Probability Measure Estimation in Positron Emission Tomography Using Loss Functions Based on Sobolev Norms

by

Alvin Kuruc

S.B., Massachusetts Institute of Technology (1979)
M.D., Yale Medical School (1984)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 1994

Signature of Author ....
Department of Mathematics
April 29, 1994

Certified by...
Mark V. Matthews
Assistant Professor, Department of Mathematics
Thesis Supervisor

Accepted by ....
David A. Vogan
ARCHIVES
Chairman, Departmental Committee on Graduate Students

Probability Measure Estimation in Positron Emission Tomography Using Loss Functions Based on Sobolev Norms

by

Alvin Kuruc

Abstract

Motivated by an idealized model of positron emission tomography, we consider the statistical problem of estimating an unknown probability measure, μ, given n independent observations distributed according to the probability measure $T\mu$, where T denotes the (scaled) Radon transform on measures. The usual approach to such a problem assumes that μ is representable by a probability density function satisfying certain smoothness conditions and uses a loss function based on metrics for probability measures which metrize convergence in total variation, e.g., the L^1 metric on the associated density functions. In contrast, we allow μ to be an arbitrary probability measure on the unit square in $\mathbb{R}^2$ and use loss functions based on L^2-Sobolev norms which metrize convergence in law. We thereby obtain results on the rate of convergence of the minimax risk as a function of n without the need for smoothness assumptions. Moreover, we argue that these metrics are more relevant to the physical problem.

Consider the loss function generated by the squared L^2 Sobolev norm of order $-\alpha$. We show that the minimax error is $O(n^{-1})$ if and only if $\alpha > 3/2$. In comparison, the minimax error given n independent observations distributed according μ itself is $O(n^{-1})$ if and only if $\alpha > 1$.

We also give several results on the estimation of integral functionals of μ. For example, let $\rho\bar{\Omega}$ denote the closed disk of radius $\rho < 1/2$ centered at the origin. The minimax risk with respect to squared error loss for the estimation of $\mu(\rho\bar{\Omega})$ is $O(n^{-1})$ given n independent observations distributed according to μ, but is bounded away from 0 given n independent observations distributed according to $T\mu$.

Thesis Supervisor: Mark V. Matthews
Title: Assistant Professor, Department of Mathematics

Acknowledgements

While reading Ph.D. theses, one might get the impression that their writers become overly sentimental as they approach graduation. Having now gone though the process myself, I am more sympathetic to their plight. I too find that I have many debts of gratitude.

I would also like to thank my friends and family members for all their encouragement. Special thanks go to my wife, Donna, who braved the rigors of graduate school with me, and my four-legged "children" Maxwell (now departed), Newton, Freud, and Russell. They were comforting and (usually) unobtrusive companions during the many hours that I had my nose in a book or computer screen.

But most of all, I would like to thank my father, Andrew W. Kuruc. For many years, he placed my needs, and those of my brother, before his own. What I have achieved, or will achieve in the future, is in large part due to the sacrifices he made for his progeny. It is to him that I dedicate this work.

Contents

Chapter 1

Introduction

1.1 Overview

The motivating physical problem for the work in this thesis is the medical imaging technique of positron emission tomography (PET). In PET, the goal is to characterize the spatial distribution of positron-emitting tracer molecules in a patient using external detectors. Shortly after a positron is emitted, it combines with an electron in an annihilation reaction. On the average, a positron travels only a very short distance between emission and annihilation, so the position of the annihilation reaction is approximately that of the positron emission. The annihilation reaction results in the emission of a pair of annihilation photons traveling in (approximately) opposite directions along a line with uniformly-distributed random spatial orientation. These photons subsequently strike detectors at approximately the same time, forming what is known as a coincident pair. From a coincident pair, one can thus infer that a positron was emitted (approximately) on the line segment between the two registering detectors. More detailed descriptions of PET may be found in, e.g., [SP87] or [MP88].

The PET problem is inherently 3-dimensional in that the positions of the positron emissions and the paths of the resulting annihilation photons are naturally modeled as points and lines, respectively, in Euclidean 3-space. However, the most common detector configurations actually detect only coincident pairs whose paths are close

to a fixed plane (and which therefore must originate from an annihilation reaction occurring near that plane). In this case, it makes sense to model the problem in 2 dimensions. In what follows, we will, for simplicity, restrict ourselves to the 2-dimensional PET problem.

To obtain a mathematical idealization of the PET problem, one may model the locations of the positron emissions as n points, $x_1, \ldots, x_n$, in Euclidean 2-space, $\mathbb{R}^2$. The observations in the PET problem may be modeled as n independent random lines, $l_1, \ldots, l_n$, in $\mathbb{R}^2$, where l_i is a line through x_i with (uniformly distributed) random spatial orientation.

To complete the specification of a statistical model for the PET problem, we need to state our assumptions about the x_i. One possibility is just to model the x_i as unknown parameters to be estimated. Alternatively, one may view the x_i as a random sample from the set of all the tracer molecules present. In the limiting case where the number of tracer molecules approaches infinity, this suggests modeling the x_i as independent random variables distributed according to a probability measure μ on a domain $D \subset \mathbb{R}^2$. (This model makes sense, of course, only if the distribution of the tracer molecules remains constant for the duration of the experiment. We will assume this to be the case.) We will use such a model in the remainder of this work.

For simplicity, we shall take D to be the closed unit square $S \subset \mathbb{R}^2$. The orientations of the l_i are assumed to be independent of the x_j and each other, so that the l_j are independent random variables distributed according to the probability measure $\nu \stackrel{\mathrm{def}}{=} T\mu$ on the set of lines in $\mathbb{R}^2$, where T is a (deterministic) function. It is shown in section 2.3 that T is proportional to the Radon transform on measures. We will call μ the probability measure of interest and call ν the probability measure of the observations. We are thus faced with the statistical inverse problem of trying to estimate the unknown probability measure μ from n independent observations distributed according to the probability measure $T\mu$. The main goal of this work is to obtain suitable notions of the intrinsic difficulty of this statistical estimation problem.

It should be emphasized that the model described here is a very idealized model of the PET problem. It ignores a number of physical effects which complicate PET in

practice. Nevertheless, it appears to capture the essence of the problem of emission tomography.

In order to put the various possible statistical approaches to the PET problem into a single theoretical framework, it is useful to introduce some standard notions from statistical decision theory. We follow [Str85].

Definition 1.1.1 A statistical experiment is defined to be an ordered triple

$$E = (X, \mathscr{A}, \{\mu_\theta : \theta \in \Theta\}),$$

where $(X, \mathscr{A})$ is a measurable space and $\{\mu_\theta : \theta \in \Theta\}$ is a set of probability measures on $(X, \mathscr{A})$ indexed by the set Θ. The set X, known as the sample space, is the set of possible observations. Under the statistical model (i.e., statistical hypothesis) indexed by θ, the observation is a random variable distributed according to the probability measure μ_θ. The set Θ is called the parameter space (cf. [Str85, def. 7.1]).

Definition 1.1.2 Let $\mathcal{D}$ be a metric space, which we will refer to as the decision space. We equip $\mathcal{D}$ with the Borel σ-algebra. A measurable function $\kappa : X \to \mathcal{D}$ is known as a (nonrandomized) decision function. A family of measurable functions $W_\theta : \mathcal{D} \to \mathbb{R}$, $\theta \in \Theta$ that are bounded from below is called a loss function. For a decision function κ, $r(\theta, \kappa) \stackrel{\text{def}}{=} E_\theta W_\theta \kappa$ is called the risk of κ with respect to the loss function W_θ, $\theta \in \Theta$. (E_θ denotes mathematical expectation with respect to the probability measure μ_θ.) The combination of a statistical experiment with a decision space and loss function is known as a statistical decision problem (cf. [Str85, defs. 33.1, 33.3, rem. 33.2]).

Definition 1.1.3 It is often useful to extend the notion of decision function to permit what are known as randomized decision functions. Roughly speaking, a randomized decision function is a random, rather than a deterministic, function of $x \in X$. The formalities may be found in, e.g., [Str85, defs. 33.1, 33.2, rem. 33.2]. In what follows, we will take the term decision function to mean a randomized decision function.

8

Definition 1.1.4 We shall be interested exclusively in statistical decision problems in which W_θ depends on θ only through some function $\phi : \Theta \to \mathcal{D}$, i.e., we will be estimating $\phi(\theta)$. Such a statistical decision problem is called a statistical estimation problem [Str85, def. 33.5]. Most frequently, we will take $\mathcal{D} = \Theta$ and ϕ equal to the identity function, i.e., we will be estimating θ. In statistical estimation problems, the decision function is called an estimator. In all instances considered here, the loss function is of the form $W_\theta = \ell \circ d(\cdot, \phi(\theta))$, where d is a metric on $\mathcal{D}$ and $\ell : [0, \infty) \to [0, \infty)$ is an increasing function (typically $\ell : x \mapsto x$ or $\ell : x \mapsto x^2$). We will say that such a loss function is based on the metric d and generated by the function $\ell \circ d$. If $\hat{\phi}$ is an estimator of $\phi(\theta)$, we will write the risk of $\hat{\phi}$ as

$$r(\theta, \hat{\phi}) = E_\theta \ell \circ d(\hat{\phi}, \phi(\theta)).$$

Roughly speaking, the goal of a statistical estimation problem is to find estimators with small risk for many of the $\theta \in \Theta$. The following definition formalizes one approach to quantifying the difficulty of a statistical estimation problem.

Definition 1.1.5 By a slight abuse of language, we define the maximum risk of an estimator $\hat{\phi}$ to be

$$\sup_{\theta \in \Theta} r(\theta, \hat{\phi}).$$

The maximum risk quantifies the performance of the estimator $\hat{\phi}$ in terms of its worst-case risk. The intrinsic difficulty of a statistical estimation problem can then be quantified in terms of the smallest maximum risk that is achievable by any estimator. Thus we define the minimax risk of a statistical estimation problem to be

$$\inf_{\hat{\phi}} \sup_{\theta \in \Theta} r(\theta, \hat{\phi}),$$

where the infimum is taken over all estimators $\hat{\phi}$.

In the PET problem described above, the sample space is the n-fold product of the

9

set of lines through S. The parameter set can be taken to index a subset, $\mathscr{P}$, of the set, $P(S)$, of probability measures on S. If the probability measure μ is indexed by θ, the probability measure of the observation under θ is given by the n-fold product of $T\mu$. The major choices in selecting a statistical model for the PET problem are the choice of $\mathscr{P}$ (along with an appropriate parameterization) and the choice of a loss function.

In most statistical models for PET, μ is assumed to belong to some predetermined finite-dimensional family of probability measures. Such models are termed "parametric". Most commonly, the domain of interest is "pixelized", i.e., divided into a finite number of squares. It is then assumed that the probability density is constant with respect to Lebesgue measure on each square (e.g., [VSK85, sec. 1.2]). Another possible parametric approach, based on a singular value decomposition of the tomography process, is described in [Bak91]. The advantage of parametric models is that one can use standard methods of parametric statistics to develop and evaluate estimators (e.g., maximum likelihood estimators and Cramér-Rao bounds). The main disadvantage of parametric models in PET is that, in practice, the actual μ is unlikely to conform to the assumed model, creating an error which falls outside the assumed statistical framework. Moreover, many natural questions about the PET problem, such as how potential spatial resolution increases with n, are difficult to formulate in a natural way using parametric models.

The preceding paragraph should not be interpreted as disparaging parametric models for PET in general. Many PET systems are constructed using a finite number of detectors with no intrinsic spatial resolution. In such a system, the observation space is inherently finite-dimensional. As a consequence, only a finite-dimensional space of models are distinguishable and a parametric model is very reasonable. (A good description of this situation may be found in [Bak91].) The only claim intended is that in PET systems with continuous spatial resolution, as in our mathematical model, parametric models may not extract all of the available information.

The limitations of parametric models may be transcended by allowing μ to lie in an infinite-dimensional family of probability measures. Such models are termed

"nonparametric". Let $\Omega \subset \mathbb{R}^2$ denote the open unit disk and $\bar{\Omega}$ its closure. Johnstone and Silverman [JS90] have recently described a statistical model where $\mathscr{P}$ was taken to be the subset of $P(\bar{\Omega})$ consisting of probability measures represented by probability density functions (with respect to normalized Lebesgue measure) which lie in a certain ellipsoid in $L^2(\bar{\Omega})$, the space of (equivalence classes of) square-integrable functions on $\bar{\Omega}$. The condition of lying in such an ellipsoid is essentially a smoothness and integrability constraint. If $\mu \in \mathscr{P}$, let f_μ denote the probability density associated with μ. Let $\hat{f}_n$ be an estimator of f_μ based on n independent observations distributed according to $T\mu$, with $\hat{\mu}_n$ denoting the probability measure corresponding to $\hat{f}_n$. The loss, $W_\mu(\hat{\mu}_n)$, was taken to be the squared L^2 distance between $\hat{f}_n$ and f_μ, resulting in a risk function given by $r(\mu, \hat{\mu}_n) = E_{(T\mu)^n} \|\hat{f}_n - f_\mu\|^2_{L^2(\bar{\Omega})}$, where $E_{(T\mu)^n}$ denotes expectation with respect to the n-fold product of $T\mu$. The overall difficulty of the problem was then assessed in terms of minimax risk. Johnstone and Silverman were able to obtain quite precise characterizations of the functional form of the dependence of the minimax risk on n. However, the results were found to depend significantly on the specific smoothness and integrability assumptions that were made. Since these assumptions would be difficult to verify in practice, the application of the results is problematic. Moreover, these assumptions were, in all cases, quite restrictive, e.g., they implied the density of interest always took values in the range $(0, 2)$.

In this thesis, we will explore a different nonparametric approach to the PET problem. We will allow μ to be an arbitrary probability measure on S, i.e., we will take $\mathscr{P} = P(S)$, and use loss functions which are based upon L^2-Sobolev metrics of order < -1 for probability measures (see definition 3.2.1). The significance of this change in loss function will be discussed in greater detail in the next section. Using this approach, we will see that it is possible to characterize the minimax risk as a function of n without the use of smoothness assumptions. In particular, we shall show that the minimax risk is $O(n^{-1})$ (i.e., bounded above by cn^{-1} for some constant c) for the loss function generated by the squared Sobolev norm of order $-\alpha$ if and only if $\alpha > 3/2$. (The significance of the $O(n^{-1})$ rate is that this is this rate is typically achieved for squared error loss in parametric models.) It is useful to

compare this rate with the one for the problem of estimating μ given n observations distributed according to μ itself. For this problem, the $O(n^{-1})$ rate is obtained if and only if $\alpha > 1$. This result, in a sense, quantifies the relative difficulty of estimating μ from observations distributed according to μ and $T\mu$. In many PET problems it is difficult to characterize the smoothness properties of the probability measure μ, so it is desirable to develop statistical methodologies, such as these, that require few assumptions about μ.

1.2 Loss functions for estimation of probability measures

In this section, we consider loss functions for the PET problem in the context of probability measure estimation problems in general. We start by recalling the definitions of some metrics and notions of convergence on spaces of probability measures.

Definition 1.2.1 For probability measures μ and μ' on a measurable space $(X, \mathscr{A})$, the variational distance between μ and μ' is defined by

$$d_v(\mu, \mu') \stackrel{\text{def}}{=} \sup_{A \in \mathscr{A}} |\mu(A) - \mu'(A)|$$

[Str85, def. 2.1]. If μ and μ' can be represented with respect to the σ-finite measure λ by the probability density functions f and f', respectively, then $d_v(\mu, \mu') = \frac{1}{2} \int_X |f - f'| \, d\lambda$. That is, $d_v(\mu, \mu')$ is proportional to the L^1 distance between f and f' [Str85, lem. 2.4]. Convergence of a sequence of probability measures with respect to variational distance will be called **strong convergence**, and any metric that metrizes strong convergence will be called a **strong metric**. Strong convergence is sometimes called convergence in total variation in the literature [Dud89, p. 228]. An important property of variational distance is that it is invariant under a bijective change of coordinates on the underlying probability space.

Definition 1.2.2 Let X be a topological space, $\mathscr{B}$ the σ-algebra of Borel subsets of X, and $C_b(X)$ the set of bounded, continuous, real-valued functions on X. The sequence $\{\mu_n\}$ of probability measures on X is said to converge weakly to the probability measure μ if $\int f \, d\mu_n \to \int f \, d\mu$ for all $f \in C_b(X)$ [HQ85, p. 361]. Any metric that metrizes weak convergence of probability measures will be called a **weak metric**.

Remark 1.2.3 There are a number of weak metrics for probability measures used in the literature. One example is the Prohorov metric. Let X be a metric space. If $B \in \mathscr{B}$, define

$$B^\epsilon \stackrel{\text{def}}{=} \{y \in X : d(x, y) < \epsilon \text{ for some } x \in B\},$$

13

i.e., B^ϵ is the set of all points in X whose distance from B is less than ϵ. Then the Prohorov metric, defined by

$$\rho(\mu,\mu') \overset{\text{def}}{=} \inf\{\epsilon > 0 : \mu(B) \leq \mu'(B^\epsilon) + \epsilon \text{ for all } B \in \mathscr{B}\},$$

is a weak metric for probability measures on X [Dud89, thm. 11.3.3]. If X is a compact Riemannian manifold of dimension d, then L^2 Sobolev metrics of order $< -d/2$ are weak metrics for probability measures on X [Gin75, thm. 2.2]. Essentially the same proof shows that L^2 Sobolev metrics of order $< -d/2$ are weak metrics for probability measures on a fixed compact set of $\mathbb{R}^d$.

Remark 1.2.4 The mode of convergence which we call weak convergence has many other names in the literature. It has been called convergence in distribution [Pol84, p. 43], convergence in law [Dud89, p. 229], weak-star convergence [Gin75, p. 1245], and vague convergence [Chu74, thm. 4.4.2].

There is a large literature on probability density estimation problems. In these problems, one is attempting to estimate an unknown probability density, $f \in \mathscr{F}$, given n independent observations distributed according to the probability density $g = Tf$, where T is some deterministic function on $\mathscr{F}$ (often T is just the identity function). These problems may be viewed as probability measure estimation problems where the unknown probability measure is assumed to be representable as a probability density, $f \in \mathscr{F}$. Most commonly, the loss functions for these density estimation problems are based on the L^1 or L^2 distances between the densities in $\mathscr{F}$. As noted in definition 1.2.1, convergence of density functions with respect to the L^1 metric is equivalent to strong convergence of the associated probability measures. Under additional conditions, (e.g., density functions uniformly bounded above and below away from 0 and a finite common dominating measure, as in [JS90]) convergence of density functions with respect to the L^2 metric is also equivalent.

In many density estimation problems, it is possible to construct consistent estimators, i.e., estimators whose risk (based on the L^1 or L^2 metric) converges to 0 as

$n \to \infty$ for each $f \in \mathcal{F}$. For example, in the case where the observations are distributed according the density of interest, there are easily checked conditions which ensure the consistency of kernel estimators with respect to the loss function generated by the L^1 norm for arbitrary density functions [Dev87, thm. 3.1]. For the more difficult problem where the observations come from the convolution of the density of interest with some noise density whose characteristic function is positive almost everywhere, it is also possible to construct estimators which are consistent with respect to the loss function generated by the L^1 norm [Dev89]. However, it is generally not possible to construct estimators whose maximum risk converges to zero as $n \to \infty$ without substantially restricting the set of possible density functions. For example, consider the problem of estimating an unknown probability density, f, on $\mathbb{R}$, from n independent observations distributed according to f. It has been shown that the minimax risk with respect to the L^1 loss function is bounded below by 1 if f is allowed to range over the set of densities with support on $[0, 1]$ which are bounded above by 2 or if f is allowed to range over the set of densities on $\mathbb{R}$ with infinitely many derivatives which are uniformly bounded by some sequence of constants [Dev87, sec 5.3]. The moral is that some combination of tail conditions and smoothness conditions are necessary to ensure a minimax error that converges to 0 as $n \to \infty$.

In what follows, we will explore a different approach to the problem of probability measure estimation. On one hand, we will allow the unknown probability measure μ to be an arbitrary probability measure on the relevant measure space. On the other hand, we will use a loss function based on a weak, rather than a strong, metric for probability measures. Using this approach, we will see that is it possible to characterize the rate at which the minimax risk converges to zero.

The obviation of smoothness constraints honors the spirit of nonparametric statistics and is thus an advantage of our approach. But what price, if any, do we pay for the change in loss functions? Strong metrics are clearly stricter than those based on weak metrics in the sense that convergence with respect to the former implies convergence with respect to the latter, but not vice versa. However, stricter does not necessarily imply more relevant in a physical application. Roughly, the difference

between the two classes of metrics is that weak metrics consider measures which live on disjoint sets which are close together in the underlying space to be close together, while strong metrics see them as far apart. An example may make this clearer. Consider measures on $\mathbb{R}$ with respect to the usual Borel σ-algebra and let δ_x denote a point mass at x. Then the sequence of measures $\{\delta_{1/n}\}$ converges to δ_0 weakly, but not strongly (note that the variational distance between $\delta_{1/n}$ and δ_0 remains fixed at 1). Which risk function is "better" in a particular application thus depends on whether or not "close counts" in that application. For general applications in PET and many other areas, it seems to the author that "close counts" and hence that loss functions based on weak metrics for probability measures are appropriate tools for studying these problems.

1.3 Notation

In this work, we shall often be concerned with estimating a probability measure on a subset of $\mathbb{R}^2$ from independent observations distributed according to a probability measure on the set of lines in $\mathbb{R}^2$, which we shall denote by $G_{1,2}$. We shall generally denote these respective probability measures by μ and ν, and (when they exist) the corresponding probability density functions by f and g. Distributions on these spaces will usually be denoted by u and v, respectively.

We will use the notations $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ to denote the set of complex numbers, real numbers, integers, and natural numbers (including 0), respectively. The notations $\mathbb{R}^+$ and $\mathbb{N}^+$ will denote the set of positive real numbers and natural numbers, respectively. The imaginary unit will be denoted by i. The notation $\mathbb{R}^d$ will be used to denote d-dimensional Euclidean space. The inner product of $x, y \in \mathbb{R}^d$ will be denoted by $x \cdot y$.

We will make frequent use of the so-called big oh and little oh notation [Apo74, sec. 8.13]. If $\{a_n\}$ and $\{b_n\}$ are two sequences with $b_n \geq 0$, we say $a_n = O(b_n)$ if there exists a constant $M > 0$ such that $|a_n| \leq M b_n$ for all n. We say $a_n = o(b_n)$ if $a_n / b_n \to 0$ as $n \to \infty$. If $I \subset \mathbb{R}^+$ and $f, g : I \to \mathbb{R}$ and $g \geq 0$, we say $f = O(g)$ if there exists a constant $M > 0$ such that $|f| \leq Mg$ on I.

We now briefly review some standard notation used for function spaces. If (X, μ) is a measure space and $1 \leq p < \infty$, we shall denote the set of measurable complex-valued functions f such that $||f||_{L^p(X,\mu)} \stackrel{\text{def}}{=} (\int_X |f|^p \, d\mu)^{1/p} < \infty$ by $\mathscr{L}^p(X, \mu)$. The usual Banach space obtained by identifying elements in $\mathscr{L}^p(X, \mu)$ whose distance is 0 with respect to $||\cdot||_{L^p(X,\mu)}$ is denoted by $L^p(X, \mu)$. The spaces $\mathscr{L}^\infty(X, \mu)$ and $L^\infty(X, \mu)$ are defined analogously in the usual way (see, e.g., [Dud89, secs. 5.1, 5.2]). The set of bounded measurable complex-valued functions on X will be denoted by $\mathscr{L}^b(X, \mu)$. The uniform norm on $\mathscr{L}^b(X, \mu)$ will be denoted by $||f||_u$, i.e., $||f||_u \stackrel{\text{def}}{=} \sup_{x \in X} |f(x)|$.

As mentioned previously, the set of probability measures on $(X, \mathscr{A})$ will be denoted by $P(X, \mathscr{A})$. The set of (finite) signed measures on $(X, \mathscr{A})$ will be denoted by $M(X, \mathscr{A})$. If X is a topological space, $\mathscr{A}$ will be understood to be the Borel σ-

algebra and we will simply write $P(X)$ and $M(X)$.

The set of continuous real-valued functions on a topological space, X, is denoted by $C(X)$. The subset of $C(X)$ whose elements have compact support will be denoted by $C_c(X)$. The subset of $C(X)$ whose elements are bounded will be denoted by $C_b(X)$. If X is a differentiable manifold, the set of infinitely differentiable functions on X will be denoted by $C^\infty(X)$.

The index contains a list of notations and terminology. The notations are listed pseudoalphabetically, e.g., $\|\cdot\|_u$ is listed under "n" for norm.

Chapter 2

The Radon Transform

In this chapter, we shall show that the observations in the mathematical model of the PET problem described in section 1.1 are distributed according to the (appropriately scaled) Radon transform of the probability measure of interest. In section 2.1, we define the Radon transform for functions on $\mathbb{R}^2$. The extension of the Radon transform to measures and distributions on $\mathbb{R}^2$ is defined in section 2.2, and its role in the PET model is discussed in section 2.3.

2.1   The Radon transform on functions

Our treatment of the Radon transform parallels that of Hertle [Her83]. More comprehensive treatments of the Radon transform may be found in [Dea83] and [Nat86]. For simplicity, we shall only consider the Radon transform on $\mathbb{R}^2$.

Essentially, the Radon transform of a real-valued function f on $\mathbb{R}^2$ is the real-valued function on the set of lines in $\mathbb{R}^2$ whose value on a given line is equal to the integral of f over that line. We shall now specify suitable coordinate systems for the set of lines in $\mathbb{R}^2$ and proceed to formalize the definition of the Radon transform.

Definition 2.1.1 Let S^1 denote the unit sphere in $\mathbb{R}^2$. We will give S^1 the structure of a differentiable manifold. (We assume familiarity with the basic notions regarding differentiable manifolds. For a reader unfamiliar with these notions, we can recommend [AMR83] as a general reference.) We define charts on S^1 as follows. Define the

map $\omega : \mathbb{R} \to S^1$ by $\theta \mapsto (\cos\theta, \sin\theta)$. For $\theta \in \mathbb{R}$, define the chart domains $\Omega_\theta \subset S^1$ by $\Omega_\theta \stackrel{\text{def}}{=} S^1 \backslash \{-\omega(\theta)\}$. There is a unique map $\omega_\theta^{-1} : \Omega_\theta \to (\theta - \pi, \theta + \pi)$ which is the left inverse to ω on $(\theta - \pi, \theta + \pi)$. It is clear that $(\Omega_\theta, \omega_\theta^{-1})$ is a chart on S^1 and that the collection of such charts gives an atlas for S^1. Note that for any $\theta \in \mathbb{R}$ there is a unique point in S^1, namely $\omega(\theta)$, that has the coordinate θ in any of the above charts. Thus while a point in S^1 may have different coordinates in different charts (differing by integer multiples of 2π), a given coordinate θ specifies the unique point $\omega(\theta) \in S^1$. It will sometimes be convenient, by a slight abuse of language, to refer to points in S^1 as $\theta \in S^1$ for some $\theta \in \mathbb{R}$. In particular, we will sometimes denote a function on S^1 by $\theta \mapsto f(\theta)$, where f is a function on $\mathbb{R}$ which is periodic with period 2π, e.g., $e^{-i2\pi\theta}$.

Definition 2.1.2 The atlas $\{(\Omega_\theta, \omega_\theta^{-1})\}_{\theta \in \mathbb{R}}$ gives rise to a global differential operator on S^1 in a natural way. Namely, the pull-back of the differential operator ∂_x on $(\theta - \pi, \theta + \pi)$ under the map ω_θ^{-1} is a differential operator on Ω_θ. It is easy to verify that these differential operators agree on the overlaps of their domains and thus give rise to a globally defined differential operator on S^1. We will denote this differential operator by ∂_θ. By duality, we then obtain the differential 1-form $d\theta$ on S^1.

Definition 2.1.3 If $f \in C^\infty(S^1)$, then it is not difficult to see that the integral of f with respect to the volume form $d\theta$, $\int_{S^1} f\, d\theta$, satisfies

$$\int_{S^1} f\, d\theta = \int_{-\pi}^{\pi} f[\omega(\theta)]\, d\theta,$$

where the integral on the right-hand side is an ordinary Riemann integral. This integral can clearly be extended to all $f \in C(S^1)$. By the Riesz representation theorem, there exists a unique Borel measure σ on S^1 such that $\int f\, d\theta = \int f\, d\sigma$ for all $f \in C(S^1)$ [AMR83, thm. 7.1.9]. We will call σ surface measure on S^1.

Definition 2.1.4 The projective space $\mathbb{P}^1$ is the quotient space of S^1 obtained by identifying each point of S^1 with its antipodal point. We will denote the natural projection map by $p : S^1 \to \mathbb{P}^1$. Since the lines through the origin generated by two

distinct elements of S^1 are identical if and only if the elements of S^1 are antipodal, we can identify $\mathbb{P}^1$ with the set of lines through the origin in the obvious way. We will give $\mathbb{P}^1$ the structure of a differentiable manifold. We define charts on $\mathbb{P}^1$ as follows. For $\theta \in \mathbb{R}$, define the chart domains $\bar{\Omega}_\theta \subset \mathbb{P}^1$ by $\bar{\Omega}_\theta \stackrel{\text{def}}{=} \mathbb{P}^1 \backslash \{p[\omega(\theta + \pi/2)]\}$. Then define $\bar{\omega}_\theta^{-1} : \bar{\Omega}_\theta \to (\theta - \pi/2, \theta + \pi/2)$ to be the (unique) left inverse to $p \circ \omega$ on $(\theta - \pi/2, \theta + \pi/2)$. It is clear that $(\bar{\Omega}_\theta, \bar{\omega}_\theta^{-1})$ is a chart on $\mathbb{P}^1$ and that the collection of such charts gives an atlas for $\mathbb{P}^1$. Note that for any $\theta \in \mathbb{R}$ there is a unique point in $\mathbb{P}^1$, namely $p[\omega(\theta)]$, that has θ as a coordinate in any of the above charts. Thus while a point in $\mathbb{P}^1$ may have different coordinates in different charts (differing by integer multiples of π), a given coordinate specifies a unique point in $\mathbb{P}^1$. $\mathbb{P}^1$ is the quotient manifold of S^1 with respect to the equivalence relation generated by p (cf. [AMR83, p. 173]).

Definition 2.1.5 We will denote the affine Grassmann manifold of lines in $\mathbb{R}^2$ by $G_{1,2}$ [Gon91]. We will give $G_{1,2}$ the structure of a vector bundle over $\mathbb{P}^1$. (A discussion of notions related to vector bundles may be found in [AMR83, sec. 3.3].) Let $\theta_0 \in \mathbb{R}$. Define $\Gamma_{\theta_0} \subset G_{1,2}$ to be the subset of lines in $\mathbb{R}^2$ which are *not* parallel to the line generated by the vector $\omega(\theta_0)$. For each line $l \in \Gamma_{\theta_0}$, there is a unique $\theta \in (\theta_0 - \pi/2, \theta_0 + \pi/2)$ such that the line generated by the vector $\omega(\theta)$ is perpendicular to l. It is easy to see that all the points on l have the same inner product with $\omega(\theta)$. Denoting this inner product by s, we assign the line l the coordinates (θ, s). Denoting the resulting map by γ_{θ_0}: $\Gamma_{\theta_0} \to (\theta_0 - \pi/2, \theta_0 + \pi/2) \times \mathbb{R}$, it is easy to verify that the pair $(\Gamma_{\theta_0}, \gamma_{\theta_0})$ is a local bundle chart on $G_{1,2}$ and that the collection of these local bundle charts for $\theta_0 \in \mathbb{R}$ is a vector bundle atlas for $G_{1,2}$. It is also easy to see that the zero section of this vector bundle consists of the set of lines through the origin in $\mathbb{R}^2$, i.e., $\mathbb{P}^1$. It will be convenient to establish the convention of referring to points in $G_{1,2}$ as $(\theta, s) \in G_{1,2}$, where the coordinates (θ, s) are understood to refer to the chart $(\Gamma_\theta, \gamma_\theta)$. Under this convention, $(\theta, s) \in G_{1,2}$ is simply the line in $\mathbb{R}^2$ through the point $s\omega(\theta)$ which is perpendicular to $\omega(\theta)$, i.e., the line in $\mathbb{R}^2$ whose points x satisfy $x \cdot \omega = s$. It is perhaps worth noting that the other possible coordinates of a point $(\theta, s) \in G_{1,2}$ under this convention are precisely those of the form $(\theta + 2k\pi, s)$

or $(\theta + (2k+1)\pi, -s)$, for integer k. In particular, note that $(-\theta, -s) = (\theta, s)$ as points in $G_{1,2}$.

Definition 2.1.6 We now define the standard double covering of $G_{1,2}$ by the product space $S^1 \times \mathbb{R}$. If $(\omega, s) \in S^1 \times \mathbb{R}$, we map (ω, s) to the line in $G_{1,2}$ which is perpendicular to the line generated by the vector ω and whose points have inner product s with ω. We will denote this map by $\pi : S^1 \times \mathbb{R} \to G_{1,2}$. It is easy to see that this map gives a double covering of $G_{1,2}$ by $S^1 \times \mathbb{R}$ and that the points $(\omega, s) \in S^1 \times \mathbb{R}$ and $(-\omega, -s) \in S^1 \times \mathbb{R}$ have the same image in $G_{1,2}$ under this covering.

Remark 2.1.7 It is not difficult to convince oneself that $G_{1,2}$, when equipped with the vector bundle structure described above, is *not* vector bundle isomorphic to the trivial vector bundle $\mathbb{P}^1 \times \mathbb{R}$ over $\mathbb{P}^1$, but rather is vector bundle isomorphic to the Möbius band, cf. [AMR83, pp. 139, 142]. The double covering of $G_{1,2}$ by $S^1 \times \mathbb{R}$ is just the usual orientable double covering of a nonorientable manifold, cf. [AMR83, p. 385].

Definition 2.1.8 For any function $g : S^1 \times \mathbb{R} \to \mathbb{R}$, define its flip, $\breve{g} : S^1 \times \mathbb{R} \to \mathbb{R}$, by

$$\breve{g}(\omega, s) \stackrel{\text{def}}{=} g(-\omega, -s),$$

its drop, $\breve{g} : G_{1,2} \to \mathbb{R}$, by

$$\breve{g}(\theta, s) \stackrel{\text{def}}{=} \{g[\omega(\theta), s] + \breve{g}[\omega(\theta), s]\}, \tag{2.1}$$

and its average, $\bar{g} : G_{1,2} \to \mathbb{R}$, by

$$\bar{g}(\theta, s) \stackrel{\text{def}}{=} \frac{1}{2} \breve{g}(\theta, s). \tag{2.2}$$

We say that g is even if $\breve{g} = g$.

Definition 2.1.9 If $g : G_{1,2} \to \mathbb{R}$, then the standard (double) covering map $\pi : S^1 \times \mathbb{R} \to G_{1,2}$ induces the even pull-back function $\tilde{g} \stackrel{\text{def}}{=} g \circ \pi : S^1 \times \mathbb{R} \to \mathbb{R}$. We shall call $\tilde{g}$ the lift of g to $S^1 \times \mathbb{R}$.

22

Definition 2.1.10 Let λ^d denote Lebesgue measure on $\mathbb{R}^d$. We will take the product measure λ^d to be the standard measure on $\mathbb{R}^d$ in the sense that $L^p(\mathbb{R}^d) \overset{\text{def}}{=} L^p(\mathbb{R}^d, \lambda^d)$. Similarly, we shall take the standard measures on $S^1 \times \mathbb{R}$ and $G_{1,2}$ to be $\sigma \times \lambda^1$ and the image measure $\frac{1}{2}(\sigma \times \lambda^1) \circ \pi^{-1}$, respectively.

Remark 2.1.11 Suppose $g \in L^1(S^1 \times \mathbb{R})$. Then we have the explicit formula

$$\int_{S^1 \times \mathbb{R}} g \, d(\sigma \times \lambda^1) = \int_{-\pi}^{\pi} \int_{\mathbb{R}} g(\omega(\theta), s) \, ds \, d\theta,$$

where the right-hand side is an ordinary double integral over a rectangle with respect to Lebesgue measure. Similarly, if $g \in L^1(G_{1,2})$, we have the explicit formula

$$
\begin{aligned}
\frac{1}{2} \int_{G_{1,2}} g \, d\left[(\sigma \times \lambda^1) \circ \pi^{-1}\right] &= \frac{1}{2} \int_{S^1 \times \mathbb{R}} \tilde{g} \, d(\sigma \times \lambda^1) \\
&= \int_{-\pi/2}^{\pi/2} \int_{\mathbb{R}} g(\theta, s) \, ds \, d\theta,
\end{aligned}
$$

since $\tilde{g}$ is even. In light of this formula, we shall often refer to the measure $\frac{1}{2}(\sigma \times \lambda^1) \circ \pi^{-1}$ on $G_{1,2}$, by an abuse of language, as $ds\, d\theta$.

Remark 2.1.12 We are now ready to define the Radon transform. In essence, the Radon transform of a real-valued function $f : \mathbb{R}^2 \to \mathbb{R}$ is the real-valued function on $G_{1,2}$ whose value on a given line $l \in G_{1,2}$ is equal to the integral of f over l. We will denote the Radon transform of f by $Rf : G_{1,2} \to \mathbb{R}$. It is, however, more standard to define the Radon transform to be the real-valued function on $S^1 \times \mathbb{R}$ which is the pullback of the Radon transform just defined with respect to the standard covering map π. The resulting Radon transform has a redundancy due to the fact that each point in $G_{1,2}$ is represented by two points in $S^1 \times \mathbb{R}$. While working on $S^1 \times \mathbb{R}$ is convenient in most (deterministic) contexts, in our probabilistic context it is sometimes better to avoid this redundancy by working directly on $G_{1,2}$. In particular, we shall later be considering samples of random variables that, from a physical point of view, live on $G_{1,2}$ rather than $S^1 \times \mathbb{R}$. There is, however, a price to be paid. Namely, to use any of the previously obtained results on the Radon transform, we will have to go back and forth between the two definitions. We shall refer to the Radon transform which

gives functions on $S^1 \times \mathbb{R}$ as the standard Radon transform and the transform which gives functions on $G_{1,2}$ simply as the Radon transform.

Definition 2.1.13 The Radon transform on $L^1(\mathbb{R}^2)$, which we shall denote by R, maps $f \in L^1(\mathbb{R}^2)$ to the function $Rf : G_{1,2} \to \mathbb{R}$ whose value at $l \in G_{1,2}$ is equal to the integral of f with respect to Lebesgue measure on l. The *standard* Radon transform on $L^1(\mathbb{R}^2)$, which we shall denote by $\tilde{R}$, maps $f \in L^1(\mathbb{R}^2)$ to the lift of Rf to $S^1 \times \mathbb{R}$, i.e., $\tilde{R}f \stackrel{\text{def}}{=} \widetilde{Rf}$. It is shown in [Her83, p. 168] that $\tilde{R}f \in L^1(S^1 \times \mathbb{R})$. It follows at once that $Rf \in L^1(G_{1,2})$. Using the coordinate system for $G_{1,2}$ described in definition 2.1.5, we can give an explicit expression for the Radon transform. We will denote the unit vector $(-\sin\theta, \cos\theta)$, obtained by rotating $\omega(\theta)$ by a angle of $\pi/2$ in the counterclockwise direction, by $\omega^\perp(\theta)$. $Rf(\theta, s)$ is equal to the integral of f over the line whose points have inner product s with $\omega(\theta)$, i.e.,

$$
\begin{aligned}
Rf(\theta, s) &= \int_{-\infty}^{\infty} f(s\omega(\theta) + t\omega^\perp(\theta))\, dt \\
&= \int_{-\infty}^{\infty} f(s\cos\theta - t\sin\theta, s\sin\theta + t\cos\theta)\, dt.
\end{aligned}
$$

We next describe the adjoint of R.

Proposition 2.1.14 *The map* $\tilde{R}^* : L^\infty(S^1 \times \mathbb{R}) \to L^\infty(\mathbb{R}^2)$ *given by*

$$
\tilde{R}^* g(x) = \int_{S^1} g(\omega, x \cdot \omega)\, d\sigma(\omega) \tag{2.3}
$$

for $g \in L^\infty(S^1 \times \mathbb{R})$ *is the adjoint of* $\tilde{R} : L^1(\mathbb{R}^2) \to L^1(S^1 \times \mathbb{R})$ *in the sense that*

$$
\int_{S^1 \times \mathbb{R}} \tilde{R}f\, g\, d(\sigma \times \lambda^1) = \int_{\mathbb{R}^2} f\, \tilde{R}^* g\, d\lambda^2
$$

for all $f \in L^1(\mathbb{R}^2)$ *and* $g \in L^\infty(S^1 \times \mathbb{R})$. *The map* $R^* : L^\infty(G_{1,2}) \to L^\infty(\mathbb{R}^2)$ *given by*

$$
\begin{aligned}
(R^* g)(x) &= \int_0^\pi g(\theta, x \cdot \omega(\theta))\, d\theta \\
&= \frac{1}{2} \int_{S^1} \tilde{g}(\omega, x \cdot \omega)\, d\sigma(\omega) \\
&= \frac{1}{2} \tilde{R}^* \tilde{g}(x)
\end{aligned} \tag{2.4}
$$

24

for $g \in L^{\infty}(G_{1,2})$ *is the adjoint of* $R : L^1(\mathbb{R}^2) \to L^1(G_{1,2})$ *in the sense that*

$$\int_{G_{1,2}} Rf\, g\, ds\, d\theta = \int_{\mathbb{R}^2} f\, R^*g\, d\lambda^2$$

for all $f \in L^1(\mathbb{R}^2)$ *and* $g \in L^{\infty}(G_{1,2})$.

Proof.   The result for $\tilde{R}$ is given in [Her83, p. 169, eq. 1.1]. To prove the result for R, suppose $f \in L^1(\mathbb{R}^2)$ and $g \in L^{\infty}(G_{1,2})$. Define $\tilde{g} \in L^{\infty}(S^1 \times \mathbb{R})$ as in definition 2.1.9. Then, using equation 2.4,

$$
\begin{aligned}
\int_{\mathbb{R}^2} f\, R^*g\, d\lambda^2 &= \frac{1}{2} \int_{\mathbb{R}^2} f\, \tilde{R}^*\tilde{g}\, d\lambda^2 \\
&= \frac{1}{2} \int_{S^1 \times \mathbb{R}} \tilde{R}f\, \tilde{g}\, d(\sigma \times \lambda^1) \\
&= \frac{1}{2} \int_{G_{1,2}} Rf\, g\, d[(\sigma \times \lambda^1) \circ \pi^{-1}] \\
&= \int_{G_{1,2}} Rf\, g\, ds\, d\theta. \quad \square
\end{aligned}
\tag{2.5}
$$

Remark 2.1.15 We note, for future reference, that $\tilde{R}^* : C_0^{\infty}(S^1 \times \mathbb{R}) \to C_0^{\infty}(\mathbb{R}^2)$ [Her83, prop. 1.2(c)]. From this, it follows easily that $R^* : C_0^{\infty}(G_{1,2}) \to C_0^{\infty}(\mathbb{R}^2)$.

We conclude this section with a simple identity which relates the adjoints of R and $\tilde{R}$.

Lemma 2.1.16 *If* $g \in L^{\infty}(S^1 \times \mathbb{R})$, *we have the identity*

$$R^*\check{g} = \tilde{R}^*g,$$

where $\check{g} \in L^{\infty}(G_{1,2})$ *is defined as in equation 2.1.*

Proof. Using equation 2.4, we have

$$
\begin{aligned}
R^*\check{g}(x) \;\; &= \frac{1}{2}\tilde{R}^*\widetilde{g}(x) \\
&= \frac{1}{2}\int_{-\pi}^{\pi} \widetilde{g}(\omega(\theta), x \cdot \omega(\theta))\, d\theta \\
&= \frac{1}{2}\int_{-\pi}^{\pi} \check{g}(\theta, x \cdot \omega(\theta))\, d\theta \\
&= \frac{1}{2}\int_{-\pi}^{\pi} g(\omega(\theta), x \cdot \omega(\theta)) + g(-\omega(\theta), -x \cdot \omega(\theta))\, d\theta \\
&= \int_{S^1} g(\omega, x \cdot \omega)\, d\sigma(\omega) \\
&= \tilde{R}^*g(x). \;\; \square
\end{aligned}
$$

2.2 The Radon transform on distributions and measures

In this section, we will develop the Radon transform on distributions and measures. We start by defining the relevant notions from the theory of distributions. The treatment here will be very brief; we can recommend [Tre67] as a general reference.

Definition 2.2.1 A multi-index is an ordered d-tuple of nonnegative integers. If $\beta = (\beta_1, \ldots, \beta_d)$ is a multi-index and $x \in \mathbb{R}^d$, we define $|\beta| \overset{\text{def}}{=} \beta_1 + \cdots \beta_n$, $x^\beta \overset{\text{def}}{=} x_1^{\beta_1} \cdots x_d^{\beta_d}$, and $\partial^\beta \overset{\text{def}}{=} \partial_{x_1}^{\beta_1} \cdots \partial_{x_d}^{\beta_d}$.

Definition 2.2.2 Let X be a d-dimensional (C^∞) differentiable manifold with atlas $\{(U_\alpha, \psi_\alpha)\}$. A function $f \in C(X)$ is said to vanish at ∞ if, for all $\epsilon > 0$, the set $\{x \in X : |f(x)| \geq \epsilon\}$ is compact [Fol84, pp. 125-6]. Define $C_0(X)$ to be the subset of $C(X)$ whose elements vanish at ∞. (In the literature, this space is sometimes denoted by C_∞ [Itô85, art. 168B].) Define $C_0^\infty(X)$ to be the subset of $C_0(X)$ whose elements have derivatives of all orders that are also in $C_0(X)$. (In the literature, this space is sometimes denoted by $\mathscr{D}$ [Itô85, art. 125N].) We equip $C_0^\infty(X)$ with the locally convex topology induced by the seminorms $\|\partial^\beta(f \circ \psi^{-1})\|_u$ for all multi-indices β and charts (U_α, ψ_α) in an atlas for X.

Definition 2.2.3 The dual space to $C_0^\infty(X)$, i.e., the space of continuous linear functionals on $C_0^\infty(X)$, is termed the space of integrable distributions and denoted by $\mathscr{D}_{L_1}'(X)$ [Itô85, art. 125N]. We identify $L^1(\mathbb{R}^d)$, $L^1(S^1 \times \mathbb{R})$, and $L^1(G_{1,2})$ with subspaces of the appropriate space of integrable distributions in the obvious way. If $g \in C_0^\infty(S^1 \times \mathbb{R})$, then $\breve{g} \in C_0^\infty(G_{1,2})$. Thus if $v \in \mathscr{D}_{L_1}'(G_{1,2})$, we can define its lift $\tilde{v} \in \mathscr{D}_{L_1}'(S^1 \times \mathbb{R})$ by $\langle \tilde{v}, g \rangle \overset{\text{def}}{=} \langle v, \breve{g} \rangle$, where the notation $\langle u, f \rangle$ indicates the operation of applying the distribution u to the test function f. If $g \in C_0^\infty(G_{1,2})$, then $\tilde{g} \in C_0^\infty(S^1 \times \mathbb{R})$. Thus if $v \in \mathscr{D}_{L_1}'(S^1 \times \mathbb{R})$, we can define its drop $\breve{v} \in \mathscr{D}_{L_1}'(G_{1,2})$ by $\langle \breve{v}, g \rangle \overset{\text{def}}{=} \langle v, \tilde{g} \rangle$. We say that $v \in \mathscr{D}_{L_1}'(S^1 \times \mathbb{R})$ is even if $\langle v, \breve{g} \rangle = \langle v, g \rangle$ for all test functions g.

Example 2.2.4 Suppose $\gamma \in L^1(S^1 \times \mathbb{R})$. Then the drop of the distribution $g \mapsto \int_{S^1 \times \mathbb{R}} g\gamma \, d(\sigma \times \lambda^1)$ on $S^1 \times \mathbb{R}$ is given by the distribution $g \mapsto \int_{S^1 \times \mathbb{R}} \check{g}\gamma \, d(\sigma \times \lambda^1)$ on $G_{1,2}$. Now

$$
\begin{aligned}
\int_{S^1 \times \mathbb{R}} \check{g}\gamma \, d(\sigma \times \lambda^1) &= \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \check{g}(\omega(\theta), s)\gamma(\omega(\theta), s) \, ds \, d\theta \\
&= \int_0^{\pi} \int_{-\infty}^{\infty} g(\theta, s)\check{\gamma}(\theta, s) \, ds \, d\theta,
\end{aligned}
$$

so the drop of the distribution $g \mapsto \int_{S^1 \times \mathbb{R}} g\gamma \, d(\sigma \times \lambda^2)$ is the distribution $g \mapsto \int_{G_{1,2}} g\check{\gamma} \, ds \, d\theta$. In other words, the drop of γ considered as a function is consistent with the drop of γ considered as a distribution.

Example 2.2.5 Suppose ν is a finite measure on $S^1 \times \mathbb{R}$. Then the drop of ν is given by the distribution $g \mapsto \int_{S^1 \times \mathbb{R}} \check{g} \, d\nu$ on $G_{1,2}$. Now

$$
\begin{aligned}
\int_{S^1 \times \mathbb{R}} \check{g} \, d\nu &= \int_{S^1 \times \mathbb{R}} g \circ \pi \, d\nu \\
&= \int_{G_{1,2}} g \, d(\nu \circ \pi^{-1}),
\end{aligned}
$$

so the drop of ν is just the image measure of ν under the map π.

Example 2.2.6 Suppose $\gamma \in L^1(G_{1,2})$. Then the lift of the distribution $g \mapsto \int_{G_{1,2}} g\gamma \, ds \, d\theta$ on $G_{1,2}$ is given by the distribution $g \mapsto \int_{G_{1,2}} \check{g}\gamma \, ds \, d\theta$ on $S^1 \times \mathbb{R}$. Now

$$
\begin{aligned}
\int_{G_{1,2}} \check{g}\gamma \, ds \, d\theta &= \int_0^{\pi} \int_{-\infty}^{\infty} \check{g}(\theta, s)\gamma(\theta, s) \, ds \, d\theta \\
&= \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} g(\omega(\theta), s)\check{\gamma}(\omega(\theta), s) \, ds \, d\theta \\
&= \int_{S^1 \times \mathbb{R}} g\check{\gamma} \, d(\sigma \times \lambda^1),
\end{aligned}
$$

so the lift of the distribution $g \mapsto \int_{G_{1,2}} g\gamma \, ds \, d\theta$ is the distribution $g \mapsto \int_{S^1 \times \mathbb{R}} g\check{\gamma} \, d(\sigma \times \lambda^1)$. In other words, the lift of γ considered as a function is consistent with the lift of γ considered as a distribution.

Definition 2.2.7 Suppose $u \in \mathscr{D}'_{L_1}(\mathbb{R}^2)$, i.e., u is a continuous linear functional on $C_0^\infty(\mathbb{R}^2)$. Then we define $Ru \in \mathscr{D}'_{L_1}(G_{1,2})$ by

$$\langle Ru, g \rangle \stackrel{\text{def}}{=} \langle u, R^* g \rangle$$

for any test function $g \in C_0^\infty(G_{1,2})$. Similarly, we define $\tilde{R}u \in \mathscr{D}'_{L_1}(S^1 \times \mathbb{R})$ by

$$\langle \tilde{R}u, g \rangle \stackrel{\text{def}}{=} \langle u, \tilde{R}^* g \rangle$$

for any test function $g \in C_0^\infty(S^1 \times \mathbb{R})$. These definitions make sense by remark 2.1.15, and, by proposition 2.1.14, extend R and $\tilde{R}$ from $L^1(\mathbb{R}^2)$ to $\mathscr{D}'_{L_1}(\mathbb{R}^2)$. Moreover, since L^1 is dense in $\mathscr{D}'_{L_1}$, these extensions are unique [Her83, rem. 1.5].

Lemma 2.2.8 *The standard Radon transform on $\mathscr{D}'_{L_1}(\mathbb{R}^2)$ is even.*

Proof.   Let $g \in C_0^\infty(S^1 \times \mathbb{R})$. For each $x \in \mathbb{R}^2$, we have

$$\begin{aligned}
\tilde{R}^* \breve{g}(x) &= \int_{S^1} g(-\omega, -x \cdot \omega) \, d\sigma(\omega) \\
&= \int_{S^1} g(\omega, x \cdot \omega) \, d\sigma(\omega) \\
&= \tilde{R}^* g(x),
\end{aligned}$$

so $\tilde{R}^* \breve{g} = \tilde{R}^* g$. It follows that

$$\begin{aligned}
\langle \tilde{R}u, \breve{g} \rangle &= \langle u, \tilde{R}^* \breve{g} \rangle \\
&= \langle u, \tilde{R}^* g \rangle \\
&= \langle \tilde{R}u, g \rangle,
\end{aligned}$$

so $\tilde{R}u$ is even. $\square$

Lemma 2.2.9 *If $u \in \mathscr{D}'_{L_1}(\mathbb{R}^2)$, then $\tilde{R}u = \widetilde{Ru}$.*

Proof. If $g \in C_0^\infty(S^1 \times \mathbb{R})$, we have, using definition 2.2.2 and lemma 2.1.16,

$$\begin{aligned}
\langle \widetilde{Ru}, g \rangle &= \langle Ru, \check{g} \rangle \\
&= \langle u, R^*\check{g} \rangle \\
&= \langle u, \tilde{R}^*g \rangle \\
&= \langle \tilde{R}u, g \rangle. \quad \Box
\end{aligned}$$

Definition 2.2.10 Let $M(X)$ denote the space of (finite) signed measures on X. $M(X)$ is a Banach subspace of $\mathscr{D}'_{L_1}(X)$ with respect to the total variation norm. That is, if μ is a signed measure on X, we can interpret μ as the integrable distribution $f \mapsto \int_X f \, d\mu$, where $f \in C_0^\infty(X)$. Thus if μ is a signed measure on $\mathbb{R}^2$, the Radon transform of μ is given by the integrable distribution

$$R\mu : g \mapsto \int_{\mathbb{R}^2} R^*g \, d\mu \tag{2.6}$$

for $g \in C_0^\infty(G_{1,2})$ and similarly for the standard Radon transform of μ. It is shown in [Her83, p. 171, ex. 1] that $\tilde{R}\mu$ is actually a signed measure. An analogous argument shows that $R\mu$ is also a signed measure. Thus, by the Riesz representation theorem [Fol84, thm. 7.17], the integrable distribution $R\mu$ can be extended to a continuous linear functional on $C_0(G_{1,2})$. Moreover, since C_0^∞ is dense in C_0 [Sch66, p. 199], this extension is unique. It is clear that this extension is obtained simply by extending equation 2.6 to $g \in C_0(G_{1,2})$. If μ is a positive finite measure on $\mathbb{R}^2$, it is easy to see that $R\mu$ and $\tilde{R}\mu$ are also positive finite measures.

Remark 2.2.11 Definition 2.2.10 gives the Radon transform of a finite positive measure as a finite positive measure which is expressed in terms of its action on functions in $C_0(G_{1,2})$. We will now work toward obtaining a more explicit expression for the Radon transform of a finite positive measure.

Definition 2.2.12 Let μ be a Borel measure on X and E a Borel subset of X. We say that μ is outer regular on E if

$$\mu(E) = \inf\{\mu(U) : U \supset E, \ U \text{ open}\},$$

30

and inner regular on E if

$$\mu(E) = \sup\{\mu(K) : K \subset E, \ K \text{ compact}\}.$$

A Borel measure is said to be a Radon measure if it is finite on compact sets, outer regular on Borel sets, and inner regular on open sets [Fol84, p. 205]. A Borel measure is said to be regular if it is outer and inner regular on all Borel sets. Any finite Borel measure on a second-countable, locally-compact Hausdorff space is a regular Radon measure [Fol84, thm. 7.8], so, in particular, any finite Borel measure on $\mathbb{R}^d$, $G_{1,2}$, or $S^1 \times \mathbb{R}$ is a Radon measure. If U is an open set of X and $f \in C_c(X)$, we say that f is subordinate to U, and write $f \prec U$, if $0 \leq f \leq 1$ and $\operatorname{supp}(f) \subset U$ ($\operatorname{supp}(f)$ denotes the support of f).

Proposition 2.2.13 *Let μ be a finite positive measure on $\mathbb{R}^2$. Let E be a Borel set of $G_{1,2}$. Then*

$$R\mu(E) = \int_{\mathbb{R}^2} R^* 1_E \, d\mu,$$

where 1_E denotes the indicator function of the set E.

Proof. The result will be proved by first showing that it holds if E is open or if E is compact. The general result will then follow from the regularity of the measure $R\mu$.

If U is a open subset of $G_{1,2}$, then, by the Riesz representation theorem for positive measures [Fol84, thm. 7.2],

$$
\begin{aligned}
R\mu(U) \ &= \sup\left\{ \int_{\mathbb{R}^2} R^* g \, d\mu : g \in C_c(G_{1,2}) \text{ and } g \prec U \right\} \\
&\leq \int_{\mathbb{R}^2} R^* 1_U \, d\mu.
\end{aligned}
\tag{2.7}
$$

To get an inequality in the opposite direction, define $\tilde{U} \overset{\text{def}}{=} \pi^{-1}(U)$ and let d denote the usual Riemannian metric on $S^1 \times R$ viewed as a cylinder. Define a sequence g_n

of functions in $C_c(S^1 \times R)$ by

$$
g_n(\omega, s) = \begin{cases} 0 & \text{if } \inf d((\omega, s), \check{U}^c) < \frac{1}{2n} \\ 2n[\inf d((\omega, s), \check{U}^c) - \frac{1}{2n}] & \text{if } \frac{1}{2n} < \inf d((\omega, s), \check{U}^c) < \frac{1}{n} \\ 1 & \text{if } \inf d((\omega, s), \check{U}^c) \geq \frac{1}{n}, \end{cases}
$$

so that $g_n \prec \check{U}$ and $\{g_n\}$ is monotonically increasing with pointwise limit function $1_{\check{U}}$. Note that, since the indicator function of $\check{U}$ is even, each g_n is even. The sequence of functions $\tilde{R}^* g_n$ converges pointwise on $\mathbb{R}^2$ to $\tilde{R}^* 1_{\check{U}}$ since, by the monotone convergence theorem, for each $x \in \mathbb{R}^2$,

$$
\begin{aligned}
\lim_{n \to \infty} \tilde{R}^* g_n(x) &= \int_{S^1} \lim_{n \to \infty} g_n(\omega, x \cdot \omega) \, d\sigma(\omega) \\
&= \int_{S^1} 1_{\check{U}}(\omega, x \cdot \omega) \, d\sigma(\omega) \\
&= \tilde{R}^* 1_{\check{U}}(x).
\end{aligned}
$$

Since each $g_n \prec \check{U}$, it is clear that each $\frac{1}{2}\check{g}_n \prec U$, hence

$$
\begin{aligned}
R\mu(U) &\geq \frac{1}{2} \int_{\mathbb{R}^2} \check{g}_n \, dR\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} R^* \check{g}_n \, d\mu \\
&= \frac{1}{4} \int_{\mathbb{R}^2} \tilde{R}^* \tilde{g}_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \tilde{R}^* g_n \, d\mu.
\end{aligned}
$$

A second application of the monotone convergence theorem then gives

$$
\begin{aligned}
R\mu(U) &\geq \frac{1}{2} \lim_{n \to \infty} \int_{\mathbb{R}^2} \tilde{R}^* g_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \lim_{n \to \infty} \tilde{R}^* g_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \tilde{R}^* 1_{\check{U}} \, d\mu \\
&= \int_{\mathbb{R}^2} R^* 1_U \, d\mu.
\end{aligned}
\tag{2.8}
$$

Comparing equations 2.7 and 2.8, we conclude that

$$R\mu(U) = \int_{\mathbb{R}^2} R^* 1_U \, d\mu.$$

If K is a compact subset of $G_{1,2}$, then, by the Riesz representation theorem for positive measures [Fol84, thm. 7.2],

$$\begin{aligned} R\mu(U) &= \inf \left\{ \int_{\mathbb{R}^2} R^* g \, d\mu(x) : g \in C_c(G_{1,2}) \text{ and } g \geq 1_K \right\} \\ &\geq \int_{\mathbb{R}^2} (R^* 1_K) \, d\mu. \end{aligned}$$

To get an inequality in the opposite direction, define $\check{K} \stackrel{\text{def}}{=} \pi^{-1}(K)$ and the sequence g_n of functions in $C_c(S^1 \times \mathbb{R})$ by

$$g_n(\omega, s) = \begin{cases} 0 & \text{if } \inf d((\omega, s), \check{K}) \geq \frac{1}{n} \\ n[\frac{1}{n} - \inf d((\omega, s), \check{K})] & \text{if } 0 < \inf d((\omega, s), \check{K}) < \frac{1}{n} \\ 1 & \text{if } s \in \check{K}, \end{cases}$$

so that $g_n \geq 1_{\check{K}}$ and $\{g_n\}$ is monotonically decreasing with pointwise limit function $1_{\check{K}}$. Again, each g_n is even. The sequence of functions $\tilde{R}^* g_n$ converges pointwise on $\mathbb{R}^2$ to $\tilde{R}^* 1_{\check{K}}$ since, by the dominated convergence theorem, for each $x \in \mathbb{R}^2$,

$$\begin{aligned} \lim_{n \to \infty} \tilde{R}^* g_n(x) &= \int_{S^1} \lim_{n \to \infty} g_n(\omega, x \cdot \omega) \, d\sigma(\omega) \\ &= \int_{S^1} 1_{\check{K}}(\omega, x \cdot \omega) \, d\sigma(\omega) \\ &= \tilde{R}^* 1_{\check{K}}(x). \end{aligned}$$

Since each $g_n \geq 1_{\check{K}}$, it is clear that each $\frac{1}{2}\check{g}_n \geq 1_K$, hence

$$
\begin{aligned}
R\mu(K) &\leq \frac{1}{2} \int_{\mathbb{R}^2} \check{g}_n \, dR\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} R^* \check{g}_n \, d\mu \\
&= \frac{1}{4} \int_{\mathbb{R}^2} \check{R}^* \tilde{\check{g}}_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \check{R}^* g_n \, d\mu.
\end{aligned}
$$

An application of the monotone convergence theorem then gives

$$
\begin{aligned}
R\mu(K) &\leq \frac{1}{2} \lim_{n \to \infty} \int_{\mathbb{R}^2} \check{R}^* g_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \lim_{n \to \infty} \check{R}^* g_n \, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \check{R}^* 1_{\check{K}} \, d\mu. \\
&= \int_{\mathbb{R}^2} R^* 1_K \, d\mu.
\end{aligned}
$$

and we conclude that

$$
R\mu(K) = \int_{\mathbb{R}^2} R^* 1_K \, d\mu.
$$

Let $\epsilon > 0$ be given. By the inner and outer regularity of $R\mu$, we can choose an open set U and a compact set K such that $K \subset E \subset U$ with

$$
R\mu(U) - \epsilon < R\mu(E) < R\mu(K) + \epsilon.
$$

We have

$$
\begin{aligned}
R\mu(E) - \int_{\mathbb{R}^2} R^* 1_E \, d\mu &\leq R\mu(E) - \int_{\mathbb{R}^2} R^* 1_K \, d\mu \\
&= R\mu(E) - R\mu(K) \\
&< \epsilon
\end{aligned}
$$

and

$$\int_{\mathbb{R}^2} R^* 1_E \, d\mu - R\mu(E) \leq \int_{\mathbb{R}^2} R^* 1_U \, d\mu - R\mu(E)$$

$$= R\mu(U) - R\mu(E)$$

$$< \epsilon,$$

so

$$\left| \int_{\mathbb{R}^2} R^* 1_E \, d\mu - R\mu(E) \right| < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, the desired result follows. $\square$

2.3 Application to PET

In this section, we will compute the probability measure for the observations in our idealization of the PET problem.

Definition 2.3.1 The set of lines in $\mathbb{R}^2$ through a point $x \in \mathbb{R}^2$ can be given the structure of a differentiable manifold by identifying it with the space $\mathbb{P}^1$ in the following way. Recall that $\mathbb{P}^1$ can be identified with the set of lines through the origin in $\mathbb{R}^2$. If l is a line through x, we identify it with the unique line through the origin, i.e., the unique point $\bar{\omega} \in \mathbb{P}^1$, which is perpendicular to l. It is easy to see that this identification gives a bijective relation between the lines in $\mathbb{R}^2$ through x and $\mathbb{P}^1$.

Definition 2.3.2 Suppose the probability distribution of the location of positron emissions in $\mathbb{R}^2$ is described by the probability measure μ. If a positron emission takes place at $x \in \mathbb{R}^2$, we will model the line which its annihilation photons travel along as a line in $\mathbb{R}^2$ through x whose orientation is randomly distributed according to the measure $\frac{1}{2\pi}\sigma \circ p^{-1}$ on $\mathbb{P}^1$. That is, we identify the lines in $\mathbb{R}^2$ through x with $\mathbb{P}^1$ as in definition 2.3.1 and give them the measure $\frac{1}{2\pi}\sigma \circ p^{-1}$. The event of a positron emission occurring at $x \in \mathbb{R}^2$ and its annihilation photons having orientation $\bar{\omega} \in \mathbb{P}^1$ is thus described by the pair $(x, \bar{\omega}) \in \mathbb{R}^2 \times \mathbb{P}^1$. Thus the joint probability distribution of x and $\bar{\omega}$ is given by the product measure $m \overset{\text{def}}{=} \mu \times \left(\frac{1}{2\pi}\sigma \circ p^{-1}\right)$ on the space $\mathbb{R}^2 \times \mathbb{P}^1$.

In the PET problem, one does not observe $(x, \bar{\omega})$, but rather only the point $\rho(x, \bar{\omega}) \in G_{1,2}$, where the function $\rho : \mathbb{R}^2 \times \mathbb{P}^1 \to G_{1,2}$ is given by

$$\rho : (x, p(\omega)) \mapsto \pi(\omega, x \cdot \omega).$$

It is easily verified that ρ is well-defined. Thus, the probability measure of the observations in the PET problem is just the image measure, $\nu \overset{\text{def}}{=} \left[\mu \times \left(\frac{1}{2\pi}\sigma \circ p^{-1}\right)\right] \circ \rho^{-1}$

induced on $G_{1,2}$ by the map ρ. If B is a Borel set of $G_{1,2}$, we have

$$
\begin{aligned}
\nu(B) \quad &\stackrel{\text{def}}{=} \left[\mu \times \left(\frac{1}{2\pi} \sigma \circ p^{-1} \right) \right] [\rho^{-1}(B)] \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{\mathbb{P}^1} 1_{\rho^{-1}(B)}(x, \bar{\omega})\, d(\sigma \circ p^{-1})(\bar{\omega})\, d\mu(x) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{S^1} 1_{\rho^{-1}(B)}(x, p(\omega))\, d\sigma(\omega)\, d\mu(x) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{S^1} 1_B[\pi(\omega, x \cdot \omega)]\, d\sigma(\omega)\, d\mu(x) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{S^1} \tilde{1}_B(\omega, x \cdot \omega)\, d\sigma(\omega)\, d\mu(x) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \tilde{R}^* \tilde{1}_B(x)\, d\mu(x) \\
&= \frac{1}{\pi} \int_{\mathbb{R}^2} R^* 1_B(x)\, d\mu(x) \\
&= \frac{1}{\pi} R\mu(B),
\end{aligned}
$$

by proposition 2.2.13. We thus see that the probability measure of the observations in the PET problem is given by $\frac{1}{\pi} R\mu$. For convenience, we shall define $T \stackrel{\text{def}}{=} \frac{1}{\pi} R$.

Remark 2.3.3 It is easy to see that if the measure μ is representable by the density $f\, d\lambda^2$, then the above agrees with the result previously obtained by Johnstone and Silverman [JS90]. Indeed, using proposition 2.1.14, we have

$$
\begin{aligned}
\nu(B) \quad &= \frac{1}{\pi} R(f\, d\lambda^2)(B) \\
&= \frac{1}{\pi} \int_{\mathbb{R}^2} R^* 1_B\, f\, d\lambda^2 \\
&= \frac{1}{\pi} \int_{G_{1,2}} 1_B\, Rf\, ds\, d\theta \\
&= \frac{1}{\pi} \int_B Rf\, ds\, d\theta,
\end{aligned}
\tag{2.9}
$$

so ν is representable as the density $\frac{1}{\pi} Rf$ with respect to the measure $ds\, d\theta$ on $G_{1,2}$. This agrees with equation (2.1) in [JS90] once the differences in the definitions of the dominating measures are taken into account.

Chapter 3

Sobolev Spaces

As stated in chapter 1, we will be analyzing the PET problem using loss functions based on Sobolev norms. In this chapter, we will collect the necessary background material on Sobolev spaces. Section 3.1 discusses the more general notion of interpolation spaces, of which Sobolev spaces are a principal example. Section 3.2 discusses Sobolev spaces on $\mathbb{R}^2$, $S^1 \times \mathbb{R}$, and $G_{1,2}$. Section 3.3 discusses modified versions of these spaces for distributions with fixed compact support. Section 3.4 gives some continuity results for the Radon transform with respect to Sobolev norms.

3.1 Interpolation spaces

Before giving the definitions of the Sobolev spaces, we review some basic notions regarding interpolation spaces. They will be used frequently in what follows. The material in this section is mostly taken from [Nat86, sec. VII.4].

The basic idea of interpolation spaces may be simply stated. One starts with a pair of Hilbert spaces $H_1 \subset H_0$, where H_1 is continuously and densely embedded in H_0. Then the interpolation spaces H_α for $\alpha \in (0,1)$ are defined so as to provide a continuum of Hilbert spaces between H_0 and H_1.

Definition 3.1.1 Suppose $H_1 \subset H_0$ are two Hilbert spaces such that H_1 is continuously and densely embedded in H_0. Such a pair is said to be an interpolation couple. It can be shown that there exists a self-adjoint, strictly-positive operator $S : H_1 \to H_0$

such that the norms $||f||_{H_1}$ and $||Sf||_{H_0}$ are equivalent on H_1. For $\beta \in [0,1]$, let S^β denote the operator S taken to the β power. (The details of the definition of S^β may be found in most treatments of the spectral theory of linear operators on Hilbert spaces, e.g., [RSN55, sec. 127] or [RS80, sec. VIII.3]). We define the interpolation space $H_\beta \stackrel{\text{def}}{=} (H_0, H_1)_\beta$ to be the domain of the operator S^β (a subspace of H_0, see e.g., [RSN55, sec. 127] or [RS80, sec. VIII.3] for details). H_β is a Hilbert space with norm $||f||_{H_\beta} \stackrel{\text{def}}{=} ||S^\beta f||_{H_0}$. It is clear that applying this procedure for $\beta = 0$ and $\beta = 1$ just recovers the original spaces H_0 and H_1. It can be shown that, while S is not uniquely determined, the space H_β is, and that the norm on H_β is determined up to equivalence.

Proposition 3.1.2 *Let H_0, H_1 and K_0, K_1 be interpolation couples and H_β and K_β the corresponding interpolation spaces. If $A : H_0 \to K_0$ is a bounded linear operator such that $A(H_1) \subset K_1$ and the restriction $A : H_1 \to K_1$ is also bounded with respect to the norms of H_1 and K_1, then A is bounded with respect to the norms of H_β and K_β for all $\beta \in [0,1]$.*

Proof.   See [LM72, ch. 1, thm. 5.1]. $\square$

Remark 3.1.3 It will be shown in remark 3.2.2 that the Sobolev spaces provide nice concrete examples of interpolation spaces.

3.2 Definitions

Definition 3.2.1 The Schwartz space $\mathscr{S}(\mathbb{R}^d)$ is defined to be the subspace of $C^\infty(\mathbb{R}^d)$ consisting of functions f such that

$$|f|_m \overset{\text{def}}{=} \sup_{|\beta|\le m} \sup_{x\in\mathbb{R}^d} (1+|x|^2)^m |\partial^\beta f(x)|$$

is finite for all $m \in \mathbb{N}$, where β is a multi-index [Tre67, ch. 10, ex. IV]. We equip $\mathscr{S}(\mathbb{R}^d)$ with the topology induced by the seminorms $|\cdot|_m$. The space of tempered distributions on $\mathbb{R}^d$ is defined to be the dual space to $\mathscr{S}(\mathbb{R}^d)$ and is denoted by $\mathscr{S}'(\mathbb{R}^d)$. The Fourier transform $\mathscr{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ is denoted by $f \mapsto \hat{f}$, where $\hat{f}$ is defined by

$$\hat{f}(\xi) \overset{\text{def}}{=} \int_{\mathbb{R}^d} e^{-i2\pi x\cdot\xi} f(x)\, dx$$

[Tre67, p. 268, def. 25.1]. The Fourier transform $\mathscr{F} : \mathscr{S}'(\mathbb{R}^d) \to \mathscr{S}'(\mathbb{R}^d)$ is defined by duality, i.e., for $u \in \mathscr{S}'(\mathbb{R}^d)$ and $f \in \mathscr{S}(\mathbb{R}^d)$, $\langle \hat{u}, f\rangle \overset{\text{def}}{=} \langle u, \hat{f}\rangle$ [Tre67, p. 275, def. 25.4]. For $\alpha \in \mathbb{R}$, the Sobolev space $L^2_\alpha(\mathbb{R}^d)$ is defined to be the subspace of $\mathscr{S}'(\mathbb{R}^d)$ whose elements u satisfy

$$\|u\|^2_{L^2_\alpha(\mathbb{R}^d)} \overset{\text{def}}{=} \int_{\mathbb{R}^d} (1+4\pi^2|\xi|^2)^\alpha |\hat{u}(\xi)|^2 \, d\xi < \infty, \tag{3.1}$$

cf. [Ste70, subsec. V.3.1]. It is implicit in the definition that the distribution $(1+4\pi^2|\xi|^2)^{\alpha/2}\hat{u}$ can be represented by an element of $L^2(\mathbb{R}^d)$. $L^2_\alpha(\mathbb{R}^d)$ is a Hilbert space with inner product

$$(u|v)_{L^2_\alpha(\mathbb{R}^d)} \overset{\text{def}}{=} \int_{\mathbb{R}^d} (1+4\pi^2|\xi|^2)^\alpha \hat{u}(\xi)\overline{\hat{v}(\xi)} \, d\xi,$$

where the bar denotes complex conjugate [Tre67, p. 330, prop. 31.7].

Remark 3.2.2 The spaces $L^2_\alpha(\mathbb{R}^d)$ and $L^2_{-\alpha}(\mathbb{R}^d)$ are duals of each other with respect to the bilinear form

$$(f,u) \mapsto \int_{\mathbb{R}^d} \hat{u}(\xi)\hat{f}(-\xi)\, d\xi$$

on $L^2_\alpha(\mathbb{R}^d) \times L^2_{-\alpha}(\mathbb{R}^d)$ [Tre67, p. 330-331]. Moreover, if $f \in \mathscr{S}$, then this bilinear form is given by $\langle u, f \rangle$, where $\langle u, f \rangle$ refers to the duality between $\mathscr{S}'$ and $\mathscr{S}$. Since, for any $\alpha \in \mathbb{R}$, $\mathscr{S}$ is dense in $L^2_\alpha(\mathbb{R}^d)$ [Tre67, prop. 31.9], it follows that

$$\|u\|_{L^2_{-\alpha}(\mathbb{R}^d)} = \sup_{f \in \mathscr{S}(\mathbb{R}^d): \|f\|_{L^2_\alpha(\mathbb{R}^d)}=1} |\langle u, f \rangle|. \qquad (3.2)$$

Remark 3.2.3 For $n \in \mathbb{Z}$, the pair $L^2_n(\mathbb{R}^d)$, $L^2_{n+1}(\mathbb{R}^d)$ form an interpolation couple. If $n \leq \alpha \leq n+1$ with $\alpha = n + \beta$, then

$$L^2_\alpha(\mathbb{R}^d) = \left(L^2_n(\mathbb{R}^d), L^2_{n+1}(\mathbb{R}^d) \right)_\beta.$$

This can be seen by choosing the interpolation operator S in definition 3.1.1 such that

$$(Su)^\wedge(\xi) = (1 + 4\pi^2 |\xi|^2)^{\beta/2} \hat{u}(\xi)$$

[Nat86, p. 202].

Remark 3.2.4 The use of the circumflex to denote the Fourier transform conflicts with its use in denoting an estimate of an unknown quantity. Since both notations are quite standard, we shall just endure this ambiguity. It should be clear from context which meaning is intended.

Remark 3.2.5 It is an unfortunate fact of life that there is no universally accepted definition of the Fourier transform. In several of the works which we shall cite, the function

$$\check{f}(\xi) \stackrel{\text{def}}{=} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-ix\cdot\xi} f(x) \, dx \qquad (3.3)$$

is defined to be the Fourier transform of f, e.g., [Her83, p. 166] and [Nat86, p. 180]. (The notation conflicts with that of definition 2.1.8, but we shall just use it in this and a related remark.) The two definitions are related by the equations

$$\check{f}(\xi) = \frac{1}{(2\pi)^{d/2}} \hat{f}(\xi/2\pi)$$

and

$$\hat{f}(\xi) = (2\pi)^{d/2} \check{f}(2\pi\xi).$$

In works which use equation 3.3 as the definition of the Fourier transform, the Sobolev norm on $L_\alpha^2(\mathbb{R}^d)$ is typically defined by

$$||u||^2_{L_\alpha^2(\mathbb{R}^d)} \overset{\text{def}}{=} \int_{\mathbb{R}^d} (1 + |\xi|^2)^\alpha |\check{u}(\xi)|^2 \, d\xi.$$

Writing this in terms of our definition of the Fourier transform, we obtain

$$
\begin{aligned}
||u||^2_{L_\alpha^2(\mathbb{R}^d)} &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (1 + |\xi|^2)^\alpha |\hat{u}(\xi/2\pi)|^2 \, d\xi \\
&= \int_{\mathbb{R}^d} (1 + |2\pi\xi|^2)^\alpha |\hat{u}(\xi)|^2 \, d\xi. \\
&= ||u||^2_{L_\alpha^2(\mathbb{R}^d)}.
\end{aligned}
$$

We see that the two definitions agree.

Remark 3.2.6 For $\alpha \in \mathbb{N}$, we can express $||f||^2_{L_\alpha^2(\mathbb{R}^2)}$ in terms of the $L^2(\mathbb{R}^2)$-norms of f and its first α partial derivatives. Indeed, we have

$$
\begin{aligned}
||f||^2_{L_\alpha^2(\mathbb{R}^2)} &= \int_{\mathbb{R}^2} (1 + 4\pi^2 |\xi|^2)^\alpha |\hat{f}(\xi)|^2 \, d\xi \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \int_{\mathbb{R}^2} [4\pi^2(\xi_1^2 + \xi_2^2)]^j |\hat{f}(\xi)|^2 \, d\xi \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \sum_{k=0}^{j} \binom{j}{k} \int_{\mathbb{R}^2} (2\pi\xi_1)^{2(j-k)} (2\pi\xi_2)^{2k} |\hat{f}(\xi)|^2 \, d\xi \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \sum_{k=0}^{j} \binom{j}{k} \int_{\mathbb{R}^2} |(2\pi\xi_1)^{j-k} (2\pi\xi_2)^k \hat{f}(\xi)|^2 \, d\xi \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \sum_{k=0}^{j} \binom{j}{k} \int_{\mathbb{R}^2} \left|(\partial_{x_1}^{j-k} \partial_{x_2}^k f)^{\hat{}}(\xi)\right|^2 \, d\xi \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \sum_{k=0}^{j} \binom{j}{k} ||\partial_{x_1}^{j-k} \partial_{x_2}^k f||^2_{L_2(\mathbb{R}^2)},
\end{aligned}
$$

(3.4)

where the derivatives are taken in the sense of tempered distributions on $\mathbb{R}^2$.

Remark 3.2.7 The set of (finite) signed measures $M(\mathbb{R}^d)$ is a subspace of $L^2_{-\alpha}(\mathbb{R}^d)$ for $\alpha > d/2$. This follows easily from the fact that the Fourier transform of a signed measure is in $C_b(\mathbb{R}^d)$, cf. [Chu74, p. 143]. In particular, we have $P(\mathbb{R}^d) \subset L^2_{-\alpha}(\mathbb{R}^d)$ for $\alpha > d/2$.

We now want to consider Sobolev spaces on $S^1 \times \mathbb{R}$. We start by considering differential operators on $S^1 \times \mathbb{R}$.

Definition 3.2.8 If $g \in C^\infty(S^1 \times \mathbb{R})$, define $g_s : S^1 \to \mathbb{R}$ by $\omega \mapsto g(\omega, s)$. We then define $\partial_\theta g(\omega, s) \stackrel{\text{def}}{=} \partial_\theta g_s(\omega)$, where the right-hand side is defined as in definition 2.1.2.

Definition 3.2.9 There is another procedure for defining differential operators on $S^1 \times \mathbb{R}$ that is used in the literature on the Radon transform. Suppose $g \in C^\infty(S^1 \times \mathbb{R})$. We can extend g to an element $\tilde{g} \in C^\infty(\mathbb{R}^2 \backslash \{0\} \times \mathbb{R})$ which is homogeneous of degree -1 by defining

$$\tilde{g}(x, s) \stackrel{\text{def}}{=} g\left(x/|x|, s/|x|\right)/|x|.$$

We then define the differential operators ∂_{x_1} and ∂_{x_2} on $S^1 \times \mathbb{R}$ to be the operators obtained by applying these operators to the extended function on $\mathbb{R}^2 \backslash \{0\} \times \mathbb{R}$ and evaluating on $S^1 \times \mathbb{R}$. If $\beta = (\beta_1, \beta_2)$ is a multi-index, we shall write ∂_x^β for the differential operator $\partial_{x_1}^{\beta_1} \partial_{x_2}^{\beta_2}$. It can be shown that the homogeneous extension of degree -1 just described is a natural extension of the Radon transform of a function [Nat86, p. 12].

Definition 3.2.10 The Schwartz space $\mathscr{S}(S^1 \times \mathbb{R})$ is defined to be the subspace of $C^\infty(S^1 \times \mathbb{R})$ consisting of functions g such that

$$|g|_m \stackrel{\text{def}}{=} \sup_{|\beta|+k \le m} \sup_{(\omega,s) \in S^1 \times \mathbb{R}} (1 + s^2)^m |\partial_x^\beta \partial_s^k g(\omega, s)|,$$

where β is a multi-index and $k \in \mathbb{N}$, is finite for all $m \in \mathbb{N}$ [HQ85, eq. 2.10]. We equip $\mathscr{S}(S^1 \times \mathbb{R})$ with the topology induced by the seminorms $|\cdot|_m$. The space of tempered distributions on $S^1 \times \mathbb{R}$, $\mathscr{S}'(S^1 \times \mathbb{R})$, is defined to be the space of continuous linear functionals on $\mathscr{S}(S^1 \times \mathbb{R})$.

43

Definition 3.2.11 On $\mathscr{S}(S^1 \times \mathbb{R})$, we take the Fourier transform to be the usual Fourier transform with respect to the second variable only, i.e.,

$$\hat{g}(\omega, \eta) \stackrel{\text{def}}{=} \int_{\mathbb{R}} e^{i2\pi\eta s} g(\omega, s)\, ds.$$

The Fourier transform on $\mathscr{S}'(S^1 \times \mathbb{R})$ is defined by duality. (It is easily verified that the Fourier transform of a function in $\mathscr{S}(S^1 \times \mathbb{R})$ is in $\mathscr{S}(S^1 \times \mathbb{R})$, so the definition makes sense.)

Definition 3.2.12 The Sobolev space $L^2_{(0,\alpha)}(S^1 \times \mathbb{R})$ is defined to be the subspace of $\mathscr{S}'(S^1 \times \mathbb{R})$ whose elements v satisfy

$$\|v\|^2_{L^2_{(0,\alpha)}} \stackrel{\text{def}}{=} \int_{S^1} \int_{\mathbb{R}} (1 + 4\pi^2\eta^2)^\alpha |\hat{v}(\omega, \eta)|^2\, d\eta\, d\sigma(\omega) < \infty,$$

where $\hat{v}(\theta, \eta)$ is defined as in definition 3.2.11. $L^2_{(0,\alpha)}(S^1 \times \mathbb{R})$ is a Hilbert space with inner product

$$(v|w)_{L^2_{(0,\alpha)}} \stackrel{\text{def}}{=} \int_{S^1} \int_{\mathbb{R}} (1 + 4\pi^2\eta^2)^\alpha \hat{v}(\omega, \eta) \overline{\hat{w}(\omega, \eta)}\, d\eta\, d\sigma(\omega).$$

Remark 3.2.13 For $\alpha \in \mathbb{N}$, we can express $\|g\|^2_{L^2_{(0,\alpha)}(S^1 \times \mathbb{R})}$ in terms of the $L^2(S^1 \times \mathbb{R})$-norms of g and its first α partial derivatives with respect to s. Indeed, we have

$$\begin{aligned}
\|g\|^2_{L^2_{(0,\alpha)}(S^1 \times \mathbb{R})} &= \int_{S^1} \int_{\mathbb{R}} (1 + 4\pi^2\eta^2)^\alpha |\hat{g}(\omega, \eta)|^2\, d\eta\, d\sigma(\omega) \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \int_{S^1} \int_{\mathbb{R}} (4\pi^2\eta^2)^j |\hat{g}(\omega, \eta)|^2\, d\eta\, d\sigma(\omega) \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \int_{S^1} \int_{\mathbb{R}} |(2\pi\eta)^j \hat{g}(\omega, \eta)|^2\, d\eta\, d\sigma(\omega) \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \int_{S^1} \int_{\mathbb{R}} |(\partial_s^j g)\hat{\ }(\omega, \eta)|^2\, d\eta\, d\sigma(\omega) \\
&= \sum_{j=0}^{\alpha} \binom{\alpha}{j} \|\partial_s^j g\|^2_{L^2(S^1 \times \mathbb{R})}
\end{aligned} \tag{3.5}$$

where the derivatives are taken in the sense of tempered distributions.

Definition 3.2.14 We define the Schwartz space $\mathscr{S}(G_{1,2})$ to be the subspace of $C^\infty(G_{1,2})$ whose elements have lifts in $\mathscr{S}(S^1 \times \mathbb{R})$. The lifting map from $\mathscr{S}(G_{1,2})$ to $\mathscr{S}(S^1 \times \mathbb{R})$ is obviously injective and we equip $\mathscr{S}(G_{1,2})$ with the relative topology induced by identifying it with its image under this map in $\mathscr{S}(S^1 \times \mathbb{R})$. The space of tempered distributions on $G_{1,2}$ is defined to be the dual of $\mathscr{S}(G_{1,2})$ and is denoted by $\mathscr{S}'(G_{1,2})$. The Sobolev space $L^2_{(0,\alpha)}(G_{1,2})$ is defined to be the subspace of $\mathscr{S}'(G_{1,2})$ whose elements have lifts in $L^2_{(0,\alpha)}(S^1 \times \mathbb{R})$. The squared norm of an element of $L^2_{(0,\alpha)}(G_{1,2})$ is defined to be 1/2 of the squared norm of its lift in $L^2_{(0,\alpha)}(S^1 \times \mathbb{R})$.

3.3 Sobolev norms for distributions with fixed compact support

In this work, we will mainly be concerned with distributions (actually probability measures) with support on a fixed compact set. For such distributions, one can construct alternative Sobolev norms in terms of their Fourier series coefficients. The resulting Sobolev spaces have the desirable property of being separable, i.e., having a countable orthonormal basis.

Definition 3.3.1 Let $rS \subset \mathbb{R}^2$ denote the closed square with sides of length r centered at the origin, i.e., $[-r/2, r/2]^2$. Let $\mathscr{E}'(\mathbb{R}^2)$ denote the space of compactly supported distributions on $\mathbb{R}^2$, i.e., the space of continuous linear functionals on $C^\infty(\mathbb{R}^2)$. (The notation $\mathscr{E}'(\mathbb{R}^2)$ stems from the fact that $C^\infty(\mathbb{R}^2)$ is sometimes denoted by $\mathscr{E}(\mathbb{R}^2)$ in the literature.) Let $\mathscr{E}'(rS)$ denote the subspace of $\mathscr{E}'(\mathbb{R}^2)$ whose elements have support contained in rS.

Definition 3.3.2 Let $C^\infty(r\mathbb{T}^2)$ denote the subspace of $C^\infty(\mathbb{R}^2)$ whose elements are periodic with period r in both coordinates. We equip $C^\infty(r\mathbb{T}^2)$ with the topology induced by the seminorms $\|\partial^\alpha f\|_u$ for all multi-indices α. $C^\infty(r\mathbb{T}^2)$ can be viewed as the set of smooth functions on the 2-torus, $r\mathbb{T}^2$, obtained by identifying points in $\mathbb{R}^2$ whose coordinates differ by integer multiples of r. Let $\mathscr{E}'(r\mathbb{T}^2)$ denote the set of continuous linear functionals on $C^\infty(r\mathbb{T}^2)$. $\mathscr{E}'(r\mathbb{T}^2)$ can be viewed as the set of distributions on $r\mathbb{T}^2$. Since $C^\infty(r\mathbb{T}^2) \subset C^\infty(\mathbb{R}^2)$, the elements of $\mathscr{E}'(rS)$ can be considered as elements of $\mathscr{E}'(r\mathbb{T}^2)$ in a natural way. For $\kappa \in \mathbb{Z}^2$, we define the Fourier series coefficients of $u \in \mathscr{E}'(r\mathbb{T}^2)$ by

$$\hat{u}(\kappa) \stackrel{\text{def}}{=} \frac{1}{r^2} \langle u, e^{-i(2\pi/r)\kappa \cdot x} \rangle.$$

In particular, if $f \in L^2(r\mathbb{T}^2)$, then it is well-known that f has the Fourier series expansion

$$f(x) = \sum_{\kappa \in \mathbb{Z}^2} \hat{f}(\kappa) e^{i(2\pi/r)\kappa \cdot x} \tag{3.6}$$

in $L^2(r\mathbb{T}^2)$. For $\alpha \in \mathbb{R}$, we define the Sobolev space $L^2_\alpha(r\mathbb{T}^2)$ to be the subset of $\mathscr{E}'(r\mathbb{T}^2)$ such that

$$\|u\|^2_{L^2_\alpha(r\mathbb{T}^2)} \overset{\text{def}}{=} r^2 \sum_{\kappa \in \mathbb{Z}^2} [1 + (2\pi/r)^2 |\kappa|^2]^\alpha |\hat{u}(\kappa)|^2 \tag{3.7}$$

is finite.

Remark 3.3.3 Results that are analogous to those given in remark 3.2.2 for $L^2_\alpha(\mathbb{R}^2)$ hold for $L^2_\alpha(r\mathbb{T}^2)$. The spaces $L^2_\alpha(r\mathbb{T}^2)$ and $L^2_{-\alpha}(r\mathbb{T}^2)$ are dual to each other with respect to the bilinear form

$$(f, u) \mapsto r^2 \sum_{\kappa \in \mathbb{Z}^2} \hat{f}(\xi) \hat{u}(-\xi)$$

on $L^2_\alpha(r\mathbb{T}^2) \times L^2_{-\alpha}(r\mathbb{T}^2)$. Moreover, if $f \in C^\infty(r\mathbb{T}^2)$, then this bilinear form is given by $\langle u, f \rangle$, where $\langle u, f \rangle$ refers to the duality between $C^\infty(r\mathbb{T}^2)$ and $\mathscr{E}'(r\mathbb{T}^2)$ [Fol84, exer. 8.39]. Since, for any $\alpha \geq 0$, $C^\infty(r\mathbb{T}^2)$ is dense in $L^2_\alpha(r\mathbb{T}^2)$, it follows that, for $\alpha \leq 0$,

$$\|u\|_{L^2_{-\alpha}(r\mathbb{T}^2)} = \sup_{f \in C^\infty(r\mathbb{T}^2) : \|f\|_{L^2_\alpha(r\mathbb{T}^2)}=1} |\langle u, f \rangle|. \tag{3.8}$$

Remark 3.3.4 For $\alpha \in \mathbb{N}$, we can express $\|f\|^2_{L^2_\alpha(r\mathbb{T}^2)}$ in terms of the $L^2(r\mathbb{T}^2)$-norms of f and its first α partial derivatives. Indeed, we have

$$
\begin{aligned}
\|f\|^2_{L^2_\alpha(r\mathbb{T}^2)} &= r^2 \sum_{\kappa \in \mathbb{Z}^2} [1 + (2\pi/r)^2 \kappa_1^2 + (2\pi/r)^2 \kappa_2^2]^\alpha |\hat{f}(\kappa)|^2 \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} r^2 \sum_{\kappa \in \mathbb{Z}^2} [(2\pi/r)^2 \kappa_1^2 + (2\pi/r)^2 \kappa_2^2]^j |\hat{f}(\kappa)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} r^2 \sum_{\kappa \in \mathbb{Z}^2} [(2\pi/r)\kappa_1]^{2(j-k)} [(2\pi/r)\kappa_2]^{2k} |\hat{f}(\kappa)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} r^2 \sum_{\kappa \in \mathbb{Z}^2} |[(2\pi/r)\kappa_1]^{j-k} [(2\pi/r)\kappa_2]^k \hat{f}(\kappa)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} r^2 \sum_{\kappa \in \mathbb{Z}^2} \left|(\partial_{x_1}^{j-k} \partial_{x_2}^k f)\check{\,}(\kappa)\right|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} \|\partial_{x_1}^{j-k} \partial_{x_2}^k g\|^2_{L_2(\mathbb{R}^2)},
\end{aligned}
\tag{3.9}
$$

where the derivatives are taken in the sense of distributions on $r\mathbb{T}^2$.

We will now show that the norm $\|\cdot\|_{L^2_\alpha(r\mathbb{T}^2)}$ is equivalent to the norm $\|\cdot\|_{L^2_\alpha(\mathbb{R}^2)}$ on $L^2(rS)$ for $\alpha \geq 0$.

Lemma 3.3.5 *The norms $\|\cdot\|_{L^2_\alpha(r\mathbb{T}^2)}$ and $\|\cdot\|_{L^2_\alpha(\mathbb{R}^2)}$ are equivalent on $L^2(rS)$ for $\alpha \geq 0$.*

Proof. First suppose that $\alpha \in \mathbb{N}$. We claim that, in this case, $\|\cdot\|_{L^2_\alpha(r\mathbb{T}^2)}$ and $\|\cdot\|_{L^2_\alpha(\mathbb{R}^2)}$ are equal on $L^2(rS)$. To prove the claim, suppose $f \in L^2(rS)$. The Fourier series expansion of f in $L^2(rS)$ is then given by equation 3.6. It follows that the distributional derivatives of f have the Fourier series expansion

$$\partial^{\beta_1}_{x_1}\partial^{\beta_2}_{x_2}f(x) = (i2\pi/r)^{\beta_1+\beta_2} \sum_{\kappa\in\mathbb{Z}^2} \kappa_1^{\beta_1}\kappa_2^{\beta_2}\hat{f}(\kappa)e^{i(2\pi/r)\kappa\cdot x}$$

on rS. From this expansion, we see that

$$\|\partial^{\beta_1}_\theta\partial^{\beta_2}_s f\|^2_{L_2(\mathbb{R}^2)} = r^2(2\pi/r)^{2(\beta_1+\beta_2)} \sum_{\kappa\in\mathbb{Z}^2} \kappa_1^{2\beta_1}\kappa_2^{2\beta_2}|\hat{f}(\kappa)|^2.$$

By equation 3.4, we thus have

$$
\begin{aligned}
\|f\|^2_{L^2_\alpha(\mathbb{R}^2)} &= \sum_{j=0}^{\alpha}\binom{\alpha}{j}\sum_{k=0}^{j}\binom{j}{k}\|\partial^{j-k}_{x_1}\partial^k_{x_2}f\|^2_{L_2(\mathbb{R}^2)} \\
&= r^2\sum_{j=0}^{\alpha}\binom{\alpha}{j}\sum_{k=0}^{j}\binom{j}{k}\sum_{\kappa\in\mathbb{Z}^2}[(2\pi/r)\kappa_1]^{2(j-k)}[(2\pi/r)\kappa_1]^{2k}|\hat{f}(\kappa)|^2 \\
&= r^2\sum_{\kappa\in\mathbb{Z}^2}\sum_{j=0}^{\alpha}\binom{\alpha}{j}(2\pi/r)^{2j}|\kappa|^{2j}|\hat{f}(\kappa)|^2 \\
&= r^2\sum_{\kappa\in\mathbb{Z}^2}[1+(2\pi/r)^2|\kappa|^2]^\alpha|\hat{f}(\kappa)|^2 \\
&= \|f\|^2_{L^2_\alpha(r\mathbb{T}^2)},
\end{aligned}
$$

which proves the claim. The result for general $\alpha \geq 0$ follows by a standard interpolation argument (cf. [Nat86, sec. VII.4]). $\square$

To get a similar result for $\alpha < 0$, we need to allow a little more room.

Lemma 3.3.6 *Suppose $\alpha > 0$ and $\epsilon > 0$. Then there exist constants $c_1(\alpha, r)$ and $c_2(\alpha, r, r + \epsilon)$ such that*

$$c_1(\alpha, r)||u||_{L^2_{-\alpha}(r\mathbb{T}^2)} \leq ||u||_{L^2_{-\alpha}(\mathbb{R}^2)} \leq c_2(\alpha, r, r + \epsilon)||u||_{L^2_{-\alpha}((r+\epsilon)\mathbb{T}^2)} \qquad (3.10)$$

for all $u \in \mathscr{E}'(rS)$.

Proof. Suppose $u \in \mathscr{E}'(rS)$. By lemma 3.3.5, there exists a constant $d(\alpha, r)$ such that $||f||_{L^2_{\alpha}(\mathbb{R}^2)} \leq d(\alpha, r)||f||_{L^2_{\alpha}(r\mathbb{T}^2)}$ for all $f \in L^2(rS)$. It follows that

$$
\begin{aligned}
||u||_{L^2_{-\alpha}(r\mathbb{T}^2)} &= \sup_{f \in C^\infty(r\mathbb{T}^2):||f||_{L^2_{\alpha}(r\mathbb{T}^2)} \leq 1} |\langle u, f \rangle| \\
&\leq d(\alpha, r) \sup_{f \in C^\infty(\mathbb{R}^2):||f||_{L^2_{\alpha}()} \leq 1} |\langle u, f \rangle| \\
&= d(\alpha, r)||u||_{L^2_{-\alpha}(\mathbb{R}^2)}.
\end{aligned}
$$

Taking $c_1(\alpha, r) = 1/d(\alpha, r)$ gives the first inequality.

To get the second inequality, let $C^\infty(rS)$ denote the subspace of $C_c^\infty(\mathbb{R}^2)$ with support on rS. Choose $\chi \in C^\infty((r + \epsilon)S)$ such that $\chi = 1$ on rS. There exists a constant $c(\alpha, \chi)$ such that, for any $f \in L^2_\alpha(\mathbb{R}^2)$, $||\chi f||_{L^2_{\alpha}(\mathbb{R}^2)} \leq c(\alpha, \chi)||f||_{L^2_{\alpha}(\mathbb{R}^2)}$ [Fol84, cor. 8.58]. Since we can choose χ depending only on r and $r + \epsilon$, we shall write $c(\alpha, \chi)$ as $c_2(\alpha, r, r + \epsilon)$. Thus, if $f \in C^\infty(\mathbb{R}^2)$ with $||f||_{L^2_{\alpha}(\mathbb{R}^2)} = 1$, $\chi f \in C^\infty((r+\epsilon)S)$, $\chi f = f$ on rS, and $||\chi f||_{L^2_{-\alpha}(\mathbb{R}^2)} \leq c_2(\alpha, r, r + \epsilon)$. Starting with equation 3.8, it follows that

$$
\begin{aligned}
||u||_{L^2_{-\alpha}(\mathbb{R}^2)} &\leq c_2(\alpha, r, r + \epsilon) \sup_{f \in C^\infty((r+\epsilon)S):||f||_{L^2_{\alpha}()} \leq 1} |\langle u, f \rangle| \\
&\leq c_2(\alpha, r, r + \epsilon) \sup_{f \in C^\infty((r+\epsilon)\mathbb{T}^2):||f||_{L^2_{\alpha}()} \leq 1} |\langle u, f \rangle| \\
&= c_2(\alpha, r, r + \epsilon)||u||_{L^2_{-\alpha}((r+\epsilon)\mathbb{T}^2)}. \quad \square
\end{aligned}
$$

Remark 3.3.7 To see why a little extra room is needed in lemma 3.3.6, consider the point masses at $(1/2, 0)$ and $(-1/2, 0)$. They are equal considered as elements of $\mathscr{E}'(\mathbb{T}^2)$, but not as elements of $\mathscr{E}'(\mathbb{R}^2)$.

As we shall be concerned principally with the Sobolev norms of positive measures, we shall frequently find it convenient to work with the Fourier sine and cosine transforms rather than directly with the Fourier transform. Suppose the distribution $u \in \mathcal{S}'(r\mathbb{T}^2)$ has the property that it maps real-valued test functions into $\mathbb{R}$. Define the Fourier sine and cosine transforms by

$$\hat{u}_s(\kappa) \stackrel{\text{def}}{=} \frac{2}{r^2}\langle u, \sin[(2\pi/r)\kappa \cdot x]\rangle$$

and

$$\hat{u}_c(\kappa) \stackrel{\text{def}}{=} \frac{2}{r^2}\langle u, \cos[(2\pi/r)\kappa \cdot x]\rangle,$$

respectively. Using the identity $\hat{u}(-\kappa) = \hat{u}^*(\kappa)$, it follows that

$$\begin{aligned}
\Re\hat{u}(\kappa) &= \frac{\hat{u}(\kappa) + \hat{u}^*(\kappa)}{2} \\
&= \frac{1}{2r^2}\langle u, e^{i(2\pi/r)\kappa\cdot x} + e^{-i(2\pi/r)\kappa\cdot x}\rangle \\
&= \frac{1}{r^2}\langle u, \cos[(2\pi/r)\kappa \cdot x]\rangle \\
&= \frac{\hat{u}_c(\kappa)}{2}
\end{aligned}$$

and, similarly,

$$\Im\hat{v}(\kappa) = \frac{\hat{u}_s(\kappa)}{2}.$$

In particular, if $f \in L^2(r\mathbb{T}^2)$, then, starting from equation 3.6, we have the expansion

$$\begin{aligned}
f(x) &= \sum_{\substack{r\in\mathbb{Z}^2 \\ 0\leq\arg\kappa<\pi}} \hat{f}(\kappa)e^{i(2\pi/r)\kappa\cdot x} + \hat{f}(-\kappa)e^{-i(2\pi/r)\kappa\cdot x} \\
&= 2\sum_{\substack{\kappa\in\mathbb{Z}^2 \\ 0\leq\arg\kappa<\pi}} \Re[\hat{f}(\kappa)e^{i(2\pi/r)\kappa\cdot x}] \\
&= \sum_{\substack{\kappa\in\mathbb{Z}^2 \\ 0\leq\arg\kappa<\pi}} \hat{f}_c(\kappa)\cos[(2\pi/r)\kappa \cdot x] - \hat{f}_s(\kappa)\sin[(2\pi/r)\kappa \cdot x].
\end{aligned} \tag{3.11}$$

We can rewrite equation 3.7 as

$$
\begin{aligned}
\|u\|_{L^2_\alpha(r\mathbb{T}^2)}^2 &= 2r^2 \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 \le \arg \kappa < \pi}} [1 + (2\pi/r)^2|\kappa|^2]^\alpha |\hat{f}(\kappa)|^2 \\
&= \frac{r^2}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 \le \arg \kappa < \pi}} [1 + (2\pi/r)^2|\kappa|^2]^\alpha [\hat{f}_c^2(\kappa) + \hat{f}_s^2(\kappa)]^2.
\end{aligned}
\tag{3.12}
$$

Let $r\bar{\Omega} \subset \mathbb{R}^2$ denote the closed disk of radius r centered at the origin. If u is an integrable distribution on $\mathbb{R}^2$ whose support is contained in $r\bar{\Omega}$, then the support of Ru is contained in $S^1 \times [-r, r]$ [Her83, thm. 2.9]. Distributions on $S^1 \times \mathbb{R}$ with support on $S^1 \times [-r, r]$ have a natural Fourier series representation and we can define an alternative Sobolev norm for them in terms of their Fourier series coefficients. The results will be seen to be quite analogous to the ones we have just obtained for $\mathbb{R}^2$.

Definition 3.3.8 Let $\mathscr{E}'(S^1 \times \mathbb{R})$ denote the space of compactly supported distributions on $S^1 \times \mathbb{R}$, i.e., the space of continuous linear functionals on $C^\infty(S^1 \times \mathbb{R})$. Let $\mathscr{E}'(S^1 \times [-r/2, r/2])$ denote the subspace of $\mathscr{E}'(S^1 \times \mathbb{R})$ whose elements have support contained in $S^1 \times [-r/2, r/2]$.

Definition 3.3.9 Let $C^\infty(S^1 \times r\mathbf{T}^1)$ denote the subspace of $C^\infty(S^1 \times \mathbb{R})$ whose elements are periodic with respect to the second variable with period r. $C^\infty(S^1 \times r\mathbf{T}^1)$ can be viewed as the set of smooth functions on the product space $S^1 \times r\mathbf{T}^1$ obtained by identifying points in $S^1 \times \mathbb{R}$ whose second coordinates differ by integer multiples of r. Let $\mathscr{E}'(S^1 \times r\mathbf{T}^1)$ denote the set of continuous linear functionals on $C^\infty(S^1 \times r\mathbf{T}^1)$. $\mathscr{E}'(S^1 \times r\mathbf{T}^1)$ can be viewed as the set of distributions on $S^1 \times r\mathbf{T}^1$. Since $C^\infty(S^1 \times r\mathbf{T}^1) \subset C^\infty(S^1 \times \mathbb{R})$, the elements of $\mathscr{E}'(S^1 \times \mathbb{R})$ can be considered as elements of $\mathscr{E}'(S^1 \times r\mathbf{T}^1)$ in a natural way. For $\kappa \in \mathbb{Z}^2$, we define the Fourier series coefficients of $v \in \mathscr{E}'(S^1 \times r\mathbf{T}^1)$ by

$$
\hat{v}(\kappa) \stackrel{\text{def}}{=} \frac{1}{2\pi r} \langle v, e^{-i[\kappa_1 \theta + (2\pi/r)\kappa_2 s]} \rangle.
$$

It is well-known that if $g \in L^2(S^1 \times r\mathbb{T}^1)$, it has the Fourier series expansion

$$g = \sum_{\kappa \in \mathbb{Z}^2} \hat{g}(\kappa) e^{i[\kappa_1 \theta + (2\pi/r)\kappa_2 s]} \tag{3.13}$$

in $L^2(S^1 \times r\mathbb{T}^1)$. We define the Sobolev space $L^2_\alpha(S^1 \times r\mathbb{T}^1)$ to be the subset of $\mathscr{E}'(S^1 \times r\mathbb{T}^1)$ such that

$$\|v\|^2_{L^2_\alpha(S^1 \times r\mathbb{T}^1)} \stackrel{\text{def}}{=} 2\pi r \sum_{\kappa \in \mathbb{Z}^2} [1 + \kappa_1^2 + (2\pi/r)^2 \kappa_2^2]^\alpha |\hat{g}(\kappa)|^2$$

is finite.

Remark 3.3.10 Results analogous to those given in remark 3.3.3 for $L^2_\alpha(r\mathbb{T}^2)$ hold for $L^2_\alpha(S^1 \times r\mathbb{T}^1)$. (Indeed, they are essentially the same space!) The spaces $L^2_\alpha(S^1 \times r\mathbb{T}^1)$ and $L^2_{-\alpha}(S^1 \times r\mathbb{T}^1)$ are dual to each other with respect to the bilinear form

$$(g, v) \mapsto 2\pi r \sum_{\kappa \in \mathbb{Z}^2} \hat{g}(\xi) \hat{v}(-\xi).$$

on $L^2_\alpha(S^1 \times r\mathbb{T}^1) \times L^2_{-\alpha}(S^1 \times r\mathbb{T}^1)$. Moreover, if $g \in C^\infty(S^1 \times r\mathbb{T}^1)$, then this bilinear form is given by $\langle v, g \rangle$, where $\langle v, g \rangle$ refers to the duality between $C^\infty(S^1 \times r\mathbb{T}^1)$ and $\mathscr{E}'(S^1 \times r\mathbb{T}^1)$. Since, for any $\alpha \geq 0$, $C^\infty(S^1 \times r\mathbb{T}^1)$ is dense in $L^2_\alpha(S^1 \times r\mathbb{T}^1)$, it follows that

$$\|v\|_{L^2_{-\alpha}(S^1 \times r\mathbb{T}^1)} = \sup_{g \in C^\infty(S^1 \times r\mathbb{T}^1) : \|g\|_{L^2_\alpha(S^1 \times r\mathbb{T}^1)} = 1} |\langle v, g \rangle|.$$

Remark 3.3.11 For $\alpha \in \mathbb{N}$, we can express $\|g\|^2_{L^2_\alpha(S^1 \times r\mathbb{T}^1)}$ in terms of the $L^2(S^1 \times \mathbb{R})$-

norms of g and its first α partial derivatives. Indeed, we have

$$
\begin{aligned}
\|g\|^2_{L^2_\alpha(S^1 \times r\mathbb{T}^1)} \ &= 2\pi r \sum_{\kappa \in \mathbb{Z}^2} [1 + \kappa_1^2 + (2\pi/r)^2 \kappa_2^2]^\alpha |\hat{g}(\kappa)|^2 \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} 2\pi r \sum_{\kappa \in \mathbb{Z}^2} [\kappa_1^2 + (2\pi/r)^2 \kappa_2^2]^j |\hat{g}(\kappa)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} 2\pi r \sum_{\kappa \in \mathbb{Z}^2} \kappa_1^{2(j-k)} [(2\pi/r)\kappa_2]^{2k} |\hat{g}(\kappa, \eta)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} 2\pi r \sum_{\kappa \in \mathbb{Z}^2} |\kappa_1^{j-k} [(2\pi/r)\kappa_2]^k \hat{g}(\kappa)|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} 2\pi r \sum_{\kappa \in \mathbb{Z}^2} \left| (\partial_\theta^{j-k} \partial_s^k g)\check{}\,(\kappa) \right|^2 \, d\xi \\
&= \sum_{j=0}^\alpha \binom{\alpha}{j} \sum_{k=0}^j \binom{j}{k} \|\partial_\theta^{j-k} \partial_s^k g\|^2_{L_2(S^1 \times \mathbb{R})},
\end{aligned}
$$

$$(3.14)$$

where the derivatives are taken in the sense of distributions.

Definition 3.3.12 We shall define $G_{1,2}(r) \subset G_{1,2}$ to be $\pi(S^1 \times [-r, r])$. The Sobolev space $L^2_\alpha(G_{1,2}(r))$ is defined to be the subspace of $\mathscr{E}'(G_{1,2})$ whose elements have lifts in $L^2_\alpha(S^1 \times \mathbb{T}^1)$. The squared norm on $L^2_\alpha(G_{1,2}(r))$ is defined to be $1/2$ of the squared norm of its lift in $L^2_\alpha(S^1 \times \mathbb{T}^1)$.

3.4 Radon transforms and Sobolev spaces

In this section, we will establish a number of rather technical results having to do with the bicontinuity of the Radon transform with respect to topologies induced by Sobolev norms. We will need to reference some results which have been stated using the standard Radon transform, $\tilde{R}$. We thus start by considering the relationship between the Sobolev norms of the Radon transform used here, R, and $\tilde{R}$.

Proposition 3.4.1 *If u is an integrable distribution on $\mathbb{R}^2$, then $\tilde{R}u \in L^2_{(0,\alpha)}(S^1 \times \mathbb{R})$ if and only $Ru \in L^2_{(0,\alpha)}(G_{1,2})$, in which case*

$$\|Ru\|_{L^2_{(0,\alpha)}(G_{1,2})} = \frac{1}{2}\|\tilde{R}u\|_{L^2_{(0,\alpha)}(S^1 \times \mathbb{R})}.$$

Proof. By definition 3.2.14, $\|Ru\|_{L^2_{(0,\alpha)}(G_{1,2})} = \frac{1}{2}\|\widetilde{Ru}\|_{L^2_{(0,\alpha)}(S^1 \times \mathbb{R})}$. The result thus follows from lemma 2.2.9. $\square$

The following proposition, due to Hertle, establishes the bicontinuity of the Radon transform with respect to appropriate Sobolev norms.

Proposition 3.4.2 *The maps*

$$R{:}L^2_\alpha(r\bar{\Omega}) \to L^2_{(0,\alpha+1/2)}(G_{1,2}(r))$$

and

$$\tilde{R}{:}L^2_\alpha(r\bar{\Omega}) \to L^2_{(0,\alpha+1/2)}(S^1 \times [-r,r])$$

are bicontinuous into for all $\alpha \in \mathbb{R}$.

Proof. The result for $\tilde{R}$ is proved in [Her83, thm. 3.1] (note that $L^2_{(0,\alpha)}(S^1 \times [-r,r])$ in our notation is equivalent to $L^2_\alpha(S^1 \times [-r,r])$ in the notation used in [Her83, p. 168]). Applying proposition 3.4.1, the result carries over to R as well. $\square$

Corollary 3.4.3 *The map $R{:}L^2(r\bar{\Omega}) \to L^2(G_{1,2}(r))$ is continuous.*

Proof. Combine proposition 3.4.2 with the fact that the natural injection

$$L^2_{(0,1/2)}(G_{1,2}(r)) \hookrightarrow L^2(G_{1,2}(r))$$

is continuous. $\square$

The Sobolev norm $\| \cdot \|_{L^2_{(0,\alpha)}(G_{1,2})}$ on $G_{1,2}$ is a convenient norm for describing continuity properties of the Radon transform. On the other hand, as we will see, statistical convergence on $G_{1,2}$ is conveniently described by the Sobolev norm $\| \cdot \|_{L^2_\alpha(G_{1,2}(r))}$. The goal of the following sequence of lemmas is to obtain analogues of proposition 3.4.2 in which the Sobolev norm $\| \cdot \|_{L^2_{(0,\alpha+1/2)}(G_{1,2})}$ is replaced by $\| \cdot \|_{L^2_{\alpha+1/2}(G_{1,2}(r+\epsilon))}$.

Lemma 3.4.4 *The map* $\tilde{R} : L^2_\alpha(r\bar{\Omega}) \to L^2_{\alpha+1/2}(S^1 \times 2r\mathbf{T}^1)$ *is continuous for all real* α *and bicontinuous into for all* $\alpha \geq -1/2$.

Proof. Suppose $\alpha \geq -1/2$. In proposition 3.4.2, it was shown that the map $\tilde{R} : L^2_\alpha(r\bar{\Omega}) \to L^2_{(0,\alpha+1/2)}(S^1 \times [-r,r])$ is bicontinuous into. It is shown below in lemma A.1.4 that the norm $\| \cdot \|_{L^2_{\alpha+1/2}(S^1 \times 2r\mathbf{T}^1)}$ is equivalent to the norm $\| \cdot \|_{L^2_{(0,\alpha+1/2)}(S^1 \times \mathbb{R})}$ on $\tilde{R}[L^2_\alpha(r\bar{\Omega})]$. It follows that the map $\tilde{R} : L^2_\alpha(r\bar{\Omega}) \to L^2_{\alpha+1/2}(S^1 \times 2r\mathbf{T}^1)$ is bicontinuous into for $\alpha \geq -1/2$. It remains to show that it is continuous for $\alpha < -1/2$. This is immediate since the natural injection $L^2_{(0,\alpha+1/2)}(S^1 \times [-r,r]) \hookrightarrow L^2_{\alpha+1/2}(S^1 \times 2r\mathbf{T}^1)$ is continuous for $\alpha \leq -1/2$ (cf. remarks 3.2.13 and 3.3.11 and apply an interpolation argument). $\square$

Definition 3.4.5 For real $\alpha < d$, we define the Riesz potential operator I^α by

$$(I^\alpha f)\hat{\ }(\xi) \stackrel{\text{def}}{=} (2\pi|\xi|)^{-\alpha} \hat{f}(\xi) \tag{3.15}$$

for functions f on $\mathbf{R}^d$ for which it makes sense (cf., [Ste70, sec. V.1]). For $f \in \mathscr{S}(\mathbb{R}^2)$, $(I^\alpha f)\hat{\ } \in L^1(\mathbf{R}^2)$, so, by the Riemann-Lebesgue lemma [Fol84, thm 8.22(f)], $I^\alpha f$ is well defined. For $f \in \mathscr{S}(\mathbf{R}^2)$ and $|\alpha| < d$, we have $I^\alpha I^{-\alpha} f = f$ [Nat86, p. 18] [Hel80,

55

thm. I.8.6]. On $\mathscr{S}(S^1 \times \mathbb{R})$, we define I^α to act on the second, or "s" variable, i.e., the operator is applied fiberwise.

Remark 3.4.6 While the definition of the Riesz potential operator given above may appear to differ from that given in [Nat86, p. 18] by a constant, they are actually the same. This is due to the fact that, as noted in remark 3.2.5, the Fourier transforms used in [Nat86] and here are defined in slightly different ways. Indeed, using the notation in remark 3.2.5, we have

$$
\begin{aligned}
(I^\alpha f)^\cup(\xi) &= (2\pi)^{-d/2}(I^\alpha f)^\wedge(\xi/2\pi) \\
&= (2\pi)^{-d/2}|\xi|^{-\alpha}\hat{f}(\xi/2\pi) \\
&= |\xi|^{-\alpha}\check{f}(\xi),
\end{aligned}
$$

which agrees with the definition of the Riesz potential operator given in [Nat86, p. 18].

Lemma 3.4.7 *The map* $I^{-1} : L^2_{\alpha+1}(\mathbb{R}^2) \to L^2_\alpha(\mathbb{R}^2)$ *is continuous and we have the estimate*

$$
\|I^{-1}f\|_{L^2_\alpha(\mathbb{R}^2)} \leq \|f\|_{L^2_{\alpha+1}(\mathbb{R}^2)}.
$$

Proof. For $f \in \mathscr{S}(\mathbb{R}^2)$, we have

$$
\begin{aligned}
\|I^{-1}f\|^2_{L^2_\alpha(\mathbb{R}^2)} &= \int_{\mathbb{R}^2}(1 + 4\pi^2|\xi|^2)^\alpha \, 4\pi^2|\xi|^2|\hat{f}(\xi)|^2 \, d\xi \\
&\leq \int_{\mathbb{R}^2}(1 + 4\pi^2|\xi|^2)^{\alpha+1}|\hat{f}(\xi)|^2 \, d\xi \\
&= \|f\|^2_{L^2_{\alpha+1}(\mathbb{R}^2)}. \quad \square
\end{aligned}
$$

Lemma 3.4.8 *The map* $\widetilde{R}^*\widetilde{R}$ *is equal to the map* $4\pi I^1$ *on* $\mathscr{S}(\mathbb{R}^2)$.

Proof. For $f \in \mathscr{S}(\mathbb{R}^2)$, we have the Radon transform inversion formula

$$
f = \frac{1}{4\pi}I^{-1}\widetilde{R}^*\widetilde{R}f
$$

[Nat86, p. 18, thm. 2.1]. Applying the map I^1 to both sides of this equation gives the identity

$$I^1 f = \frac{1}{4\pi} \tilde{R}^* \tilde{R} f,$$

i.e., $\tilde{R}^* \tilde{R} = 4\pi I^1$ on $\mathscr{S}(\mathbb{R}^2)$. $\square$

Corollary 3.4.9 *For $f \in \mathscr{S}(\mathbb{R}^2)$, we have the identity*

$$f = \frac{1}{4\pi} \tilde{R}^* \tilde{R} I^{-1} f.$$

Proof. For $f \in \mathscr{S}(\mathbb{R}^2)$, we have

$$
\begin{aligned}
f &= I^1 I^{-1} f. \\
&= \frac{1}{4\pi} \tilde{R}^* \tilde{R} I^{-1} f. \quad \square
\end{aligned}
$$

Lemma 3.4.10 *Let $\epsilon > 0$ be given. The maps $R : L^2_\alpha(r\bar{\Omega}) \to L^2_{\alpha+1/2}(G_{1,2}(r+\epsilon))$ and $\tilde{R} : L^2_\alpha(r\bar{\Omega}) \to L^2_{\alpha+1/2}(S^1 \times 2(r+\epsilon)\mathbf{T}^1)$ are continuous for all real α and bicontinuous into for $\alpha < -1/2$.*

Proof. It suffices to establish the result for $\tilde{R}$, since the result for R will then follow from proposition 3.4.1. Since $L^2_\alpha(r\bar{\Omega}) \subset L^2_\alpha((r+\epsilon)\bar{\Omega})$, the result for $\tilde{R}$ was established in lemma 3.4.4, except for the continuity of the inverse map for $\alpha < -1/2$. To show that the inverse map is continuous for $\alpha < -1/2$, suppose $\alpha < -1/2$ and $u \in L^2_\alpha(\bar{\Omega})$. Starting with equation 3.2 and using corollary 3.4.9, we can write

$$
\begin{aligned}
\|u\|_{L^2_\alpha(\mathbb{R}^2)} &= \frac{1}{4\pi} \sup_{f \in \mathscr{S}(\mathbb{R}^2) : \|f\|_{L^2_{-\alpha}(\mathbb{R}^2)} \leq 1} \left| \langle u, \tilde{R}^* \tilde{R} I^{-1} f \rangle \right| \\
&= \frac{1}{4\pi} \sup_{f \in \mathscr{S}(\mathbb{R}^2) : \|f\|_{L^2_{-\alpha}(\mathbb{R}^2)} \leq 1} \left| \langle \tilde{R} u, \tilde{R} I^{-1} f \rangle \right|.
\end{aligned}
$$

Applying lemma 3.4.7, it follows that

$$\|u\|_{L^2_\alpha(\mathbb{R}^2)} \leq \frac{1}{4\pi} \sup_{\phi \in L^2_{-(\alpha+1)}(\mathbb{R}^2) : \|\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)} \leq 1} \left| \langle \tilde{R} u, \tilde{R} \phi \rangle \right|,$$

where $\langle \tilde{R}u, \tilde{R}\phi \rangle$ refers to the duality between $L^2_{\alpha+1}(\mathbb{R}^2)$ and $L^2_{-(\alpha+1)}(\mathbb{R}^2)$.

We now want to employ the portion of lemma 3.4.4 regarding the continuity of $\tilde{R}$. To do so, we must first restrict the set over which the supremum is taken in the above inequality to functions whose support is contained in a fixed compact set. Let $\chi \in C_c^\infty(\mathbb{R}^2)$ such that $\chi = 1$ on $r\bar{\Omega}$ and $\chi = 0$ on $\mathbb{R}^2 \backslash (r + \epsilon)\Omega$. Then there exists a constant $d_1(\alpha, \chi)$ such that $\|\chi\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)} \leq d_1(\alpha, \chi)\|\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)}$ for all $\phi \in L^2_{-(\alpha+1)}(\mathbb{R}^2)$ [Fol84, thm. 8.57]. Since χ can be chosen to depend only on r and ϵ, we will write $d_1(\alpha, r, \epsilon) \stackrel{\text{def}}{=} d_1(\alpha, \chi)$. Thus if $\phi \in L^2_{-(\alpha+1)}(\mathbb{R}^2)$ with $\|\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)} \leq 1$, $\|\chi\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)} \leq d_1(\alpha, r, \epsilon)$ and $\phi = \chi\phi$ on $r\bar{\Omega}$. It follows that

$$\|u\|_{L^2_\alpha(\mathbb{R}^2)} \leq \frac{d_1(\alpha, r, \epsilon)}{4\pi} \sup_{\phi \in L^2_{-(\alpha+1)}((r+\epsilon)\bar{\Omega}):\|\phi\|_{L^2_{-(\alpha+1)}}(\mathbb{R}^2)\leq 1} \left|\langle \tilde{R}u, \tilde{R}\phi \rangle\right|.$$

By lemma 3.4.4, there exists a constant $d_2(\alpha, r + \epsilon)$ such that

$$\|\tilde{R}\phi\|_{L^2_{-(\alpha+1/2)}(S^1 \times 2(r+\epsilon)\mathbb{T}^1)} \leq d_2(\alpha, r + \epsilon)\|\phi\|_{L^2_{-(\alpha+1)}(\mathbb{R}^2)}$$

for all $\phi \in L^2_{-(\alpha+1)}(r\bar{\Omega})$. It follows that

$$\begin{aligned}
\|u\|_{L^2_\alpha(\mathbb{R}^2)} &\leq \frac{d_1(\alpha, r, \epsilon)d_2(\alpha, r + \epsilon)}{4\pi} \sup_{g^{S^1 \times 2(r+\epsilon)\mathbb{T}^1}:\|g\|_{L^2_{-(\alpha+1/2)}(S^1 \times 2(r+\epsilon)\mathbb{T}^1)}\leq 1} \left|\langle \tilde{R}u, g \rangle\right| \\
&= \frac{d_1(\alpha, r, \epsilon)d_2(\alpha, r + \epsilon)}{4\pi}\|\tilde{R}u\|_{L^2_{-(\alpha+1/2)}(S^1 \times 2(r+\epsilon)\mathbb{T}^1)}. \quad \square
\end{aligned}$$

Chapter 4

Lower Bounds on Minimax Risk of Measure Estimation

In parametric problems where the parameter space is an open subset of $\mathbb{R}^d$ and the loss function is generated by the squared Euclidean norm (resp. the Euclidean norm), the minimax risk is typically $O(n^{-1})$ (resp. $O(n^{-1/2})$), where n is the number of independent observations (cf. [Mil83, ch. VII]). In section 4.1, we shall see that a rate of $O(n^{-1})$ is not achievable in the PET problem for loss functions generated by squared Sobolev norms of order $\geq -3/2$. In section 4.2, we shall derive a slightly tighter lower bound for squared Sobolev norms of order between -1 and $-3/2$. Finally, in section 4.3, we show that the minimax risk is infinite for squared Sobolev norms of order > -1.

4.1 Loss functions induced by Sobolev norms of order $\geq -3/2$

Let S denote the unit square centered at the origin, i.e., $[-1/2 \times 1/2]^2$. Our goal in this section is to show that the minimax risk for estimating $\mu \in P(S)$ from n independent observations distributed according to $T\mu$ with respect to the loss function generated by $\| \cdot \|^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ for $\alpha \leq 3/2$ is *not* $O(n^{-1})$. We begin by considering the case where

the loss function is generated by $|| \cdot ||_{L^2_{-\alpha}(\mathbb{R}^2)}$.

Theorem 4.1.1 *The minimax risk for estimating $\mu \in P(S)$ from n independent observations distributed according to $T\mu$ is not $O(n^{-1/2})$ with respect to the loss function generated by $|| \cdot ||_{L^2_{-\alpha}(\mathbb{R}^2)}$ for $\alpha \leq 3/2$.*

Proof. We shall prove the analogous result where $|| \cdot ||_{L^2_{-\alpha}(\mathbb{R}^2)}$ is replaced by $|| \cdot ||_{L^2_{-\alpha}(\mathbb{T}^2)}$. The stated result will then follow from lemma 3.3.6.

We will consider a sequence of restrictions of the estimation problem in question where the unknown measure μ is assumed to belong to a finite-dimensional subset of $P(S)$. For each of these finite-dimensional (parametric) problems, we will compute a lower bound for the asymptotic minimax risk of the form $c_m n^{-1/2}$, where c_m is a constant. We will show that a finite-dimensional subproblem can be chosen to make c_m arbitrarily large, which implies that no bound of the form $cn^{-1/2}$ exists for the original problem.

Let $k \in \mathbb{N}^+$. We will consider finite-dimensional parametric families of probability measures that are indexed by a real vector, d, of dimension $\#\{\kappa \in \mathbb{Z}^2 : 0 < |\kappa| \leq m\}$ which is composed of two subvectors, a and b, of dimension $\frac{1}{2}\#\{\kappa \in \mathbb{Z}^2 : 0 < |\kappa| \leq m\}$ ($\#\{\cdot\}$ denotes the cardinality of the set $\{\cdot\}$). For $\kappa \in \mathbb{Z}^2$, define $\arg \kappa \overset{\text{def}}{=} \arg(\kappa_1 + i\kappa_2)$, where arg on the right-hand side is just the usual argument of a complex variable. Define the probability measure μ_d with support on S by $\mu_d \overset{\text{def}}{=} f_d \, d\lambda^2$, where $f_d : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$f_d(x) = 1_S[1 + \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 < |\kappa| \leq m \\ 0 \leq \arg \kappa < \pi}} a_\kappa \cos(2\pi\kappa \cdot x) + b_\kappa \sin(2\pi\kappa \cdot x)],$$

where a_κ and b_κ denote components of a and b, respectively. Comparison with equation 3.11 shows that a_κ and b_κ are the Fourier cosine and sine transform components of f_d, respectively. It is easily verified that if the real vector d is contained in a sufficiently small neighborhood N_m of 0, then we indeed have a parametric family of probability measures. (We need to choose N_m sufficiently small so that $f_d \geq 0$

for all $d \in N_m$). Consider the problem of estimating the vector d given n independent observations distributed according to the probability measure $\nu_d \stackrel{\text{def}}{=} T f_d \, d\theta \, ds$ on $G_{1,2}$. Let $r(d, \hat{d}_n)$ denote the risk of the estimator $\hat{d}_n$ with respect to the loss function $\|\mu_{\hat{d}} - \mu_d\|_{L^2_{-\alpha}(\mathbb{T}^2)}$. It is easy to see from equation 3.12 that

$$\|\mu_{\hat{d}} - \mu_d\|_{L^2_{-\alpha}(\mathbb{T})} = \{\frac{1}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 < |\kappa| \leq m \\ 0 \leq \arg \kappa < \pi}} (1 + 4\pi^2 |\kappa|^2)^{-\alpha}[(\hat{a}_\kappa - a_\kappa)^2 + (\hat{b}_\kappa - b_\kappa)^2]\}^{1/2}$$

($\hat{a}_\kappa$ and $\hat{b}_\kappa$ are, of course, just the components of $\hat{d}$ corresponding to a_κ and b_κ, respectively). Let F_{a_κ} and F_{b_κ} denote the diagonal elements of the Fisher information matrix for the estimation of d from one observation distributed according to ν_d that correspond to the components a_κ and b_κ, respectively. By the asymptotic minimax lower bound [Mil83, p. 146, thm. VII.2.6][Str85, thm. 83.11, rem. 83.12],

$$\liminf_{n \to \infty} n^{1/2} \sup_{a \in N_m} r(d, \hat{d}_n) \geq [\frac{1}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 < |\kappa| \leq m \\ 0 \leq \arg \kappa < \pi}} (1 + 4\pi^2 |\kappa|^2)^{-\alpha}(F_{a_\kappa}^{-1} + F_{b_\kappa}^{-1})]^{1/2}, \qquad (4.1)$$

It is shown below, in lemma 4.1.3, that F_{a_κ} and F_{b_κ}, when evaluated at $d = 0$, are $\leq |\kappa|^{-1}(3 + \log|\kappa|/\pi)$. It follows that

$$\liminf_{n \to \infty} n^{1/2} \sup_{d \in N_m} r(d, \hat{d}_n) \geq [\frac{1}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0 < |\kappa| \leq m}} (1 + 4\pi^2 |\kappa|^2)^{-\alpha} |\kappa| (3 + \log|\kappa|/\pi)^{-1}]^{1/2}. \qquad (4.2)$$

Now, by definition,

$$\liminf_{n \to \infty} n^{1/2} \sup_{d \in N_m} r(d, \hat{d}_n) = \liminf_{n \to \infty} n^{1/2} \sup_{d \in N_m} \|\hat{\mu}_d - \mu_d\|_{L^2_{-\alpha}(\mathbb{T}^2)},$$

where $\hat{\mu}_d$ is any estimator of μ_d which has Fourier coefficients of 0 for $|\kappa| > m$. But since μ_d has Fourier coefficients of 0 for $|\kappa| > m$, such a restriction does not increase

the minimax risk. We thus have

$$\liminf_{n\to\infty} n^{1/2} \sup_{\mu\in P(S)} ||\hat\mu - \mu||_{L^2_{-\alpha}(\mathbb{T}^2)}$$

$$\geq \liminf_{n\to\infty} n^{1/2} \sup_{d\in N_m} ||\hat\mu - \mu_d||_{L^2_{-\alpha}(\mathbb{T}^2)}$$

$$\geq [\frac{1}{2} \sum_{\substack{\kappa\in\mathbb{Z}^2 \\ 0<|\kappa|\leq m}} (1 + 4\pi^2|\kappa|^2)^{-\alpha}|\kappa|(3 + \log|\kappa|/\pi)^{-1}]^{1/2}.$$

Finally, we note that the series in the last line of this inequality converges as $m \to \infty$ if and only if $\alpha > 3/2$. Since this is fairly obvious for $\alpha \neq 3/2$, we shall just demonstrate that the series diverges for the "borderline" case $\alpha = 3/2$. We have

$$\sum_{\substack{\kappa\in\mathbb{Z}^2 \\ |\kappa|\geq e^{3\pi}}} (1 + 4\pi^2|\kappa|^2)^{-3/2}|\kappa|(3 + \log|\kappa|/\pi)^{-1}$$

$$\geq \sum_{\substack{\kappa\in\mathbb{Z}^2 \\ |\kappa|\geq e^{3\pi}}} (16\pi^2|\kappa|^2)^{-3/2}|\kappa|(2\log|\kappa|/\pi)^{-1}$$

$$= 2^{-7}\pi^{-2} \sum_{\substack{\kappa\in\mathbb{Z}^2 \\ |\kappa|\geq e^{3\pi}}} \frac{1}{|\kappa|^2 \log|\kappa|}$$

$$\geq 2^{-7}\pi^{-2} \int_{\substack{x\in\mathbb{R}^2 \\ |x|\geq e^{3\pi}+1}} \frac{1}{|x+1|^2 \log|x+1|}$$

$$= 2^{-6}\pi^{-1} \int_{e^{3\pi}+1}^{\infty} \frac{r\,dr}{(r+1)^2 \log(r+1)}$$

$$= 2^{-6}\pi^{-1} \int_{e^{3\pi}+2}^{\infty} \frac{r-1}{r^2 \log r}\,dr$$

$$\geq 2^{-7}\pi^{-1} \int_{e^{3\pi}+2}^{\infty} \frac{dr}{r \log r}$$

$$= 2^{-7}\pi^{-1} \log(\log r)|_{e^{3\pi}+2}^{\infty}$$

$$= \infty.$$

We conclude that for any $\alpha \leq 3/2$ and $c > 0$, there exists a finite-dimensional restriction of the PET estimation problem whose minimax risk with respect to the loss function $||\mu_{\hat d} - \mu_d||_{L^2_{-\alpha}(\mathbb{T}^2)}$ is bounded below by $cn^{-1/2}$ and hence the minimax risk for the original problem is not $O(n^{-1/2})$. $\square$

Corollary 4.1.2 *The minimax risk for estimating $\mu \in P(S)$ from n independent ob-*

servations distributed according to $T\mu$ is not $O(n^{-1})$ with respect to the loss function generated by $\|\cdot\|^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ for $\alpha \leq 3/2$.

Proof. Apply Jensen's inequality. (For a statement of this inequality, see, e.g., [Dud89, 10.2.6] or [Rud87, thm. 3.3].) $\square$

It remains to prove the lemma that was used in the proof of theorem 4.1.1.

Lemma 4.1.3 *Let F_{a_κ} and F_{b_κ} denote the diagonal elements of the Fisher information matrix corresponding to the components a_κ and b_κ, respectively, for the estimation of d from an observation distributed according to ν_d. Then F_{a_κ} and F_{b_κ}, when evaluated at $d = 0$, satisfy the inequalities*

$$F_{a_\kappa} \leq |\kappa|^{-1}(3 + \log|\kappa|/\pi)$$

and

$$F_{b_\kappa} \leq |\kappa|^{-1}(3 + \log|\kappa|/\pi).$$

Proof. We shall just give the proof for F_{a_κ}. The proof for F_{b_κ} is essentially the same.

The probability density function of the observation is given by

$$g_d \stackrel{\text{def}}{=} T f_d$$
$$= \frac{1}{\pi}\{\bar{R}(1_S) + \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ 0<|\kappa|\leq m \\ 0\leq \arg \kappa < \pi}} a_\kappa R[1_S \cos(2\pi\kappa \cdot x)] + b_\kappa R[1_S \sin(2\pi\kappa \cdot x)]\}$$

with respect to the measure $d\theta\, ds$ on $G_{1,2}$. Recall that F_{a_κ}, evaluated at $d = 0$, is given by

$$F_{a_\kappa} = E_{g_0}(\partial_{a_\kappa} \log g_d)^2,$$

where the expectation is taken over the subset of $G_{1,2}$ where $R1_S > 0$ [Mil83, p. 104].

We have

$$\partial_{a_\kappa} \log g_d = \frac{\partial_{a_\kappa} g_d}{g_d}$$
$$= \frac{R[1_S \cos(2\pi\kappa \cdot x)]}{g_d}.$$

Thus F_{a_κ}, when evaluated at $d = 0$, is given by

$$F_{a_\kappa} = \frac{1}{\pi} \int_{\phi-\pi/2}^{\phi+\pi/2} \int_{s:R1_S(\theta,s)>0} \left(\frac{R[1_S \cos(2\pi\kappa \cdot x)](\theta,s)}{R1_S(\theta,s)} \right)^2 R1_S(\theta,s) \, ds \, d\theta, \qquad (4.3)$$

where ϕ is any real number.

We will now obtain some more explicit expressions for $R1_S$ and $R[1_S \cos(2\pi\kappa \cdot x)]$. Denote the length of the intersection of the line $(\theta, s) \in G_{1,2}$ with S by $\ell(\theta, s)$. We have

$$R1_S(\theta, s) = \ell(\theta, s) \qquad (4.4)$$

and

$$R[1_S \cos(2\pi\kappa \cdot x)](\theta, s) = \int_{t:s\omega(\theta)+t\omega^\perp(\theta)\in S} \cos\{2\pi\kappa \cdot [s\omega(\theta) + t\omega^\perp(\theta)]\} \, dt$$
$$= \int_{t:s\omega(\theta)+t\omega^\perp(\theta)\in S} \cos[2\pi s\kappa \cdot \omega(\theta) + 2\pi t\kappa \cdot \omega^\perp(\theta)] \, dt.$$

There exists a unique $\phi \in [0, \pi)$ such that $\kappa/|\kappa| = \omega(\phi)$. We can thus rewrite the last equation as

$$R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)$$
$$= \int_{t:s\omega(\theta)+t\omega^\perp(\theta)\in S} \cos[2\pi s|\kappa|\omega(\phi) \cdot \omega(\theta) + 2\pi t|\kappa|\omega(\phi) \cdot \omega^\perp(\theta)] \, dt$$
$$= \int_{t:s\omega(\theta)+t\omega^\perp(\theta)\in S} \cos[2\pi s|\kappa| \cos(\theta - \phi) + 2\pi t|\kappa| \cos(\theta - \phi - \pi/2)] \, dt \qquad (4.5)$$
$$= \int_{t:s\omega(\theta)+t\omega^\perp(\theta)\in S} \cos[2\pi s|\kappa| \cos(\theta - \phi) + 2\pi t|\kappa| \sin(\theta - \phi)] \, dt.$$

We will need to use some estimates for the quantity $|R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)|$. On one hand, we have the obvious estimate

$$|R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)| \le \ell(\theta, s). \qquad (4.6)$$

64

To obtain another estimate, note, from equation 4.5, that $R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)$ is given by the integral of a cosine function of period $1/|\kappa||\sin(\theta - \phi)|$ over an interval. It is easy to see that the magnitude of an integral of a cosine function of period t over an interval cannot exceed t/π. We thus obtain the estimate

$$|R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)| \leq \frac{1}{\pi|\kappa||\sin(\theta - \phi)|}.$$

Using the inequality $|\sin\theta| \geq (2/\pi)|\theta|$, which is valid for $|\theta| \leq \pi/2$, we obtain the simpler estimate

$$|R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)| \leq \frac{1}{2|\kappa||\theta - \phi|} \tag{4.7}$$

for $|\phi - \theta| \leq \pi/2$.

We will now use the estimates on $|R[1_S \cos(2\pi\kappa \cdot x)](\omega, s)|$ to obtain an estimate of F_{a_κ}. We can decompose the right-hand side of equation 4.3 into

$$
\begin{aligned}
F_{a_\kappa} \\
= \frac{1}{\pi} \int_{0 \leq |\theta - \phi| \leq 1/|\kappa|} \int_{s:\ell(\theta,s)>0} \left(\frac{R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)}{R1_S(\theta, s)} \right)^2 R1_S(\theta, s)\, ds\, d\theta, \\
+ \frac{1}{\pi} \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:0<\ell(\theta,s)\leq 1/|\kappa||\theta-\phi|} \left(\frac{R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)}{R1_S(\theta, s)} \right)^2 R1_S(\theta, s)\, ds\, d\theta \\
+ \frac{1}{\pi} \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:\ell(\theta,s)>1/|\kappa||\theta-\phi|} \left(\frac{R[1_S \cos(2\pi\kappa \cdot x)](\theta, s)}{R1_S(\theta, s)} \right)^2 R1_S(\theta, s)\, ds\, d\theta.
\end{aligned}
\tag{4.8}
$$

Using equation 4.6, the first term on the right-hand side of equation 4.8 is bounded above by

$$
\begin{aligned}
\frac{1}{\pi} & \int_{0 \leq |\theta - \phi| \leq 1/|\kappa|} \int_{s:\ell(\theta,s)>0} \ell(\theta, s)\, ds\, d\theta \\
& \leq \frac{1}{\pi} \int_{0 \leq |\theta - \phi| \leq 1/|\kappa|} \int_{-1}^{1} \sqrt{1 - s^2}\, ds\, d\theta \\
& = \frac{1}{\pi} \cdot \frac{2}{|\kappa|} \cdot \frac{\pi}{2} \\
& = \frac{1}{|\kappa|},
\end{aligned}
\tag{4.9}
$$

where we used the result

$$\int_0^1 \sqrt{1 - s^2} = \frac{\pi}{4}$$

[Sel72, p. 427, eq. 200]. Using equation 4.6 again, the second term on the right-hand side of equation 4.8 is bounded above by

$$
\begin{aligned}
\frac{1}{\pi} & \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:0 < \ell(\theta,s) \leq 1/|\kappa||\theta - \phi|} \ell(\theta, s) \, ds \, d\theta \\
& \leq \frac{1}{\pi|\kappa|} \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:0 < \ell(\theta,s) \leq 1/|\kappa||\theta - \phi|} \frac{1}{|\theta - \phi|} \, ds \, d\theta \\
& \leq \frac{1}{\pi|\kappa|} \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \frac{2}{|\kappa||\theta - \phi|^2} \, d\theta \\
& = \frac{4}{\pi|\kappa|^2} \int_{1/|\kappa|}^{\pi/2} \frac{1}{\theta^2} \, d\theta. \\
& = \frac{4}{\pi|\kappa|^2}(|\kappa| - 2/\pi) \\
& < \frac{4}{\pi|\kappa|},
\end{aligned}
\tag{4.10}
$$

where, in the second inequality, we used the fact that the set $\{s : 0 < \ell(\theta, s) \leq 1/|\kappa||\theta - \phi|\}$ consists of two intervals of length less than $1/|\kappa||\theta - \phi|$. Using equation 4.7, the third term on the right-hand side of equation 4.8 is bounded above by

$$
\begin{aligned}
\frac{1}{\pi} & \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:\ell(\theta,s) > 1/|\kappa||\theta - \phi|} \frac{1}{4|\kappa|^2|\theta - \phi|^2\ell(\theta, s)} ds \, d\theta \\
& \leq \frac{1}{\pi} \int_{1/|\kappa| \leq |\theta - \phi| \leq \pi/2} \int_{s:\ell(\theta,s) > 1/|\kappa||\theta - \phi|} \frac{1}{4|\kappa||\theta - \phi|} ds \, d\theta \\
& < \frac{1}{\pi|\kappa|} \int_{1/|\kappa|}^{\pi/2} \frac{1}{\theta} \, d\theta \\
& = \frac{1}{\pi|\kappa|}[\log(\pi/2) + \log|\kappa|] \\
& < \frac{1}{\pi|\kappa|}(1/2 + \log|\kappa|).
\end{aligned}
\tag{4.11}
$$

Using the estimates obtained in equations 4.9, 4.10, and 4.11 in equation 4.8, we get

$$
\begin{aligned}
F_{a_\kappa} & < \frac{1}{|\kappa|} + \frac{4}{\pi|\kappa|} + \frac{1}{2\pi|\kappa|} + \frac{\log|\kappa|}{\pi|\kappa|} \\
& = \frac{2\pi + 9}{2\pi|\kappa|} + \frac{\log|\kappa|}{\pi|\kappa|} \\
& < |\kappa|^{-1}(3 + \log|\kappa|/\pi). \quad \square
\end{aligned}
$$

Remark 4.1.4 Corollary 4.1.2 concerns estimators of μ that are based on n independent observations distributed according to $T\mu$. It is useful to contrast this result

66

with one concerning estimators of μ that are based on n independent observations distributed according to μ itself. In this case, a (much simpler) analogue of lemma 4.1.3 shows that the corresponding F_{a_κ} and F_{b_κ} are uniformly bounded below by a positive constant. Using this result, one may derive a result analogous to corollary 4.1.2 stating that the minimax risk is not $O(n^{-1})$ for $\alpha \leq 1$.

4.2 Loss functions induced by Sobolev norms of order $-3/2$ to -1

In section 4.1, we showed that the minimax risk in the PET problem with respect to the loss function $||\hat{\mu} - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ is not $O(n^{-1})$ for $\alpha \leq 3/2$. In this section, we will obtain some tighter lower bounds for $1 \leq \alpha < 3/2$.

Theorem 4.2.1 *Let* $1 \leq \alpha < \beta < 3/2$. *The minimax risk with respect to the loss function* $||\hat{\mu} - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ *for estimating* $\mu \in P(S)$ *from* n *independent observations distributed according to* $T\mu$ *is bounded below by* $O(n^{-2(\beta+1)/5})$.

Proof. We shall prove the analogous result where $|| \cdot ||_{L^2_{-\alpha}(\mathbb{R}^2)}$ is replaced by $|| \cdot ||_{L^2_{-\alpha}(\mathbb{T}^2)}$. The stated result will then follow from lemma 3.3.6.

We will apply the procedure used in [JS90, sec. 5]. This procedure gives a lower bound for the estimation of a probability measure of interest from n independent observations distributed according a probability measure which is a linear function of the probability measure of interest. The possible probability measures of interest are assumed to be representable by probability density functions with respect to some common measure on the observation space. These density functions are assumed to belong to a subset $\mathscr{F}$ of an inner product space H. $\mathscr{F}$ is assumed to be a translate of a set, H_0, that is balanced about the origin. If the probability density of interest is $f \in \mathscr{F}$, the probability measure of the observations is assumed to be represented by the probability density function Bf with respect to a common measure, which we shall denote by ζ. These density functions are assumed to lie in an inner product space, K. The function B is assumed to extend to a linear operator $B : H \to K$.

Define $\mathscr{G} \stackrel{\text{def}}{=} B\mathscr{F}$ and let $I(g, g')$ denote the Kullback-Leibler information number, $\int \log(g/g')g \, d\zeta$. It is also assumed that there exists a constant c such that $I(g, g') \leq c||g - g'||^2_K$ for all $g, g' \in \mathscr{G}$. This condition will be satisfied if the densities in $\mathscr{G}$ are uniformly bounded above and below away from 0.

For a finite-dimensional subspace $M \subset H$, denote the dimension of M by $|M|$ and the ball of radius δ centered at $0 \in M$ by $B_M(\delta)$. The norm of the restriction of B to

M is denoted by $||B||_M \overset{\text{def}}{=} \sup_{f \in M} \{||Bf||_K / ||f||_H\}$. Define $\mathcal{M}_\delta \overset{\text{def}}{=} \{M : B_M(\delta) \subset H_0\}$ and the modulus function

$$\sigma(\delta) \overset{\text{def}}{=} \delta \inf\{||B||_M / |M|^{1/2} : M \in \mathcal{M}_\delta\}.$$

The function σ is strictly increasing and we define the function τ to be the left-continuous inverse of σ. Proposition 5.1 of [JS90] states that, under the conditions given in the preceding paragraphs, there exist constants d_1 and d_2 such that

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} E_f ||\hat{f}_n - f||_H^2 \geq d_1 \tau^2(d_2 n^{-1/2}), \tag{4.12}$$

where the infimum is taken over all estimators $\hat{f}$ that are based on n independent observations distributed according to Bf. (The result in [JS90] is actually stated for Hilbert spaces H and K, but proof works for inner product spaces as well.)

To apply this result, we consider a restriction of the PET problem where the probability measure of interest is assumed to be representable by a probability density function on S with respect to Lebesgue measure whose values are contained in $[1/2, 3/2]$. We can then identify the set of allowable probability measures of interest with the subset $\mathcal{F} \subset L^2(S)$ whose values are contained in $[1/2, 3/2]$. We can write $\mathcal{F}$ as $1_S + H_0$, where 1_S is the indicator function of S and H_0 is the subset of $L^2_{-\alpha}(\mathbf{T}^2)$ consisting of functions in $L^2(S)$ whose values are contained in $[-1/2, 1/2]$. H_0 is clearly balanced about the origin. We will take H to be $L^2(S)$ equipped with the norm induced by $L^2_{-\alpha}(\mathbf{T}^2)$.

If the probability measure of interest is represented by $f \in H_0$, the probability measure of the observations is given by the probability measure which is represented by the probability density function Tf with respect to the measure $ds\, d\theta$ on $G_{1,2}$. In order to get probability density functions that are uniformly bounded above and below away from zero, we will instead represent the probability measures of the observations by probability density functions with respect to the measure $\zeta \overset{\text{def}}{=} T1_S\, ds\, d\theta$ on $G_{1,2}$.

Define $K = L^2(\zeta)$ and the linear map $B : H \to K$ by

$$Bf(\theta, s) = \begin{cases} Rf(\theta, s)/R1_S(\theta, s) & T1_S(\theta, s) \neq 0 \\ 0 & T1_S(\theta, s) = 0 \end{cases}.$$

Then, if the probability measure of interest is represented by $f \in H_0$, the probability measure of the observations is represented by Bf with respect to the measure ζ on $G_{1,2}$.

For $\lambda \geq 1$, we define M_λ to be the subspace of H spanned by the functions $\cos(2\pi\kappa \cdot x)$ and $\sin(2\pi\kappa \cdot x)$ for $\kappa \in \mathbb{Z}^2$ with $\lambda/2 \leq |\kappa| \leq \lambda$ and $0 \leq \arg \kappa < \pi$. For $\delta > 0$, define $\lambda(\delta) \stackrel{\text{def}}{=} d_2(\alpha)\delta^{-1/(\alpha+1)}$, where $d_2(\alpha) \stackrel{\text{def}}{=} (3\pi)^{-\alpha/(\alpha+1)}2^{-4/(\alpha+1)}$. It is shown below, in lemma 4.2.2, that, if $\lambda(\delta) \geq 1$, $M_{\lambda(\delta)} \in \mathcal{M}_\delta$. Thus, if δ is sufficiently small, $\sigma(\delta) \leq \delta\|B\|_{M_{\lambda(\delta)}}/|M_{\lambda(\delta)}|^{1/2}$. It is shown below, in lemma 4.2.3, that, if $\lambda(\delta) \geq 2$,

$$\|B\|_{M_{\lambda(\delta)}} = O(\lambda^{\beta-1/2}(\delta)).$$

Since, by lemma 4.2.4 below, $|M_\lambda| \geq c_1\lambda^2$, where $c_1 \stackrel{\text{def}}{=} 1/[4(\sqrt{2}+1)^2]$, we have, for δ sufficiently small,

$$\begin{aligned} \sigma(\delta) &\leq \delta\frac{O(\lambda^{\beta-1/2}(\delta))}{\sqrt{c_1}\lambda(\delta)} \\ &= \delta O(\lambda^{\beta-3/2}(\delta)) \\ &= \delta O(\delta^{(3/2-\beta)/(\alpha+1)}) \\ &\leq \delta O(\delta^{(3/2-\beta)/(\beta+1)}) \\ &= O(\delta^{5/2(\beta+1)}). \end{aligned}$$

It follows that, for sufficiently small ϵ, there exists a constant d such that

$$\tau(\epsilon) \geq d\epsilon^{2(\beta+1)/5},$$

hence

$$\tau^2(\epsilon) \geq d^2\epsilon^{4(\beta+1)/5}.$$

We conclude that

$$\tau^2(n^{-1/2}) \geq d^2 n^{-2(\beta+1)/5}.$$

for sufficiently large n. The desired result now follows by applying equation 4.12 $\square$

It remains to prove the lemmas which were used in the proof of theorem 4.2.1.

Lemma 4.2.2 *Define* $\lambda = \lambda(\delta) = d_2(\alpha)\delta^{-1/(\alpha+1)}$, *where*

$$d_2(\alpha) \overset{\text{def}}{=} (3\pi)^{-\alpha/(\alpha+1)} 2^{-4/(\alpha+1)}.$$

Then, if $\lambda(\delta) \geq 1$, $M_\lambda \in \mathcal{M}_\delta$.

Proof. Suppose $f \in B_{M_\lambda}(\delta)$. Since $f \in M_\lambda$, we can write f as

$$f(x) = 1_S \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ \lambda/2 \leq |\kappa| \leq \lambda \\ 0 \leq \arg \kappa < \pi}} a_\kappa \cos(2\pi\kappa \cdot x) + b_\kappa \sin(2\pi\kappa \cdot x).$$

By equation 3.12,

$$\|f\|^2_{L^2_{-\alpha}(\mathbb{T}^2)} = \frac{1}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ \lambda/2 \leq |\kappa| \leq \lambda \\ 0 \leq \arg \kappa < \pi}} (1 + 4\pi^2|\kappa|^2)^{-\alpha}(a_\kappa^2 + b_\kappa^2).$$

Since $f \in B_{M_\lambda}(\delta)$, its coefficients must satisfy the inequality

$$\frac{1}{2} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ \lambda/2 \leq |\kappa| \leq \lambda \\ 0 \leq \arg \kappa < \pi}} (1 + 4\pi^2|\kappa|^2)^{-\alpha}(a_\kappa^2 + b_\kappa^2) < \delta^2$$

and, *a fortiori*,

$$\frac{1}{2}(3\pi)^{-2\alpha}\lambda^{-2\alpha} \sum_{\substack{\kappa \in \mathbb{Z}^2 \\ \lambda/2 \leq |\kappa| \leq \lambda \\ 0 \leq \arg \kappa < \pi}} a_\kappa^2 + b_\kappa^2 < \delta^2.$$

71

Solving the equation $\lambda = d_2(\alpha)\delta^{-1/(\alpha+1)}$ for δ gives $\delta = (\lambda/d_2(\alpha))^{-(\alpha+1)}$. Substituting this expression for δ into the last display gives the sequence of inequalities

$$\frac{1}{2}(3\pi)^{-2\alpha}\lambda^{-2\alpha} \sum_{\substack{\kappa \in Z^2 \\ \lambda/2 \le |\kappa| \le \lambda \\ 0 \le \arg \kappa < \pi}} a_\kappa^2 + b_\kappa^2 < (\lambda/d_2(\alpha))^{-2(\alpha+1)},$$

$$32\lambda^2 \sum_{\substack{\kappa \in Z^2 \\ \lambda/2 \le |\kappa| \le \lambda \\ 0 \le \arg \kappa < \pi}} a_\kappa^2 + b_\kappa^2 < 1/4.$$

Using the general inequality

$$\left(\sum_{i=1}^n x_i\right)^2 \le n \sum_{i=1}^n x_i^2 \tag{4.13}$$

and lemma 4.2.4 below, it now follows that

$$\sum_{\substack{\kappa \in Z^2 \\ \lambda/2 \le |\kappa| \le \lambda \\ 0 \le \arg \kappa < \pi}} |a_\kappa| + |b_\kappa| \le 1/2. \tag{4.14}$$

It is clear that this last condition insures that $f \in H_0$. $\square$

Lemma 4.2.3 *For $\lambda \ge 2$ and $1 \le \alpha < \beta < 3/2$*

$$\|B\|_{H_\lambda} = O(\lambda^{\beta-1/2}).$$

Proof. M_λ is spanned by the functions of the form $1_S \cos(2\pi\kappa \cdot x)$ and $1_S \sin(2\pi\kappa \cdot x)$ with $\lambda/2 \le |\kappa| \le \lambda$. It therefore suffices to that $\|Bf\|_K/\|f\|_H = O(\lambda^{\beta-1/2})$ for functions f of this form. We will prove this for f of the form $1_S \cos(2\pi\kappa \cdot x)$. The proof for f of the form $1_S \sin(2\pi\kappa \cdot x)$ is essentially the same.

Define $\ell(\theta, s)$ to be the length of the intersection of the line $(\theta, s) \in G_{1,2}$ with S

and $g_0 \overset{\text{def}}{=} T1_S$. Then

$$
\begin{aligned}
\|Bf\|_K^2 &= \|g/g_0\|_{L^2(\zeta)}^2 \\
&= \frac{1}{\pi} \int_{\phi-\pi/2}^{\phi+\pi/2} \int_{s:\ell(\theta,s)>0} \left(\frac{R[1_S \cos(2\pi\kappa \cdot x)](\theta,s)}{R1_S(\theta,s)} \right)^2 R1_S(\theta,s)\, ds d\theta.
\end{aligned}
$$

where θ is any real number. The last quantity on the right-hand side of this equation is just the quantity F_{a_κ} in theorem 4.1.1 (see equation 4.3). Using the upper bound for this quantity obtained in that proof, we get

$$
\begin{aligned}
\|Bf\|_K^2 &\leq |\kappa|^{-1}(3 + \log|\kappa|/\pi) \\
&\leq 2|\lambda|^{-1}(3 + \log|\lambda|/\pi) \\
&= O(\lambda^{-1}\log\lambda).
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\|f\|_H^2 &= \frac{1}{2}(1 + 4\pi^2|\kappa|^2)^{-\alpha} \\
&\geq \frac{1}{2}(3\pi)^{-2\alpha}\lambda^{-2\alpha}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{\|Bf\|_K}{\|f\|_H} &\leq \frac{O(\lambda^{-1/2}\log^{1/2}\lambda)}{2^{-1/2}(3\pi)^{-\alpha}\lambda^{-\alpha}} \\
&= O(\lambda^{\alpha-1/2}\log^{1/2}\lambda) \\
&= O(\lambda^{\beta-1/2}). \quad \square
\end{aligned}
$$

Lemma 4.2.4 *If $\lambda \geq 1$, then*

$$
c_1\lambda^2 \leq |M_\lambda| \leq c_2\lambda^2, \tag{4.15}
$$

where $c_1 \overset{\text{def}}{=} 1/[4(\sqrt{2}+1)^2]$ and $c_2 \overset{\text{def}}{=} 18$.

Proof. For $\lambda \in \mathbb{R}$, let $\bar{\lambda}$ and $\underline{\lambda}$ denote the smallest integer $\geq \lambda$ and the largest integer $\leq \lambda$, respectively. Now $|M_\lambda| \geq 4$ for all $\lambda \geq 1$, since $\lambda/2 \leq |(\pm\underline{\lambda}, 0)|, |(0, \pm\underline{\lambda})| \leq \lambda$.

73

Thus for $\lambda < 4(\sqrt{2} + 1)$,

$$
\begin{aligned}
|M_\lambda| &\geq 4 \\
&= \frac{16(\sqrt{2} + 1)^2}{4(\sqrt{2} + 1)^2} \\
&> \frac{1}{4(\sqrt{2} + 1)^2} \lambda^2.
\end{aligned}
$$

On the other hand, for $\lambda \geq 4(\sqrt{2} + 1)$

$$
\begin{aligned}
|M_\lambda| &= \#\{\kappa \in \mathbb{Z}^2 : \lambda/2 \leq |\kappa| \leq \lambda\} \\
&= \#\{\kappa \in \mathbb{Z}^2 : |\kappa| \leq \lambda\} - \#\{\kappa \in \mathbb{Z}^2 : |\kappa| < \lambda/2\} \\
&\geq \#\{\kappa \in \mathbb{Z}^2 : |\kappa_1|, |\kappa_2| \leq \lambda/\sqrt{2}\} - \#\{\kappa \in \mathbb{Z}^2 : |\kappa_1|, |\kappa_2| < \overline{\lambda/2}\} \\
&= (2\underline{\lambda/\sqrt{2}} + 1)^2 - (2\overline{\lambda/2} - 1)^2 \\
&\geq (\sqrt{2}\lambda - 1)^2 - (\lambda + 1)^2 \\
&= 2\lambda^2 - 2\sqrt{2}\lambda + 1 - (\lambda^2 + 2\lambda + 1) \\
&= \lambda^2 - 2(\sqrt{2} + 1)\lambda \\
&\geq \lambda^2/2.
\end{aligned}
$$

This gives the left-hand side of equation 4.15. The right-hand side of equation 4.15 holds since

$$
\begin{aligned}
|M_\lambda| &= \#\{\kappa \in \mathbb{Z}^2 : \lambda/2 \leq |\kappa| \leq \lambda\} \\
&\leq \#\{\kappa \in \mathbb{Z}^2 : |\kappa| \leq \bar{\lambda}\} \\
&\leq \#\{\kappa \in \mathbb{Z}^2 : |\kappa_1|, |\kappa_2| \leq \bar{\lambda}\} \\
&= (2\bar{\lambda} + 1)^2 \\
&= 4\bar{\lambda}^2 + 4\bar{\lambda} + 1 \\
&\leq 9\bar{\lambda}^2 \\
&\leq 18\lambda^2.
\end{aligned}
$$

Remark 4.2.5 While the results in this section provide a lower bound on the rate of convergence of the minimax risk, I would conjecture that they are, in fact, not very tight. I would make the further conjecture that, at least under the slightly stronger regularity conditions used in section 6.4, the actual minimax rates are $O(n^{-2(\alpha-1)})$. This conjecture is based upon preliminary calculations of the performance of estimators for estimating the Fourier coefficients of the measure of interest which are

74

constructed along lines discussed in chapter 6.

Remark 4.2.6 The procedure in [JS90, sec. 5] that we used in this section stems from a technique pioneered by Birgé [Bir83]. In essence, it involves constructing large finite sets of allowable probability densities that are at a distance of $\geq \delta$ from each other with respect to the appropriate Sobolev norm, but whose Radon transforms are at a distance $\leq \eta$ with respect to the L^2 norm. The best results are obtained by choosing sets which satisfy such conditions whose cardinality is as large as possible. Since one is dealing with Sobolev norms, it is natural to specify the probability densities by their Fourier coefficients. One then needs to make sure that the functions specified are actually probability density functions, i.e., that they are nonnegative. This is, in part, why the densities in our proof are constrained to have values in $[1/2, 3/2]$. In the preceding proof, we have used the inequality given by 4.13 and condition given by equation 4.14 to ensure the satisfaction of this constraint. It is clear that this is a pretty heavy-handed approach; there are many densities which satisfy the constraint, but not these conditions. However, it appears to be very difficult to characterize the maximal cardinality of sets satisfying all the necessary conditions. As a result, we conjecture that the results are suboptimal.

4.3 Loss functions induced by Sobolev norms of order > -1

In this section, we will show that the minimax risk for estimating $\mu \in P(S)$ from n independent observations distributed according to $T\mu$ with respect to the loss function $\|\hat{\mu}_n - \mu\|_{L^2_{-\alpha}(\mathbb{R}^2)}$ for $\alpha < 1$ is infinite. If fact, we shall prove the stronger statement that this is true for estimating $\mu \in P(S)$ from n independent observations distributed according to μ.

The idea of the proof can be briefly sketched. Suppose $\hat{\mu}_n$ is an estimator of μ based on n independent observations distributed according to μ and that $\sup_{\mu \in P(S)} E_\mu \|\hat{\mu}_n - \mu\|_{L^2_{-\alpha}(\mathbb{R}^2)} < \infty$. If $\phi \in L^2_\alpha(\mathbb{R}^2)$, denote the functional $P(S) \to \mathbb{R}$ given by $\mu \mapsto \int_{\mathbb{R}^2} \phi \, d\mu$ by $\dot{\phi}$. Then $\dot{\phi}(\hat{\mu}_n)$ is an estimator of the functional $\dot{\phi}(\mu)$ which satisfies

$$
\begin{aligned}
\sup_{\mu \in P(S)} E_{\mu^n} |\dot{\phi}(\hat{\mu}_n) - \dot{\phi}(\mu)| &= \sup_{\mu \in P(S)} E_{\mu^n} |(\hat{\mu}_n - \mu)(\phi)| \\
&\le \|\phi\|_{L^2_\alpha(\mathbb{R}^2)} \sup_{\mu \in P(S)} E_{\mu^n} \|\hat{\mu}_n - \mu\|_{L^2_{-\alpha}(\mathbb{R}^2)}.
\end{aligned}
\tag{4.16}
$$

For each $\alpha < 1$, we will construct an $\phi \in L^2_\alpha(\mathbb{R}^2)$ such that

$$
\inf_{\dot{\phi}_n} \sup_{\mu \in P(S)} E_{\mu^n} |\hat{\dot{\phi}}_n - \dot{\phi}(\mu)| = \infty.
$$

Equation 4.16 will then imply that

$$
\inf_{\hat{\mu}_n} \sup_{\mu \in P(S)} E_{\mu^n} \|\hat{\mu}_n - \mu\|_{L^2_{-\alpha}(\mathbb{R}^2)} = \infty.
$$

Theorem 4.3.1 *Let $\alpha < 1$ be given. There exists a function $\phi \in L^2_\alpha(\mathbb{R}^2)$ such that*

$$
\inf_{\dot{\phi}_n} \sup_{\mu \in P(S)} E_{\mu^n} |\hat{\dot{\phi}}_n - \dot{\phi}(\mu)| = \infty,
$$

where the infimum is taken over all estimators $\hat{\dot{\phi}}_n$ of $\dot{\phi}(\mu)$ based on n independent observations distributed according to μ.

Proof. If ν and ν' are probability measures on the measurable space X, let $H(\nu, \nu')$ denote the Hellinger distance between ν and ν'. If ν and ν' are represented by the densities g and g', respectively, with respect to some σ-finite measure λ, then $H(\nu, \nu')$ is given by

$$H(\nu, \nu') = \frac{1}{\sqrt{2}} \|\sqrt{g} - \sqrt{g'}\|_{L^2(X,\lambda)}$$

[Str85, def. 2.7]. Define the modulus of continuity of the functional $\dot{\phi}$ over $P(S)$ by

$$\omega(\epsilon) \stackrel{\text{def}}{=} \sup\{|\dot{\phi}(\mu) - \dot{\phi}(\mu')| : H(\mu', \mu) \leq \epsilon, \mu, \mu' \in P(S)\}.$$

We will use the lower bound

$$\inf_{\hat{\phi}_n} \sup_{\mu \in P(S)} E_{\mu^n} |\hat{\phi}_n - \dot{\phi}(\mu)| \geq c\omega(n^{-1/2}) \tag{4.17}$$

for some $c > 0$ [DL91]. The proof will proceed by constructing a suitable $\phi \in L^2_\alpha(\mathbb{R}^2)$ and showing that the modulus of continuity $\omega(\epsilon)$ is infinite for all $\epsilon > 0$. Note that, since $L^2_{\alpha'}(\mathbb{R}^2) \subset L^2_\alpha(\mathbb{R}^2)$ for $\alpha' < \alpha$, it suffices to prove the theorem for α satisfying $0 < \alpha < 1$. We shall therefore assume, without loss of generality, that $0 < \alpha < 1$.

We start by constructing a function $\phi \in L^2_\alpha(\mathbb{R}^2)$. Define the function $\chi \in C^\infty_c(\mathbb{R}^2)$ by

$$\chi(x) = \begin{cases} e^{-1/(1-16|x|^2)} & |x| \leq 1/4 \\ 0 & |x| > 1/4 \end{cases},$$

cf. [Fol84, exer. 8.3]. Define $\lambda \stackrel{\text{def}}{=} (\alpha - 1)/2$ and the function $\phi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi(x) = \begin{cases} \chi(x)|x|^\lambda & \text{if } |x| \neq 0 \\ 0 & \text{if } |x| = 0 \end{cases}.$$

It is shown below in lemma 4.3.4 that $\phi \in L^2_\alpha(\mathbb{R}^2)$.

We now want to show that the modulus of continuity of the functional $\dot{\phi}$ is infinite for all $\epsilon > 0$. Since the modulus of continuity is obviously an increasing function, it suffices to prove that it is infinite for $0 < \epsilon < 1$. Let $0 < \epsilon < 1$ be given. Define the

measure μ to be the uniform probability measure with respect to Lebesgue measure on the annulus of inner radius $1/4$ and outer radius $1/2$ centered at the origin. For $0 < \eta < 1/4$, define the measure μ_η to be the sum of $(1 - \epsilon^2)\mu$ and the positive measure of total mass ϵ^2 whose density is uniform with respect to Lebesgue measure on the disk of radius η and 0 outside this disk. We obviously have $||\mu - \mu_\eta||_v = \epsilon^2$. Now, for any probability measures μ and μ',

$$H^2(\mu', \mu) \leq ||\mu' - \mu||_v \tag{4.18}$$

[Str85, lem. 2.15], so $H(\mu, \mu_\eta) \leq \epsilon$.

We now compute the difference between the values of the functional $\dot{\phi}$ on μ and μ_η. We have

$$\begin{aligned} \dot{\phi}(\mu_\eta) - \dot{\phi}(\mu) &= \dot{\phi}(\mu_\eta) \\ &\geq \epsilon^2 e^{-1/(1-16\eta^2)} \eta^\lambda. \end{aligned}$$

The last expression approaches infinity as $\eta \to 0$. Thus, for any ϵ, $M > 0$, by taking η sufficiently small, we have $H(\mu, \mu_\eta) \leq \epsilon$ and $|\phi(\mu) - \phi(\mu_\eta)| > M$. This says that the modulus of continuity $\omega(\epsilon)$ is equal to infinity for $\epsilon > 0$. $\square$

Corollary 4.3.2 *Let $\alpha < 1$ be given. Then*

$$\inf_{\dot{\phi}_n} \sup_{\mu \in P(S)} E_{(T\mu)^n} |\hat{\phi}_n - \dot{\phi}(\mu)| = \infty,$$

where the infimum is taken over all estimators $\hat{\phi}_n$ of $\dot{\phi}(\mu)$ based on n independent observations distributed according to $T\mu$.

Proof. It is clear that, given observations distributed according to μ, one could construct a derived experiment in which the observations would be distributed according to $T\mu$. Thus, given any estimator based on observations distributed according to $T\mu$, one could construct a (randomized) estimator with the same risk that is based on observations distributed according to μ. This reasoning is usually expressed by saying that the experiment with observations distributed according to $T\mu$ is a randomization

of the experiment with observations distributed according to μ [Str85, sec. 55]. Thus the minimax risk for the experiment with observations distributed according to $T\mu$ is bounded below by the minimax risk for the experiment with observations distributed according to μ. $\square$

Corollary 4.3.3 *Let $\alpha < 1$ be given. Then*

$$\inf_{\hat{\mu}_n} \sup_{\mu \in P(S)} E||\hat{\mu}_n - \mu|| = \infty,$$

and

$$\inf_{\hat{\mu}_n} \sup_{\mu \in P(S)} E||\hat{\mu}_n - \mu||^2 = \infty,$$

where the infimum is taken over all estimators $\hat{\mu}_n$ of μ based on n independent observations distributed according to μ or $T\mu$.

Proof. To prove the first equation, let ϕ be as in theorem 4.3.1. Assume, to obtain a contradiction, that $\inf_{\hat{\mu}_n} \sup_{\mu \in P(S)} E||\hat{\mu}_n - \mu|| < \infty$. Then, by equation 4.16,

$$\inf_{\hat{\phi}_n} \sup_{\mu \in P(D)} E|\hat{\phi}_n - \phi(\mu)| \leq ||\phi||_{L^2_\alpha(\mathbb{R}^2)} \inf_{\hat{\mu}_n} \sup_{\mu \in P(S)} E||\hat{\mu}_n - \mu||_{L^2_{-\alpha}(\mathbb{R}^2)}$$

$$< \infty,$$

which contradicts theorem 4.3.1 and corollary 4.3.2. This proves the first equation and the second follows from the first by Jensen's inequality. $\square$

It remains to prove the lemma that was used in the proof of theorem 4.3.1.

Lemma 4.3.4 *Let $0 < \alpha < 1$ be given and let $\lambda = (\alpha - 1)/2$. Define the function $\chi : \mathbb{R}^2 \to \mathbb{R}$ by*

$$\chi(x) = \begin{cases} e^{-1/(1-16|x|^2)} & |x| \leq 1/4 \\ 0 & |x| > 1/4 \end{cases}.$$

Then the function $\phi : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\phi(x) = \begin{cases} \chi(x)|x|^\lambda & \text{if } |x| \neq 0 \\ 0 & \text{if } |x| = 0 \end{cases}$$

79

is in $L_\alpha^2(\mathbb{R}^2)$.

Proof.   We have to show that the quantity

$$\|\phi\|_{L_\alpha^2(\mathbb{R}^2)}^2 = \int_{\mathbb{R}^2} (1 + 4\pi^2 |\xi|^2)^\alpha \left|\hat{\phi}(\xi)\right|^2 d\xi \tag{4.19}$$

is finite. Let $\psi_\beta : \mathbb{R}^2 \to \mathbb{R}$ denote the function given by

$$\psi_\lambda(x) = \begin{cases} |x|^\beta & \text{if } |x| \neq 0 \\ 0 & \text{if } |x| = 0 \end{cases}.$$

Then, if $\beta \notin 2\mathbb{Z}$, the Fourier transform of ψ_β is given by

$$\hat{\psi}_\beta(\xi) = d_1(\beta)\psi_{-2-\beta}(\xi),$$

where

$$d_1(\beta) \stackrel{\text{def}}{=} \frac{2^{1+\beta}\Gamma((2+\beta)/2)}{\Gamma(-\beta/2)}$$

and the finite part of ψ_γ is understood when $\gamma < -2$ [Itô85, app. A, tab. 11.II]. Using this notation, we can write

$$\phi = \chi\psi_\lambda.$$

Now $\chi \in C_c^\infty(\mathbb{R}^2)$, hence $\chi \in \mathscr{O}_M(\mathbb{R}^2)$, where $\mathscr{O}_M(\mathbb{R}^2)$ denotes the space of smooth functions slowly increasing at infinity [Tre67, p. 275, def. 25.3]. Also ψ_λ, being the sum of a function with compact support and a function in $L^\infty(\mathbb{R}^2)$, is in $\mathscr{S}'(\mathbb{R}^2)$ [Tre67, p. 274]. It thus follows that the Fourier transform of ϕ is given by the convolution of $\hat{\chi}$ and $\hat{\psi}_\lambda$:

$$\begin{aligned} \hat{\phi}(\xi) &= \hat{\chi} * \hat{\psi}_\lambda(\xi) \\ &= d_1(\lambda)\hat{\chi} * \psi_{-2-\lambda}(\xi) \end{aligned}$$

[Tre67, p. 319, thm. 30.4]. Since $\psi_{-2-\lambda} \in \mathscr{S}'(\mathbb{R}^2)$ and $\hat{\chi}(\xi) \in \mathscr{S}(\mathbb{R}^2)$, $\hat{\phi} \in C^\infty(\mathbb{R}^2)$ [Tre67, p. 317, thm. 30.2]. To prove the desired result, it therefore suffices to show that the integrand in equation 4.19 is dominated by an integrable function for $|\xi| > 2$.

Since $\hat{\chi} \in \mathscr{S}$, for each $k \in \mathbb{N}$ there exists a constant $d_2(k)$ such that, for $\xi \neq 0$,

$$
\begin{aligned}
\hat{\chi}(\xi) &\leq d_2(k)|\xi|^{-k} \\
&= d_2(k)\psi_{-k}(\xi).
\end{aligned}
$$

We can write

$$
\begin{aligned}
\hat{\phi}(\xi) &= d_1(\lambda)(\hat{\chi} * 1_{|\cdot|\leq 1}\psi_{-2-\lambda} + 1_{|\cdot|\leq 1}\hat{\chi} * 1_{|\cdot|>1}\psi_{-2-\lambda} \\
&\quad + 1_{|\cdot|>1}\hat{\chi} * 1_{|\cdot|>1}\psi_{-2-\lambda})(\xi).
\end{aligned}
\tag{4.20}
$$

For $|\xi| > 2$, we have the estimates

$$
\begin{aligned}
\hat{\chi} * 1_{|\cdot|\leq 1}\psi_{-2-\lambda}(\xi) &\leq d_2(3)\psi_{-3} * 1_{|\cdot|\leq 1}\psi_{-2-\lambda}(\xi) \\
&= d_2(3)\int_{|\eta|\leq 1} |\xi - \eta|^{-3}|\eta|^{-2-\lambda}\, d\eta \\
&< d_2(3)(|\xi| - 1)^{-3}\int_{|\eta|\leq 1} |\eta|^{-2-\lambda}\, d\eta \\
&< 2\pi d_2(3)(|\xi|/2)^{-3}\int_0^1 r^{-1-\lambda}\, dr \\
&= 16\pi d_2(3)|\xi|^{-3} \\
&< 16\pi d_2(3)|\xi|^{-2-\lambda},
\end{aligned}
\tag{4.21}
$$

$$
\begin{aligned}
1_{|\cdot|\leq 1}\hat{\chi} * 1_{|\cdot|>1}\psi_{-2-\lambda}(\xi) &\leq d_2(1)1_{|\cdot|\leq 1}\psi_{-1} * 1_{|\cdot|>1}\psi_{-2-\lambda}(\xi) \\
&= d_2(1)\int_{|\eta|>1} 1_{|\xi-\eta|\leq 1}|\xi - \eta|^{-1}|\eta|^{-2-\lambda}\, d\eta \\
&< d_2(1)(|\xi| - 1)^{-2-\lambda}\int 1_{|\xi-\eta|\leq 1}|\xi - \eta|^{-1}\, d\eta \\
&< 2\pi d_2(1)(|\xi|/2)^{-2-\lambda}\int_0^1 dr \\
&= 16\pi d_2(1)|\xi|^{-2-\lambda},
\end{aligned}
\tag{4.22}
$$

and

$$1_{|\cdot|>1}\hat{\chi} * 1_{|\cdot|>1}\psi_{-2-\lambda}(\xi)$$

$$\leq d_2(3)1_{|\cdot|>1}\psi_{-3} * 1_{|\cdot|>1}\psi_{-2-\lambda}(\xi)$$

$$= d_2(3)\left(\int_{1<|\eta|\leq|\xi|/2} 1_{|\xi-\eta|>1}|\xi-\eta|^{-3}|\eta|^{-2-\lambda}\,d\eta\right.$$

$$\left.+\int_{|\eta|>|\xi|/2} 1_{|\xi-\eta|>1}|\xi-\eta|^{-3}|\eta|^{-2-\lambda}\,d\eta\right)$$

$$< d_2(3)\left(8|\xi|^{-3}\int_{|\eta|\leq|\xi|/2}|\eta|^{-2-\lambda}\,d\eta\right. \tag{4.23}$$

$$\left.+2^{2+\lambda}|\xi|^{-2-\lambda}\int_{|\eta|>|\xi|/2} 1_{|\xi-\eta|>1}|\xi-\eta|^{-3}\,d\eta\right)$$

$$\leq 2\pi d_2(3)\left[8|\xi|^{-3}\int_0^{|\xi|/2} r^{-1-\lambda}\,dr + 2^{2+\lambda}|\xi|^{-2-\lambda}\int_1^\infty r^{-2}\,dr\right]$$

$$= 2\pi d_2(3)\left[2^{3+\lambda}|\xi|^{-3-\lambda} + 2^{2+\lambda}|\xi|^{-2-\lambda}\right]$$

$$< 2^{5+\lambda}\pi d_2(3)|\xi|^{-2-\lambda}.$$

Substituting the estimates given by equations 4.21, 4.22, and 4.23 into equation 4.20 gives, for $|\xi| > 2$,

$$\hat{\phi}(\xi) < d_1(\lambda)\left[16\pi d_2(3) + 16\pi d_2(1) + 2^{5+\lambda}\pi d_2(3)\right]|\xi|^{-2-\lambda}$$

$$= 16\pi d_1(\lambda)\left[d_2(3) + d_2(1) + 2^{1+\lambda}d_2(3)\right]|\xi|^{-2-\lambda}.$$

Defining

$$d_3(\lambda) = 16\pi d_1(\lambda)\left[d_2(3) + d_2(1) + 2^{1+\lambda}d_2(3)\right],$$

we see that

$$|\hat{\phi}(\xi)| < d_3(\lambda)|\xi|^{-2-\lambda}$$

for $|\xi| > 2$. Thus, for $|\xi| > 2$

$$(1 + 4\pi^2|\xi|^2)^\alpha\left|\hat{\phi}(\xi)\right|^2 < d_3^2(\lambda)(1 + 4\pi^2|\xi|^2)^\alpha|\xi|^{-4-2\lambda}$$

$$< (3\pi)^{2\alpha}d_3^2(\lambda)|\xi|^{-4-2\lambda+2\alpha}$$

$$= (3\pi)^{2\alpha}d_3^2(\lambda)|\xi|^{-4-(\alpha-1)+2\alpha}$$

$$= (3\pi)^{2\alpha}d_3^2(\lambda)|\xi|^{-3+\alpha}.$$

The last function in this chain of inequalities is integrable over the region $|\xi| > 2$,

82

which gives the desired result. $\square$

Chapter 5

Upper Bounds on Minimax Risk of Measure Estimation

In chapter 4, we showed that the minimax risk in the PET problem is not $O(n^{-1})$ with respect to loss functions generated by squared Sobolev norms of order $\geq -3/2$. In this chapter, we will show that the minimax risk *is* $O(n^{-1})$ with respect to loss functions generated by squared Sobolev norms of order $< -3/2$.

5.1 Loss functions induced by Sobolev norms of order $< -3/2$

The main theorem in this section shows that the minimax risk for estimating $\mu \in P(S)$ with respect to the loss function generated by the squared Sobolev norm $\| \cdot \|_{L^2_{-\alpha}(\mathbb{R}^2)}^2$ for $\alpha > 3/2$ is $O(n^{-1})$. It is proved by constructing a sequence of estimators, $\{\hat{\mu}_n\}$, whose maximum risk over $P(S)$ is $O(n^{-1})$.

Definition 5.1.1 If $x_1, \ldots, x_n$ are points of the measurable space $(X, \mathscr{A})$, we define the empirical probability measure generated by $x_1, \ldots, x_n$ to be the probability measure on $(X, \mathscr{A})$ consisting of the sum of n point masses of measure $1/n$ located at $x_1, \ldots, x_n$. Let $P(G_{1,2}(r))$ denote the set of probability measures on $G_{1,2}$ with support on $G_{1,2}(r)$.

Theorem 5.1.2 *For* $\alpha > 3/2$, *there exists a sequence of estimators* $\hat{\mu}_n$, *based on* n *independent observations distributed according to* $T\mu$, *such that*

$$\sup_{\mu \in P(S)} E_{(T\mu)^n} ||\hat{\mu}_n - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)} = O(n^{-1}). \tag{5.1}$$

Proof. We start by constructing the sequence of estimators. Fix n and let ν_n denote the empirical probability measure generated from n independent observations that are distributed according to the probability measure $\nu = T\mu$. We first construct an auxiliary estimate, $\hat{\nu}_n$, of ν by choosing some $\hat{\nu}_n \in T[P(S)]$ which satisfies

$$||\hat{\nu}_n - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} \leq \inf_{\check{\nu} \in T[P(S)]} ||\check{\nu} - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} + n^{-1}.$$

(It is, of course, obvious that such an estimate exists. Unfortunately, we cannot construct our estimate by simply choosing some $\hat{\nu}_n \in T[P(S)]$ which satisfies

$$||\hat{\nu}_n - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} = \inf_{\check{\nu} \in T[P(S)]} ||\check{\nu} - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}$$

since there is no guarantee that such a $\hat{\nu}_n$ exists, i.e., that the infimum can be achieved.) We then estimate μ by $\hat{\mu}_n \stackrel{\text{def}}{=} T^{-1}\hat{\nu}_n$.(The right-hand side of this equation makes sense since T is injective on $P(S)$ and $\hat{\nu}_n \in T[P(S)]$.)

We will proceed by first showing that $\sup_{\nu \in T[P(S)]} E_{\nu^n} ||\hat{\nu}_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} = O(n^{-1})$ and then using lemma 3.4.10 to show that $\sup_{\mu \in P(S)} E_{(T\mu)^n} ||\hat{\mu}_n - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)} = O(n^{-1})$. By the triangle inequality and the easy inequality $2ab < a^2 + b^2$, we have

$$||\hat{\nu}_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}$$
$$\leq 2(||\hat{\nu}_n - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} + ||\nu_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))})$$
$$\leq 2(\inf_{\check{\nu} \in T[P(S)]} ||\check{\nu} - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} + n^{-1} + ||\nu_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}) \tag{5.2}$$
$$\leq 4||\nu_n - \nu||_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} + 2n^{-1}$$

(the last inequality follows since $\nu \in T[P(S)]$, hence

$$\inf_{\hat{\nu} \in T[P(S)]} ||\hat{\nu} - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} \leq ||\nu - \nu_n||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}).$$

Taking the expectation of equation (5.2) and using lemma 5.1.3 below, we find that

$$\sup_{\nu \in T[P(S)]} E_{\nu^n} ||\hat{\nu}_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}$$
$$\leq \sup_{\nu \in T[P(S)]} 4 E_{\nu^n} ||\nu_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))} + 2n^{-1/2}$$
$$= O(n^{-1}).$$

To complete the proof, we note that lemma 3.4.10 shows that there exists a constant c such that

$$E_{(T\mu)^n} ||\hat{\mu}_n - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)} \leq c^2 E_{(T\mu)^n} ||R\hat{\mu}_n - R\mu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}$$
$$= c^2 \pi^2 E_{(T\mu)^n} ||\hat{\nu}_n - \nu||^2_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}.$$

It follows that

$$\sup_{\mu \in P(S)} E_{(T\mu)^n} ||\hat{\mu}_n - \mu||^2_{L^2_{-\alpha}(\mathbb{R}^2)} \leq c^2 \pi^2 \sup_{\nu \in T[P(S)]} E_{\nu^n} ||\hat{\nu}_n - \nu||_{L^2_{-(\alpha-1/2)}(G_{1,2}(1))}$$
$$= O(n^{-1}). \quad \square$$

It remains to give the proof of lemma 5.1.3, which was used in the above proof.

Lemma 5.1.3 *Let ν_n be the empirical distribution obtained from n independent observations distributed according to ν. Suppose $\alpha > 1$. Then*

$$\sup_{\nu \in P(G_{1,2}(\sqrt{2}/2))} E_{\nu^n} ||\nu_n - \nu||^2_{L^2_{-\alpha}(G_{1,2}(1))} = O(n^{-1}).$$

Proof.   Let $\nu \in P(G_{1,2}(\sqrt{2}/2))$ and let $\check{\nu}$ denote the lift of ν to a distribution on $S^1 \times \mathbb{R}$ as in definition 2.2.2. Clearly $\check{\nu}$ has support on $S^1 \times [-\sqrt{2}/2, \sqrt{2}/2]$, hence on $S^1 \times [-1, 1]$.

It is well-known that the set of exponential functions $\{\frac{1}{\sqrt{4\pi}} e^{-i(\kappa_1\theta + \pi\kappa_2 s)}\}_{\kappa \in \mathbb{Z}^2}$ is an

86

orthonormal basis for $L_0^2(S^1 \times 2\mathbb{T}^1)$. Since the map

$$U_\alpha : f \mapsto \mathscr{F}^{-1}((1 + \kappa_1^2 + \pi^2\kappa_2^2)^{\alpha/2}\mathscr{F}(f)),$$

where $\mathscr{F}$ denotes the Fourier transform, is an isometric isomorphism of $L_0^2(S^1 \times 2\mathbb{T}^1)$ onto $L_{-\alpha}^2(S^1 \times 2\mathbb{T}^1)$ (cf. [Tre67, p. 330, prop. 31.8]), the set of distributions

$$\{U_\alpha(\frac{1}{\sqrt{4\pi}}e^{-i(\kappa_1\theta+\pi\kappa_2 s)})\}_{\kappa\in\mathbb{Z}^2} = \{\frac{1}{\sqrt{4\pi}}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{\alpha/2}e^{-i(\kappa_1\theta+\pi\kappa_2 s)}\}_{\kappa\in\mathbb{Z}^2}$$

is an orthonormal basis for $L_{-\alpha}^2(S^1 \times 2\mathbb{T}^1)$. For any distribution $v \in L_{-\alpha}^2(S^1 \times 2\mathbb{T}^1)$, we have

$$\|v\|_{L_{-\alpha}^2(S^1\times 2\mathbb{T}^1)}^2 = 4\pi \sum_{\kappa\in\mathbb{Z}^2}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha}|\hat{v}(\kappa)|^2.$$

Since

$$\begin{aligned}
[(\nu_n - \nu)^\sim]^\wedge(\kappa) &= \frac{1}{\sqrt{4\pi}}\int_{S^1\times 2\mathbb{T}^1}e^{-i(\kappa_1\theta+\pi\kappa_2 s)}\,d(\nu_n - \nu)^\sim \\
&= \frac{1}{\pi}\int_{G_{1,2}}\cos(\kappa_1\theta + \pi\kappa_2 s)\,d(\nu_n - \nu),
\end{aligned}$$

it follows that

$$\begin{aligned}
&\|(\nu_n - \nu)^\sim\|_{L_{-\alpha}^2(S^1\times 2\mathbb{T}^1)}^2 \\
&= 4\sum_{\kappa\in\mathbb{Z}^2}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha}\left(\int_{G_{1,2}}\cos(\kappa_1\theta + \pi\kappa_2 s)\,d(\nu_n - \nu)\right)^2.
\end{aligned} \tag{5.3}$$

Now if ν is a probability measure on $G_{1,2}$ and g is a measurable function on $G_{1,2}$ such that $|g| \le 1$, $\int g\,d\nu_1$ (ν_1 denotes the empirical measure generated by one observations distributed according to ν) is a random variable with mean $\int g\,d\nu$ and variance

$$\int_{G_{1,2}} g^2\,d\nu - \left(\int_{G_{1,2}} g\,d\nu\right)^2.$$

Since the observations generating the empirical measure ν_n are independent, it follows

that $\int g \, d\nu_n - \int g \, d\nu$ is a zero-mean random variable with variance

$$\frac{1}{n}\left[\int_{G_{1,2}} g^2 \, d\nu - \left(\int g \, d\nu\right)^2\right] \leq \frac{1}{n}\int_{G_{1,2}} g^2 \, d\nu$$

$$\leq n^{-1}.$$

Applying this result to equation 5.3, we thus have

$$E_{\nu^n}\|(\nu_n - \nu)^{\tilde{}}\|^2_{L^2_{-\alpha}(S^1 \times 2\mathbb{T}^1)} \leq 4n^{-1}\sum_{\kappa \in \mathbb{Z}^2}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha}. \qquad (5.4)$$

To show that the infinite sum on the right-hand side of this inequality is finite, view the infinite sum as the integral of the function whose value is equal to $(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha}$ on the unit square centered at $\kappa \in \mathbb{Z}^2$. It follows that

$$\begin{aligned}
\sum_{\kappa \in \mathbb{Z}^2}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha} &= \sum_{\kappa \in \mathbb{Z}^2}\int_{\kappa + S \subset \mathbb{R}^2}(1 + \kappa_1^2 + \pi^2\kappa_2^2)^{-\alpha}dx \\
&< 1 + \sum_{\kappa \in \mathbb{Z}^2\backslash\{0\}}\int_{\kappa + S \subset \mathbb{R}^2}|\kappa|^{-2\alpha}dx \\
&\leq 1 + \int_{|x|\geq 1/2}|x|^{-2\alpha}\,dx \\
&= 1 + 2\pi\int_{1/2}^{\infty}r^{1-2\alpha}\,dr \\
&= 1 + \frac{2\pi}{2(1-\alpha)}r^{2(1-\alpha)}|^{\infty}_{r=1/2} \\
&= 1 + \frac{\pi 2^{2(\alpha-1)}}{(\alpha - 1)}.
\end{aligned}$$

Since ν is an arbitrary element of $G_{1,2}(\sqrt{2}/2)$, we conclude that

$$\sup_{\nu \in G_{1,2}(\sqrt{2}/2)} E_{\nu^n}\|(\nu_n - \nu)^{\tilde{}}\|^2_{L^2_{-\alpha}(S^1 \times 2\mathbb{T}^1)} = O(n^{-1}).$$

The desired result now follows by definition 3.2.14. $\square$

Remark 5.1.4 We cannot take $\alpha \leq 1$ in lemma 5.1.3 since $\|\nu_n - \nu\|_{L^2_{-\alpha}(G_{1,2}(1))}$ will not then, in general, even be finite.

Remark 5.1.5 Theorem 5.1.2 concerns estimators of μ that are based on n inde-

pendent observations distributed according to $T\mu$. It is useful to contrast this result with one concerning estimators of μ that are based on n independent observations distributed according to μ itself. In this case, there exists a sequence of estimators $\hat{\mu}_n$ such that $\sup_{\mu \in P(S)} E_{\mu^n} \|\hat{\mu}_n - \mu\|^2_{L^2_{-\alpha}(\mathbb{R}^2)} = O(n^{-1})$ for $\alpha > 1$. Indeed, an argument similar to the one given in the proof of lemma 5.1.3 shows that the empirical distribution μ_n is one such estimator.

Remark 5.1.6 Combining corollary 4.1.2 and theorem 5.1.2, we get the result that the minimax risk for estimating $\mu \in P(S)$ from n independent observations distributed according to $T\mu$ is $O(n^{-1})$ with respect to the loss function generated by $\|\cdot\|^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ if and only if $\alpha > 3/2$. Combining remarks 4.1.4 and 5.1.5, we get the companion result that the minimax risk for estimating $\mu \in P(S)$ from n independent observations distributed according to μ is $O(n^{-1})$ with respect to the loss function generated by $\|\cdot\|^2_{L^2_{-\alpha}(\mathbb{R}^2)}$ if and only if $\alpha > 1$.

Chapter 6

Estimation of Integral Functionals

Up until this point, we have focused on the problem of estimating the unknown probability measure μ in the PET problem. In this chapter, we will consider the problem of estimating certain integral functionals of the unknown probability measure μ. Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a bounded measurable function. The function ϕ induces a linear integral functional $\dot{\phi} : M(S) \to \mathbb{R}$ defined by $\mu \mapsto \int \phi \, d\mu$. We shall consider the problem of estimating the quantity $\dot{\phi}(\mu)$, which we will refer to as the integral functional generated by ϕ.

6.1 Functionals generated by functions in L_α^2

In this section, we will consider the estimation of integral functionals generated by functions in $L_\alpha^2(\mathbb{R}^2)$ for $\alpha > 3/2$. The Sobolev embedding theorem [Fol84, thm. 8.54] shows that $L_\alpha^2(\mathbb{R}^2) \subset C_0(\mathbb{R}^2)$ when $\alpha > 1$, so the functions in $L_\alpha^2(\mathbb{R}^2)$ are, in particular, continuous and bounded. We shall see that the minimax risk with respect to squared error loss for the estimation of these functionals is $O(n^{-1})$.

If $\phi \in \mathscr{L}^b(\mathbb{R}^2)$ and $\hat{\mu}_n$ is an estimator of μ, then an obvious estimator of $\dot{\phi}(\mu)$ is given by $\hat{\dot{\phi}}_n \stackrel{\text{def}}{=} \dot{\phi}(\hat{\mu}_n)$. Let $\alpha > 3/2$. By theorem 5.1.2, there exists an estimator $\hat{\mu}_n$ such that $\sup_{\mu \in P(S)} E_{(T\mu)^n} \|\hat{\mu}_n - \mu\|^2_{L^2_{-\alpha}(\mathbb{R}^2)} = O(n^{-1})$. If $\phi \in L_\alpha^2(\mathbb{R}^2)$, then the

estimate $\dot\phi(\hat\mu_n)$ of $\dot\phi(\mu)$ satisfies

$$
\begin{aligned}
\sup_{\mu\in P(S)} E_{(T_\mu)^n}|\hat{\dot\phi}_n - \phi(\mu)|^2 &= \sup_{\mu\in P(S)} E_{(T_\mu)^n}|(\hat\mu_n - \mu)(\phi)|^2 \\
&\leq \|\phi\|^2_{L^2_\alpha(\mathbb{R}^2)} \sup_{\mu\in P(S)} E_{(T_\mu)^n}\|\hat\mu_n - \mu\|^2_{L^2_{-\alpha}(\mathbb{R}^2)} \qquad (6.1) \\
&\leq \|\phi\|^2_{L^2_\alpha(\mathbb{R}^2)} O(n^{-1}).
\end{aligned}
$$

Examination of the bound given by equation (6.1) gives some insight into the meaning of the result obtained in section 5.1. For example, if ϕ is a nonnegative function, $\dot\phi(\mu)$ is a weighted (by the magnitude of ϕ) average of the mass contained in the support of ϕ. One can thus think of ϕ as representing a generalized pixel. Now a natural quantity that one might want to estimate in the PET problem is the mass contained in some measurable set $Q \subset \mathbb{R}^2$, i.e., the fraction of the tracer in some (ordinary) pixel. This quantity could be represented as the integral functional 1_Q, where 1_Q denotes the indicator function of Q. However, note that equation (6.1) says nothing about how well one might be able to estimate this functional, since, in general, $1_Q \notin L^2_\alpha(\mathbb{R}^2)$ (recall that the functions in $L^2_\alpha(\mathbb{R}^2)$ for $\alpha > 3/2$ are continuous). To get an approximate answer to how much mass is contained in Q, one would have to apply equation (6.1) to some function in $L^2_\alpha(\mathbb{R}^2)$ which is close in some sense to 1_Q. For example, for some $\epsilon > 0$, we might choose ϕ_ϵ such that $0 \leq \phi_\epsilon(x) \leq 1$, $\phi_\epsilon(x) = 1$ for $x \in Q$, and $\phi_\epsilon(x) = 0$ for any x whose distance from Q is $\geq \epsilon$. It is easy to see that as ϵ becomes smaller, $\|\phi_\epsilon\|_{L^2_\alpha(\mathbb{R}^2)}$, and hence the bound in equation (6.1), must become larger.

The fact that our theory says nothing about how well one can estimate the mass in an ordinary pixel is disappointing. However, it reflects an intrinsic property of the estimation problem. In fact, in section 6.2, we shall see that the minimax risk for the estimation of the mass in some very benign-looking ordinary pixels is bounded away from 0 as $n \to \infty$.

Remark 6.1.1 We conclude this section with a few words on our use of the term "generalized pixel". We use this term in a way which is fairly close to the use of

the term generalized function for a distribution. We identify an ordinary pixel, i.e., a set $Q \subset \mathbb{R}^2$, with its indicator function, 1_Q, and thence with the integral linear functional i_Q on $M(S)$. It is then natural to use the term generalized pixel for any linear functional on $M(S)$ and to identify generalized pixels which are integral functionals with the functions that generate them. In [Bak91], the term generalized pixel is used for an element of a basis set for the image (lying in a suitable Hilbert space) whose coefficient is to be estimated. That is, in [Bak91], a generalized pixel is an element in the image space, while, here, a generalized pixel is a linear functional on the image space. There is no real conflict between the definitions since a generalized pixel in the sense of [Bak91] can always be viewed as a generalized pixel in the sense used here by the usual identification of an element of a Hilbert space with an element in its dual.

6.2 Functionals generated by indicator functions

Let Q be a Borel subset of S. From a physical point of view, it is very natural to want to estimate $\mu(Q) = \hat{1}_Q(\mu)$. It is a somewhat surprising fact that for some very benign-looking sets Q the minimax risk for estimating $\mu(Q)$ is bounded away from zero as $n \to \infty$. For example, the following proposition shows that the minimax risk for estimating $\mu(\rho\bar{\Omega})$ is bounded away from 0 when $0 < \rho < 1/2$.

Proposition 6.2.1 *Let $\rho\bar{\Omega}$ denote the closed disk of radius $0 < \rho < 1/2$ centered at the origin. There exists $c > 0$, independent of n, such that*

$$\inf_{\hat{1}_{\rho\bar{\Omega}}} \sup_{\mu \in P(S)} E_{(T\mu)^n} |\hat{1}_{\rho\bar{\Omega}} - \hat{1}_{\rho\bar{\Omega}}(\mu)| \geq c,$$

where the infimum is taken over all estimators $\hat{1}_{\rho\bar{\Omega}}$ based on n independent observations distributed according to $T\mu$.

Proof. Define the modulus of continuity of the functional $\hat{1}_{\rho\bar{\Omega}}$ over $P(S)$ by

$$\omega(\epsilon) \stackrel{\text{def}}{=} \sup\{|\hat{1}_{\rho\bar{\Omega}}(\mu) - \hat{1}_{\rho\bar{\Omega}}(\mu')| : H\,(T\mu, T\mu') \leq \epsilon, \mu, \mu' \in P(S)\},$$

where H denotes the Hellinger distance as defined at the top of the proof of theorem 4.3.1. We will use the lower bound

$$\inf_{\hat{1}_{\rho\bar{\Omega}}} \sup_{\mu \in P(S)} E_{(T\mu)^n} |\hat{1}_{\rho\bar{\Omega}}(\mu) - \hat{1}_{\rho\bar{\Omega}}(\mu)| \geq c\omega(n^{-1/2}) \tag{6.2}$$

for some $c > 0$ [DL91]. We claim that $\omega(\epsilon) = 1$ for $\epsilon > 0$. To prove the claim, let $\epsilon > 0$ be given. Let μ_r denote the probability measure on $\mathbb{R}^2$ which is uniformly distributed on the circle of radius r centered at the origin with respect to the measure $d\theta$. In appendix A.2, it is shown that $\nu_r \stackrel{\text{def}}{=} T\mu_r$ is given by the probability density

$$\frac{d\nu_r}{ds\,d\theta}(\theta, s) = \begin{cases} \frac{1}{\pi^2\sqrt{r^2-s^2}} & \text{if } |s| \leq r \\ 0 & \text{if } |s| > r \end{cases}$$

on $G_{1,2}$. For $r > \rho$, we have

$$
\begin{aligned}
\|\nu_r - \nu_\rho\|_v &= \left\|\frac{d\nu_r}{ds\,d\theta} - \frac{d\nu_\rho}{ds\,d\theta}\right\|_{L^1(G_{1,2})} \\
&= \frac{2}{\pi}\int_0^\rho \left(\frac{1}{\sqrt{\rho^2 - s^2}} - \frac{1}{\sqrt{r^2 - s^2}}\right) ds + \frac{2}{\pi}\int_\rho^r \frac{1}{\sqrt{r^2 - s^2}}\, ds \\
&= \frac{2}{\pi}\int_0^\rho \frac{ds}{\rho\sqrt{1 - (s/\rho)^2}} - \frac{2}{\pi}\int_0^\rho \frac{ds}{r\sqrt{1 - (s/r)^2}} \\
&\quad + \frac{2}{\pi}\int_0^r \frac{ds}{r\sqrt{1 - (s/r)^2}} - \frac{2}{\pi}\int_0^\rho \frac{ds}{r\sqrt{1 - (s/r)^2}} \\
&= \frac{2}{\pi}\int_0^1 \frac{ds}{\sqrt{1 - s^2}} - \frac{2}{\pi}\int_0^{\rho/r} \frac{ds}{\sqrt{1 - s^2}} \\
&\quad + \frac{2}{\pi}\int_0^1 \frac{ds}{\sqrt{1 - s^2}} - \frac{2}{\pi}\int_0^{\rho/r} \frac{ds}{\sqrt{1 - s^2}} \\
&= \frac{4}{\pi}\int_0^1 \frac{ds}{\sqrt{1 - s^2}} - \frac{4}{\pi}\int_0^{\rho/r} \frac{ds}{\sqrt{1 - s^2}} \\
&= 2 - \tfrac{2}{\pi}\arcsin\left(\tfrac{\rho}{r}\right),
\end{aligned}
\tag{6.3}
$$

which $\to 0$ as $r \searrow \rho$. By equation 4.18, it follows that $\lim_{r \searrow \rho} H(T\mu_r, T\mu_\rho) = 0$. On the other hand $|\dot{\mathbf{1}}_{\rho\tilde\Omega}(\mu_r) - \dot{\mathbf{1}}_{\rho\tilde\Omega}(\mu_\rho)| = 1$ for any $r > \rho$. This proves the claim that $\omega(\epsilon) = 1$. Substituting this result into equation 6.2 completes the proof. $\square$

Corollary 6.2.2 *There exists $c > 0$, independent of n, such that*

$$
\inf_{\dot{\mathbf{1}}_{\rho\tilde\Omega}}\, \sup_{\mu \in P(S)}\, E_{(T\mu)^n}|\hat{\dot{\mathbf{1}}}_{\rho\tilde\Omega} - \dot{\mathbf{1}}_{\rho\tilde\Omega}(\mu)|^2 \geq c.
$$

where the infimum is taken over all estimators $\hat{f}_n$ based on n independent observations distributed according to $T\mu$.

Proof. Apply Jensen's inequality.

Remark 6.2.3 This result contrasts sharply with the situation for the problem of estimating $\mu(\rho\tilde\Omega)$ given n independent observations distributed according to μ. In fact, let μ_n denote the empirical measure constructed from n independent observations distributed according to μ. Then it is shown below that for any Borel subset Q of

S the estimate $1_Q(\mu_n)$ has a maximum risk of $O(n^{-1})$, cf. [Gin75, thm. 3.4]. We thus see that corollary 6.2.2 reflects an inherent loss of spatial resolution when only observations distributed according to $T\mu$ are available.

Proposition 6.2.4 *If $\mu \in P(S)$ and Q is a Borel subset of S,*

$$E_{\mu^n}|\mu_n(Q) - \mu(Q)|^2 \leq 2^{-2}n^{-1}.$$

Proof. The random variable $\mu_n(Q)$ is equal to n^{-1} times the sum of n independent random variables with the probability distribution of the random variable $\mu_1(Q)$, where μ_1 is the empirical distribution function constructed from one observation. The latter random variable is obtained by evaluating the function 1_Q at a random sample drawn according to the probability measure μ. Thus

$$E_\mu \mu_1(Q) = \mu(Q)$$

and

$$
\begin{aligned}
E_\mu \mu_1^2(Q) &= \int_{\mathbb{R}^2} 1_Q^2 \, d\mu \\
&= \int_{\mathbb{R}^2} 1_Q \, d\mu \\
&= \mu(Q).
\end{aligned}
$$

Hence

$$
\begin{aligned}
E_{\mu^n}\mu_n(Q) &= E_\mu \mu_1(Q) \\
&= \mu(Q)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{var}_{\mu^n}\,\mu_n(Q) &= n^{-1}\{E_\mu \mu_1^2(Q) - [E_\mu \mu_1(Q)]^2\} \\
&= n^{-1}[\mu(Q) - \mu^2(Q)] \\
&\leq 2^{-2}n^{-1}.
\end{aligned}
$$

Thus $\mu_n(Q)$ is an unbiased estimator of $\mu(Q)$ with variance $\leq 2^{-2}n^{-1}$. $\square$

6.3 Functionals generated by smooth, rapidly decreasing functions

In this section, we will consider the estimation of the integral functional generated by a function $\phi \in \mathscr{S}(\mathbb{R}^2)$. Starting with the Radon transform inversion formula given in [Nat86, p. 18, thm. 2.1] and using lemma 2.1.16, we can rewrite $\phi \in \mathscr{S}(\mathbb{R}^2)$ as

$$\begin{aligned} \phi &= \frac{1}{4\pi} \tilde{R}^* I^{-1} \tilde{R}\phi \\ &= \frac{1}{4\pi} R^* (I^{-1} \tilde{R}\phi)^\vee. \end{aligned}$$

(recall that I^{-1} acts on the second or "s" variable of functions on $S^1 \times \mathbb{R}$). Writing the function $\phi \mapsto \frac{1}{4\pi}(I^{-1}\tilde{R}\phi)^\vee$ as the right inverse R^{-*}, we thus have

$$\begin{aligned} \int_{\mathbb{R}^2} \phi \, d\mu &= \int_{\mathbb{R}^2} R^* R^{-*} \phi \, d\mu \\ &= \int_{G_{1,2}} R^{-*} \phi \, dR\mu \\ &= \int_{G_{1,2}} T^{-*} \phi \, dT\mu, \end{aligned}$$

where $T^* \stackrel{\text{def}}{=} \pi R^{-*}$, so the functional $\dot{\phi}$ can be expressed as $\mu \mapsto \int_{G_{1,2}} T^{-*}\phi \, dT\mu$. That is, defining $\psi \stackrel{\text{def}}{=} T^{-*}\phi$, the functional $\dot{\phi}$ on $P(S)$ is equivalent to the functional $\dot{\psi}$ on $T[P(S)]$ in the sense that $\dot{\phi}(\mu) = \dot{\psi}(T\mu)$.

The quantity $\dot{\psi}(T\mu)$ is obviously the expected value of the random variable constructed by evaluating the function ψ at a random sample drawn according to the probability measure $T\mu$. Thus an obvious estimate for $\dot{\phi}(\mu)$ is the sample mean of the derived observations obtained by evaluating the function ψ at the original observations. We can write this estimator as $\hat{\phi}_n \stackrel{\text{def}}{=} \dot{\psi}(\nu_n)$, where ν_n is the empirical distribution function constructed from the first n observations.

Let us now compute the performance of this estimator. The first thing to note is that $\hat{\phi}_n$ is equal to n^{-1} times the sum of n independent random variables with the same distribution as $\hat{\phi}_1 \stackrel{\text{def}}{=} \int_{G_{1,2}} \psi \, d\nu_1$, where ν_1 is the empirical distribution function

96

constructed from one observation. We have

$$E_{T\mu}\hat{\phi}_1 \quad = \dot{\psi}(T\mu)$$
$$= \dot{\phi}(\mu)$$

and

$$E_{T\mu}\hat{\phi}_1^2 = (\psi^2)\dot{\,}(T\mu)$$

Hence

$$E_{(T\mu)^n}\hat{\phi}_n = \dot{\phi}(\mu)$$

and

$$\mathrm{var}_{(T\mu)^n}\hat{\phi}_n \quad = n^{-1}[E_\mu\hat{\phi}_1^2 - (E_\mu\hat{\phi}_1)^2]$$
$$= n^{-1}\left[\dot{\psi}^2(T\mu) - [\dot{\phi}(\mu)]^2\right].$$

Remark 6.3.1 We conclude this section with a brief comment that is relevant to obtaining a numerical approximation to the function $I^{-1}\tilde{R}\phi$. By the so-called projection-slice theorem [Nat86, p. 11, thm. 1.1],

$$\widehat{\tilde{R}\phi}(\omega,\eta) = \hat{\phi}(\eta\omega),$$

hence $(I^{-1}\tilde{R}\phi)\dot{\,}(\omega,\eta) = |\eta|\hat{\phi}(\eta\omega)$. Thus $I^{-1}\tilde{R}\phi(\omega,\cdot)$ may be computed by taking the inverse Fourier transform of $|\eta|\hat{\phi}(\eta\omega)$.

6.4 Asymptotically locally minimax estimation of functionals generated by smooth, rapidly decreasing functions

Throughout this work, we have been concerned with characterizing the minimax risk for estimating quantities in the PET problem as a function of n. Up until this point, the results have been expressed in terms of the *order* of the minimax risk as $n \to \infty$. That is, we have been characterizing these rates up a constant. In practice, one wants to use the minimax risk as a guide to system design and a benchmark for algorithm performance. For these purposes, one needs to know the exact rate, i.e., one needs to know the constant. In this section, we will develop a local theory which yields exact asymptotic minimax rates for estimation of functionals generated by functions in $\mathscr{S}(\mathbb{R}^2)$.

To develop this theory with a minimum of fuss, we will use slightly stronger assumptions on the set $\mathscr{P}$ of allowable probability measures μ. Instead of taking $\mathscr{P} = P(S)$, we will take $\mathscr{P}$ to be the set of probability measures whose elements can be represented by probability density functions with respect to Lebesgue measure on S that are bounded above and below away from 0. In our opinion, this restriction on $\mathscr{P}$ does not seriously restrict the applicability of our results. The assumption that the density function is bounded above corresponds to the physical condition that the concentration of tracer in the subject is bounded above. Since this upper bound can be arbitrarily large, we consider this condition to be quite benign. The assumption that the density function is bounded below away from 0 is perhaps more problematic, but since this lower bound can be chosen to be arbitrarily small, we do not believe it alters the essence of the problem. Physically, one can think of it as postulating some positive level of background radiation. Alternatively, one could alter the experiment by adding some artificial observations that mimic those that would be obtained from a low-intensity uniform distribution.

The approach we will take is based upon differential structures. We follow [Pfa82]

in our treatment of these structures.

Definition 6.4.1 Let $\mathscr{P}$ be a set of mutually absolutely continuous probability measures on a measurable space $(X, \mathscr{A})$ and fix $\mu_0 \in \mathscr{P}$. A path in $\mathscr{P}$ is defined to be a map from $(0, \epsilon)$ (for some $\epsilon > 0$) to $\mathscr{P}$ and is denoted by $t \mapsto \mu_t$. By an abuse of language, we will refer to such a map as the path μ_t. For technical reasons, we will also admit paths defined only on a sequence of points in $\mathbb{R}^+$ converging to 0, cf. [Pfa82, rem 1.1.6]. For brevity, if the measure μ can be represented by the density function ζ with respect to the measure λ, we shall simply say that μ has the λ-density ζ. The path μ_t is said to be differentiable (in the strong sense) at μ_0 with derivative $h \in L^2(\mu_0)$ if the μ_0-density of μ_t, ζ_t, can be represented as

$$\zeta_t = 1 + t(h + r_t),$$

where $\|r_t\|_{L^2(\mu_0)} \to 0$ as $t \to 0$ [Pfa82, def. 1.1.1, eq. 1.1.5]. Let $L^2_*(\mu_0)$ denote the subspace of $L^2(\mu_0)$ whose elements h satisfy $\int h \, d\mu_0 = 0$. It is easy to show that a derivative of $\mathscr{P}$ at μ_0 must lie in $L^2_*(\mu_0)$, see [Pfa82, p. 23]. The tangent cone of $\mathscr{P}$ at μ_0, denoted by $T(\mu_0, \mathscr{P})$, is defined to be the subset of $L^2_*(\mu_0)$ whose elements are derivatives of paths in $\mathscr{P}$ at μ_0. It can be shown that $T(\mu_0, \mathscr{P})$ is closed [Pfa82, p. 25]. In the cases considered here, $T(\mu_0, \mathscr{P})$ will always turn out to be a linear subspace of $L^2_*(\mu_0)$, hence we will refer to it as the tangent space of $\mathscr{P}$ at μ_0.

Definition 6.4.2 We next introduce the notion of the canonical gradient of a functional on $\mathscr{P}$. Let $\kappa : \mathscr{P} \to \mathbb{R}$ be a functional. A function $\kappa^{\cdot} \in L^2_*(\mu_0)$ is said to be a gradient of κ at μ_0 for $\mathscr{P}$ if for every $h \in T(\mu_0, \mathscr{P})$ and every path μ_t in $\mathscr{P}$ with derivative h,

$$\kappa(\mu_t) - \kappa(\mu_0) = t(\kappa^{\cdot}|h)_{L^2(\mu_0)} + o(t).$$

If a gradient of κ exists, then κ is said to be differentiable [Pfa82, def. 4.1.1]. While a gradient is not necessarily unique, there is a unique gradient in $T(\mu_0, \mathscr{P})$. This unique gradient in $T(\mu_0, \mathscr{P})$ is called the canonical gradient and will be denoted by κ^*. It can be obtained by projecting any gradient onto $T(\mu_0, \mathscr{P})$ [Pfa82, sec. 4.3.2].

Definition 6.4.3 To avoid confusion with the use of T in this section to denote tangent spaces, we shall define $A \overset{\text{def}}{=} \frac{1}{\pi} R$ and use it instead of T to denote the scaled Radon transform.

We will consider the local asymptotic minimax risk at $\mu_0 = f_0 \, d\lambda^2 \in \mathscr{P}$. We define $\nu_0 = A\mu_0 = Af_0 \, ds \, d\theta$. The first step is to characterize the tangent spaces at μ_0 and ν_0.

Proposition 6.4.4

$$T(\mu_0, \mathscr{P}) = L^2_*(\mu_0).$$

Proof. The proof will mimic that of [Pfa82, ex. 2.1.1]. Since $T(\mu_0, \mathscr{P})$ is defined as a subset of $L^2_*(\mu_0)$, we have to show that each $h \in L^2_*(\mu_0)$ is in $T(\mu_0, \mathscr{P})$. Let $h \in L^2_*(\mu_0)$ be given and consider the path whose μ_0-density is given by

$$\zeta_t = 1 + t(h + r_t),$$

where

$$r_t \overset{\text{def}}{=} -h 1_{\{th < -1/2\}} + \int_{\mathbb{R}^2} h 1_{\{th < -1/2\}} \, d\mu_0.$$

It is readily verified that this is indeed a path in $\mathscr{P}$. Moreover,

$$
\begin{aligned}
\|r_t\|^2_{L^2(\mu_0)} &= \int_{\mathbb{R}^2} \left(-h 1_{\{th < -1/2\}} + \int_{\mathbb{R}^2} h 1_{\{th < -1/2\}} \, d\mu_0 \right)^2 d\mu_0 \\
&\leq \int_{\mathbb{R}^2} h^2 1_{\{th < -1/2\}} \, d\mu_0
\end{aligned}
$$

and the last quantity $\to 0$ as $t \to 0$. $\square$

Proposition 6.4.5 *Define the linear operator* $A_{\mu_0} : L^2(\mu_0) \to L^2(\nu_0)$ *by* $\zeta \mapsto \frac{A(\zeta f_0)}{Af_0}$. *Then*

$$T(\nu_0, A\mathscr{P}) = A_{\mu_0} L^2_*(\mu_0).$$

100

Proof. Suppose $\mu = \zeta\mu_0 = \zeta f_0 \, d\lambda^2 \in \mathscr{P}$. Then

$$
\begin{aligned}
A\mu &= A(\zeta f_0) \, ds \, d\theta \\
&= \frac{A(\zeta f_0)}{A f_0} \nu_0,
\end{aligned}
$$

so the Radon transform of the measure with μ_0-density ζ is represented by the ν_0-density $A_{\mu_0}\zeta$. We start by showing that $A_{\mu_0} L_*^2(\mu_0) \subset T(\mu_0, A\mathscr{P})$. Let $h \in L_*^2(\mu_0)$. Since $T(\mu_0, \mathscr{P}) = L_*^2(\mu_0)$, there exists a path of μ_0-densities ζ_t in $\mathscr{P}$ such that

$$
\zeta_t = 1 + t(h + r_t),
$$

where $\|r_t\|_{L^2(\mu_0)} \to 0$. Thus

$$
A_{\mu_0}\zeta t = A_{\mu_0} 1 + t(A_{\mu_0} h + A_{\mu_0} r_t).
$$

Since A_{μ_0} is a bounded operator on $L^2(\mu_0)$ (using corollary 3.4.3), $\|A_{\mu_0} r_t\|_{L^2(\nu_0)} \to 0$, so $A_{\mu_0} h$ is in the tangent space of $\mathscr{P}$ at ν_0. This shows that $A_{\mu_0} L_*^2(\mu_0) \subset T(\nu_0, A\mathscr{P})$. To prove the opposite inclusion, suppose η is in the tangent space at ν_0. Then there is a path in $A\mathscr{P}$ whose ν_0-density is given by

$$
A_{\mu_0}\zeta_t = A_{\mu_0} 1 + t(\eta + r_t),
$$

where ζ_t is a path of μ_0-densities in $\mathscr{P}$ and $\|r_t\|_{L^2(\nu_0)} \to 0$. (Note that $A_{\mu_0} 1 = 1$.) Solving this equation for η gives

$$
\begin{aligned}
\eta &= \frac{A_{\mu_0}\zeta_t - A_{\mu_0} 1}{t} - r_t \\
&= A_{\mu_0}\left(\frac{\zeta_t - 1}{t}\right) - r_t,
\end{aligned}
$$

so η is in the closure of the range of A_{μ_0}. But $A_{\mu_0} L_*^2(\mu_0)$, being the image of a closed set under a continuous map, is closed. Thus, $A_{\mu_0} L_*^2(\mu_0) \subset T(\nu_0, A\mathscr{P})$ and we conclude that $T(\mu_0, A\mathscr{P}) = A_{\mu_0} L_*^2(\mu_0)$. $\square$

We are now ready to compute the local asymptotic minimax lower bound. Sup-

pose $\phi \in \mathscr{S}(\mathbb{R}^2)$. In section 6.3, we established that the problem of estimating the functional $\dot{\phi}(\mu)$ is equivalent to the problem of estimating the functional $\dot{\psi}(T\mu)$, where $\psi = T^{-*}\phi$. A gradient of this functional at $\nu_0 = A\mu_0$ is given by $\psi - \dot{\psi}(\nu_0)$ [Pfa82, prop. 5.1.5]. Letting p_{μ_0} denote the projection onto $A_{\mu_0} L_*^2(\mu_0) \subset L^2(\nu_0)$, it follows that the canonical gradient of this functional at $\nu_0 = A\mu_0$ is given by $\dot{\psi}^* = p_{\mu_0}[\psi - \dot{\psi}(\nu_0)]$. We will consider neighborhoods of ν_0 of the form

$$\bar{B}(r, \nu_0) \stackrel{\text{def}}{=} \{\nu \in A\mathscr{P} : \nu = (1 + \xi)\nu_0 \text{ with } \|\xi\|_{L_*^2(\nu_0)} \leq r\}.$$

Using the local asymptotic minimax bound as formulated by [Mil83, prop. XII.2.7], we get the result that

$$\lim_{c \to \infty} \lim_{n \to \infty} \inf_{\hat{\psi}_n} \sup_{\nu \in B(cn^{-1/2}, \nu_0)} n E_\nu[\hat{\psi}_n - \dot{\psi}(\nu)]^2 \geq \|\dot{\psi}^*\|_{L^2(\nu_0)}^2. \tag{6.4}$$

(Note: there is apparently a typographical error in the statement of proposition XII.2.7 in [Mil83], with the $\lim_{n \to \infty}$ as in equation 6.4 being omitted. It is, however, present throughout the proof.)

Definition 6.4.6 An estimator $\hat{\psi}_n$ is said to be locally asymptotically minimax at $\nu_0 \in A\mathscr{P}$ if

$$\lim_{c \to \infty} \lim_{n \to \infty} \sup_{\nu^n \in B(cn^{-1/2}, \nu_0)} n E_\nu[\hat{\psi}_n - \dot{\psi}(\nu)]^2 = \lim_{c \to \infty} \lim_{n \to \infty} \inf_{\hat{\psi}_n} \sup_{\nu^n \in B(cn^{-1/2}, \nu_0)} n E_\nu[\hat{\psi}_n - \dot{\psi}(\nu)]^2,$$

i.e., if it achieves the lower bound given in equation 6.4. If $\hat{\psi}_n$ is locally asymptotically minimax at each $\nu_0 \in A\mathscr{P}$, $\hat{\psi}_n$ is said to be locally asymptotically minimax, cf. [Mil83, def. VII.2.4].

It is easy to construct a locally asymptotically minimax estimator for $\dot{\psi}(\nu)$ at $\nu_0 \in A\mathscr{P}$.

Proposition 6.4.7 *The estimator*

$$\hat{\psi}_n \stackrel{\text{def}}{=} \dot{\psi}(\nu_0) + \dot{\psi}^*(\nu_n) \tag{6.5}$$

is locally asymptotically minimax for $\dot{\psi}(\nu)$ at $\nu_0 \in A\mathscr{P}$.

Proof. Suppose $\xi \in L^2_*(\nu_0)$ (note that $\xi \in L^2_*(\nu_0)$ is a necessary condition for $(1 + \xi)\nu_0$ to be a probability measure). The mean of $\hat{\psi}_n$ is given by

$$
\begin{aligned}
E_{[(1+\xi)\nu_0]^n}\hat{\psi}_n &= \dot{\psi}(\nu_0) + \int_{G_{1,2}} p_{\mu_0}[\psi - \dot{\psi}(\nu_0)](1 + \xi)\,d\nu_0 \\
&= \dot{\psi}(\nu_0) + \int_{G_{1,2}} p_{\mu_0}[\psi - \dot{\psi}(\nu_0)]\xi\,d\nu_0 \\
&= \dot{\psi}(\nu_0) + \int_{G_{1,2}} [\psi - \dot{\psi}(\nu_0)]\xi\,d\nu_0 \\
&= \dot{\psi}(\nu_0) + \int_{G_{1,2}} \psi\xi\,d\nu_0 \\
&= \dot{\psi}[(1 + \xi)\nu_0],
\end{aligned}
$$

where the second equality follows since $p_{\mu_0}[\psi - \dot{\psi}(\nu_0)] \in A_{\mu_0}L^2_*(\mu_0) \subset L^2_*(\nu_0)$, and the third and fourth equalities follow since $\xi \in L^2_*(\nu_0)$. The variance of $\hat{\psi}_n$ is given by

$$
\mathrm{var}_{[(1+\xi)\nu_0]^n}\,\hat{\psi}_n = n^{-1}\int_{G_{1,2}} [\dot{\psi}^* - (\psi|\xi)_{L^2(\nu_0)}]^2(1 + \xi)\,d\nu_0. \tag{6.6}
$$

Thus the risk at ν_0 is equal to the lower bound given in equation 6.4. The result now follows from the continuity of the risk of this estimator with respect to the $L^2_*(\nu_0)$-distance on ν_0-densities. $\square$

The trouble with proposition 6.4.7 is, of course, that the estimator depends on the unknown measure μ_0. One approach to building a practical estimator might be to use some auxiliary estimator of μ_0 in conjunction with the estimator in proposition 6.4.7, cf. [Mil83, sec. VII.3]. We will content ourselves here with constructing a simple linear estimator which, while suboptimal, is an improvement over the naive estimator proposed in section 6.3. The idea is that the estimator in section 6.3 would be optimal if ν were allowed to be an arbitrary measure on $G_{1,2}$. However, we have some additional information about ν, namely that it is in the range of the Radon transform. This suggests replacing the function ψ in the estimator in section 6.3 with its projection onto a suitable subspace.

Definition 6.4.8 Define p'_{μ_0} to be the projection onto the set $A_{\mu_0} L^2(\mu_0)$ in $L^2(\nu_0)$. The following lemma show that this projection operator is actually independent of $\mu_0 \in \mathscr{P}$.

Lemma 6.4.9 *If* $\mu_0, \mu_1 \in \mathscr{P}$, *then* $p'_{\mu_0} = p'_{\mu_1}$.

Proof. We use the fact that $p'_\mu \xi$ is uniquely determined by the condition

$$\int_{G_{1,2}} p'_\mu \xi \, \eta \, d\nu = \int_{G_{1,2}} \xi \, \eta \, d\nu$$

for all $\eta \in A_\mu L^2_*(\mu)$, where $\nu = T\mu$ [Pfa82, p. 69, prop. 4.2.1(iii)]. Let $\eta = A_{\mu_1} h \in A_{\mu_1} L^2_*(\mu_0)$, $\mu_0 = f_0 \, ds \, d\theta$, and $\mu_1 = f_1 \, ds \, d\theta$. We have

$$
\begin{aligned}
\int_{G_{1,2}} p'_{\mu_0} \xi \, \eta \, d\nu_1 &= \int_{G_{1,2}} p'_{\mu_0} \xi \frac{A(hf_1)}{Af_1} Af_1 \, ds \, d\theta \\
&= \int_{G_{1,2}} p'_{\mu_0} \xi \frac{A(h\frac{f_1}{f_0} f_0)}{Af_0} Af_0 \, ds \, d\theta \\
&= \int_{G_{1,2}} p'_{\mu_0} \xi A_{\mu_0}(hf_1/f_0) \, d\nu_0 \\
&= \int_{G_{1,2}} \xi A_{\mu_0}(hf_1/f_0) \, d\nu_0 \\
&= \int_{G_{1,2}} \xi \frac{A(h\frac{f_1}{f_0} f_0)}{Af_0} \frac{Af_0}{Af_1} Af_1 \, ds \, d\theta \\
&= \int_{G_{1,2}} \xi \frac{A(hf_1)}{Af_1} \, d\nu_1 \\
&= \int_{G_{1,2}} \xi \eta \, d\nu_1,
\end{aligned}
$$

which proves the claim. $\square$

Thus, taking μ_0 to be the uniform distribution on S, for example, the estimator

$$\hat{\phi}_n \stackrel{\text{def}}{=} \int_{G_{1,2}} p'_{\mu_0} \psi \, d\nu_n \tag{6.7}$$

104

has, by an argument similar to that given in section 6.3, at $\mu = \zeta\mu_0$, mean

$$
\begin{aligned}
E_{(A\mu)^n}\,\hat{\dot{\phi}}_n &= \int_{G_{1,2}} p'_{\mu_0}\,\psi A_{\mu_0}\zeta\,d\nu_0 \\
&= \int_{G_{1,2}} \psi A_{\mu_0}\zeta\,d\nu_0 \\
&= \dot{\phi}(\mu).
\end{aligned}
$$

and variance

$$
\text{var}_{(A\mu)^n}\,\hat{\dot{\phi}}_n = n^{-1}\left[\int_{G_{1,2}} (p'_{\mu_0}\psi)^2\,dA\mu - [\dot{\phi}(\mu)]^2\right]. \tag{6.8}
$$

It would be of interest to compare the quantities in equations 6.6 and 6.8 numerically to see how performance is lost by this approach.

Remark 6.4.10 The reader may understandably wonder about the somewhat complicated notion of optimality proposed in definition 6.4.6. Some motivation for the definition and a discussion of why a simpler definition of optimality will not work may be found in [Mil83, sec. VII.2]. A defense of the use in definition 6.4.6 of neighborhoods whose radii shrink at a rate proportional to $n^{-1/2}$ is given in [Mil83, sec. XI.2].

Appendix A

Some Auxiliary Results

In this appendix, we shall state and prove some auxiliary results that are used in the body of this work. In section A.1, we discuss an alternative way of defining Sobolev spaces on $S^1 \times \mathbb{R}$ that is used in the literature and relate it to some of the Sobolev norms used here. In section A.2, we compute the Radon transform of a circle.

A.1   An alternative Sobolev space definition

Definition A.1.1 For $\alpha \in \mathbb{N}$, define the norm $|| \cdot ||_{L^2_\alpha(S^1 \times \mathbb{R})}$ for functions $g \in L^2(S^1 \times \mathbb{R})$ by

$$||g||^2_{L^2_\alpha(S^1 \times \mathbb{R})} \stackrel{\text{def}}{=} \sum_{|\beta|+k \leq \alpha} ||\partial_x^\beta \partial_s^k g||^2_{L^2(S^1 \times \mathbb{R})},$$

where the differential operator ∂_x^β is defined as in definition 3.2.9 and the derivatives are taken in the sense of tempered distributions on $S^1 \times \mathbb{R}$. The space $\bar{L}^2_\alpha(S^1 \times \mathbb{R})$ is defined to be the subspace of $L^2(S^1 \times \mathbb{R})$ for which this norm exists and is finite [Nat86, p. 45]. For arbitrary $\alpha \geq 0$, $L^2_\alpha(S^1 \times \mathbb{R})$ is defined by interpolation. That is, if $\alpha = n + \beta$ with $n \in \mathbb{N}$ and $0 < \beta < 1$, then

$$L^2_\alpha(S^1 \times \mathbb{R}) \stackrel{\text{def}}{=} \left(L^2_n(S^1 \times \mathbb{R}), L^2_{n+1}(S^1 \times \mathbb{R}) \right)_\beta.$$

In [Nat86, thm. II.5.2], the norm $||\cdot||_{L^2(S^1\times\mathbb{R})}$ is shown to be equivalent to the norm $||\cdot||_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}$ on $\tilde{R}[C_c^\infty(\Omega)]$ for $\alpha\geq 0$. (Note that $||\cdot||_{H^\alpha(Z)}$ as defined in [Nat86, p. 42] is the same as $||\cdot||_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}$ defined here.) We shall now extend this result to $\tilde{R}[L_\alpha^2(r\bar{\Omega})]$.

Lemma A.1.2 *For* $\alpha\geq 0$, *the norms* $||\cdot||_{L_\alpha^2(S^1\times\mathbb{R})}$ *and* $||\cdot||_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}$ *are equivalent on* $\tilde{R}[L_{\alpha-1/2}^2(r\bar{\Omega})]$.

Proof. The proof proceeds along the lines of the proof of [Nat86, thm. II.5.2]. We start by noting that the well-known identity

$$\partial_x^\beta Rf = -1^{|\beta|}\partial_s^{|\beta|}R(x^\beta f)$$

[Nat86, eq. II.1.5] extends, in a distributional sense, to the case where $f \in \mathscr{E}'(\mathbb{R}^2)$ [Her83, p. 173]. Suppose $u \in L_{\alpha-1/2}^2(r\bar{\Omega})$ and $\alpha \in \mathbb{N}$. Then, using remark 3.2.13,

$$
\begin{aligned}
||Ru||^2_{L_\alpha^2(S^1\times\mathbb{R})} &= \sum_{|\beta|+k\leq\alpha} ||\partial_x^\beta\partial_s^k Ru||^2_{L^2(S^1\times\mathbb{R})}\\
&= \sum_{|\beta|+k\leq\alpha} ||\partial_s^{|\beta|+k}R(x^\beta u)||^2_{L^2(S^1\times\mathbb{R})}\\
&= \sum_{|\beta|\leq\alpha}\sum_{l=|\beta|}^\alpha ||\partial_s^l R(x^\beta u)||^2_{L^2(S^1\times\mathbb{R})}\\
&\leq \sum_{|\beta|\leq\alpha}\sum_{l=0}^\alpha \binom{\alpha}{j} ||\partial_s^l R(x^\beta u)||^2_{L^2(S^1\times\mathbb{R})}\\
&= \sum_{|\beta|\leq\alpha} ||R(x^\beta u)||^2_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}.
\end{aligned}
$$

By proposition 3.4.2, there exist constants d_1 and d_2, depending only on α and r, such that

$$||Ru||^2_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})} \leq d_1||u||^2_{L_{\alpha-1/2}^2(\mathbb{R}^2)}$$

and

$$||u||^2_{L_{\alpha-1/2}^2(\mathbb{R}^2)} \leq d_2||Ru||^2_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}$$

for all $u \in L^2_{\alpha-1/2}(r\bar\Omega)$. It follows that

$$\|Ru\|^2_{L^2_\alpha(S^1\times\mathbb{R})} \le d_1 \sum_{|\beta|\le\alpha} \|x^\beta u\|^2_{L^2_{\alpha-1/2}(\mathbb{R}^2)}.$$

Since u has compact support, the effect of multiplying by x^β can be achieved by multiplying by some function $\phi_\beta \in C^\infty_c(\mathbb{R}^2)$. Thus there exist constants $d_3(\beta)$, depending only on α, r, and β, such that $\|x^\beta u\|^2_{L^2_{\alpha-1/2}(\mathbb{R}^2)} \le d_3(\beta)\|u\|^2_{L^2_{\alpha-1/2}(\mathbb{R}^2)}$ for all $u \in L^2_{\alpha-1/2}(r\bar\Omega)$. It follows that

$$
\begin{aligned}
\|Ru\|^2_{L^2_\alpha(S^1\times\mathbb{R})} &\le d_1 \sum_{|\beta|\le\alpha} d_3(\beta)\|u\|^2_{L^2_{\alpha-1/2}(\mathbb{R}^2)} \\
&\le d_1 d_2 \sum_{|\beta|\le\alpha} d_3(\beta)\|Ru\|^2_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})}.
\end{aligned}
$$

Going the other way, again using remark 3.2.13, we have

$$
\begin{aligned}
\|Ru\|^2_{L^2_{(0,\alpha)}(S^1\times\mathbb{R})} &= \sum_{l=0}^\alpha \binom{\alpha}{j} \|\partial^l_s Ru\|^2_{L^2(S^1\times\mathbb{R})} \\
&\le \sup_{0\le j\le\alpha} \binom{\alpha}{j} \sum_{l=0}^\alpha \|\partial^l_s Ru\|^2_{L^2(S^1\times\mathbb{R})} \\
&\le \sup_{0\le j\le\alpha} \binom{\alpha}{j} \sum_{|\beta|+k\le\alpha} \|\partial^\beta_x \partial^k_s Ru\|^2_{L^2(S^1\times\mathbb{R})} \\
&= \sup_{0\le j\le\alpha} \binom{\alpha}{j} \|Ru\|^2_{L^2_\alpha(S^1\times\mathbb{R})}.
\end{aligned}
$$

This proves the result for $\alpha \in \mathbb{N}$. The result for general $\alpha \ge 0$ follows by an interpolation argument which is essentially the same as the one given in the proof of [Nat86, thm. 5.2]. We omit the details. $\square$

Lemma A.1.3 *For $\alpha \ge 0$, the norms $\|\cdot\|_{L^2_\alpha(S^1\times\mathbb{R})}$ and $\|\cdot\|_{L^2_\alpha(S^1\times 2r\mathbb{T})}$ are equivalent on $L^2(S^1 \times [-r,r])$.*

Proof. The statement is trivial for $\alpha = 0$. To prove the statement for $\alpha = 1$, note that, by lemma A.1.5 below,

$$[\partial_\theta g(\omega(\theta), s)]^2 = [-\sin\theta \partial_{x_1} g(\omega(\theta), s) + \cos\theta \partial_{x_2}(\omega(\theta), s)]^2.$$

Using the inequality

$$(a + b)^2 \leq 2(a^2 + b^2),$$

it follows that

$$\begin{aligned}
[\partial_\theta g(\omega(\theta), s)]^2 &\leq 2\{\sin^2\theta[\partial_{x_1} g(\omega(\theta), s)]^2 + \cos^2\theta[\partial_{x_2} g(\omega(\theta), s)]^2\} \\
&\leq 2\{[\partial_{x_1} g(\omega(\theta), s)]^2 + [\partial_{x_2} g(\omega(\theta), s)]^2\},
\end{aligned}$$

hence

$$\|\partial_\theta g\|^2_{L^2(S^1 \times \mathbb{R})} \leq 2(\|\partial_{x_1} g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_{x_2} g\|^2_{L^2(S^1 \times \mathbb{R})}).$$

By remark 3.3.11, it follows that, for all $g \in L^2(S^1 \times [-r, r])$,

$$\begin{aligned}
&\|g\|^2_{L^2_1(S^1 \times 2r\mathbb{T}^1)} \\
&= \|g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_\theta g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_s g\|^2_{L^2(S^1 \times \mathbb{R})} \\
&\leq 2\left(\|g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_{x_1} g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_{x_2} g\|^2_{L^2(S^1 \times \mathbb{R})} + \|\partial_s g\|^2_{L^2(S^1 \times \mathbb{R})}\right) \\
&= 2\|g\|^2_{L^2_1(S^1 \times \mathbb{R})},
\end{aligned}$$

hence $\|g\|_{L^2_1(S^1 \times 2r\mathbb{T}^1)} \leq \sqrt{2}\|g\|_{L^2_1(S^1 \times \mathbb{R})}$.

To get an inequality in the opposite direction, note that, again by lemma A.1.5,

$$[\partial_{x_1} g(\omega(\theta), s)]^2 = \{-\sin\theta \partial_\theta g(\omega(\vartheta), s) - \cos\theta[s\partial_s g(\omega(\theta), s) + g(\omega(\theta), s)]\}^2$$

and

$$[\partial_{x_2} g(\omega(\theta), s)]^2 = \{\cos\theta \partial_\theta g(\omega(\theta), s) - \sin\theta[s\partial_s g(\omega(\theta), s) + g(\omega(\theta), s)]\}^2.$$

Using the inequality

$$(a + b + c)^2 \leq 3(a^2 + b^2 + c^2),$$

it follows that

$$[\partial_{x_1} g(\omega(\theta), s)]^2$$
$$\leq 3(\sin^2 \theta [\partial_\theta g(\omega(\theta), s)]^2 + s^2 \cos^2 \theta [\partial_s g(\omega(\theta), s)]^2 + \cos^2 \theta [g(\omega(\theta), s)]^2)$$
$$\leq 3([\partial_\theta g(\omega(\theta), s)]^2 + r^2 [\partial_s g(\omega(\theta), s)]^2 + [g(\omega(\theta), s)]^2),$$

and

$$[\partial_{x_2} g(\omega(\theta), s)]^2$$
$$\leq 3(\cos^2 \theta [\partial_\theta g(\omega(\theta), s)]^2 + s^2 \sin^2 \theta [\partial_s g(\omega(\theta), s)]^2 + \sin^2 \theta [g(\omega(\theta), s)]^2).$$
$$\leq 3([\partial_\theta g(\omega(\theta), s)]^2 + r^2 [\partial_s g(\omega(\theta), s)]^2 + [g(\omega(\theta), s)]^2),$$

hence

$$||\partial_{x_1} g||^2_{L^2(S^1 \times \mathbb{R})} \leq 3(||\partial_\theta g||^2_{L^2(S^1 \times \mathbb{R})} + r^2 ||\partial_s g||^2_{L^2(S^1 \times \mathbb{R})} + ||g||^2_{L^2(S^1 \times \mathbb{R})}),$$

and

$$||\partial_{x_2} g||^2_{L^2(S^1 \times \mathbb{R})} \leq 3(||\partial_\theta g||^2_{L^2(S^1 \times \mathbb{R})} + r^2 ||\partial_s g||^2_{L^2(S^1 \times \mathbb{R})} + ||g||^2_{L^2(S^1 \times \mathbb{R})}).$$

It follows that, for all $g \in L^2(S^1 \times [-r, r])$,

$$||g||^2_{L^2_1(S^1 \times \mathbb{R})} = ||g||^2_{L^2(S^1 \times \mathbb{R})} + ||\partial_{x_1} g||^2_{L^2(S^1 \times \mathbb{R})} + ||\partial_{x_2} g||^2_{L^2(S^1 \times \mathbb{R})} + ||\partial_s g||^2_{L^2(S^1 \times \mathbb{R})}$$
$$\leq 7 \max(1, r^2) \left(||g||^2_{L^2(S^1 \times \mathbb{R})} + ||\partial_\theta g||^2_{L^2(S^1 \times \mathbb{R})} + ||\partial_s g||^2_{L^2(S^1 \times \mathbb{R})} \right)$$
$$= 7 \max(1, r^2) ||g||^2_{L^2_1(S^1 \times 2r\mathbb{T}^1)},$$

hence $||g||_{L^2_1(S^1 \times \mathbb{R})} \leq \sqrt{7} \max(1, r) ||g||_{L^2_1(S^1 \times 2r\mathbb{T}^1)}$. The result for $\alpha = 1$ follows immediately from these estimates. The result for $\alpha \in \mathbb{N}$ follows by applying a similar argument inductively. The result for arbitrary $\alpha \geq 0$ follows by a standard interpolation argument using proposition 3.1.2. $\square$

Lemma A.1.4 *For $\alpha \geq 0$, the norms $|| \cdot ||_{L^2_{(0,\alpha)}(S^1 \times \mathbb{R})}$ and $|| \cdot ||_{L^2_\alpha(S^1 \times 2r\mathbb{T})}$ are equivalent*

on $\tilde{R}[L^2_{\alpha-1/2}(r\bar{\Omega})]$.

Proof. Combine lemmas A.1.2 and A.1.3, using the fact that $\tilde{R}[L^2_{\alpha-1/2}(r\bar{\Omega})] \subset L^2(S^1 \times [-r, r])$. □

It remains to prove the lemma that was used in the proof of lemma A.1.3.

Lemma A.1.5 *The differential operators on $S^1 \times \mathbb{R}$ defined in definitions 3.2.8 and 3.2.9 are related by the equations*

$$\partial_\theta = -\sin\theta\partial_{x_1} + \cos\theta\partial_{x_2},$$

$$\partial_{x_1} = -\sin\theta\partial_\theta + \cos\theta(s\partial_s + 1),$$

and

$$\partial_{x_2} = \cos\theta\partial_\theta g - \sin\theta(s\partial_s + 1).$$

Proof. Let $g \in C^\infty(S^1 \times \mathbb{R})$ and $\tilde{g}$ its extension to $\mathbb{R}^2\backslash\{0\} \times \mathbb{R}$ as described in definition 3.2.9. Let $(\omega(\theta_0), s_0) = ((\cos\theta_0, \sin\theta_0), s_0) \in S^1 \times \mathbb{R}$. Since the curve $\theta \mapsto ((\cos\theta, \sin\theta), s_0)$ in $\mathbb{R}^2\backslash\{0\} \times \mathbb{R}$ has the tangent vector $((-\sin\theta_0, \cos\theta_0), 0)$ at $(\omega(\theta_0), s_0)$,

$$\partial_\theta\tilde{g}(\omega(\theta_0), s_0) = -\sin\theta_0\partial_{x_1}\tilde{g}(\omega(\theta_0), s_0) + \cos\theta_0\partial_{x_2}\tilde{g}(\omega(\theta_0), s_0).$$

Define the differential operator ∂_r on $\mathbb{R}^2\backslash\{0\} \times \mathbb{R}$ by

$$\partial_r\tilde{g} = \cos\theta_0\partial_{x_1}\tilde{g} + \sin\theta_0\partial_{x_2}\tilde{g}.$$

It is easy to see that $\partial_r\tilde{g}(\omega(\theta_0), s_0)$ is obtained by differentiating $\tilde{g}$ along the curve $r \mapsto (r(\cos\theta_0, \sin\theta_0), s_0)$ and evaluating at $r = 1$. Solving for ∂_{x_1} and ∂_{x_2} in terms of ∂_θ and ∂_r, we get

$$\partial_{x_1}\tilde{g}(\omega(\theta_0), s_0) = -\sin\theta_0\partial_\theta\tilde{g}(\omega(\theta_0), s_0) + \cos\theta_0\partial_r\tilde{g}(\omega(\theta_0), s_0)$$

and

$$\partial_{x_2}\tilde{g}(\omega(\theta_0), s_0) = \cos\theta_0 \partial_\theta \tilde{g}(\omega(\theta_0), s_0) + \sin\theta_0 \partial_r \tilde{g}(\omega(\theta_0), s_0).$$

Computing $\partial_r \tilde{g}(\omega(\theta_0), s_0)$ explicitly, we have

$$
\begin{aligned}
\partial_r \tilde{g}(\omega(\theta_0), s_0) &= \frac{d}{dr}\left[\tilde{g}(r(\cos\theta_0, \sin\theta_0), s_0)\right]\Big|_{r=1} \\
&= \frac{d}{dr}\left(\frac{g((\cos\theta_0, \sin\theta_0), s_0/r)}{r}\right)\Big|_{r=1} \\
&= \frac{r\partial_s g((\cos\theta_0, \sin\theta_0), s_0/r) \cdot (-s_0 r^{-2}) - g((\cos\theta_0, \sin\theta_0), s_0/r)}{r^2}\Big|_{r=1} \\
&= -s_0 \partial_s g(\omega(\theta_0), s_0) - g(\omega(\theta_0), s_0).
\end{aligned}
$$

It follows that

$$\partial_{x_1} g(\omega(\theta_0), s_0) = -\sin\theta_0 \partial_\theta g(\omega(\theta_0), s_0) - \cos\theta_0[s_0 \partial_s g(\omega(\theta_0), s_0) + g(\omega(\theta_0), s_0)]$$

and

$$\partial_{x_2} g(\omega(\theta_0), s_0) = \cos\theta_0 \partial_\theta g(\omega(\theta_0), s_0) - \sin\theta_0[s_0 \partial_s g(\omega(\theta_0), s_0) + g(\omega(\theta_0), s_0)]. \quad \square$$

A.2   The Radon transform of a circle

In this section, we will compute the Radon transform of a circle of radius r centered at the origin.

Proposition A.2.1 *Let μ denote the probability measure on $\mathbb{R}^2$ that is concentrated on the circle of radius $r > 0$ centered at the origin and is uniformly distributed with respect to $d\theta$. Then*

$$T\mu = \begin{cases} \frac{1}{\pi\sqrt{r^2 - s^2}}\, d\theta\, ds & \text{if } |s| \leq r \\ 0\, d\theta\, ds & \text{if } |s| > r \end{cases} .$$

Proof.   Define $E \subset G_{1,2}$ by $E \overset{\text{def}}{=} \pi(S^1 \times [-s, s])$. Using proposition 2.2.13, we have

$$
\begin{aligned}
\nu(E) &= \int_{\mathbb{R}^2} R^* 1_E\, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \tilde{R}^* \tilde{1}_E\, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \tilde{R}^* 1_{S^1 \times [-s,s]}\, d\mu \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} 1_{S^1 \times [-s,s]}(\omega(\theta), x \cdot \omega(\theta))\, d\theta\, d\mu(x). \\
&= \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} 1_{S^1 \times [-s,s]}(\omega(\theta), r\omega(\theta') \cdot \omega(\theta))\, d\theta\, d\theta' \\
&= \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} 1_{[-s,s]}(r\omega(\theta') \cdot \omega(\theta))\, d\theta\, d\theta',
\end{aligned}
$$

where we write x as $r\omega(\theta')$. For fixed r and θ', we need to determine for which values of θ the condition $-s \leq r\omega(\theta') \cdot \omega(\theta) \leq s$ is satisfied. Now $-s \leq r\omega(\theta') \cdot \omega(\theta) \leq s$ if and only if $-s/r \leq \cos(\theta' - \theta) \leq s/r$. If $r < s$, then the condition $-s/r \leq \cos(\theta' - \theta) \leq s/r$ is satisfied for all θ. Thus, in this case

$$
\begin{aligned}
\nu(E) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\theta\, d\theta' \\
&= \pi.
\end{aligned}
$$

If $r \geq s$, then the condition $-s/r \leq \cos(\theta' - \theta) \leq s/r$ is satisfied if and only if $\arccos(s/r) \leq |\theta' - \theta| \leq \arccos(-s/r)$. It follows that

$$
\begin{aligned}
\nu(E) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{\arccos(s/r)}^{\arccos(-s/r)} d\theta \, d\theta' \\
&= \arccos(-s/r) - \arccos(s/r) \\
&= \pi - 2\arccos(s/r) \\
&= 2\arcsin(s/r).
\end{aligned}
$$

From considerations of symmetry, it is clear that $R\mu$ can be represented by a density function f with respect to $ds \, d\theta$ that depends only on s and is symmetric on s. Such a function must satisfy

$$
\begin{aligned}
\int_0^{\pi} \int_{-s}^{s} f(s) \, ds \, d\theta &= 2\pi \int_0^{s} f(s) \, ds \\
&= \begin{cases} 2\arcsin(s/r) & \text{if } s \leq r \\ \pi & \text{if } s > r \end{cases}.
\end{aligned}
$$

Using the fact that

$$
\frac{d}{ds} \arcsin\left(\frac{s}{r}\right) = \frac{1}{\sqrt{r^2 - s^2}},
$$

we get the desired result. $\square$

Bibliography

[AMR83] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Addison-Wesley Publishing Co., Reading, MA, 1983.

[Apo74] T. M. Apostol. *Mathematical Analysis*. Addison-Wesley Publishing Co., Reading, MA, second edition, 1974.

[Bak91] J. R. Baker. *Spatially Variant Tomographic Imaging: Estimation, Identification, and Optimization*. PhD thesis, University of California at Berkeley, Berkeley, CA, 1991.

[Bir83] L. Birgé. Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete*, 65:181–237, 1983.

[Chu74] K. L. Chung. *A Course in Probability Theory*. Probability and Mathematical Statistics Series. Academic Press, New York, second edition, 1974.

[Dea83] S. R. Deans. *The Radon Transform and some of its Applications*. John Wiley & Sons, New York, 1983.

[Dev87] L. Devroye. *A Course in Density Estimation*. Birkhäuser, Boston, 1987.

[Dev89] L. Devroye. Consistent deconvolution in density estimation. *Can. J. Statist.*, 17:235–239, 1989.

[DL91] D. L. Donoho and R. C. Liu. Geometrizing rates of convergence, II. *Ann. Statist.*, 19:633–667, 1991.

[Dud89] R. M. Dudley. *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA, 1989.

[Fol84] G. B. Folland. *Real Analysis*. John Wiley & Sons, New York, 1984.

[Gin75] E. Giné. Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Probab.*, 3:1243–1266, 1975.

[Gon91] F. B. Gonzales. On the range of the Radon d-plane transform and its dual. *Trans. Amer. Math. Soc.*, 327:601–61, 1991.

[Hel80] S. Helgason. *The Radon Transform*. Birkhauser, Boston, 1980.

[Her83]   A. Hertle. Continuity of the Radon transform and its inverse on Euclidean space. *Math. Z.*, 184:165–192, 1983.

[HQ85]   M. G. Hahn and E. T. Quinto. Distances between measures from 1-dimensional projections as implied by continuity of the inverse Radon transform. *Z. Wahrsch. verw. Gebiete*, 70:361–380, 1985.

[Itô85]   K. Itô. *Encyclopedic Dictionary of Mathematics*. MIT Press, Cambridge, MA, second edition, 1985.

[JS90]   I. M. Johnstone and B. W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, 18:251–280, 1990.

[LM72]   J. L. Lions and E. Magenes. *Non-homogeneous Boundary Value Problems and Applications*, volume I. Springer-Verlag, New York, 1972.

[Mil83]   P. W. Millar. The minimax principle in asymptotic statistical theory. In P. L. Hennequin, editor, *École d'été de probabilités de Saint-Flour XI-1982*, number 976 in Lecture Notes in Mathematics, pages 75–265. Springer-Verlag, New York, 1983.

[MP88]   M. A. Mandelkern and M. E. Phelps. Methods and instrumentation for positron emission tomography. In A. Gottschalk, P. B. Hoffer, and E. J. Potchen, editors, *Diagnostic Nuclear Medicine*, Golden's Diagnostic Radiology Series, pages 128–149. Williams and Wilkins, Baltimore, second edition, 1988.

[Nat86]   F. Natterer. *The Mathematics of Computed Tomography*. John Wiley & Sons, New York, 1986.

[Pfa82]   J. Pfanzagl. *Contributions to a General Asymptotic Statistical Theory*. Number 13 in Lecture Notes in Statistics. Springer-Verlag, New York, 1982.

[Pol84]   D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.

[RS80]   M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*. Academic Press, New York, 1980.

[RSN55]   F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Frederick Ungar Publishing Co., New York, 1955.

[Rud87]   W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, third edition, 1987.

[Sch66]   L. Schwartz. *Théorie des Distributions*. Hermann, Paris, 1966.

[Sel72]   S. M. Selby, editor. *CRC Standard Mathematical Tables*. The Chemical Rubber Co., Cleveland, OH, 20th edition, 1972.

[SP87]   J. A. Sorenson and M. E. Phelps. *Physics in Nuclear Medicine.* Grune & Stratton, Inc., Orlando, FL, second edition, 1987.

[Ste70] E. M. Stein. *Singular Integrals and Differentiability Properties of Integrals.* Princeton University Press, Princeton, NJ, 1970.

[Str85] H. Strasser. *Mathematical Theory of Statistics.* Walter de Gruyter, New York, 1985.

[Tre67] F. Treves. *Topological Vector Spaces, Distributions, and Kernels.* Academic Press, New York, 1967.

[VSK85] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Stat. Assn.*, 80:8–20, 1985.

Index

projective space, 20

R

$\mathbb{R}$, 17

$\mathbb{R}^+$, 17

$\mathbb{R}^d$, 17

$\tilde{R}$, 24

$\tilde{R}^*$, 24

R, 24

R^*, 24

$r\mathbb{T}^2$, 46

rS, 46

$r\bar{\Omega}$, 51

$r\mathbb{T}$, 51

Radon measure, 31

Radon transform

    on distributions, 29

    on functions, 24

randomization, 78

randomized decision function, 8

Riesz potential operator

    for functions on $\mathbb{R}^d$, 55

    for functions on $S^1 \times \mathbb{R}$, 56

risk, 8

S

S, 7

S^1, 19

σ, 20

$\mathscr{S}'(\mathbb{R}^d)$, 40

$\mathscr{S}(\mathbb{R}^d)$, 40

$\mathscr{S}'(S^1 \times \mathbb{R})$, 43

$\mathscr{S}(S^1 \times \mathbb{R})$, 43

$\mathscr{S}'(G_{1,2})$, 45

$\mathscr{S}(G_{1,2})$, 45

sample space, 8

Schwartz space

    on $\mathbb{R}^d$, 40

    on $G_{1,2}$, 45

    on $S^1 \times \mathbb{R}$, 43

singular value decomposition, 10

smoothness, 11, 15

Sobolev space

    on $\mathbb{R}^d$, 40

    on $G_{1,2}$, 45

    on $G_{1,2}(r)$, 53

    on $r\mathbb{T}^2$, 47

    on $S^1 \times \mathbb{R}$, 44

    on $S^1 \times r\mathbb{T}^1$, 52

standard covering map, 22

standard Radon transform

    on distributions, 29

    on functions, 24

statistical decision problem, 8

statistical decision theory, 8

statistical estimation problem, 9

statistical experiment, 8

strong convergence , 13

surface measure, 20

T