Omega)$ consisting of functions of specified smoothness quantified by an asymptotic decay rate in the coefficients with respect to the Fourier basis to a countable dimensional cube.

It turns out that dealing with a class of $L^2(\Omega)$ functions of a specified smoothness or equivalently to the countable dimensional cube with a specified decay rate in the interval lengths provides a systematic way to approximate the probability structure on the infinite dimensional domain by probability structures on the finite dimensional domains obtained by "truncation" or ending the product after a specified point. The error in these approximations can be controlled by increasing the dimension.

The countable dimensional cube is a metric space with the natural metric induced by the ℓ_2-norm. Because of the decaying edge lengths, the topology induced by this metric is the product topology, for which the basis is cylindrical open set whose "face" is an open set in the product of intervals for first n dimensions $n \in \mathbb{N}$ and whose "side" are the product of intervals in the remaining dimensions.

For notational convenience, we center and scale the countable dimensional cube C to $[0,1]^\infty$ and instead consider a metric with the decay built in. Let this map be $p : C \to [0,1]^\infty$. Then p is one-to-one and onto. With the ℓ_2-norm on C, we define a metric d on $[0,1]^\infty$ by $d(\lambda, \mu) = ||p^{-1}(\lambda) - p^{-1}(\mu)||_2$, and denote the metric space $([0,1]^\infty, d)$. The metric d poses the same decay rate as edges of C, which induces the product topology on $[0,1]^\infty$. For any $n \in \mathbb{N}$, $[0,1]^\infty$ as a topology space is a product space of $[0,1]^n$, endowed with the topology induced from $\mathbb{R}^n$, and $[0,1]^\infty_n = \prod_{k=n+1}^\infty [0,1]$, endowed with the product topology, written as $[0,1]^\infty = [0,1]^n \times [0,1]^\infty_n$

We consider the SIP for $Q : [0,1]^\infty \to \mathcal{D}$, $\mathcal{D} \subset \mathbb{R}^m$. The product decomposition allows the definition of $Q_{\hat{\lambda}^n} : [0,1]^n \to \mathcal{D}$ as $Q_{\hat{\lambda}^n} = Q(\cdot, \hat{\lambda}^n)$ for $\hat{\lambda}^n \in [0,1]^\infty_n$. When $n \geq m$, the SIP for $Q_{\hat{\lambda}^n} : [0,1]^n \to \mathcal{D}$ falls into the previous formulation hence can be solved by Algorithm 2. Later we show that the solution of the SIP for $Q : [0,1]^\infty \to \mathcal{D}$ can be approximated by the solution of the SIP for $Q_{\hat{\lambda}^n} : [0,1]^n \to \mathcal{D}$ under some conditions. The convergence rate depends on both the decay rate and the properties of map Q. In this regard, taking a countable dimensional cube to be $\mathbf{\Lambda}$

in the SIP amounts to regularizing the domain where the regularization strength is controlled by the decay rate, in contrast to the usual approach to inverse problem, where the map Q is regularized.

In this chapter, we present a formulation of the SIP for $\mathbf{\Lambda} = ([0,1]^\infty, d)$. In Section 4.2, we present a technical result regarding the SIP for a domain that can be expressed as a product space of two spaces, i.e. $\mathbf{\Lambda} = \mathbf{\Lambda}_1 \times \mathbf{\Lambda}_2$. In Section 4.3, we give the conditions for the density to exist and be continuous with respect to $\mathbf{\Lambda} = \mathbf{\Lambda}_1 \times \mathbf{\Lambda}_2$. In Section 4.4, we discuss the topology, Borel $\sigma-$algebra and "volume" measure of $([0,1]^\infty, d)$. In Section 4.5, we show how to adapt the previous proof of continuity to accommodate the infinite dimensional setting. This version of continuity theorem is then applied to show the existence of solution. Other implications of continuity are also discussed.

In this chapter, $\mathbf{\Lambda}$ is still the domain of mathematical map Q, but represents different sets from before. In Section 4.2 and Section 4.3, it denotes a general space that can be expressed as a product space of two spaces, i.e. $\mathbf{\Lambda} = \mathbf{\Lambda}_1 \times \mathbf{\Lambda}_2$. In Section 4.4 and Section 4.5, $\mathbf{\Lambda} = ([0,1]^\infty, d)$.

4.2 The SIP for a product space domain

We consider the case where the domain of the SIP is a product measure space $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ of two measure spaces $(\mathbf{\Lambda}_1, \mathcal{B}_{\mathbf{\Lambda}_1}, \mu_{\mathbf{\Lambda}_1})$ and $(\mathbf{\Lambda}_2, \mathcal{B}_{\mathbf{\Lambda}_2}, \mu_{\mathbf{\Lambda}_2})$. In this section, we assume $Q : \mathbf{\Lambda} \to \mathcal{D}$ is measurable. Consequently, $Q_{\lambda_2}(\cdot) = Q(\cdot, \lambda_2) : \mathbf{\Lambda}_1 \to \mathcal{D}$ is also measurable for all $\lambda_2 \in \mathbf{\Lambda}_2$ (Folland, 2013). We let $(\mathcal{D} \times \mathbf{\Lambda}_2, \mathcal{B}_{\mathcal{D}\times\mathbf{\Lambda}_2}, \mu_{\mathcal{D}\times\mathbf{\Lambda}_2})$ be the product measure space of $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$ and $(\mathbf{\Lambda}_2, \mathcal{B}_{\mathbf{\Lambda}_2}, \mu_{\mathbf{\Lambda}_2})$.

Next theorem states how measure induced by Q can be computed from measures induced by Q_{λ_2}.

Theorem 4.2.1. *Let $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$ and $\tilde{\mu}_{\mathcal{D},\lambda_2} = Q_{\lambda_2}\mu_{\mathbf{\Lambda}_1}$ for $\lambda_2 \in \mathbf{\Lambda}_2$. Then for $A \in \mathcal{B}_{\mathcal{D}}, \tilde{\mu}_{\mathcal{D}}(A) = \int_{\mathbf{\Lambda}_2} \tilde{\mu}_{\mathcal{D},\lambda_2}(A) d\mu_{\mathbf{\Lambda}_2}$.*

Moreover, if $d\tilde{\mu}_{\mathcal{D},\lambda_2}(y) = \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y, \lambda_2) d\mu_{\mathcal{D}}(y)$, where $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y, \lambda_2)$ is uniquely defined $\mu_{\mathcal{D}}$-a.e. for each λ_2, and $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y, \lambda_2)$ is measurable on $\mathcal{D} \times \mathbf{\Lambda}_2$, then $\tilde{\mu}_{\mathcal{D}}$ is absolutely continuous with respect to $\mu_{\mathcal{D}}$ and has density $\tilde{\rho}_{\mathcal{D}}(y) = \int_{\mathbf{\Lambda}_2} \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y, \lambda_2) d\mu_{\mathbf{\Lambda}_2}$ uniquely defined up to a $\mu_{\mathcal{D}}$-null set.

Proof. For $A \in \mathcal{B}_{\mathcal{D}}$,

$$
\begin{aligned}
\int_A d\tilde{\mu}_{\mathcal{D}} &= \int_{Q^{-1}(A)} d\mu_{\mathbf{\Lambda}} && \text{definition of } \tilde{\mu}_{\mathcal{D}} \\
&= \int_{\mathbf{\Lambda}_2} \int_{\mathbf{\Lambda}_1} \mathbb{1}_{(Q_{\lambda_2}^{-1}(A))} d\mu_{\mathbf{\Lambda}_1} d\mu_{\mathbf{\Lambda}_2} && \text{Tonelli's Theorem} \\
&= \int_{\mathbf{\Lambda}_2} \tilde{\mu}_{\mathcal{D},\lambda_2}(A) d\mu_{\mathbf{\Lambda}_2} \\
&= \int_{\mathbf{\Lambda}_2} \int_A \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2) d\mu_{\mathcal{D}} d\mu_{\mathbf{\Lambda}_2} \\
&= \int_A \int_{\mathbf{\Lambda}_2} \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2) d\mu_{\mathbf{\Lambda}_2} d\mu_{\mathcal{D}} && \text{Tonelli's Theorem}
\end{aligned}
$$

□

The minimal condition guaranteeing $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ to be measurable is difficult to find, but continuity can be used as a sufficient condition.

Proposition 4.2.2. *If $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ is continuous a.e. on $\mathcal{D}\times\mathbf{\Lambda}_2$, then $\tilde{\rho}_{\mathcal{D}}(y) = \int_{\mathbf{\Lambda}_2} \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2) d\mu_{\mathbf{\Lambda}_2}$.*

Proof. When $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ is continuous a.e. on $\mathcal{D}\times\mathbf{\Lambda}_2$, it is measurable with respect to $\mathcal{B}_{\mathcal{D}\times\mathbf{\Lambda}_2}$. □

4.3 The continuity of density in a product space

In Chapter 2, we discuss the existence and continuity of $\tilde{\rho}_{\mathcal{D}}$ for the SIP. We can give analogous conditions to ensure $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ exist and be continuous on $\mathbf{\Lambda}_2\times\mathcal{D}$. Next theorem states the conditions for $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ to exist.

Theorem 4.3.1. *Let $\mathbf{\Lambda}_1 \subset \mathbb{R}^n$, $\mathcal{D} \subset \mathbb{R}^m$, and $m \le n$. Given a $\lambda_2 \in \mathbf{\Lambda}_2$, if $Q_{\lambda_2} : \mathbf{\Lambda}_1 \to \mathcal{D}$ is C^1 map and $J_{Q_{\lambda_2}}$ is full rank on $\mathbf{\Lambda}_1$, then $d\tilde{\mu}_{\mathcal{D},\lambda_2} = \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2) d\mu_{\mathcal{D}}$, with*

$$
\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2) = \int_{Q_{\lambda_2}^{-1}(y)} \frac{1}{\sqrt{(J_{Q_{\lambda_2}} J_{Q_{\lambda_2}}^T)}} ds. \tag{4.1}
$$

which is an integration over manifold $Q_{\lambda_2}^{-1}$ in $\mathbf{\Lambda}_1$, and $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ is uniquely defined $\mu_{\mathcal{D}}$-a.e..

Proof. This is implied by Theorem 2.1.1. □

To study the continuity property of $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$, we extend the domain $\mathbf{\Lambda}_1$ to an open set $U_{\mathbf{\Lambda}_1}$ containing $\mathbf{\Lambda}_1$. Write $\lambda \in U_{\mathbf{\Lambda}_1} \times \mathbf{\Lambda}_2$ as $\lambda = (\lambda_1, \lambda_2)$ where $\lambda_1 \in U_{\mathbf{\Lambda}_1}$ and $\lambda_2 \in \mathbf{\Lambda}_2$. We denote the Jacobian of Q with respect to λ_1 as J_{Q,λ_1}, that is $J_{Q,\lambda_1}(\lambda_1,\lambda_2) = J_{Q_{\lambda_2}}(\lambda_1)$.

Let Q be a map from $U_{\mathbf{\Lambda}_1} \times \mathbf{\Lambda}_2$ to $\mathcal{D}$, where $U_{\mathbf{\Lambda}_1}$ is an open set in $\mathbb{R}^n$, $\mathbf{\Lambda}_2$ is a metric space, $\mathcal{D}$ is a subset of $\mathbb{R}^m$ and $n \geq m$. Assume J_{Q,λ_1} exist and is continuous on $U_{\mathbf{\Lambda}_1} \times \mathbf{\Lambda}_2$. Further assume $\mathbf{\Lambda}_1 \subset U_{\mathbf{\Lambda}_1}$ is a compact subset of $\mathbb{R}^n$, and J_{Q,λ_1} is full rank on $U_{\mathbf{\Lambda}_1} \times \mathbf{\Lambda}_2$. We define $\tilde{\mu}_{\mathcal{D},\lambda_2} = Q_{\lambda_2}\mu_{\mathbf{\Lambda}_1}$ then $d\tilde{\mu}_{\mathcal{D},\lambda_2} = \tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)d\mu_{\mathcal{D}}$ for every $\lambda_2 \in \mathbf{\Lambda}_2$ by Theorem 4.3.1. Next theorem shows conditions analogous to Theorem 2.4.2 guarantees $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ to be continuous on $\mathcal{D} \times \mathbf{\Lambda}_2$.

Theorem 4.3.2. *Let $Q : U_{\mathbf{\Lambda}_1} \times \mathbf{\Lambda}_2 \to \mathcal{D}$ be a continuous map. Suppose $\partial\mathbf{\Lambda}_1$ is locally determined by $B(\lambda_1) = 0$, where $B : U_{\mathbf{\Lambda}_1} \to \mathbb{R}^m$ is a piecewise C^1 map. Given $y_0, \lambda_{2,0}$, such that $\begin{pmatrix} J_{Q_{\lambda_{2,0}}} \\ J_B \end{pmatrix}$ is full rank on $\partial\mathbf{\Lambda}_1 \cap Q^{-1}_{\lambda_{2,0}}(y_0)$ or $\partial\mathbf{\Lambda}_1 \cap Q^{-1}_{\lambda_{2,0}}(y_0) = \varnothing$, then $\tilde{\rho}_{\mathcal{D},\mathbf{\Lambda}_2}(y,\lambda_2)$ is continuous in a neighborhood of $y_0, \lambda_{2,0}$.*

This theorem can be proved by altering a few key steps in the proofs of Theorem 2.4.1 and Theorem 2.4.2. In Theorem 2.4.1, we require the submatrix of J_Q consisting of its first m columns to be full rank, a special case allowing uniform parameterization. Let $Q'(\lambda, y) = Q(\lambda) - y$. Theorem 2.4.1 is proved by following several vital steps given below.

1. We use the Implicit Function Theorem with respect to $Q'(\lambda, y) = 0$, to derive a uniform parameterization of the manifold $Q^{-1}(y) \cap \mathbf{\Lambda}$ in terms of its projection to subspace of last $n - m$ coordinates for y in a neighborhood of y_0. So $\tilde{\rho}_{\mathcal{D}}$ can be expressed as an integration of a function involving $J_{Q'}(\lambda, y)$ over a domain in the subspace of last $n - m$ coordinates.

2. The boundary of the integration domain is the projection of $Q^{-1}(y) \cap \partial\mathbf{\Lambda}$ to subspace of last $n - m$ coordinates. We can apply the Implicit Function Theorem with respect to

$$Q^*(\lambda, y) = \begin{pmatrix} Q'(\lambda, y) \\ B(\lambda) \end{pmatrix} = 0$$

to get piecewise parameterization of the boundary of the integration domain. That means the integration domain changes continuously with respect to y;

3. The integrand is continuous since $J_{Q'}(\lambda, y)$ is continuous on $U_{\mathbf{\Lambda}} \times \mathcal{D}$.

Theorem 2.4.2 is proved by decomposing $\mathbf{\Lambda}$ to subdomains determined by balls covering $\mathbf{\Lambda}$. In each of the subdomains, $Q^{-1}(y)$ admits uniform parameterization, and the boundary of each subdomain is determined by the sphere of the ball and the orginal boundary $\partial\mathbf{\Lambda}$, which are both smooth manifolds.

By redefining Q' as $Q'(\lambda_1, \lambda_2, y) = Q(\lambda_1, \lambda_2) - y$ and change λ to λ_1, y to (λ_2, y), and y_0 to $(\lambda_{2,0}, y_0)$ of the proof of Theorem 2.4.1, we come up with a proof of the special case of Theorem 4.3.2 where $\mathbf{\Lambda}_1 \cap Q^{-1}_{\lambda_{2,0}}(y_0)$ can be uniformly parameterized. Then, by doing the same subdomain trick as in the proof of Theorem 2.4.2, we prove Theorem 4.3.2 for the general case.

4.4 The countable dimensional cube $([0,1]^\infty, d)$

In this section, we are concerned with the metrics, topology, Borel σ-algebra and the measure of domain $\mathbf{\Lambda} = ([0,1]^\infty, d)$.

4.4.1 Decaying metric and product topology

For $([0,1]^\infty, d)$, we prove the decaying metric induces the product topology.

Theorem 4.4.1. *The product topology on $[0,1]^\infty$ can be induced by the metric d defined as $d(\lambda, \mu) = (\sum_{i=1}^\infty a_i^p |\lambda_i - \mu_i|^p)^{\frac{1}{p}}$, where $1 \le p < \infty$, and $\{a_i\}$ is a sequence of positive numbers with $\sum_{i=1}^\infty a_i^p < \infty$.*

Proof. Recall that the basis of product topology has the form $V = \prod_{i=1}^\infty U_i$ where $U_i \subset [0,1]$ is open for all i and $U_i = [0,1]$ for i outside for some finite subset of $\{1, 2, 3, \dots\}$. We refer to the family of all such V as $\mathcal{V}$.

We only need to prove that given a $V \in \mathcal{V}$, for all $\lambda \in V$, there exists ϵ, such that $B_d(\lambda, \epsilon) \subset V$; and for all $B_d(\lambda, \epsilon)$, there exists a $V \in \mathcal{V}$, $\lambda \in V$ such that $V \subset B_d(\lambda, \epsilon)$.

On the one side, by the construction of set V, there exists n such that $V = U \times \hat{\mathbf{\Lambda}}^n$, where U is an open set in $\mathbf{\Lambda}^n$, and $\mathbf{\Lambda} = \mathbf{\Lambda}^n \times \hat{\mathbf{\Lambda}}^n$. Then for all $\lambda \in V$ where $\lambda = (\lambda^n, \hat{\lambda}^n)$, $\lambda^n \in U$, $\hat{\lambda}^n \in \hat{\mathbf{\Lambda}}^n$.

Define $d^n : \mathbf{\Lambda}^n \times \mathbf{\Lambda}^n \to \mathbb{R}^+$ as $d^n(\lambda^n, \mu^n) = (\sum_{i=1}^n a_i^p |\lambda_i - \mu_i|^p)^{\frac{1}{p}}$ for $\lambda^n, \mu^n \in \mathbf{\Lambda}^n$. d^n is a well defined metric on $\mathbf{\Lambda}^n$. Since $\lambda^n \in U$, there exists an ϵ, $B_{d^n}(\lambda^n, \epsilon) \subset U$. Consequently, $B_{d^n}(\lambda^n, \epsilon) \times \hat{\mathbf{\Lambda}}^n \subset U \times \hat{\mathbf{\Lambda}}^n = V$. Since $B_d(\lambda, \epsilon) \subset B_{d^n}(\lambda^n, \epsilon) \times \hat{\mathbf{\Lambda}}^n$, $B_d(\lambda, \epsilon) \subset V$.

On the other side, given any $B_d(\lambda, \epsilon)$, there exists an m, such that $\sum_{i=m+1}^{\infty} a_i^p < \frac{\epsilon^p}{2}$. Then for all $y \in V = B_{d^m}(\lambda^m, \epsilon/2) \times \hat{\mathbf{\Lambda}}^m$,

$$
\begin{aligned}
d(\lambda, \mu) &= (\sum_{i=1}^{\infty} a_i^p |\lambda_i - \mu_i|^p)^{\frac{1}{p}} \\
&< (\sum_{i=1}^{m} a_i^p |\lambda_i - \mu_i|^p + \sum_{i=m+1}^{\infty} a_i^p)^{\frac{1}{p}} \\
&< ((\epsilon/2)^p + \frac{\epsilon^p}{2})^{\frac{1}{p}} \\
&= \epsilon(\frac{1}{2^p} + \frac{1}{2})^{\frac{1}{p}} \leq \epsilon
\end{aligned}
$$

Then $V \subset B_d(\lambda, \epsilon)$. □

4.4.2 The Borel $\sigma-$algebra and measure

In order to formulate the SIP with domain $\mathbf{\Lambda} = ([0,1]^\infty, d)$, we need to define σ-algebra $\mathcal{B}_{\mathbf{\Lambda}}$ and measure $\mu_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$. Analogous to the SIP with finite dimensional domain, the measure $\mu_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ should properly reflect intuition of "volume". Specifically, $\mu_{\mathbf{\Lambda}}(\mathbf{\Lambda}) = 1$, and it should equal the Lebesgue measure on a finite dimensional face of a cylindrical set, thus $\mu_{\mathbf{\Lambda}}(A \times [0,1]_n^\infty) = \mu(A)$, where A is a measurable set in $\mathbb{R}^n$ and μ is Lebesgue measure on $\mathbb{R}^n$. But $\mu_{\mathbf{\Lambda}}$ may lack other properties, like rotation invariance, of the Lebesgue measure. We use a construction adapted from Da Prato (2006), using cylindrical sets to generate the Borel σ-algebra and desired measure.

Define cylindrical subsets $I_{n,A}$ of $[0,1]^\infty$, where $n \in \mathbb{N}$ and $A \in \mathcal{B}_{\mathbf{\Lambda}^n}$ as:

$$
I_{n,A} = \{\lambda = (\lambda_k) \in [0,1]^\infty : (\lambda_1, \ldots, \lambda_n) \in A\}.
$$

Denote the family of all such cylindrical sets as $\mathcal{C}$.

The set $I_{n,A}$ can be also expressed as

$$I_{n,A} = I_{n+k,A\times[0,1]^k}, k, n \in \mathbb{N}.$$

Using this identity we can show that $\mathcal{C}$ is an algebra. In fact, if $I_{n,A}$ and $I_{m,B}$ are two cylindrical sets we have

$$\begin{aligned} I_{n,A} \cup I_{m,B} &= I_{m+n,A\times[0,1]^m} \cup I_{m+n,B\times[0,1]^n} \\ &= I_{m+n,A\times[0,1]^m \cup B\times[0,1]^n}, \end{aligned}$$

and $I_{n,A}^c = I_{n,A^c}$.

We make the concept of "face" of cylindrical sets formal. We call a set $A \in \mathcal{B}_{\mathbf{\Lambda}^\ell}$ as the face of cylindrical set I if

$$I = A \times [0,1]_\ell^\infty.$$

Therefore A and $A \times [0,1]^k$ for positve integer k are faces of $I_{n,A}$.

Define the minimal face of I as $F(I)$, such that $F(I)$ is the face of I and $F(I) \in \mathcal{B}_{\mathbf{\Lambda}^n}$ where n is the smallest possible integer.

The σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ coincides with $\mathcal{B}_{\mathbf{\Lambda}}$ since any ball (with respect to the decaying metric defined in Theorem 4.4.1) is a countable intersection of cylindrical sets.

Finally, we define the product pre-measure

$$\mu(I_{n,A}) = (\mu_1 \times \cdots \times \mu_n)(A), I_{n,A} \in \mathcal{C}$$

It is easy to see μ is additive. We only need to verify that μ is $\sigma-$additive on $\mathcal{C}$. The Caratheodory Extension Theorem implies that μ can be uniquely extended to a measure on the product $\sigma-$algebra $\mathcal{B}_{[0,1]^\infty}$.

Theorem 4.4.2. *μ is $\sigma-$additive on $\mathcal{C}$ and it possesses a unique extension to a measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$.*

Proof. To prove the $\sigma-$additivity of μ, we show continuity of μ at $\varnothing$. This is equivalent to proving that if $\{E_j\}$ is a decreasing sequence on $\mathcal{C}$ such that

$$\mu(E_j) \geq \epsilon, j \in \mathbb{N}$$

for some $\epsilon > 0$, we have

$$\bigcap_{j=1}^{\infty} E_j \neq \varnothing$$

To show this, we consider the sections of E_j,

$$E_j(\alpha) = \{\lambda \in [0,1]_1^{\infty} : (\alpha, \lambda) \in E_j\}, \alpha \in [0,1].$$

Let

$$F_j^{(1)} = \{\alpha \in [0,1] : \mu^{(1)}(E_j(\alpha)) \geq \frac{\epsilon}{2}\}, j \in \mathbb{N},$$

where $\mu^{(n)} = \prod_{k=n+1}^{\infty} \mu_k, n \in \mathbb{N}$. Then Fubuni's Theorem implies

$$\begin{aligned}
\mu(E_j) &= \int_{[0,1]} \mu^{(1)}(E_j(\alpha))\mu_1(d\alpha) \\
&= \int_{F_j^{(1)}} \mu^{(1)}(E_j(\alpha))\mu_1(d\alpha) + \int_{[F_j^{(1)}]^c} \mu^{(1)}(E_j(\alpha))\mu_1(d\alpha) \\
&\leq \mu_1(F_j^{(1)}) + \frac{\epsilon}{2}
\end{aligned}$$

Therefore,

$$\mu_1(F_j^{(1)}) \geq \frac{\epsilon}{2}$$

Since μ_1 is Lebesgue measure on $([0,1], \mathcal{B}_{[0,1]})$, it is continuous at $\varnothing$. Therefore, since the sequence $\{F_j^{(1)}\}$ is decreasing, there exists $\overline{\alpha_1} \in [0,1]$, such that

$$\mu^1(E_j(\overline{\alpha_1})) \geq \frac{\epsilon}{2}, j \in \mathbb{N},$$

Consequently, we have

$$E_j(\overline{\alpha_1}) \neq \varnothing.$$

Now set

$$E_j(\overline{\alpha_1}, \alpha_2) = \{\lambda_2 \in [0,1]_2^\infty : (\overline{\alpha_1}, \alpha_2, \lambda) \in E_j\}, j \in \mathbb{N}, \alpha_2 \in [0,1]$$

and

$$F_j^{(2)} = \{\alpha_2 \in [0,1] : \mu^{(2)}(E_j(\overline{\alpha_1}, \alpha_2)) \geq \frac{\epsilon}{4}\}, j \in \mathbb{N}.$$

Then, again by Fubini's theorem, we have

$$\begin{aligned}
\mu^1(E_j(\overline{\alpha_1})) &= \int_{[0,1]} \mu^{(2)}(E_j(\overline{\alpha_1}, \alpha_2))\mu_2(d\alpha_2) \\
&= \int_{F_j^{(2)}} \mu^{(2)}(E_j(\overline{\alpha_1}, \alpha_2))\mu_2(d\alpha_2) + \int_{[F_j^{(2)}]^c} \mu^{(2)}(E_j(\overline{\alpha_1}, \alpha_2))\mu_2(d\alpha_2) \\
&\leq \mu_2(F_j^{(2)}) + \frac{\epsilon}{4}.
\end{aligned}$$

Therefore,

$$\mu_2(F_j^{(2)}) \geq \frac{\epsilon}{4}$$

Since $(F_j^{(2)})$ is decreasing, there exists $\overline{\alpha_2} \in \mathbb{R}$ such that

$$\mu^2(E_j(\overline{\alpha_1}, \overline{\alpha_2})) \geq \frac{\epsilon}{4}, j \in \mathbb{N}$$

and consequently we have

$$E_j(\overline{\alpha_1}, \overline{\alpha_2}) \neq \varnothing$$

Arguing in a similar way as before we see that there exists a sequence $(\overline{\alpha_k}) \in [0,1]^\infty$ such that

$$E_j(\overline{\alpha_1},\dots,\overline{\alpha_n}) \neq \varnothing$$

where

$$E_j(\alpha_1,\dots,\alpha_n) = \{\lambda \in [0,1]_n^\infty : (\alpha_1,\dots,\alpha_n,x) \in E_j\}, n \in \mathbb{N}$$

This implies

$$(\alpha_n) \in \bigcap_{j=1}^{\infty} E_j.$$

Therefore $\bigcap_{j=1}^{\infty} E_j$ is not empty. Thus we have proved that μ is $\sigma-$additive on $\mathcal{C}$ and consequently on $\mathcal{B}_{\mathbf{\Lambda}}$. $\square$

4.5 Convergence of the solution on the dimension truncated domain

4.5.1 Solution of the SIP on domain $([0,1]^\infty, d)$

The fact that the domain $\mathbf{\Lambda} = ([0,1]^\infty, d)$ admits product decomposition $\mathbf{\Lambda} = [0,1]^\infty = [0,1]^n \times [0,1]_n^\infty = \mathbf{\Lambda}^n \times \hat{\mathbf{\Lambda}}^n$ for any $n \in \mathbb{N}$ gives us great flexibility in choosing an appropriate n to show the existence of the solution $\rho_{\mathbf{\Lambda}}$.

Theorem 4.5.1. *Let $Q : \mathbf{\Lambda} \to \mathcal{D}$, and $\mathcal{D} \subset \mathbb{R}^m$. If there exists $M \geq m$ with $\mathbf{\Lambda} = \mathbf{\Lambda}^M \times \hat{\mathbf{\Lambda}}^M$, such that $Q^M_{\hat{\lambda}^M}(\cdot) = Q(\cdot, \hat{\lambda}^M) : \mathbf{\Lambda}^M \to \mathcal{D}$ is continuously differentiable on $\mathbf{\Lambda}$ and has full rank $J_{Q^M_{\hat{\lambda}^M}}$ for all $\hat{\lambda}^M \in \hat{\mathbf{\Lambda}}^M$. Define $\tilde{\mu}_{\mathcal{D},\hat{\lambda}^M} = Q_{\hat{\lambda}^M}\mu_{\mathbf{\Lambda}^M}$, then $d\tilde{\mu}_{\mathcal{D},\hat{\lambda}^M} = \tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M) d\mu_{\mathcal{D}}$ with*

$$\tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M) = \int_{Q^{-1}_{\hat{\lambda}^M}(y)} \frac{1}{\sqrt{(J_{Q_{\hat{\lambda}^M}} J^T_{Q_{\hat{\lambda}^M}})}} ds. \tag{4.2}$$

$\tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M)$ is uniquely defined $\mu_{\mathcal{D}}$-a.e..

Proof. This is a special case of Theorem 4.3.1 where $\mathbf{\Lambda}_2 = \hat{\mathbf{\Lambda}}^M$. $\square$

Just like in Section 4.3, to study the continuity property of $\tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)$, we extend $\mathbf{\Lambda}^M$ to slightly larger open set $U_{\mathbf{\Lambda}^M} \subset \mathbb{R}^M$ containing $\mathbf{\Lambda}^M$. We denote the Jacobian of Q with respect to λ^M as J_{Q,λ^M}. That is $J_{Q,\lambda^M}(\lambda^M,\hat{\lambda}^M) = J_{Q_{\hat{\lambda}^M}}(\lambda^M)$.

Let Q be a map from $U_{\mathbf{\Lambda}^M} \times \hat{\mathbf{\Lambda}}^M$ to $\mathcal{D}$, where $U_{\mathbf{\Lambda}^M}$ is an open set in $\mathbb{R}^M$ containing $\mathbf{\Lambda}^M$, $\mathcal{D}$ is a subset of $\mathbb{R}^m$ and $M \geq m$. Assume J_{Q,λ^M} exist and is continuous on $U_{\mathbf{\Lambda}^M} \times \hat{\mathbf{\Lambda}}^M$. Further assume J_{Q,λ^M} is full rank on $\mathbf{\Lambda}$. We define $\tilde{\mu}_{\mathcal{D},\hat{\lambda}^M} = Q_{\hat{\lambda}^M}\mu_{\mathbf{\Lambda}^M}$, then $d\tilde{\mu}_{\mathcal{D},\hat{\lambda}^M} = \tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)d\mu_{\mathcal{D}}$ for every $\hat{\lambda}^M \in \hat{\mathbf{\Lambda}}^M$ by Theorem 4.5.1. We then have the following result concerning continuity.

Theorem 4.5.2. *Let $B : U_{\mathbf{\Lambda}^M} \to \mathbb{R}^m$ be a piecewise C^1 map, such that $\partial\mathbf{\Lambda}^M$ is locally determined by $B(\lambda^M) = 0$, $\lambda^M \in U_{\mathbf{\Lambda}^M}$. If $\begin{pmatrix} J_{Q^M_{\hat{\lambda}^M}} \\ J_B \end{pmatrix}$ is full rank on $\partial\mathbf{\Lambda}^M \cap Q^{-1}_{\hat{\lambda}^M}(y)$ or $\partial\mathbf{\Lambda}_1 \cap Q^{-1}_{\hat{\lambda}^M}(y) = \varnothing$ for $\mu_{\hat{\mathbf{\Lambda}}^M\times\mathcal{D}}$-almost all $(\hat{\lambda}^M, y)$, $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)$ is continuous $\mu_{\hat{\mathbf{\Lambda}}^M\times\mathcal{D}}$-a.s. on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$.*

Proof. This is a direct result of Theorem 4.3.2 with $\mathbf{\Lambda}_2$ replaced by $\hat{\mathbf{\Lambda}}^M$. □

Corollary 4.5.3. *Given all the conditions in Theorem 4.5.2, $\tilde{\rho}_{\mathcal{D}}$ exists and has the expression as $\tilde{\rho}_{\mathcal{D}}(y) = \int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)d\mu_{\hat{\mathbf{\Lambda}}^M}$.*

Proof. This is implied by Proposition 4.2.2. □

Corollary 4.5.4. *Given $P_{\mathcal{D}}$ on $\mathcal{D}$, which has a density $\rho_{\mathcal{D}}$ with respect to $\mu_{\mathcal{D}}$, the solution density of the SIP with respect to $\mu_{\mathbf{\Lambda}}$ under the Uniform Ansatz with respect to $\mu_{\mathbf{\Lambda}}$, is*

$$\rho_{\mathbf{\Lambda}}(\lambda) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}.$$

Proof. This is a direct result of Theorem 2.2.2. □

The continuity of $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)$ on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$ has more implications than measurability and existence of $\tilde{\rho}_{\mathcal{D}}$; for example, it also leads to the continuity of $\tilde{\rho}_{\mathcal{D}}$.

Theorem 4.5.5. *Under the conditions in Theorem 4.5.1, if $\rho_{\mathcal{D}}$ defined in Corollary 4.5.4 is continuous a.e., the solution density $\rho_{\mathbf{\Lambda}}(\lambda)$ is continuous a.e. on $\mathbf{\Lambda}$.*

Proof. Since $\tilde{\rho}_{\mathcal{D}}(y) = \int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M) d\mu_{\hat{\mathbf{\Lambda}}^M}$, when $|y_1 - y_2| < \delta$,

$$\left|\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_1, \hat{\lambda}^M) - \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_2, \hat{\lambda}^M)\right| < \epsilon,$$

then

$$\left|\int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_1, \hat{\lambda}^M) d\mu_{\hat{\mathbf{\Lambda}}^M} - \int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_2, \hat{\lambda}^M) d\mu_{\hat{\mathbf{\Lambda}}^M}\right| < \epsilon.$$

Thus $|\tilde{\rho}_{\mathcal{D}}(y_1) - \tilde{\rho}_{\mathcal{D}}(y_2)| < \epsilon$, i.e. $\tilde{\rho}_{\mathcal{D}}$ is continuous a.e.. In addition, when both $\rho_{\mathcal{D}}$ and Q are continuous a.e., $\rho_{\mathbf{\Lambda}}$ is continuous a.e.. □

Since $\mathbf{\Lambda} = \mathbf{\Lambda}^N \times \hat{\mathbf{\Lambda}}^N$ for any N, for $\lambda = (\lambda^N, \hat{\lambda}^N) \in \mathbf{\Lambda}$, we write

$$\rho_{\mathbf{\Lambda}}(\lambda) = \rho_{\mathbf{\Lambda}}((\lambda^N, \hat{\lambda}^N)) \equiv \rho_{\mathbf{\Lambda}}(\lambda^N, \hat{\lambda}^N),$$

to get rid of repetition of parenthesis without causing confusion.

Note that the notation $\hat{\lambda}^N$ (or λ^N) has two layers of meaning: a point in $\hat{\mathbf{\Lambda}}^N$, and there is one $\lambda^N \in \mathbf{\Lambda}^N$, such that $\lambda = (\lambda^N, \hat{\lambda}^N) \in \mathbf{\Lambda}$. To distinguish variables from spaces $\mathbf{\Lambda}^N$ and the space $\hat{\mathbf{\Lambda}}^N$, we write $\lambda_0^N \in \mathbf{\Lambda}^N$ and $\hat{\lambda}^N \in \hat{\mathbf{\Lambda}}^N$, so $(\lambda_0^N, \hat{\lambda}^N) \in \mathbf{\Lambda}$. The marginal density of $\hat{\lambda}^N \in \hat{\mathbf{\Lambda}}^N$ is

$$\begin{aligned}
\rho_{\hat{\mathbf{\Lambda}}^N}(\hat{\lambda}^N) &= \int_{\mathbf{\Lambda}^N} \rho_{\mathbf{\Lambda}}(\lambda_0^N, \hat{\lambda}^N) d\mu_{\mathbf{\Lambda}^N}(\lambda_0^N) \\
&= \int_{\hat{\mathbf{\Lambda}}^N} \int_{\mathbf{\Lambda}^N} \rho_{\mathbf{\Lambda}}(\lambda_0^N, \hat{\lambda}^N) d\mu_{\mathbf{\Lambda}^N}(\lambda_0^N) d\mu_{\hat{\mathbf{\Lambda}}^N}(\hat{\lambda}_0^N) \\
&= \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}}(\lambda_0^N, \hat{\lambda}^N) d\mu_{\mathbf{\Lambda}}(\lambda_0).
\end{aligned}$$

When we evaluate $\rho_{\hat{\mathbf{\Lambda}}^N}(\hat{\lambda}^N)$, we treat $\hat{\lambda}^N$ as a point in $\hat{\mathbf{\Lambda}}^N$ without considering its complementary part λ^N.

In next theorem, we prove that if $\rho_{\mathbf{\Lambda}}$ is uniformly continuous on $\mathbf{\Lambda}$, $\rho_{\hat{\mathbf{\Lambda}}^N}$ converges to uniform distribution as $N \to \infty$.

Theorem 4.5.6. *If $\rho_{\mathbf{\Lambda}}$ is uniformly continuous on $\mathbf{\Lambda}$, given $\epsilon > 0$, there exists N, such that for all $n > N$, $|\rho_{\hat{\mathbf{\Lambda}}^n}(\hat{\lambda}^n) - 1| < \epsilon$ for any $\lambda \in \mathbf{\Lambda}$.*

Proof. Since $\rho_{\mathbf{\Lambda}}$ is uniformly continuous, there exists N, such that for all $n > N$ for any $\lambda_0, \lambda \in \mathbf{\Lambda}$, $d((\lambda_0^n, \hat{\lambda}^n), \lambda_0) < \delta$, then $|\rho_{\mathbf{\Lambda}}(\lambda_0^n, \hat{\lambda}^n) - \rho_{\mathbf{\Lambda}}(\lambda_0)| < \epsilon$. So

$$\left| \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}}(\lambda_0^n, \hat{\lambda}^n) d\mu_{\mathbf{\Lambda}}(\lambda_0) - \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}}(\lambda_0) d\mu_{\mathbf{\Lambda}}(\lambda_0) \right| < \epsilon,$$

i.e. $|\rho_{\hat{\mathbf{\Lambda}}^n}(\hat{\lambda}^n) - 1| < \epsilon$. □

4.5.2 Convergence of the solution

To our knowledge, there is no sampling method in an infinite dimensional space taking the cylindrical sets, the generating sets of the Borel σ-algebra, into account. At the same time, both the theoretical and computational aspects of the SIP with respect to $Q^N_{\hat{\lambda}_0^N} : \mathbf{\Lambda}^N \to \mathcal{D}$ have been thoroughly investigated. In this part, we study whether the SIP with respect to $Q : \mathbf{\Lambda} \to \mathcal{D}$ can be approximated by the solution of SIP with respect to $Q^N_{\hat{\lambda}_0^N} : \mathbf{\Lambda}^N \to \mathcal{D}$. This is true under conditions in Theorem 4.5.1 and Theorem 4.5.2, and we refer to such an approach as approximation by "the SIP with truncated domain", and N the truncation number.

As is implied by Corollary 4.5.4, the solution density with respect to $Q : \mathbf{\Lambda} \to \mathcal{D}$ under the Uniform Ansatz exists. Furthermore, when $N > M$ for M in Theorem 4.5.1, the solution density with respect to $Q^N_{\hat{\lambda}_0^N} : \mathbf{\Lambda}^N \to \mathcal{D}$ under the Uniform Ansatz exists as well.

To see this, we first look at the decomposition $\mathbf{\Lambda} = \mathbf{\Lambda}^M \times \hat{\mathbf{\Lambda}}^M$. For $N > M$, $\hat{\mathbf{\Lambda}}^M$ can be further decomposed as $\hat{\mathbf{\Lambda}}^M = \mathbf{\Lambda}^{N-M} \times \hat{\mathbf{\Lambda}}^N$. Define $\tilde{\mu}_{\mathcal{D},\hat{\lambda}_0^N} = Q_{\hat{\lambda}_0^N} \mu_{\mathbf{\Lambda}^N}$ then $d\tilde{\mu}_{\mathcal{D},\hat{\lambda}_0^N} = \tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}_0^N}(y, \hat{\lambda}_0^N) d\mu_{\mathcal{D}}$ with

$$\tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(y, \hat{\lambda}_0^N) = \int_{\mathbf{\Lambda}^{N-M}} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, (\lambda^{N-M}, \hat{\lambda}_0^N)) d\mu_{\mathbf{\Lambda}^{N-M}}(\lambda^{N-M}),$$

for $\lambda^{N-M} \in \mathbf{\Lambda}^{N-M}$.

Then the solution density on $\mathbf{\Lambda}^N$ under the Uniform Ansatz is given by Theorem 2.2.2 as

$$\rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}(\lambda^N) = \frac{\rho_{\mathcal{D}}(Q(\lambda^N, \hat{\lambda}_0^N))}{\tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(Q(\lambda^N, \hat{\lambda}_0^N), \hat{\lambda}_0^N)}.$$

Given $N_0 \leq N$, $\mathbf{\Lambda}^N$ can be further decomposed as $\mathbf{\Lambda}^N = \mathbf{\Lambda}^{N_0} \times \mathbf{\Lambda}^{N-N_0}$. We expect $\rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}(\lambda^{N_0}, \lambda_1^{N-N_0})$ to be close to $\rho_{\mathbf{\Lambda}}(\lambda)$ if N_0 is big enough as long as $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M)$ is continuous on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$.

Theorem 4.5.7. *If $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M)$ is continuous on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$, given λ, for all $\lambda_0 \in \hat{\mathbf{\Lambda}}^M$ and $N \geq N_0$,*

$$\begin{aligned} \rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}(\lambda^{N_0}, \lambda_1^{N-N_0}) &= \frac{\rho_{\mathcal{D}}(Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N))}{\tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N), \hat{\lambda}_0^N)} \\ &\to \rho_{\mathbf{\Lambda}}(\lambda) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))}, \end{aligned}$$

as $N_0 \to \infty$.

Proof. Since Q and $\rho_{\mathcal{D}}$ are both continuous, $\rho_{\mathcal{D}}(Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N)) \to \rho_{\mathcal{D}}(Q(\lambda))$ as $N \to \infty$. We only need to prove $\tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N), \hat{\lambda}_0^N) \to \tilde{\rho}_{\mathcal{D}}(Q(\lambda))$.

Since $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M)$ is continuous on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$, it is uniformly continuous. Therefore for all ϵ, there exists δ_ϵ, such that when $d((y_1, \hat{\lambda}_1^M), (y_2, \hat{\lambda}_2^M)) < \delta_\epsilon$,

$$\left| \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_1, \hat{\lambda}_1^M) - \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y_2, \hat{\lambda}_2^M) \right| < \epsilon.$$

Since Q is uniformly continuous on $\mathbf{\Lambda}$, there exists N_1^*, such that for all $N_0 > N_1^*$,

$$d(Q(\lambda^{N_0}, \hat{\lambda}^{N_0}), Q(\lambda^{N_0}, \hat{\lambda}_0^{N_0})) < \delta.$$

Then there exists N_2^*, such that for all $N_0 > N_2^*$,

$$d((Q((\lambda^{N_0}, \lambda^{N-N_0}), \hat{\lambda}^N), \hat{\lambda}'^{N-M}, \hat{\lambda}'^N), (Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N), \hat{\lambda}'^{N-M}, \hat{\lambda}_0^N)) < \delta_\epsilon,$$

for all $\lambda', \lambda_0 \in \hat{\mathbf{\Lambda}}^M$, so

$$\left| \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(Q((\lambda^{N_0}, \lambda^{N-N_0}), \hat{\lambda}^N), (\hat{\lambda}'^{N-M}, \hat{\lambda}'^N)) - \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(Q((\lambda^{N_0}, \lambda_1^{N-N_0}), \hat{\lambda}_0^N), (\hat{\lambda}'^{N-M}, \hat{\lambda}_0^N)) \right| < \epsilon.$$

Consequently

$$\Bigg| \int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(Q((\lambda^{N_0},\lambda^{N-N_0}),\hat{\lambda}^N),(\hat{\lambda}'^{N-M},\hat{\lambda}'^N))d\mu_{\hat{\mathbf{\Lambda}}^M}(\hat{\lambda}'^M) \\ - \int_{\hat{\mathbf{\Lambda}}^M} \tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(Q((\lambda^{N_0},\lambda_1^{N-N_0}),\hat{\lambda}_0^N),(\hat{\lambda}'^{N-M},\hat{\lambda}_0^N))d\mu_{\hat{\mathbf{\Lambda}}^M}(\hat{\lambda}'^M)\Bigg| < \epsilon,$$

thus

$$\left|\tilde{\rho}_{\mathcal{D}}(Q(\lambda)) - \tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(Q((\lambda^{N_0},\lambda_1^{N-N_0}),\hat{\lambda}_0^N),\hat{\lambda}_0^N)\right| < \epsilon.$$

□

The next result shows the density of SIP with truncated domain is close to the density of SIP with original domain at the same point.

Theorem 4.5.8. *If $\tilde{\rho}^M_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y,\hat{\lambda}^M)$ is continuous on $\mathcal{D}\times\hat{\mathbf{\Lambda}}^M$, then for all $\hat{\lambda}_0 \in \hat{\mathbf{\Lambda}}^M$,*

$$\rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}(\lambda^N) = \frac{\rho_{\mathcal{D}}(Q(\lambda^N,\hat{\lambda}_0^N))}{\tilde{\rho}^N_{\mathcal{D},\hat{\mathbf{\Lambda}}^N}(Q(\lambda^N,\hat{\lambda}_0^N),\hat{\lambda}_0^N)} \to \rho_{\mathbf{\Lambda}}(\lambda) = \frac{\rho_{\mathcal{D}}(Q(\lambda))}{\tilde{\rho}_{\mathcal{D}}(Q(\lambda))},$$

as $N\to\infty$.

Proof. Set $N_0 = N$, and it follows from Theorem 4.5.7. □

In Theorem 4.5.7, we show that given λ, $\rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}((\lambda^{N_0},\lambda_1^{N-N_0})) \to \rho_{\mathbf{\Lambda}}(\lambda)$, as $N_0\to\infty$. If this convergence holds uniformly for all $\lambda\in\mathbf{\Lambda}$, we can obtain similar results to Theorem 4.5.6 for the SIP with dimension truncated domain as follows. Define the marginal distribution of $\lambda_1^{N-N_0} \in \mathbf{\Lambda}^{N-N_0}$ as

$$\rho^N_{\mathbf{\Lambda}^{N-N_0},\hat{\lambda}_0^N}(\lambda_1^{N-N_0}) = \int_{\mathbf{\Lambda}^{N_0}} \rho^N_{\mathbf{\Lambda}^N,\hat{\lambda}_0^N}(\lambda^{N_0},\lambda_1^{N-N_0})d\mu_{\mathbf{\Lambda}^{N_0}}(\lambda^{N_0}),$$

we have the following result.

Theorem 4.5.9. *If given $\epsilon > 0$, there exists N_0 such that for any $N \geq N_0$,*

$$\left|\rho^{N}_{\mathbf{\Lambda}^N,\hat{\lambda}^N_0}(\lambda^{N_0},\lambda^{N-N_0}_1) - \rho_{\mathbf{\Lambda}}(\lambda)\right| < \epsilon,$$

for all $\lambda \in \mathbf{\Lambda}$, then $\rho^{N}_{\mathbf{\Lambda}^{N-N_0},\hat{\lambda}^N_0}(\lambda^{N-N_0}_1) \to 1$ as $N_0 \to \infty$.

Proof. If $|\rho^{N}_{\mathbf{\Lambda}^N,\hat{\lambda}^N_0}(\lambda^{N_0},\lambda^{N-N_0}_1) - \rho_{\mathbf{\Lambda}}(\lambda)| < \epsilon$, for all $\lambda \in \mathbf{\Lambda}^N$,

$$\begin{aligned}\left|\rho^{N}_{\mathbf{\Lambda}^{N-N_0},\hat{\lambda}^N_0}(\lambda^{N-N_0}_1) - 1\right| &= \left|\int_{\mathbf{\Lambda}} \rho^{N}_{\mathbf{\Lambda}^N,\hat{\lambda}^N_0}((\lambda^{N_0},\lambda^{N-N_0}_1))d\mu_{\mathbf{\Lambda}}(\lambda) - \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}}(\lambda)d\mu_{\mathbf{\Lambda}}(\lambda)\right| \\ &\leq \int_{\mathbf{\Lambda}} \left|\rho^{N}_{\mathbf{\Lambda}^N,\hat{\lambda}^N_0}((\lambda^{N_0},\lambda^{N-N_0}_1)) - \rho_{\mathbf{\Lambda}}(\lambda)\right| d\mu_{\mathbf{\Lambda}} < \epsilon.\end{aligned}$$

□

This means when the truncation number is high enough, the marginal distribution $\rho^{N}_{\mathbf{\Lambda}^{N-N_0},\hat{\lambda}^N_0}$ will converge to uniform distribution just like $\rho_{\hat{\mathbf{\Lambda}}^n}$.

Chapter 5

Numerical analysis for the solution of the Infinite Dimensional SIP

In this chapter, we use results from the last chapter as the basic setup. Specifically, we assume there is a continuous map $Q : \mathbf{\Lambda} \to \mathcal{D}$, where $\mathbf{\Lambda} = ([0,1]^\infty, d)$ and $\mathcal{D} \subset \mathbb{R}^m$. We define $\tilde{\mu}_\mathcal{D} = Q\mu_\mathbf{\Lambda}$, and assume $d\tilde{\mu}_\mathcal{D} = \tilde{\rho}_\mathcal{D} d\mu_\mathcal{D}$. We also assume that the probability measure $P_\mathcal{D}$ has a density $\rho_\mathcal{D}$ with respect to $\mu_\mathcal{D}$. Theorem 2.2.2 yields the solution density $\rho_\mathbf{\Lambda} = \dfrac{\rho_\mathcal{D}(Q(\lambda))}{\tilde{\rho}_\mathcal{D}(Q(\lambda))}$ on $\mathbf{\Lambda}$.

One way to solve this SIP is to solve the SIP with the truncated dimension, i.e. applying Algorithm 2 with respect to $Q_{\hat{\lambda}_0^k} : \mathbf{\Lambda}^L \to \mathcal{D}$. Letting the truncation number k vary yields an new algorithm as Algorithm 3 with respect to the original $Q : \mathbf{\Lambda} \to \mathcal{D}$.

Algorithm 3: Approximation to inverse probability.

Generate approximating sets $\{I_i\}_{i=1}^{M}$ of $\mathcal{D}$

Choose k then use $[0,1]^k$ or $\mathbf{\Lambda}^k$, as substitute domain, and $Q_{\hat{\lambda}_0^k}$ as substitue map for any $\hat{\lambda}_0^k$

Generate approximating sets $\{b_j\}_{j=1}^{N}$ of $[0,1]^k$, associate every b_j with a point λ_j^k inside

Fix and normalize the simple function approximation $\rho_{\mathcal{D},M} = \sum_{i=1}^{M} p_i \mathbb{1}_{I_i}(q)$

Initiate an $M \times N$ matrix V

foreach $j = 1, \cdots, N$ **do**

    foreach $i = 1, \cdots, M$ **do**

        if $Q_{\hat{\lambda}_0^k}(\lambda_j^k) \in I_i$ **then**

            $V_{ij} = \mu_{\mathbf{\Lambda}^k}(b_j)$

        else

            $V_{ij} = 0$

        end

    end

end

foreach $j = 1, \cdots, N$ **do**

    Set $P(b_j)$ to $\sum_{i=1}^{M} p_i (V_{ij} / \sum_{j=1}^{N} V_{ij})$

end

This approach poses questions in terms of approximating sets and density functions on $\mathbf{\Lambda}$. For example, how can $A \in \mathcal{B}_{\mathbf{\Lambda}}$ be approximated by a cylindrical set with face in $\mathcal{B}_{\mathbf{\Lambda}^k}$.

We address issues concerning the approximation of sets and density functions that only arise in the settings of $\mathbf{\Lambda} = ([0,1]^\infty, d)$ in Sections 1 and 2. In Section 3, we prove a theorem showing probabilities computed by Algorithm 3 converges to the true probability.

5.1 Approximation of sets

As a starting point, we restate Theorem D of Page 56, in Section 13 of Halmos (1974) as

Theorem 5.1.1. *For a measure space $(\mathbf{\Omega}, \mathcal{B}_{\mathbf{\Omega}}, \mu_{\mathbf{\Omega}})$, where $\mu_{\mathbf{\Omega}}$ is a σ-finite measure, suppose $\mathcal{B}_{\mathbf{\Omega}}$ is generated by $\mathcal{C}$, that is $\sigma(\mathcal{C}) = \mathcal{B}_{\mathbf{\Omega}}$. Given an event $A \in \sigma(\mathcal{C})$ such that $\mu_{\mathbf{\Omega}}(A) < \infty$, for $\epsilon > 0$, we can find a set $C \in \mathcal{C}$, such that $\mu_{\mathbf{\Omega}}(A \triangle C) < \epsilon$.*

Theorem 5.1.1 implies that a set A in $\mathcal{B}_{\mathbf{\Lambda}}$ can be approximated by cylindrical set in the sense that given a sequence of positive numbers $\{\epsilon_\ell\}$ where $\epsilon_\ell \to 0$, we can find a sequence of cylindrical sets $\{A^\ell\}$ such that $\mu_{\mathbf{\Lambda}}(A \triangle A^\ell) < \epsilon_j$. Assuming the minimal face of A^ℓ is A'^ℓ, we want to approximate $P_{\mathbf{\Lambda}}(A)$ by approximating probabilities of A'^ℓ.

However, in Section 3.3, we only address the approximation of measurable set with negligible boundary in $\mathbf{\Lambda}^N$, so we only focus on a set which can be approximated by a sequence of cylindrical sets with negligible boundaries. This does not mean that the set A itself has negligible boundary. For example,

$$A = \prod_{k=1}^{\infty} \left[0, \frac{1 - 1/(k+1)}{1 - 1/(k+2)}\right],$$

is measurable and has a measure of $1/2$, and can be approximated by

$$\{A^\ell = \prod_{k=1}^{\ell} \left[0, \frac{1 - 1/(k+1)}{1 - 1/(k+2)}\right] \times [0,1]^\infty\},$$

and

$$\mu_{\mathbf{\Lambda}^\ell}\left(\partial\left(\prod_{k=1}^{\ell} \left[0, \frac{1 - 1/(k+1)}{1 - 1/(k+2)}\right]\right)\right) = 0,$$

so this set is within the framework of the proposed algorithm. However, boundary of A is not measure 0. One can show that $\bar{A} = A$, $\mathring{A} = \phi$, so $\partial A = \bar{A} \setminus \mathring{A} = A$.

Next we examine a special case of Borel σ-algebra being approximated by cylindrical sets. As a preparation, we need the following Lemma.

Lemma 5.1.2. *For a given ϵ, there exists δ, such that if $E \in \mathcal{B}_{\mathcal{D}}$ and $\mu_{\mathcal{D}}(E) < \delta$, $\mu_{\mathbf{\Lambda}}(Q^{-1}(E)) < \epsilon$. Furthermore, $\mu_{\mathbf{\Lambda}}(Q^{-1}(E)) = 0$ if $\mu_{\mathcal{D}}(E) = 0$.*

Proof. This is a direct result of $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$ and $d\tilde{\mu}_{\mathcal{D}} = \tilde{\rho}_{\mathcal{D}} d\mu_{\mathcal{D}}$. $\square$

Next, we show $Q^{-1}(E)$, the preimage of $E \in \mathcal{B}_{\mathcal{D}}$ under Q, can be approximated by cylindrical set with a given form.

Theorem 5.1.3. *For $E \in \mathcal{B}_{\mathcal{D}}$ of which $\mu_{\mathcal{D}}(E) > 0$ and $\mu_{\mathcal{D}}(\partial E) = 0$,*

$$\lim_{\ell \to \infty} \mu_{\mathbf{\Lambda}} \left(({Q^{\ell}_{\hat{\lambda}^{\ell}_0}}^{-1}(E) \times \hat{\mathbf{\Lambda}}^{\ell}) \triangle Q^{-1}(E) \right) = 0$$

for any $\lambda_0 \in \mathbf{\Lambda}$.

Proof. Let $E^{\delta} := (E + \delta B(0,1)) \cap \mathcal{D}$ denote the Minkowski sum of E and $\delta B(0,1)$ restricted to $\mathcal{D}$, where $B(0,1)$ is the unit ball in $\mathbb{R}^m$. Let $E_{\delta} = ((E^c)^{\delta})^c$. Since $\mu_{\mathcal{D}}(\partial E) = 0$, for a given $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that $\mu_{\mathcal{D}}(E^{2\delta(\epsilon)} \backslash E_{2\delta(\epsilon)})$ is sufficiently small so that $\mu_{\mathbf{\Lambda}}(Q^{-1}(E^{2\delta(\epsilon)}) \backslash Q^{-1}(E_{2\delta(\epsilon)})) < \epsilon$.

First we show that there exists a truncation number L^0, such that for all $L \geq L^0$, $E \in \mathcal{B}_D$ and $\lambda_0 \in \mathbf{\Lambda}$,

$$\left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E_{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^L \subset Q^{-1}(E) \subset \left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E^{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^L.$$

Because $Q : \mathbf{\Lambda} \to \mathcal{D}$ is continuous so uniformly continuous, there exists an L^0, such that for all $L \geq L^0$, with $|Q^L_{\hat{\lambda}^L_1}(\lambda^L) - Q^L_{\hat{\lambda}^L_2}(\lambda^L)| < \delta(\epsilon)$ for any $\lambda, \lambda_1, \lambda_2$, so for any $\lambda_0 \in \mathbf{\Lambda}$, $Q\left(\left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E_{\delta(\epsilon)}) \times \{\hat{\lambda}^L\} \right) \subset E$ for any $\hat{\lambda}^L$. Hence

$$\bigcup_{\hat{\lambda}^L \in \hat{\mathbf{\Lambda}}^L} \left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E_{\delta(\epsilon)}) \times \{\hat{\lambda}^L\} = \left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E_{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^L \subset Q^{-1}(E).$$

On the other side, since for all $\hat{\lambda}^L \in \hat{\mathbf{\Lambda}}^L$, $\left(Q^L_{\hat{\lambda}^L} \right)^{-1} (E) \subset \left(Q^L_{\hat{\lambda}^L_0} \right)^{-1} (E^{\delta(\epsilon)})$,

$$Q^{-1}(E) = \bigcup_{\hat{\lambda}^L \in \hat{\mathbf{\Lambda}}^L} \left(Q^L_{\hat{\lambda}^L} \right)^{-1} (E) \times \{\hat{\lambda}^L\} \subset {Q^L_{\hat{\lambda}^L_0}}^{-1} (E^{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^L.$$

Since $E_{2\delta(\epsilon)} = (E_{\delta(\epsilon)})_{\delta(\epsilon)}$ and $E^{2\delta(\epsilon)} = (E^{\delta(\epsilon)})^{\delta(\epsilon)}$, we have $Q^{-1}(E_{2\delta(\epsilon)}) \subset \left(Q^N_{\hat{\lambda}^N_0}\right)^{-1}(E_{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^N \subset Q^{-1}(E) \subset \left(Q^N_{\hat{\lambda}^N_0}\right)^{-1}(E^{\delta(\epsilon)}) \times \hat{\mathbf{\Lambda}}^N \subset Q^{-1}(E^{2\delta(\epsilon)})$.

For a positive sequence $\{\epsilon_i\}$ decreasing to 0, we can find a positive decreasing sequence $\{\delta(\epsilon_i)\}$ converges to 0 as well. Suppose $\{L^0_{\delta(\epsilon_i)}\}$ is the sequence of truncation number constructed as above. We define a new sequence $\{\delta_\ell\}$ such that

$$\delta_\ell = \begin{cases} +\infty & \text{if } \ell < L^0_{\delta_{\epsilon_1}} \\ \max\{\delta(\epsilon_i), L^0_{\delta(\epsilon_i)} \leq \ell\}\} & \text{if } \ell \geq L^0_{\delta_{\epsilon_1}} \end{cases},$$

so that

$$C^1_\ell = \left(Q^\ell_{\hat{\lambda}^\ell_0}\right)^{-1}(E_{\delta_\ell}) \times \hat{\mathbf{\Lambda}}^\ell, C^2_\ell = \left(Q^\ell_{\hat{\lambda}^\ell_0}\right)^{-1}(E^{\delta_\ell}) \times \hat{\mathbf{\Lambda}}^\ell$$

satisfy $C^1_\ell \subset Q^{-1}(E) \subset C^2_\ell$ and $\mu_{\mathbf{\Lambda}}(C^2_\ell \setminus C^1_\ell) \to 0$ as $\ell \to \infty$. The statement of the theorem holds since $C^1_\ell \subset {Q^\ell_{\hat{\lambda}^\ell_0}}^{-1}(E) \times \hat{\mathbf{\Lambda}}^\ell \subset C^2_\ell$. □

5.2 Approximation of densities

In this section, we investigate how for a given $\mathcal{I} = \{I_i\}_{i=1}^M$ on $\mathcal{D}$, the $\rho_{\mathbf{\Lambda},M}$ induced by Q defined as in Theorem 3.3.1,

$$\rho_{\mathbf{\Lambda},M} = \sum_{i=1}^M p_i \mathbb{1}_{Q^{-1}(I_i)}(\lambda), \quad p_i = \frac{P_{\mathcal{D}}(I_i)}{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_i))}, \tag{5.1}$$

can be approximated by a simple function induced by $Q^\ell_{\hat{\lambda}^\ell_0}$. Define a simple function on $\mathbf{\Lambda}^\ell$ as

$$\rho'_{\mathbf{\Lambda}^\ell,M} = \sum_{i=1}^M p'_i \mathbb{1}_{\left(Q^\ell_{\hat{\lambda}^\ell_0}\right)^{-1}(I_i)}(\lambda^\ell), \quad p'_i = \frac{P_{\mathcal{D}}(I_i)}{\mu_{\mathbf{\Lambda}^\ell}\left(\left(Q^\ell_{\hat{\lambda}^\ell_0}\right)^{-1}(I_i)\right)}. \tag{5.2}$$

We omit λ_0 in the notation $\rho'_{\mathbf{\Lambda}^\ell,M}$, since its properties do not depend on the choice of $\hat{\lambda}_0^\ell \in \hat{\mathbf{\Lambda}}$. So when we use $\rho'_{\mathbf{\Lambda}^\ell,M}$, we assume $Q^\ell_{\hat{\lambda}_0^N}$ has been defined according to an arbitrarily chosen $\hat{\lambda}_0^\ell$. At the same time, we define $\rho_{\mathbf{\Lambda}^\ell,M}$ as density on $\mathbf{\Lambda}$ by

$$\rho_{\mathbf{\Lambda}^\ell,M}(\lambda) = \rho'_{\mathbf{\Lambda}^\ell,M}(\lambda^\ell),$$

or we can define it as

$$\rho_{\mathbf{\Lambda}^\ell,M} = \sum_{i=1}^{M} p'_i \mathbb{1}_{{Q^\ell_{\hat{\lambda}_0^\ell}}^{-1}(I_i)\times\hat{\mathbf{\Lambda}}^\ell}(\lambda), \quad p'_i = \frac{P_{\mathcal{D}}(I_i)}{\mu_{\mathbf{\Lambda}}({Q^\ell_{\hat{\lambda}_0^\ell}}^{-1}(I_i)\times\hat{\mathbf{\Lambda}}^\ell)}. \tag{5.3}$$

The next result shows $\rho_{\mathbf{\Lambda}^\ell,M}$ converges to $\rho_{\mathbf{\Lambda},M}$ a.e..

Theorem 5.2.1. *Given a partition $\mathcal{I} = \{I_i\}_{i=1}^M$ of $\mathcal{D}$,*

a.$\rho_{\mathbf{\Lambda}^\ell,M} \to \rho_{\mathbf{\Lambda},M}$ a.e. as $\ell \to \infty$.

b. $\{\rho_{\mathbf{\Lambda}^\ell,M}\}$ is uniformly integrable.

Proof. Since $\mu_{\mathcal{D}}(\bigcup_{i=1}^M \partial I_i) = 0$, $\mu_{\mathbf{\Lambda}}(\bigcup_{i=1}^M Q^{-1}(\partial I_i)) = 0$ by Theorem 5.1.2, we only consider $\lambda \in \mathbf{\Lambda} \setminus \bigcup_{i=1}^M Q^{-1}(\partial I_i)$.

For $\lambda \in Q^{-1}(I_j)$ and $\lambda \notin Q^{-1}(\partial I_j)$, there exists an L such that for all $\ell > L$, $\lambda \in \left(Q^\ell_{\hat{\lambda}_0^\ell}\right)^{-1}(I_j)\times\hat{\mathbf{\Lambda}}^\ell$, so $\rho_{\mathbf{\Lambda}^\ell,M}(\lambda) = p_j^\ell$ when $\ell > L$, where

$$p_j^\ell = \frac{P_{\mathcal{D}}(I_j)}{\mu_{\mathbf{\Lambda}^\ell}\left(\left(Q^\ell_{\hat{\lambda}_0^\ell}\right)^{-1}(I_j)\right)} = \frac{P_{\mathcal{D}}(I_j)}{\mu_{\mathbf{\Lambda}}\left(\left(Q^\ell_{\hat{\lambda}_0^\ell}\right)^{-1}(I_j)\times\hat{\mathbf{\Lambda}}^n\right)}.$$

Since $\mu_{\mathbf{\Lambda}}\left(\left(\left(Q^\ell_{\hat{\lambda}_0^\ell}\right)^{-1}(I_j)\times\hat{\mathbf{\Lambda}}^\ell\right)\triangle Q^{-1}(I_j)\right) \to 0$, when $\ell \to \infty$,

$$p'_j \to \frac{P_{\mathcal{D}}(I_j)}{\mu_{\mathbf{\Lambda}}(Q^{-1}(I_j))} = p_j.$$

which means $\rho_{\mathbf{\Lambda}^\ell,M}(\lambda) \to \rho_{\mathbf{\Lambda},M}$. Therefore $\rho_{\mathbf{\Lambda}^\ell,M} \to \rho_{\mathbf{\Lambda},M}$ a.e. as $\ell \to \infty$.

For a fixed M, when ℓ is sufficiently large, we have $|\rho_{\mathbf{\Lambda}^\ell,M}(\lambda) - \rho_{\mathbf{\Lambda},M}(\lambda)| < \epsilon/\delta$ for $\lambda \in \mathbf{\Lambda}' \subset \mathbf{\Lambda}$ where $\mu_{\mathbf{\Lambda}}(\mathbf{\Lambda} \setminus \mathbf{\Lambda}') < \epsilon/(\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta)$, and $|\max \rho_{\mathbf{\Lambda}^\ell,M} - \max \rho_{\mathbf{\Lambda},M}| < \epsilon/\delta$. So,

$$
\begin{aligned}
\int_E \rho_{\mathbf{\Lambda}^\ell,M} d\mu_{\mathbf{\Lambda}} &= \int_{E\cap\mathbf{\Lambda}'} \rho_{\mathbf{\Lambda}^\ell,M} d\mu_{\mathbf{\Lambda}} + \int_{E\setminus\mathbf{\Lambda}'} \rho_{\mathbf{\Lambda}^\ell,M} d\mu_{\mathbf{\Lambda}} \\
&\leq \int_{E\cap\mathbf{\Lambda}'} (\rho_{\mathbf{\Lambda},M} + \epsilon/\delta) d\mu_{\mathbf{\Lambda}} + \frac{\epsilon}{\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta} (\max \rho_{\mathbf{\Lambda},M} + \epsilon/\delta) \\
&\leq \int_E \rho_{\mathbf{\Lambda},M} + \epsilon + \epsilon < 3\epsilon.
\end{aligned}
$$

for all $E \in \mathcal{B}_{\mathbf{\Lambda}}$ satisfying $\mu_{\mathbf{\Lambda}}(E) \leq \delta$. Thus, the sequence $\{\rho_{\mathbf{\Lambda}^\ell,M}\}$ is uniformly integrable.

5.3 Approximation of probabilities

To discuss the convergence properties of Algorithm 3, we denote the computed $P(b_j)$ as $P_{\mathbf{\Lambda}^k,M,N}(b_j)$, and reserve $P_{\mathbf{\Lambda}}$ to denote the true probability measure. For $A \in \mathcal{B}_{\mathbf{\Lambda}}$ with $P_{\mathbf{\Lambda}}(\partial A) = 0$, define A_N as

$$
A_N = \bigcup_{\lambda_j \in A, 1 \leq j \leq N} b_j,
$$

then use

$$
P_{\mathbf{\Lambda},M,N}(A_N) = \sum_{\lambda_j \in A, 1 \leq j \leq N} P_{\mathbf{\Lambda},M,N}(b_j)
$$

as an approximation of $P_{\mathbf{\Lambda}}(A)$. The next theorem shows that when $\{I_i\}_{i=1}^M$ and $\{b_j\}_{j=1}^N$ become fine, $P_{\mathbf{\Lambda},M,N}(A_N)$ converges to $P_{\mathbf{\Lambda}}(A)$. $\square$

Theorem 5.3.1. *Given an event $A \in \mathcal{B}_{\mathbf{\Lambda}}$, which can be approximated by a sequence of cylindrical sets with negligible boundaries, i.e. there exists $\{A^\ell\}$, $A^\ell = A'^\ell \times \hat{\mathbf{\Lambda}}^{k_\ell}$, $A'^\ell \in \mathcal{B}_{\mathbf{\Lambda}^{k_\ell}}$, $\mu_{\mathbf{\Lambda}^{k_\ell}}(\partial A'^\ell) = 0$ such that $\lim_{\ell\to\infty} \mu_{\mathbf{\Lambda}}\left(A \triangle A^\ell\right) = 0$. There exists a sequence of approximations $P_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}(A'^\ell_{N_n})$ using simple function approximations and requiring only calculations of volumes in $\mathbf{\Lambda}^{k_\ell}$ based on partition sequence $\{\mathcal{I}_m\}_{m=1}^\infty = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^\infty$ on $\mathcal{D}$ and $\{\mathcal{T}_{\ell,n}\}_{n=1}^\infty = \{\{b_{k_\ell,j}\}_{j=1}^{N_n}\}_{n=1}^\infty$ on $\mathbf{\Lambda}^{k_\ell}$ that satisfies*

$$
\lim_{m\to\infty} \lim_{\ell\to\infty} \lim_{n\to\infty} P_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}(A'^\ell_{N_n}) = \int_A \rho_{\mathbf{\Lambda}} d\mu_{\mathbf{\Lambda}},
$$

if $\lim_{m\to\infty}\max_{1\le i\le M_m} diam(I_i) = 0$ *and* $\lim_{n\to\infty}\max_{1\le j\le N_n} diam(b_{k_\ell,j}) = 0$.

Proof. The proof of this theorem is closely related to proof of Theorem 3.3.1. First of all, given a sequence of partitions $\{\mathcal{I}_m\}_{m=1}^\infty = \{\{I_i\}_{i=1}^{M_m}\}_{m=1}^\infty$ on $\mathcal{D}$, we define $\rho_{\mathbf{\Lambda},M_m}$ according to Theorem 3.3.1, $\rho'_{\mathbf{\Lambda}^{k_\ell},M_m}$ as (5.1), $\rho_{\mathbf{\Lambda}^{k_\ell},M_m}$ as (5.2). For a sequence of partitions $\{\mathcal{T}_{\ell,n}\}_{n=1}^\infty = \{\{b_{k_\ell,j}\}_{j=1}^{N_n}\}_{n=1}^\infty$ on $\mathbf{\Lambda}^{k_\ell}$, we associate each $b_{k_\ell,j}$ with a point $\lambda_j^{k_\ell} \in b_{k_\ell,j}$, and construct

$$\rho'_{\mathbf{\Lambda}^{k_\ell},M,N_n}(\lambda^{k_\ell}) = \sum_{j=1}^{N_n} p_j \mathbb{1}_{b_{k_\ell,j}}(\lambda^{k_\ell}), \quad p_j = \frac{P_{\mathcal{D}}(I_i)}{\sum_{\lambda_o^{k_\ell}\in\left(Q^{k_\ell}_{\hat{\lambda}_0^{k_\ell}}\right)^{-1}(I_i)} \mu_{\mathbf{\Lambda}^{k_\ell}}(b_{k_\ell,o})} \text{ if } \lambda_j^{k_\ell} \in {Q^{k_\ell}_{\hat{\lambda}_0^{k_\ell}}}^{-1}(I_i), \tag{5.4}$$

and

$$\begin{aligned} P_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}(A'^\ell_{N_n}) &= \sum_{\lambda_j^{k_\ell}\in A'^\ell} P_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}(b_{k_\ell,j}) \\ &= \int_{A'^\ell_{N_n}} \rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} d\mu_{\mathbf{\Lambda}^{k_\ell}} \end{aligned}$$

As with the proof for the finite dimensional case, the proof consists of 3 steps. First we show a.e. convergence; second we show uniform integrability; finally we combine a.e. convergence and uniform integrability to show the convergence of the measures.

As in the proof of Theorem 3.3.1, $\rho_{\mathbf{\Lambda},M_m} \to \rho_{\mathbf{\Lambda}}$ a.e. when $m\to\infty$, and $\rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} \to \rho'_{\mathbf{\Lambda}^{k_\ell},M_m}$ a.e. when $n\to\infty$. By Theorem 5.2.1, $\rho_{\mathbf{\Lambda}^{k_\ell},M_m} \to \rho_{\mathbf{\Lambda},M_m}$ a.e. as $\ell\to\infty$.

For uniform integrability, in the proof of Theorem 3.3.1, $\{\rho_{\mathbf{\Lambda},M_m}\}$ is shown to be uniformly integrable, and $\{\rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}\}$ for fixed m and ℓ are shown to be uniformly integrable as well. Lemma 5.2.1 also shows $\{\rho_{\mathbf{\Lambda}^{k_\ell},M_m}\}$ for fixed m are uniformly integrable.

Convergence of measure can be proved in 3 steps as below: First, the Vitali convergence theorem implies

$$\int_{\mathbf{\Lambda}^{k_\ell}} |\rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}\mathbb{1}_{A'^\ell} - \rho'_{\mathbf{\Lambda}^{k_\ell},M_m}\mathbb{1}_{A'^\ell}| d\mu_{\mathbf{\Lambda}^{k_\ell}} \to 0 \text{ as } n\to\infty;$$

for fixed m and ℓ. By the uniform integrability of $\rho_{\mathbf{\Lambda}^{k_\ell},M_m,N_n}$,

$$\int_{\mathbf{\Lambda}^{k_\ell}} |\rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} \mathbb{1}_{A'^\ell_{N_n}} - \rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} \mathbb{1}_{A'^\ell}| d\mu_{\mathbf{\Lambda}} = \int_{A'^j_{N_n} \triangle A'^j} \rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} d\mu_{\mathbf{\Lambda}^{k_\ell}} \to 0 \text{ as } n \to \infty.$$

So we have

$$\int_{\mathbf{\Lambda}^{k_\ell}} |\rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} \mathbb{1}_{A'^\ell_{N_n}} - \rho'_{\mathbf{\Lambda}^{k_\ell},M_m} \mathbb{1}_{A'^\ell}| d\mu_{\mathbf{\Lambda}^{k_\ell}} \to 0 \text{ as } n \to \infty. \tag{5.5}$$

Since $\rho'_{\mathbf{\Lambda}^{k_\ell},M_m}(\lambda^{k_\ell}) = \rho_{\mathbf{\Lambda}^{k_\ell},M_m}(\lambda)$

$$\int_{\mathbf{\Lambda}^{k_\ell}} \rho'_{\mathbf{\Lambda}^{k_\ell},M_m} \mathbb{1}_{A'^\ell} d\mu_{\mathbf{\Lambda}^{k_\ell}} = \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}^{k_\ell},M_m} \mathbb{1}_{A^\ell} d\mu_{\mathbf{\Lambda}}.$$

Second, by the Vitali convergence theorem,

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda}^{k_\ell},M_m} \mathbb{1}_{A^\ell} - \rho_{\mathbf{\Lambda},M_m} \mathbb{1}_{A^\ell}| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } \ell \to \infty,$$

and by properties of the sequence $\{A^\ell\}$

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m} \mathbb{1}_{A^\ell} - \rho_{\mathbf{\Lambda},M_m} \mathbb{1}_{A}| d\mu_{\mathbf{\Lambda}} = \int_{A^\ell \triangle A} \rho_{\mathbf{\Lambda},M_m} d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } \ell \to \infty.$$

Combining these two yields

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda}^{k_\ell},M_m} \mathbb{1}_{A^\ell} - \rho_{\mathbf{\Lambda},M_m} \mathbb{1}_{A}| d\mu_{\mathbf{\Lambda}} \to 0. \tag{5.6}$$

Last, the Vitali convergence theorem implies,

$$\int_{\mathbf{\Lambda}} |\rho_{\mathbf{\Lambda},M_m} \mathbb{1}_{A} - \rho_{\mathbf{\Lambda}} \mathbb{1}_{A}| d\mu_{\mathbf{\Lambda}} \to 0 \text{ as } m \to \infty. \tag{5.7}$$

Combing (5.5), (5.6), and (5.7), we have

$$\lim_{m\to\infty} \lim_{\ell\to\infty} \lim_{n\to\infty} \left| \int_{\mathbf{\Lambda}^{k_\ell}} \rho'_{\mathbf{\Lambda}^{k_\ell},M_m,N_n} \mathbb{1}_{A'^j_{N_n}} d\mu_{\mathbf{\Lambda}^{k_\ell}} - \int_{\mathbf{\Lambda}} \rho_{\mathbf{\Lambda}} \mathbb{1}_{A} d\mu_{\mathbf{\Lambda}} \right| \to 0.$$

$\square$

Chapter 6

Numerical examples

In this chapter, we present an example of the SIP for Q consisting of a set of linear functionals on $L^2([0,1])$.

By using Fourier basis, we establish a isometric isomorphism between $L^2([0,1])$ and the space of Fourier coefficients ℓ^2. Any $\lambda \in L^2([0,1])$ is isomorphic to a function on $[0,1)$, which we extend to a periodic function $\bar{\lambda} : \mathbb{R} \to \mathbb{R}$ by $\bar{\lambda}(t) = \lambda(t), t \in [0,1)$ and $\bar{\lambda}(t+1) = \bar{\lambda}(t)$. For simplicity, we use λ to denote its extension $\bar{\lambda}$ as well.

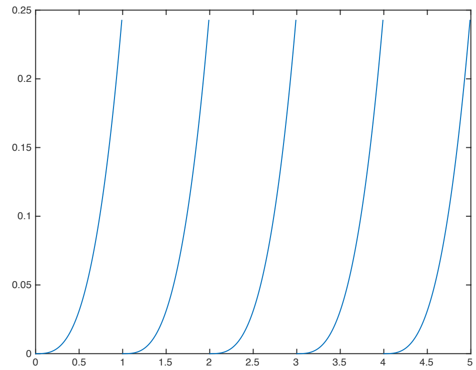
Consider the following functions in $L^2([0,1])$:

$$\begin{aligned}
\lambda_1(t) &= \frac{1}{4}t^3, \\
\lambda_2(t) &= (1-t)t, \\
\lambda_3(t) &= \frac{5}{2}t(1-t)(0.5-t), \\
\lambda_4(t) &= 32t^2(1-t)^2.
\end{aligned}$$

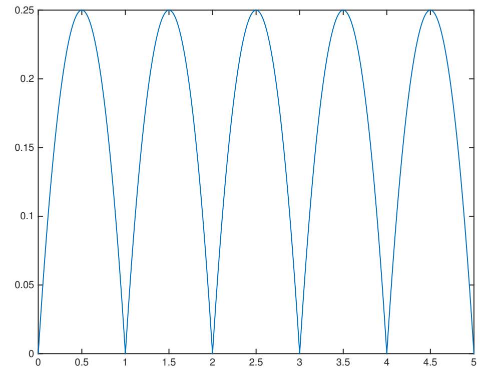
Although all of them seem arbitrarily smooth, when extended to periodic functions on $\mathbb{R}$, they have very different behaviors at points $0, \pm 1, \pm 2, \cdots$ as shown in Figure 6.1. Specifically, after such extension, λ_1 is not continuous, λ_2 is a C^0 function, λ_3 is a C^1 function, λ_4 is a C^2 function. So their Fourier coefficients have a decay rate of $\frac{1}{n}, \frac{1}{n^2}, \frac{1}{n^3}, \frac{1}{n^4}$ respectively.

We consider the SIP to detect a function in $L^2([0,1])$, which can be treated as an SIP for a sequence in ℓ^2. We set the QoI as the function values at 5 points $(0.1, 0.2, 0.4, 0.5, 0.7)$, so the map $Q : \mathbf{\Lambda} \to \mathcal{D}$ is defined to be

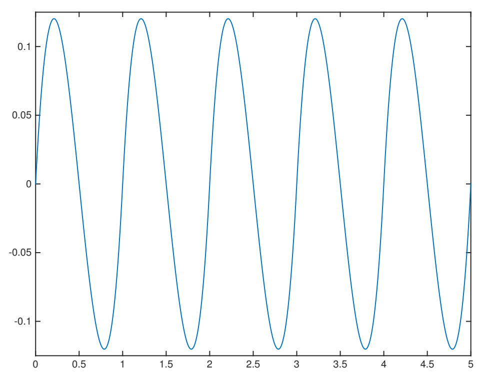
$$Q(\lambda) = (\lambda(0.1), \lambda(0.2), \lambda(0.4), \lambda(0.5), \lambda(0.7)).$$



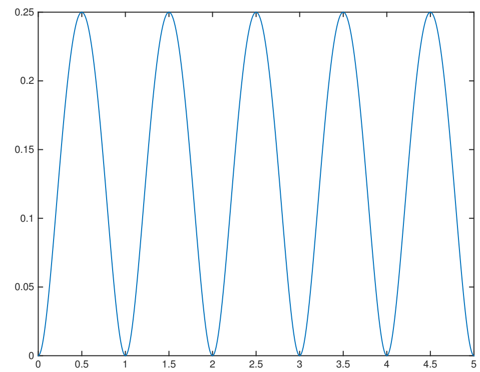
(a) $\lambda_1(t) = \frac{1}{4}t^3$.



(b) $\lambda_2(t) = t(1-t)$.



(c) $\lambda_3(t) = \frac{5}{2}t(1-t)(0.5-t)$.



(d) $\lambda_4(t) = 32t^2(1-t)^2$.

Figure 6.1: Periodic functions whose Fourier coefficients have different decay rates

We take $\lambda_2(t) = t(1-t)$ as the true function, which is equivalent to taking the Fourier coefficients

$$\lambda_2 = (a_0, a_i, b_i) \text{ where } a_0 = \frac{1}{3}, a_i = -\frac{cos(i\pi)}{i^2\pi^2}, b_i = 0, i = 1, 2, 3, \cdots$$

as the true point in ℓ^2. Then

$$Q(\lambda_2) = (0.08547066, 0.16300031, 0.24245401, 0.24758289, 0.20740456)$$

is the true value of $Q(\lambda)$. We let $\mathcal{D}$ be given by a cube centered at $Q(\lambda_2)$ with edge length being 0.2. The probability measure to be inverted is a uniform distribution on $\mathcal{D}$.

Since Q is continuously differentiable, and for any $\mathbf{\Lambda} = ([0,1]^\infty, d)$ and $M \geq 5$, $J_{Q_{\hat{\lambda}^M}}(\lambda^M)$ does not depend on $\hat{\lambda}^M$, we can show $\tilde{\rho}_{\mathcal{D},\hat{\mathbf{\Lambda}}^M}(y, \hat{\lambda}^M)$ is continuous on $\mathcal{D} \times \hat{\mathbf{\Lambda}}^M$ a.e. by Theorem 2.4.3. Consequently, all convergence results apply, and solution to the infinite dimensional SIP can be approximated by the SIP with truncated dimension. All subsequent computations are carried out in BET package McDougall (2016).

We now formulate the SIP to solve for the probability distribution on $\mathbf{\Lambda}$ using the specified distribution on $\mathcal{D}$. To study the effect of decay, we consider domains of Fourier coefficients with decay powers from 1 to 4 for the SIP

$$\mathbf{\Lambda}^{(-p)} = \{(a_0, a_i, b_i) : a_0 \in [-\frac{1}{2}, \frac{1}{2}], a_i \in [-\frac{1}{2i^p}, \frac{1}{2i^p}], b_i \in [-\frac{1}{2i^p}, \frac{1}{2i^p}], i = 1, 2, 3, \cdots\}$$

for $p = 1, 2, 3, 4$.

Since all such SIP can be approximated by the SIP with truncated dimension, we set the truncation number $M = 9$ and then only consider a_0, a_i, b_i, $i = 1, 2, 3, 4$. Abusing the notation, we denote the truncated cube as $\mathbf{\Lambda}^{(-p)}$, that is

$$\mathbf{\Lambda}^{(-p)} = \{(a_0, a_i, b_i) : a_0 \in [-\frac{1}{2}, \frac{1}{2}], a_i \in [-\frac{1}{2i^p}, \frac{1}{2i^p}], b_i \in [-\frac{1}{2i^p}, \frac{1}{2i^p}], i = 1, 2, 3, 4\}$$

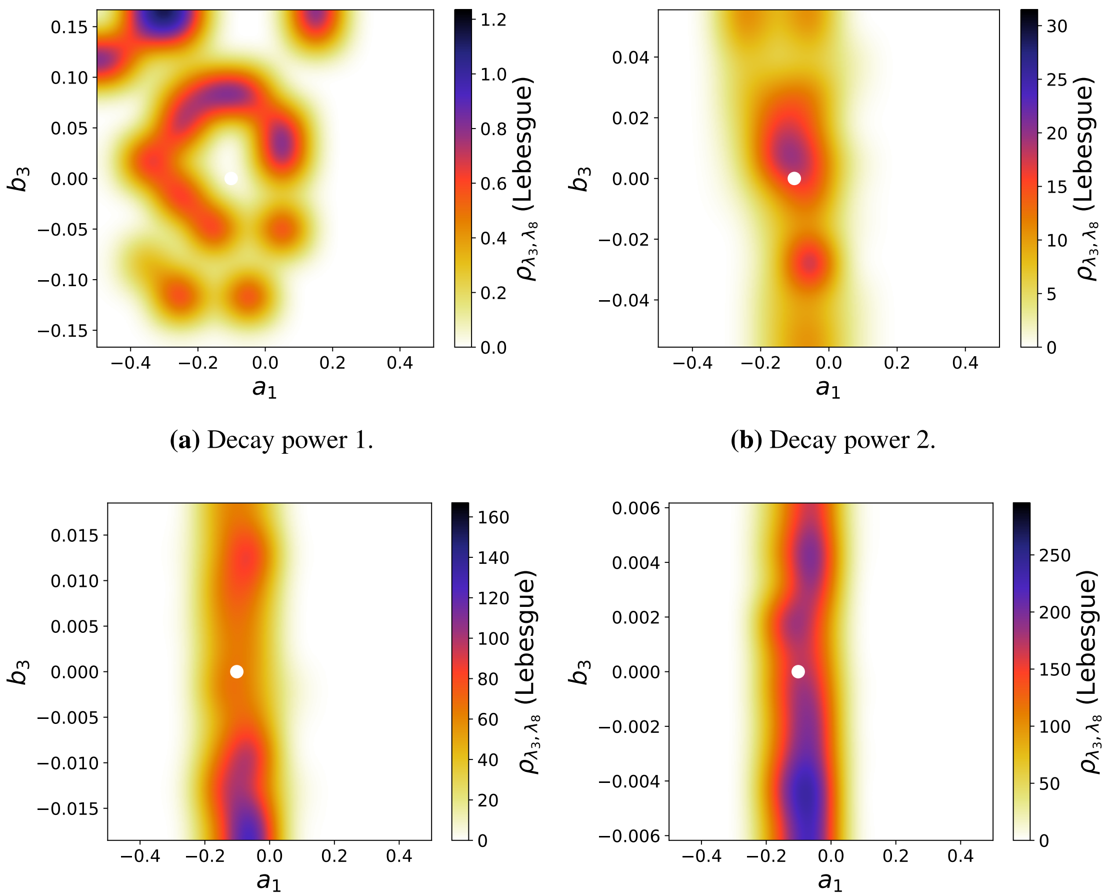
for $p = 1, 2, 3, 4$, as the domain for the SIP we actually solve. The true value on the truncated domain are then

$$\lambda_2 = (a_0, a_i, b_i) \text{ where } a_0 = \frac{1}{3}, a_i = -\frac{cos(i\pi)}{i^2\pi^2}, b_i = 0, i = 1, 2, 3, 4.$$

To solve these SIP, we simply let the partition on the range be the cube itself, and the partition on the domains be a Voronoi Tesselation generated by random sample of size 20000. The numerical solutions to each of the SIP are probability estimates for cells centered at every sample point in Voronoi tessellation on the domain, although we do not construct the Voronoi tessellation explicitly.

To compare the solutions, we show a two dimensional marginal density of a_1 and b_3 in Figure 6.2. While the edges of domains have different sizes, we scale them to the same size for comparison. When the decay power is increasing, the 2-dimensional marginal plots show less multimodal behavior and tend to be more spread out over the domain of b_3.

We draw sample points from the high probability region from different domains, and plot the functions they represent in Figure 6.3. The effect of the decay is then reflected in the fluctuation of the functions: the higher the decay rate, the less the fluctuation of the functions from the high probability region.



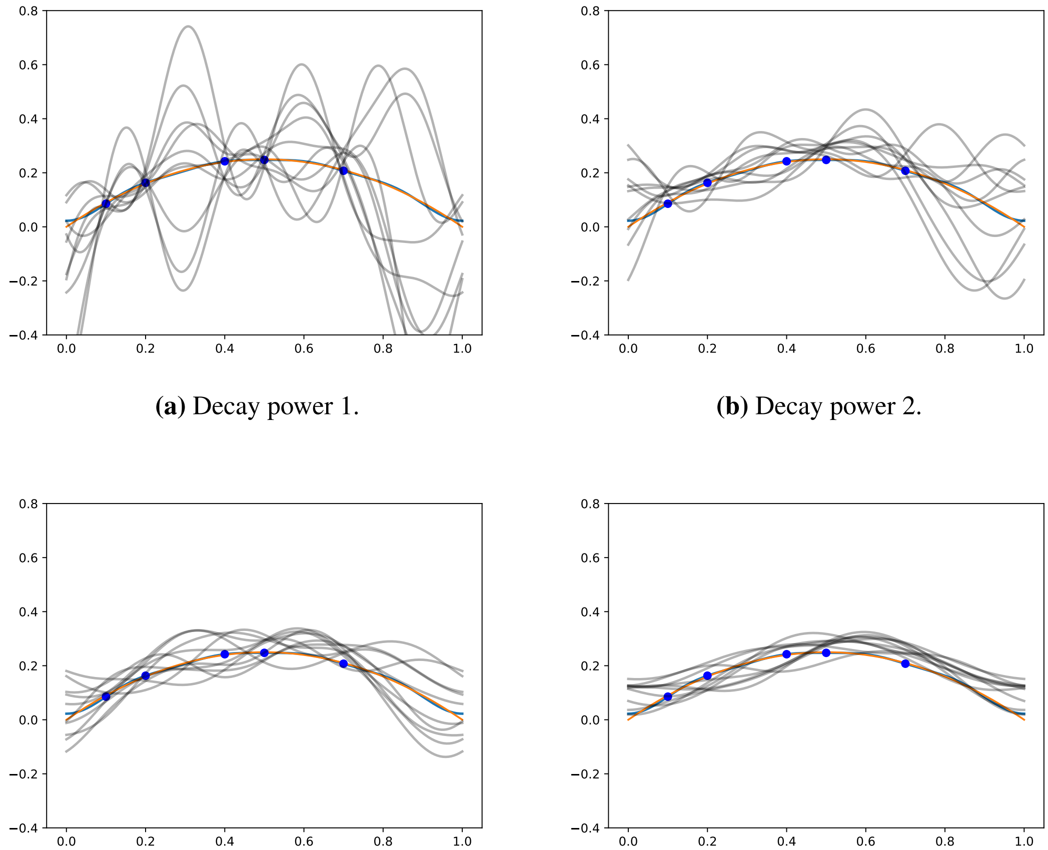
(a) Decay power 1.

(b) Decay power 2.

(c) Decay power 3.

(d) Decay power 4.

Figure 6.2: Marginal density of a_1, b_3 of solutions of SIP with domains of different decay powers.



(a) Decay power 1.

(b) Decay power 2.

(c) Decay power 3.

(d) Decay power 4.

Figure 6.3: Orange curve represents the true function $\lambda_2(t) = t(1-t)$, blue curve represents its Fourier expansion approximation up to truncation number 9, blue dots are the true values $Q(\lambda_2)$, grey curves are 10 samples from the high probability regions from solutions of the SIP.

Chapter 7

Review of probabilistic inverse problem methods

This chapter is a review about all inverse problem methods which solve model inputs probabilistically. It is a self-contained chapter begins with reviewing some of the pertinent materials from Chapter 1.

7.1 Introduction

Mathematical models, for example differential equations implemented by computer code using numerical methods, are widely used to characterize physical processes in systems in science and engineering disciplines. The solution to differential equations characterize behavior of the system being modeled. However the interest is typically not in the solution, but in Quantities of Interest (QoI) computed as functionals of the solution. Thus the model induces a complex, nonlinear, multiscale, computationally expensive measurable map from the parameter domain to the range defined as the space of QoI. Solutions of these models are usually determined by a set of parameters having physical meanings. Thus, parameter values are of interest for the information about the physical system being modeled. Parameter values are also critical for predicting system behavior. The procedure of inferring unknown parameters of mathematical models from QoI observations is a fundamental inverse problem that has many variations in mathematics and statistics. In this paper we just use “inverse problem" as a broad term for all such procedures.

Uncertainty and stochastic variation associated with parameters due to inherent stochasticity and/or incomplete knowledge is ubiquitous in physical modelling, and is a central topic of the discipline of Uncertainty Quantification (UQ). The main approach to represent and quantify such uncertainty uses probability theory, treats parameters as random variables and describes the uncertainty using probability distributions.

In recent years, we have developed a formulation of inverse problems called the Stochastic Inverse Problem (SIP) for computing the probability distribution of stochastic parameters from stochastic ovservations of QoI. We assume that there is an unknown probability distribution governing behavior of the parameters which results in a distribution describing the variations of the QoI. If the distribution on the QoI is given the SIP is to determnine the distribution on the parameters. This is complicated becasue the inverse problem for the model has set-valued solutions in general. This approach has a complete theoretical foundation based in measure theory, differential geometry, stochastic geometry, and numerical analysis, as well as an implementation that has allowed application to complex engineering problems.

The SIP begins with the assumption that a probability measure is given on the range of the mathematical model, while in practice only data are available and estimating the probability distribution from data is a well-studied task in statistics. As for many applications, the ideal approach would be applying the SIP coupled with statistical methods.

One of the most popular statistical methods is parametric statistical models based on the assumption that the true distribution comes from a family of distributions indexed by finite dimensional parameters. The inputs to the mathematical models are also often called parameters as we have done in the preceding discussions. To distinguish the two, we name the parameters to mathematical models as “inputs”, and reserve “parameter” to denote the vectors indexing statistical models.

Bayesian methods, such as Bayesian Inversion (Stuart, 2010) and Bayesian calibration(Kennedy and O’Hagan, 2001), are popular partly because Bayes rule automatically outputs the posterior distribution, up to its functional form, as the solution once the prior distribution and likelihood (statistical model) are specified. Random samples drawn supposedly from the posterior distribution via algorithms like MCMC are often used in applications. While Bayesian Inversion and Bayesian calibration both employ Bayes rule to yield posterior distributions of inputs, they have different assumptions and goals.

This article is divided into five parts. The next section, Section 2 describes the Deterministic Inverse Problem. Section 3 discusses the formulation and computational considerations of the SIP. Section 4 describes the basic formulation of Bayesian inversion, and Bayesian calibration with some discussions of their implications. It also contains a subsection explaining how Bayesian Inversion and SIP are related, with a simple example. Section 5 shows how SIP can be coupled with statistical methods.

7.2 Inverse problem

7.2.1 Basic setup

Differential equations induce an operator $Q_0 : \mathbf{\Lambda} \to S$ from the parameter space $\mathbf{\Lambda}$ to the solution function space S. Usually, the interest lies not in the solution itself, but in QoI, computed by functionals $Q_1 : S \to \mathcal{D}$. Then the composite operator $Q = Q_1 \circ Q_0 : \mathbf{\Lambda} \to \mathcal{D}$ maps the unknown parameter space $\mathbf{\Lambda}$ to space of observable QoI $\mathcal{D}$. We assume $\mathbf{\Lambda} \subset \mathbb{R}^n$. Since the collected data are always finite dimensional, we assume the observation space $\mathcal{D} \subset \mathbb{R}^m$.

In order to work with the probability distributions, we consider measurable spaces $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ and $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, where $\mathcal{B}_{\mathbf{\Lambda}}$ and $\mathcal{B}_{\mathcal{D}}$ are both Borel σ-algebras induced by the corresponding Borel σ-algebras in $\mathbb{R}^n$ and $\mathbb{R}^m$. Different approaches discussed in the following sections all rely on such measure structures.

7.2.2 The Deterministic Inverse Problem

Given a point $d \in \mathcal{D}$, the problem of finding input points $\lambda^* \in \mathbf{\Lambda}$ such that $Q(\lambda^*) = d$ is the Deterministic Inverse Problem. We assume J_Q has full rank. The existence and uniqueness of the inverse solution depend on the relationship between n and m. When $n > m$, the problem is underdetermined in which the solution, if it exists, is an $n-m$ dimensional manifold; when $n = m$, the solution is a unique point; when $n < m$, the problem is overdetermined, and the solution exists only if d belongs to an n dimensional submanifold in $\mathcal{D}$.

When λ^* does not exist, a least square estimator is often used instead. Namely, for a given norm $||\cdot||_{\mathcal{D}}$, we compute

$$\underset{\lambda}{\operatorname{argmin}}\, ||d - Q(\lambda)||^2_{\mathcal{D}},$$

and interpret λ^* as a "best" possible solution. Note that the least square estimate is the solution of the deterministic inverse problem, even when the exact solution λ^* exists, since $||d-Q(\lambda^*)||^2_{\mathcal{D}} = 0$. However, this approach does not solve the problem of non-uniqueness when $n < m$. For $n > m$, even when the solution exists and is unique, it is generally difficult to find. A common modification that alleviates numerical instability, and yields a unique solution in the underdetermined case is the so-called regularized least square, which has the form,

$$\underset{\lambda}{\operatorname{argmin}} \, ||d - Q(\lambda)||^2_{\mathcal{D}} + ||\lambda - \lambda_0||^2_{\mathbf{\Lambda}},$$

where a point λ_0 is specified along with a norm $||\cdot||_{\mathbf{\Lambda}}$ (Tikhonov and Arsenin, 1977).

The weight of regularization term can be adjusted so that the regularized least square estimator is unique and has nice numerical properties. However, the connection between the regularized least square estimator and the least square estimator and/or the set-valued inverse in the underdetermnined case is unclear.

7.2.3 The Stochastic Inverse Problem

Studying of forward stochastic problem, which concerns the definition and computing for probability distribution of QoI induced by a given probability distribution of the model inputs (Breidt et al., 2011; Butler et al., 2012, 2014), motivates the SIP, which is the direct inverse of the forward stochastic problem. In the Stochastic Forward Problem (SFP), a probability measure $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ induces a probability measure $P_{\mathcal{D}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ in the sense that for $A \in \mathcal{B}_{\mathcal{D}}$, $P_{\mathbf{\Lambda}}(Q^{-1}(A)) = P_{\mathcal{D}}(A)$, which is denoted $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$. The SIP is the direct inverse to the SFP in which the existence and computation of $P_{\mathbf{\Lambda}}$ such that $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$ given $P_{\mathcal{D}}$ is studied.

As with the Deterministic Inverse Problem, when $n \geq m$, solution exists but is not unique; when $n = m$, Q is a diffeomorphism, and there is a unique solution; when $n = \infty$ or $n > m$, the solution is not unique. We do not consider the overdetermined case. In that case generalizations of a least square estimator, for example, defining the solution as the one inducing a measure with the least Kullback-Leibler divergence with $P_{\mathcal{D}}$, are possible.

We assume $\mathcal{D} = Q(\mathbf{\Lambda})$, thus $\mathcal{D}$ is the image of Q on $\mathbf{\Lambda}$. As a result, each value $d \in \mathcal{D}$ corresponds to a set of values in Λ, denoted as $Q^{-1}(d)$. Using results from differential geometry, $Q^{-1}(d)$ is a $n-m$ dimensional smooth manifold, called generalized contour, and $\mathbf{\Lambda}$ can be decomposed as a union of non-intersecting generalized contours indexed by points in $\mathcal{D}$. If we define the equivalence relation on $\mathbf{\Lambda}$ as being mapped to the same point on $\mathcal{D}$, a generalized contour is an equivalence class under this equivalence relation, which induces an equivalence decomposition of $\mathbf{\Lambda}$. We can index the space of equivalence classes in $\mathbf{\Lambda}$ by $\mathcal{D}$.

The Disintegration Theorem(Chang and Pollard, 1997), which is analogous to a non-linear generalization of product measure theorem, plays a central role in the SIP. It asserts that there exists a family of probability measures $\{P_N(\cdot; d)\}$, each of which is concentrated on the generalized contour $Q^{-1}(d)$, such that

$$P_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(d) \cap A} dP_N(\lambda; d) dP_{\mathcal{D}}(d).$$

The disintegration of $P_{\mathbf{\Lambda}}$ reveals that $P_{\mathcal{D}}$ entirely determines the distributions of generalized contours, but provides no information as to $\{P_N(\cdot; d)\}$. So different families of conditional probability distributions $\{P_N(\cdot; d)\}$ yield different solutions of the SIP. The choice of $\{P_N(\cdot; d)\}$ is totally subjective and hence called "Ansatz". A common Ansatz is associated with the disintegration of Lebesgue measure $\mu_{\mathbf{\Lambda}}$:

$$\mu_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(d) \cap A} d\mu_N(\lambda; d) d\tilde{\mu}_{\mathcal{D}}(d),$$

where $\tilde{\mu}_{\mathcal{D}} = Q\mu_{\mathbf{\Lambda}}$. Under reasonable regularity conditions on Q and $\mathbf{\Lambda}$, $\mu_{\mathcal{D}}$ is absolutely continuous with respect to $\tilde{\mu}_{\mathcal{D}}$. So if $P_{\mathcal{D}}$ is absolutely continuous with respect to $\mu_{\mathcal{D}}$, it is absolutely continuous with respect to $\tilde{\mu}_{\mathcal{D}}$. The Uniform Ansatz (Butler et al., 2014) assumes $d\mu_N(\lambda; d) = dP_N(\lambda; d)$, and uniquely define $P_{\mathbf{\Lambda}}$ up to an almost sure equivalence on $\mathbf{\Lambda}$ with the density

$$p_{\mathbf{\Lambda}}(\lambda) = \frac{d\mu_{\mathcal{D}}(Q(\lambda))}{d\tilde{\mu}_{\mathcal{D}}(Q(\lambda))} p_{\mathcal{D}}(Q(\lambda)),$$

with respect to $\mu_{\mathbf{\Lambda}}$. Furthermore, any probability measure $P^0_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ which has a density with respect to $\mu_{\mathbf{\Lambda}}$ can provide a different family of conditional distribution by disintegration, and in turn determine a different $P_{\mathbf{\Lambda}}$ as a solution to the SIP under the Uniform Ansat. We refer to $P^0_{\mathbf{\Lambda}}$ as a base measure.

Now that the solution probability measure is uniquely defined, it needs to be computed numerically. A partition of $\mathbf{\Lambda}$ of which the elements are generating sets for Borel $\sigma-$algebra of $\mathbf{\Lambda}$, along with their probability estimates (Butler et al., 2014) is taken as a representation of $P_{\mathbf{\Lambda}}$. The partition can be a regular grid, while a Voronoi tessellation generated by Monte Carlo samples from a known probability distribution is more robust in high dimensional settings. The latter approach resembles importance sampling, as its probability estimates for each of Voronoi cells can be interpreted as weights in importance sampling.

7.3 Bayesian Inversion

The underpinning assumption of Bayesian Inversion have two ingredients: the inputs are fixed yet unknown, and the observed data point is an outcome of true output perturbed by additive random noise(Marzouk et al., 2007; Stuart, 2010). So it is necessary to distinguish the space of observed QoI, $\bar{\mathcal{D}}$, from the space of true QoI, $\mathcal{D}$. Again, we assume $(\bar{\mathcal{D}}, \mathcal{B}_{\bar{\mathcal{D}}})$ is a measurable space.

Let random variable Λ distributed as the prior distribution $P_{\mathbf{\Lambda}}$, so $Y = Q(\Lambda)$ is the random variable taking values in $\mathcal{D}$. By assumption, the observed random variable $\bar{Y}$ follows a parametric distribution $\{P_{\bar{\mathcal{D}}|\mathcal{D}}(\cdot|y)\}$ when $Y = y$, and each $P_{\bar{\mathcal{D}}|\mathcal{D}}$ as a measure on $(\bar{\mathcal{D}}, \mathcal{B}_{\bar{\mathcal{D}}})$ is absolutely continuous with respect to a measure ν on $(\bar{\mathcal{D}}, \mathcal{B}_{\bar{\mathcal{D}}})$, i.e.

$$p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{y}|y) = \frac{dP_{\bar{\mathcal{D}}|\mathcal{D}}(\cdot|y)}{d\nu}(\bar{y}).$$

Treated as a function of y with $\bar{y}$ fixed, $p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{y}|y)$ is the likelihood.

It is also assumed that the conditional distribution for $\bar{Y}$ given $\Lambda = \lambda$ is

$$P_{\bar{\mathcal{D}}|\mathbf{\Lambda}}(\cdot|\lambda) = P_{\bar{\mathcal{D}}|\mathcal{D}}(\cdot|Q(\lambda)).$$

So

$$p_{\bar{\mathcal{D}}|\mathbf{\Lambda}}(\bar{y}|\lambda) = \frac{dP_{\bar{\mathcal{D}}|\mathcal{D}}(\cdot|Q(\lambda))}{d\nu}(\bar{y}) = p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|Q(\lambda)).$$

When $p_{\bar{\mathcal{D}}|\mathbf{\Lambda}}(\bar{d}|\lambda)$ is a measurable function on $\bar{\mathcal{D}} \times \mathbf{\Lambda}$, Bayes Theorem (Schervish, 2012) implies that the posterior distribution of Λ satisfies:

$$\frac{dP_{\mathbf{\Lambda}|\bar{\mathcal{D}}}}{dP_{\mathbf{\Lambda}}}(\lambda|\bar{y}) = \frac{p_{\bar{\mathcal{D}}|\mathbf{\Lambda}}(\bar{y}|\lambda)}{\int_{\mathbf{\Lambda}} p_{\bar{\mathcal{D}}|\mathbf{\Lambda}}(\bar{y}|\lambda) dP_{\mathbf{\Lambda}}} = \frac{p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{y}|Q(\lambda))}{\int_{\mathbf{\Lambda}} p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{y}|Q(\lambda)) dP_{\mathbf{\Lambda}}}.$$

A common choice of $p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{y}|y)$ follows from $\bar{Y} = Y + \eta$ where Y and η are independent, and η follows a specific distribution. It is further assumed that i.i.d. additive errors are responsible for the discrepancy between each true QoI Y_i and its measurement $\bar{Y}_i$; i.e.,

$$\bar{Y}_i = Q(\Lambda_i) + \eta_i, \quad \eta_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0,\ \sigma^2\right), \quad 1 \leq i \leq m.$$

Frequently, it is assumed that $\bar{Y} - Q(\Lambda) \sim N(0, \sigma^2 \Sigma_0)$ with specified σ^2 and Σ_0 (Marzouk et al., 2007).

As an example, we assume a Gaussian prior for λ centered at λ_0, the posterior distribution of Λ satisfies:

$$p_{\mathbf{\Lambda}|\bar{\mathcal{D}}}(\lambda|\bar{d}) \propto \exp\left(-\frac{1}{2}||\bar{d} - Q(\lambda)||^2_{\mathcal{D}} - \frac{1}{2}||\lambda - \lambda_0||^2_{\mathbf{\Lambda}}\right).$$

This implies that the MAP estimator is equivalent to the regularized least square estimator. In general the solution depends on the choice of prior, where Bayesian theory does not provide a prior.

The primary goal of solving Bayesian Inversion is to represent the entire posterior distribution, which usually does not belong to any standard parametric distribution. This can either be done by drawing random samples from the distribution via algorithms like MCMC; or approximating the target probability density by variational methods.

7.4 The Bayesian Inversion posterior distribution as a solution to the SIP

If $\Lambda \sim P_{\mathbf{\Lambda}}$ and $Y = Q(\mathbf{\Lambda})$, it follows that $Y \sim P_{\mathcal{D}}$ where $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$. We refer to $P_{\mathcal{D}}$ as the induced prior. Given a conditional distribution $\{p_{\bar{\mathcal{D}}|\mathcal{D}}(\cdot|y)\}$, the posterior distribution $P_{\mathcal{D}|\bar{\mathcal{D}}}$ of Y satisfies

$$\frac{dP_{\mathcal{D}|\bar{\mathcal{D}}}}{dP_{\mathcal{D}}}(d|\bar{d}) = \frac{p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|d)}{\int_{\mathcal{D}} p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|d)dP_{\mathcal{D}}},$$

given $\bar{Y} = \bar{y}$.

Recall that the first step of the formulation of the SIP is to uniquely define a probability measure on $\mathbf{\Lambda}$ as the solution with the Uniform Ansatz. One natural question arises as to what is the posterior distribution of the Bayesian Inversion from the perspective of the SIP for the case $n \geq m$. Next theorem asserts that it can be shown as a solution to the SIP given $P_{\mathcal{D}|\bar{\mathcal{D}}}$ on $\mathcal{D}$.

Theorem 7.4.1. *When $n > m$, the posterior distribution $P_{\mathbf{\Lambda}|\bar{\mathcal{D}}}(\cdot|\bar{d})$ of Λ, is a solution to the SIP given $P_{\mathcal{D}|\bar{\mathcal{D}}}(\cdot|\bar{d})$, the posterior distribution of Y on $\mathcal{D}$, with the Uniform Ansatz with respect to the prior $P_{\mathbf{\Lambda}}$. A direct result is $P_{\mathcal{D}|\bar{\mathcal{D}}}(\cdot|\bar{d}) = QP_{\mathbf{\Lambda}|\bar{\mathcal{D}}}(\cdot|\bar{d})$.*

Proof. By the Disintegration Theorem (Chang and Pollard, 1997), given $A \in \mathcal{B}(\mathbf{\Lambda})$, $P_{\mathbf{\Lambda}}$ admits the following disintegration:

$$P_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(d)\cap A} dP_N(\lambda; d)dP_{\mathcal{D}}(d).$$

The posterior distribution $P_{\mathcal{D}|\bar{\mathcal{D}}}$ satisfies

$$\frac{dP_{\mathcal{D}|\bar{\mathcal{D}}}}{dP_{\mathcal{D}}}(d|\bar{d}) = \frac{p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|d)}{\int_{\mathcal{D}} p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|d)dP_{\mathcal{D}}}.$$

Assuming there is a $\hat{P}_{\mathbf{\Lambda}}$ on $\mathbf{\Lambda}$ such that its induced measure on $\mathcal{D}$ is $P_{\mathcal{D}|\bar{\mathcal{D}}}$, then:

$$\hat{P}_{\mathbf{\Lambda}}(A) = \int_{Q(A)} \int_{Q^{-1}(d)\cap A} d\hat{P}_N(\lambda; d)dP_{\mathcal{D}|\bar{\mathcal{D}}}(d|\bar{d}).$$

Under the Uniform Ansatz $d\hat{P}_N(\lambda; d) = dP_N(\lambda; d)$, we get

$$\frac{d\hat{P}_{\mathbf{\Lambda}}}{dP_{\mathbf{\Lambda}}}(\lambda|\bar{d}) = \frac{p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|Q(\lambda))}{\int_{\mathbf{\Lambda}} p_{\bar{\mathcal{D}}|\mathcal{D}}(\bar{d}|Q(\lambda))dP_{\mathbf{\Lambda}}},$$

which means $\hat{P}_{\mathbf{\Lambda}} = P_{\mathbf{\Lambda}|\bar{\mathcal{D}}}(\cdot|\bar{d})$ is the solution to the Bayesian Inversion problem. $\square$

7.5 Example

We assume that $\mathbf{\Lambda} = \{(\lambda_1, \lambda_2) : |\lambda_1| < 1, |\lambda_2| < 1\}$, and the map is $Q(\lambda_1, \lambda_2) = 2\lambda_1^2 + \lambda_2^3$. We also assume a truncated Gaussian prior on $\mathbf{\Lambda}$, which can be sampled by first sampling $(\lambda_1, \lambda_2)^T \in \mathbb{R}^2$ by

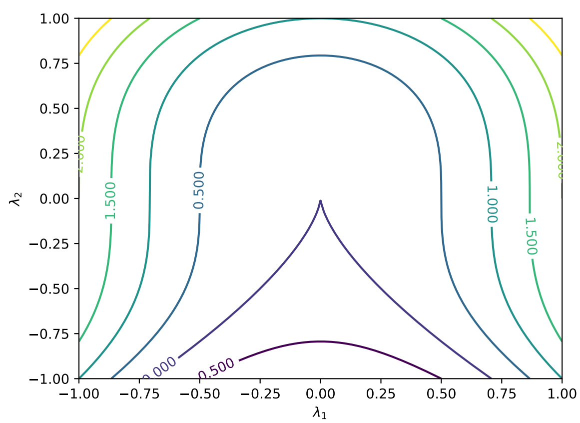


Figure 7.1: Contour plot of Q.

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0.125 \\ 0.125 & 0.3125 \end{pmatrix}\right),$$

then remove samples outside $\mathbf{\Lambda}$.

We assume $\bar{d} = 2$ and a Gaussian noise model with $\sigma^2 = 1$. So the likelihood is proportional to

$$\exp\{-\frac{1}{2}(Q(\lambda_1, \lambda_2) - 2)^2\}.$$

Then, the posterior has the form

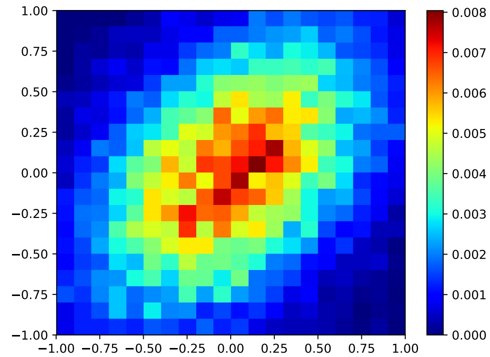
$$p_{\mathbf{\Lambda}|\bar{\mathcal{D}}}(\lambda|\bar{d}) \propto \exp\{-\frac{1}{2}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}^T \begin{pmatrix}0.25 & 0.125\\ 0.125 & 0.3125\end{pmatrix}^{-1}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix} \\ -\frac{1}{2}\left(2\lambda_1^2+\lambda_2^3-2\right)^2\}1_{\lambda_1^2+\lambda_2^2<1}.$$

While MCMC algorithms are widely used in Bayesian methods, we use an importance sampling approach due to its easy implementation and close relation to the SIP numerical algorithms. Specifically, we generate random samples from the prior distribution, and then use the likelihood as weights of the samples. By Bayes rule, we know that these weighted samples can represent the posterior distribution.

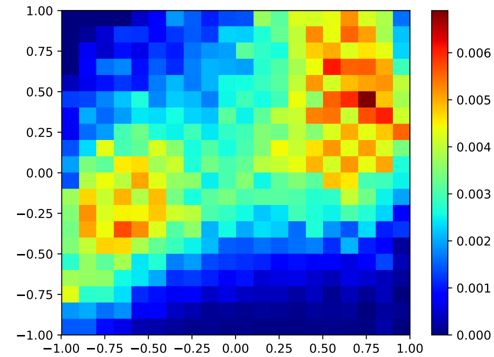
Although the solution to the Bayesian Inversion is uniquely defined, there are infinitely many probability measures on $\mathbf{\Lambda}$ capable of inducing the posterior distribution on $\mathcal{D}$. To illustrate this and compare with results for Bayesian Inversion, we solve the SIP using the Uniform Ansatz with respect to the Lebesgue measure and the prior distribution given the posterior distribution on $\mathcal{D}$. As shown in Figure 7.3, while two probability measures are very different, they induce the same probability measure on $\mathcal{D}$. The solution using the Lebesgue base measure has contours which match those in Figure 7.1 due to the Uniform Ansatz. The solution using prior base measure is the same as the posterior for the Bayesian Inversion as implied by Theorem 7.4.1.

7.6 Bayesian Calibration

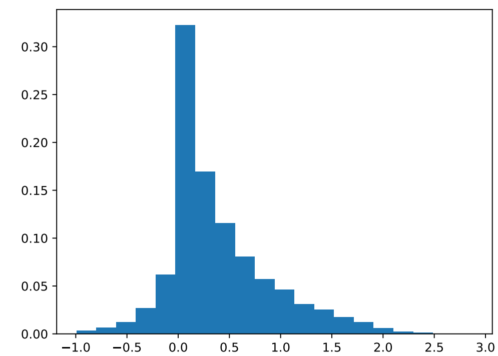
There is another popular approach towards modelling the inputs of deterministic model as random variables named Bayesian Calibration pioneered by Kennedy and O'Hagan (2001). Calibration models use Gaussian processes to represent both the real physical process and mathematical model represented by computer code. Then it aims to find a best-fitting parameter λ "in the sense of representing the data faithfully according to error structure that is specified for the residuals"(Kennedy and O'Hagan, 2001). The processes often have spatial or temperal domain.



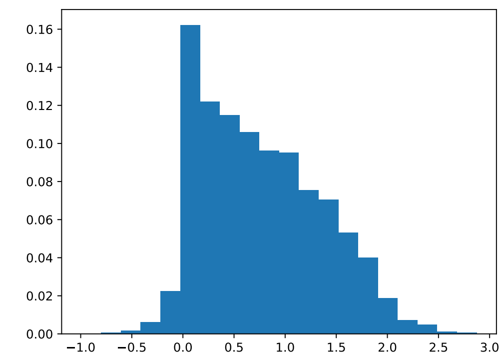
(a) Prior distribution on $\mathbf{\Lambda}$.



(b) Posterior distribution on $\mathbf{\Lambda}$.

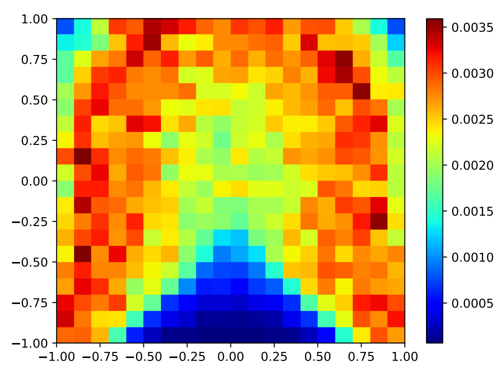


(c) Induced prior distribution on $\mathcal{D}$

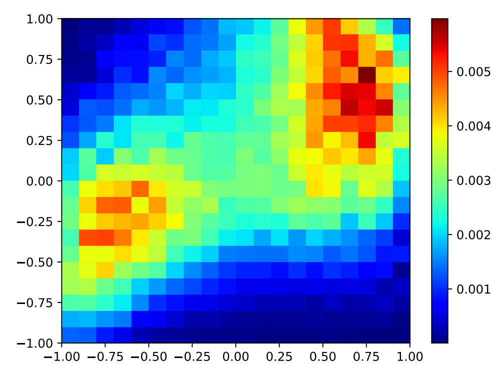


(d) Induced posterior distribution on $\mathcal{D}$.

Figure 7.2: 2d histogram for prior and posterior on $\mathbf{\Lambda}$, and 1d histogram for prior and posterior on $\mathcal{D}$.



(a) Solution using uniform base measure.



(b) Solution using prior base measure.

Figure 7.3: 2d histogram for solutions of the SIP

For physical process, the data are real observations; for computer code process, the data are generated by computer program. Bayesian hierarchical model provides a flexible tool to incorporate mean and covariance parameters for both processes and the inputs. With a multiple of parameters, Bayesian hierarchical model enjoys strong expressive power but may suffer from identifiable or consistency problems (Tuo and Jeff Wu, 2016).

7.7 Comment

The above theorem and example illustrate how the SIP can be employed to connect two posterior distributions from the Bayesian Inversion. However, the SIP and Bayesian Inversion are fundamentally different approaches reflecting distinct beliefs. The SIP begins with the idea that there is inherent stochasticity of the inputs responsible for the stochasticity of the QoI. In Bayesian Inversion, however, the posterior distribution is totally determined by the noise model with assumed variance and subjective belief summarized by prior distribution. The Bayesian Inversion is usually formulated for a one point problem, i.e. there are no replications in each QoI. Should the replications are given, the variance of noise can be estimated and uncertainty of the inputs can be hugely reduced.

So in Bayesian Inversion, more replications should result in a more concentrated posterior distribution of Λ, indicating less uncertainty; while for the SIP, ideally more data lead to a more accurate estimation of an irreducible probability distribution of Y and hence of Λ. To faithfully reflect the motivation of SIP, we need to apply a full statistical modeling approach to the data with replications, as is described in next section.

7.8 SIP coupled with statistical models

Consider the problem of estimating probability distribution from a random sample of i.i.d. random vectors whose observations are not available. Rather, there is a deterministic map from the space of random vector values, refered to as input space, to an observational space, thus inducing a new set of i.i.d. random vectors whose observations can be collected.

We assume a random vector Λ taking values on $\mathbf{\Lambda}$ with distribution $P_{\mathbf{\Lambda}}$, and it induces random vector $Y = Q(\Lambda)$ taking values on $\mathcal{D}$ with distribution $P_{\mathcal{D}}$, then $P_{\mathcal{D}} = QP_{\mathbf{\Lambda}}$. Given $p_{\mathcal{D}}$ as the density of $P_{\mathcal{D}}$ with respect to $\mu_{\mathcal{D}}$, denoted as $Y \sim p_{\mathcal{D}}$, we want to know the distribution of Λ. Then under usual conditions of the SIP, $\Lambda \sim p_{\mathbf{\Lambda}}$, where

$$p_{\mathbf{\Lambda}}(\lambda) = \frac{d\mu_{\mathcal{D}}(Q(\lambda))}{d\tilde{\mu}_{\mathcal{D}}(Q(\lambda))} p_{\mathcal{D}}(Q(\lambda)).$$

A common task of statistics is to estimate a probability distribution function by random samples drawn from it. When data are abundant, nonparametric statistical methods like histogram often yield reasonable results. On the contrary when data are relatively scarce and can be comfortably assumed to drawn from a family of probability distributions indexed by parameters, parametric statistical modelling are more efficient.

The SIP and statistical models, each addressing one aspect of the proposed problem, should be combined to solve it as a whole. One possibility is to apply them stepwise: first to fit a statistical model to the data on $\mathcal{D}$ yielding a probability distribution of the observations, then find the probability measure as the solution of the SIP on $\mathbf{\Lambda}$. Another possibility is to directly make inference for the statistical models on $\mathbf{\Lambda}$ induced by statistical models on $\mathcal{D}$. Here we compare the two approaches in a parametric statistical modelling framework.

We first consider a one variable problem where there is one random variable Λ taking values on $\mathbf{\Lambda}$ and $Y = Q(\Lambda)$. Assume a parametric model indexed by $\theta \in \mathbf{\Theta} \subset \mathbb{R}^k$ denoted as $\mathcal{P}_{\mathcal{D}|\mathbf{\Theta}} = \{p_{\mathcal{D}|\mathbf{\Theta}}(\cdot|\theta) : \theta \in \mathbf{\Theta}\}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, which induces a parametric model $\mathcal{P}_{\mathbf{\Lambda}|\mathbf{\Theta}}$ on $\mathbf{\Lambda}$ by

$$p_{\mathbf{\Lambda}|\mathbf{\Theta}}(\lambda|\theta) = \frac{d\mu_{\mathcal{D}}(Q(\lambda))}{d\tilde{\mu}_{\mathcal{D}}(Q(\lambda))} p_{\mathcal{D}|\Theta}(Q(\lambda)|\theta), \quad \theta \in \Theta,$$

under the Uniform Ansatz. Assume y is a realization of Y, then treated as a function of θ, the likelihood function on $\mathcal{D}$ is just

$$\mathcal{L}(\theta|y) = p_{\mathcal{D}|\mathbf{\Theta}}(y|\theta).$$

Since $Y = Q(\Lambda)$, the realization of Λ satisfies $\lambda \in Q^{-1}(y)$, yielding the likelihood on $\mathbf{\Lambda}$ as

$$\mathcal{L}(\theta|\lambda) = \frac{d\mu_{\mathcal{D}}(Q(\lambda))}{d\tilde{\mu}_{\mathcal{D}}(Q(\lambda))} p_{\mathcal{D}|\mathbf{\Theta}}(Q(\lambda)|\theta) = \frac{d\mu_{\mathcal{D}}(y)}{d\tilde{\mu}_{\mathcal{D}}(y)} p_{\mathcal{D}|\mathbf{\Theta}}(y|\theta) \propto \mathcal{L}(\theta|y).$$

Thus the likelihood on $\mathcal{D}$ is proportional to the likelihood on $\mathbf{\Lambda}$.

Where there are i.i.d. $\{\Lambda_\ell\}_{\ell=1}^L$ and $Y_\ell = Q(\Lambda_\ell)$, $\{Y_\ell\}_{\ell=1}^L$ are also i.i.d.. Further assume $Y_\ell \overset{\text{iid}}{\sim} p_{\mathcal{D}|\mathbf{\Theta}}(\cdot|\theta), 1 \le \ell \le L, \theta \in \mathbf{\Theta}$, given their realizations $\{y_\ell\}_{\ell=1}^L$, define $\mathbf{y} = (y_\ell)_{\ell=1}^L$, the likelihood based on the statistical models on $\mathcal{D}$ is

$$\mathcal{L}(\theta|\mathbf{y}) = \prod_{\ell=1}^{L} p_{\mathcal{D}|\mathbf{\Theta}}(y_\ell|\theta).$$

The realizations of $\{\Lambda_\ell\}_{\ell=1}^L$ are unknown $\{\lambda_\ell\}_{\ell=1}^L$ with $Q(\lambda_\ell) = y_\ell$. We define $\boldsymbol{\lambda} = (\lambda_\ell)_{\ell=1}^L$, so under the Uniform Ansatz the likelihood function on $\mathbf{\Lambda}$ is:

$$\mathcal{L}(\theta|\boldsymbol{\lambda}) = \prod_{\ell=1}^{L} \frac{d\mu_{\mathcal{D}}(Q(\lambda_\ell))}{d\tilde{\mu}_{\mathcal{D}}(Q(\lambda_\ell))} p_{\mathcal{D}|\mathbf{\Theta}}(Q(\lambda_\ell)|\theta) = \prod_{\ell=1}^{L} \frac{d\mu_{\mathcal{D}}(y_\ell)}{d\tilde{\mu}_{\mathcal{D}}(y_\ell)} \prod_{\ell=1}^{L} p_{\mathcal{D}|\mathbf{\Theta}}(y_\ell|\theta) \propto \mathcal{L}(\theta|\mathbf{y}).$$

Under both scenarios, the implication of the Uniform Ansatz can be summarized as this: the location of the assumed λ on the generalized contour $Q^{-1}(y)$ does not affect the the value of likelihood function for $Q(\lambda) = y$.

There is a famous principle in statistics called "likelihood principle". One statement of this is "all information about θ obtainable from an experiment is contained in the likelihood function for θ given x"(Lindsey, 1999). Based on the fact that $\mathcal{L}(\theta|\boldsymbol{\lambda})$ and $\mathcal{L}(\theta|\mathbf{y})$ only differ by a factor, we come to the following result.

Lemma 7.8.1 (Invariance of θ). *Statistical inference procedures satisfying likelihood principle produce same results concerning θ on $\mathcal{D}$ and on $\mathbf{\Lambda}$ with the Uniform Ansatz.*

Many popular statistical inference procedures, including maximum likelihood and Bayesian statistics, follow likelihood principle. Broadly speaking, many nonparametric procedures, such as histogram, can be interpreted as a result of maximum likelihood estimation by the method of sieves (Geman and Hwang, 1982). These methods will lead to the same inference of θ for statistical models on $\mathcal{D}$ and $\mathbf{\Lambda}$. As the probability distributions on $\mathbf{\Lambda}$ and $\mathcal{D}$ indexed by the same parameter θ have one-to-one correspondence in the SIP framework under the Uniform Ansatz, the two different approaches of combining SIP and statistical models are equivalent in the sense explained below.

Theorem 7.8.2 (Invariance of probability distribution of Λ)**.** *Two approaches of combining the SIP under the Uniform Ansatz and statistical models satisfying likelihood principle:*

1. *First fit the statistical model to the data on $\mathcal{D}$ and obtain the corresponding probability distribution on $\mathcal{D}$, then solve the SIP using the Uniform Ansatz to get the probability distribution on $\mathbf{\Lambda}$;*

2. *Since the statistical model on $\mathcal{D}$ induces a statistical model on $\mathbf{\Lambda}$ by the SIP under the Uniform Ansatz, directly make statistical inference on $\mathbf{\Lambda}$ and generate the corresponding probability distribution on $\mathbf{\Lambda}$.*

Since Bayesian method is one such invariant procedure, Theorem 7.8.1 implies the posterior distributions of θ generated by the statistical models on $\mathcal{D}$ and $\mathbf{\Lambda}$ are the same given the same prior distribution; and Theorem 7.8.2 implies the posterior predictive distributions of Λ generated by the two mentioned procedures are the same as well.

Unlike the Bayesian Inversion or the Bayesian Calibration, the prior in this approach directly specified to the statistical parameter space Θ rather than the mathematical input space $\mathbf{\Lambda}$. In statistical terminology, the parameter Θ is indeed the hyperparameter for Λ. So the uncertainty expressed by the prior and posterior distributions are in terms of the statistical parameters. So more data result in a more accurate estimation of probability distribution of $\mathcal{D}$ under the Uniform Ansatz, rather than a more accurate estimation of $\mathcal{D}$ itself.

The choice of appropriate statistical methods is vital for every statistical application, and the SIP coupled with statistical model is no exception. Specifically, the statistical models should be able to capture the variations of QoI. The choice of parameter estimation method, among maximum likelihood method, Bayesian method etc., is secondary. A poorly chosen statistical model would not produce a satisfactory fit regardless of parameter estimation method.

Bibliography

KC Border. Notes on the implicit function theorem. *California Institut of Technology, Division of the Humanities and Social Sciences, link: www. hss. calthch. edu/kcb/Notes/ITF. pdf*, 2013.

J Breidt, T Butler, and D Estep. A measure-theoretic computational method for inverse sensitivity problems i: Method and analysis. *SIAM journal on numerical analysis*, 49(5):1836–1859, 2011.

Andrew M Bruckner, Judith B Bruckner, and Brian S Thomson. *Real analysis*. ClassicalRealAnalysis. com, 1997.

T Butler, D Estep, and J Sandelin. A computational measure theoretic approach to inverse sensitivity problems ii: a posteriori error analysis. *SIAM journal on numerical analysis*, 50(1):22–45, 2012.

T Butler, D Estep, S Tavener, C Dawson, and J Westerink. A measure-theoretic computational method for inverse sensitivity problems iii: Multiple quantities of interest. *SIAM Journal on Uncertainty Quantification*, pages 1–27, 2014.

Joseph T Chang and David Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3): 287–317, 1997.

John B Conway. *A course in functional analysis*, volume 96. Springer Science & Business Media, 1994.

Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Springer, 2006.

Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

Stuart Geman and Chii-Ruey Hwang. Nonparametric maximum likelihood estimation by the method of sieves. *The Annals of Statistics*, pages 401–414, 1982.

Loukas Grafakos. *Classical fourier analysis*, volume 2. Springer, 2008.

Paul R Halmos. *Measure theory*, volume 18. Springer, 1974.

Morris W. Hirsch. *Differential Topology*. Springer, 1976.

Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

Estate Khmaladze and N Toronjadze. On the almost sure coverage property of voronoi tessellation: the âĎİ 1 case. *Advances in Applied Probability*, 33(4):756–764, 2001.

JK Lindsey. Some statistical heresies. *The Statistician*, pages 1–40, 1999.

Youssef M Marzouk, Habib N Najm, and Larry A Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, 2007.

Lindley Graham; Steven Mattis; Scott Walsh; Troy Butler; Michael Pilosov; Damon McDougall. Butler, estep, tavener method, 2016. URL http://ut-chg.github.io/BET/.

Walter Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 1987.

Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

Arthur Sard. Hausdorff measure of critical images on banach manifolds. *American Journal of Mathematics*, 87(1):158–174, 1965.

Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

Henrik Schlichtkrull. Differentiable manifolds, lecture notes for geometry 2. *University of Copenhagen, Department of Mathematics, link: www.math.ku.dk/ jakobsen/geom2/manusgeom2.pdf*, 2013.

Elias M Stein and Rami Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2005.

Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

A.N. Tikhonov and V.I.A. Arsenin. *Solutions of ill-posed problems*. Scripta series in mathematics. Winston, 1977. ISBN 9780470991244. URL https://books.google.com/books?id=ECrvAAAAMAAJ.

Rui Tuo and CF Jeff Wu. A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795, 2016.

Wikipedia. Smith-volterra-cantor set — Wikipedia, the free encyclopedia, 2017. URL https://en.wikipedia.org/w/index.php?title=Smith%E2%80%93Volterra%E2%80%93Cantor_set&oldid=774204876. [Online; accessed 2-May-2018].

Jenny Wilson. The definition of a manifold and first examples. *Stanford University, Department of Mathematics, link: https://web.stanford.edu/ jchw/WOMPtalk-Manifolds.pdf*, 2012.