Topics in Statistics and Convex Geometry: Rounding, Sampling, and Interpolation

Adam M. Gustafson

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Hariharan Narayanan, Chair

Michael D. Perlman

Sébastien Bubeck

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Topics in Statistics and Convex Geometry: Rounding, Sampling, and Interpolation

Adam M. Gustafson

Chair of the Supervisory Committee:
Title of Chair Hariharan Narayanan
Department of Statistics

We consider a few aspects of the interplay between convex geometry and statistics. We consider three problems of interest: how to bring a convex body specified by a self-concordant barrier into a suitably "rounded" position using an affine-invariant random walk; how to design a rapidly-mixing affine-invariant random walk with maximal volume ellipsoids; and how to perform interpolation in multiple dimensions given noisy observations of the original function on a finite set.

We begin with an overview of some background information on convex bodies which recur in this dissertation, discussing polytopes, the Dikin ellipsoid, and John's ellipsoid in particular. We also discuss cutting plane methods and Vaidya's algorithm, which we employ in subsequent analysis. We then review Markov chains on general state spaces to motivate designing rapidly mixing geometric random walks, and the means by which the mixing time may be analyzed.

Using these results regarding convex bodies and general state space Markov chains, along with recently developed concentration inequalities on general state space Markov chains, we employ an affine-invariant random walk using Dikin ellipsoids to provably bring a convex body specified by a self-concordant barrier into an approximately "rounded" position. We also design a random walk using John's ellipsoids and derive its mixing time.

Departing somewhat from these themes, we also discuss regression in multiple dimensions over classes of continuously differentiable functions. We provide a cutting plane algorithm for the case of first-order differentiable functions with Lipschitz gradients. We additionally consider the more general case of higher-order continuously differentiable functions, proving Whitney's extension theorem for this case, and outlining a quadratic program to perform regression for this setting.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation for the patience and guidance of Hariharan Narayanan, his research advisor, and Michael D. Perlman, his academic advisor. He would also like to thank Jon Wakefield for his wisdom and advice.

DEDICATION

To my father, David, in loving memory, for pushing me to achieve.

To my mother, Deborah, for her unending love and support.

To my sister, Danielle, for her compassion.

To my step-father, Stephen, for his encouragement.

To my friend, Kumi, for always lifting my spirits.

Chapter 1

INTRODUCTION

This dissertation considers a few aspects of the interplay between convex geometry and statistics. We consider three problems of interest: how to bring a convex body specified by a self-concordant barrier into a suitably "rounded" position using an affine-invariant random walk; how to design a rapidly-mixing affine-invariant random walk with maximal volume ellipsoids; and how to perform interpolation in multiple dimensions given noisy observations of the original function on a finite set.

Chapter 2 introduces background information on convex bodies that recur in this dissertation. We give a basic overview some convex bodies, such as polytopes and ellipsoids. We discuss a few manners in which convex bodies may be specified, and we also discuss cutting-plane methods for minimizing convex functions using one of these specifications. We finally discuss two types of ellipsoids in detail, namely the Dikin ellipsoid and John's ellipsoid, and the means by which they provide a suitable approximation for certain convex bodies.

Similarly, Chapter 3 introduces background information on Markov chains which recur in this dissertation. Geometric random walks provide the means by which approximately uniform random samples from such convex bodies may be taken, a necessary step in bringing a convex body into a rounded position. This chapter discusses how to construct a random walk with a specified stationary distribution on a convex body, and basic tools for analyzing the mixing time of such random walks.

Chapter 4 provides an algorithm to bring a convex body into a suitably rounded position via an affine transformation. The algorithm makes use of samples drawn from an affine-invariant random walk known as the Dikin walk, which uses samples drawn from Gaussian proposal distributions with covariances specified by scaled Dikin ellipsoids. The proof that this algorithm works rely on recent concentration inequalities for general state space Markov chains.

Chapter 5 proposes another affine-invariant random walk known as John's walk, which utilizes the uniform distribution on scaled John's ellipsoids to draw samples. The mixing time of this random walk is derived, and an algorithm to provide suitably approximate John's ellipsoids using cutting-plane methods are

Departing somewhat from the theme of the previous chapters, we discuss interpolation in multiple dimensions for the class of first-order differentiable functions with Lipschitz gradients in Chapter 6. We discuss noiseless and noisy variants of the problem, the latter which refer to as the regression problem. A cutting-plane method for solving this regression problem is discussed. Sample complexity arguments yielding the choice of the complexity parameter for the regression problem are given. Empirical results of the methods of this chapter are provided.

Finally, in Chapter 7 we conclude with describing an algorithm to extend the methods of Chapter 6 to the class of functions whose m-th order derivatives are continuous. We give a Calderón-Zygmund decomposition and discuss its properties. We provide a proof of Whitney's extension theorem using this decomposition, and show that the function interpolation problem may be solved via a linear program. We also show that its noisy variant may be solved via a quadratic program. We leave the implementation of these methods, including the choice of the complexity parameter, for future work.

We owe the reader a comment on notation used throughout this dissertation. We use lowercase such as a, b to either indicate vectors in the Euclidean space $\mathbb{R}^n$ or constants in $\mathbb{R}$, and the distinction will be obvious depending on the context. We use the notation $a^\top b$ or $\langle a, b \rangle$ to indicate the standard inner product for vectors $a, b \in \mathbb{R}^n$. Given a set $S \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, we denote the gradient of a function $f : S \to \mathbb{R}$ at x as $\nabla f(x) \in \mathbb{R}^n$ and its linear form $Df(x)[h_1] = \nabla f(x)^\top h_1$. Similarly, we denote the Hessian as $\nabla^2 f(x)$ and its bilinear form $D^2 f(x)[h_1, h_2] = h_1^\top \nabla^2 f(x) h_2$. In chapters 6 and 7, we use ∂ in place of D. We use the script notation $\mathcal{K}, \mathcal{X}$, etc., to denote convex bodies in $\mathbb{R}^n$, whereas we use standard uppercase fonts C, S, etc., to denote subsets of $\mathbb{R}^n$. Additionally matrices will use uppercase notation such as A, B, etc. Again the distinction between sets and matrices will be obvious depending on the context. The trace of a square matrix M is denoted $\mathrm{tr}(M)$, and its determinant is denoted $\det(M)$. The notation $\mathrm{diag}(d_1, \ldots, d_m)$ indicates the diagonal matrix with entries $d_1, \ldots, d_m$ on its diagonal. We use $\mathrm{int}(S)$ to denote the interior of a set $S \subset \mathbb{R}^n$, and ∂S to denote its boundary. Linear operators mapping between Banach spaces will be denoted with the typeface, P, I, etc. The notation $\widetilde{O}$ refers to "big O" notation in which polylogarithmic factors and terms depending on the error parameters are suppressed. Finally, the constant $\omega < 2.373$ is the current value of the fast matrix multiplication constant (Williams, 2011; Davie and Stothers, 2013; Le Gall, 2014).

Chapter 2

SOME PROPERTIES OF CONVEX GEOMETRY AND CONVEX OPTIMIZATION

In this chapter we give an overview of properties from convex geometry and convex optimization that we make repeated use of in Chapters 4, 5, and 6.

2.1 Convex Sets and Convex Functions

We enumerate some standard definitions, properties, and theorems regarding convex sets and convex functions in $\mathbb{R}^n$ (see, for example, (Boyd and Vandenberghe, 2004) and (Bubeck, 2015)). We additionally discuss means by which we may specify convex sets and functions.

Definition 2.1.1 (Convex Set and Convex Function)**.** *A set $S \subset \mathbb{R}^n$ is said to be convex if given any $x, y \in S$ and any $\alpha \in [0, 1]$, we have*

$$\alpha x + (1 - \alpha)y \in S.$$

In other words, the set S contains all of its line segments. A function $f : S \to \mathbb{R}$ is said to be convex if S is convex and given any $x, y \in S$ and any $\alpha \in [0, 1]$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

In other words, the function f between any two points lies below the chord connecting the function's values at those points.

We will also require the more specific notion of a convex body.

Definition 2.1.2 (Convex Body)**.** *A convex body $\mathcal{K} \subset \mathbb{R}^n$ is a closed and bounded convex set with non-empty interior.*

A basic property of convex sets that we require is the separation theorem.

Theorem 2.1.3 (Separation Theorem)**.** *Let $S \subset \mathbb{R}^n$ be a closed convex set, and assume that $x_0 \in \mathbb{R}^n \setminus S$. There exists $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$a^\top x_0 > b$$

and for all $x \in S$,

$$a^\top x \leq b.$$

The hyperplane $\left\{x \in \mathbb{R}^n \mid a^\top x = b\right\}$ is said to "separate" x_0 from S.

This result implies the supporting hyperplane theorem.

Theorem 2.1.4 (Supporting Hyperplane Theorem)**.** *Let $S \subset \mathbb{R}^n$ be a non-empty convex set, and let $x_0 \in \partial S$ (i.e., the boundary of S). There exists $a \in \mathbb{R}^n \setminus \{0\}$ such that*

$$a^\top (x - x_0) \leq 0$$

for all $x \in S$. The set $\left\{x \in \mathbb{R}^n \mid a^\top x = a^\top x_0\right\}$ is said to be a "supporting hyperplane" for S.

One basic theorem we require is the first-order condition for convexity (Boyd and Vandenberghe, 2004).

Theorem 2.1.5 (First-Order Condition)**.** *Suppose $f : S \to \mathbb{R}^n$ is differentiable. Then f is convex if and only if S is convex and for all $y \in S$ and all $x \in \operatorname{int} S$ we have*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

The affine function of y given by $f(x) + \nabla f(x)^\top (y - x)$ is a first-order Taylor approximation of f near x. For convex functions, the first-order approximation is a global underestimator of the function. We may generalize this condition for non-differentiable functions as follows.

Definition 2.1.6 (Subgradient and Subdifferential)**.** *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$, where f is not necessarily convex. The vector $g \in \mathbb{R}^n$ is a "subgradient" of f at x if for all $y \in S$,*

$$f(y) \geq f(x) + g^\top (y - x).$$

The set of all subgradients of f at x is called the "subdifferential" of f at x, and is denoted $\partial f(x)$.

One may show (see, e.g., (Bubeck, 2015) proposition 1.1) that given f is convex, for any $x \in \operatorname{int} S$, the subdifferential $\partial f(x) \neq \emptyset$. If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. Additionally, consider minimizing f on S. If we are given a test point x and a subgradient g_x at x, then any $y \in S$ such that $g_x^\top (y - x) > 0$ implies that $f(y) > f(x)$ by the first-order condition. Thus the subgraident g_x provides a means of localizing a minimizer of f by stating that it must lie in the half-space $\left\{y \in S \mid g_x^\top (y - x) \leq 0\right\}$.

Now we define a few convex sets which frequently occur in this dissertation.

Definition 2.1.7 (Polyhedron and Polytope)**.** *A "polyhedron" P is an intersection of half-spaces,*

$$\begin{aligned} P &= \bigcap_{i=1}^{m} \left\{x \in \mathbb{R}^n \mid a_i^\top x \leq b_i\right\} \\ &= \left\{x \in \mathbb{R}^n \mid a_i^\top x \leq b_i,\, i = 1, \ldots, m\right\}. \end{aligned}$$

If P is closed and bounded, and then we refer to P as a polytope.

Half-spaces are clearly convex, and it is trivial to verify that convexity is preserved under intersections. Thus polyhedrons are convex. If a polytope is full-dimensional, we use the notation $\mathcal{P}$ in the sequel.

Definition 2.1.8 ((Convex) Cone)**.** *A set $C \subset \mathbb{R}^n$ is called a cone if for every $x \in C$ and every $\alpha \geq 0$, we also have $\alpha x \in C$. A set C is a convex if for any $\alpha_1, \alpha_2 \geq 0$ and any $x_1, x_2 \in C$, we have $\alpha_1 x_1 + \alpha_2 x_2 \in C$.*

The space of positive semidefinite matrices form a closed convex cone, which we denote S_+^n. Its interior is an open convex cone, the set of positive definite matrices, S_{++}^n. Finally, we define ellipsoids.

Definition 2.1.9. *Let S_{++}^n denote the cone of positive definite matrices, and let $E \in S_{++}^n$. An "ellipsoid" $\mathcal{E} \subset \mathbb{R}^n$ is a convex body of the form*

$$\mathcal{E} = \left\{x \in \mathbb{R}^n \mid (x - x_c)^\top E^{-1} (x - x_c) \leq r^2\right\},$$

where x_c denotes its center, and r its radius. Equivalently,

$$\mathcal{E} = \left\{ x_c + E^{1/2} y \mid |y| \leq r \right\}.$$

We employ ellipsoids which adapt to the local geometry of a convex set $\mathcal{K} \subset \mathbb{R}^n$ in Chapters 4 and 5.

Given a convex body $\mathcal{K} \subset \mathbb{R}^n$, there are three types of oracles which characterize the body which are mentioned in this dissertation. The simplest oracle is a *membership oracle*, which given a test point $y \in \mathbb{R}^n$, indicates whether $y \in \mathcal{K}$. A slightly more complicated oracle is a *separation oracle*, which either certifies that $y \in \mathcal{K}$ or provides a hyperplane which separates y from $\mathcal{K}$ as in theorem 2.1.3. A third way to specify the body is to use a *barrier* function. We will be interested in *self-concordant barriers*, which we describe next.

2.1.1 Self-Concordant Barriers

Informally, a barrier is a continuous, convex function defined on the interior of a convex set which tends to infinity as one approaches the boundary of that set. Additionally, self-concordance is a regularity condition in that the second derivative at a point along any direction is large compared to its first and third derivatives along the same direction. A formal definition from (Nesterov and Nemirovsky, 1994) follows.

Definition 2.1.10 (Regular self-concordant barrier)**.** *Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex set. For any function $F : \operatorname{int} \mathcal{K} \to \mathbb{R}$ having continuous derivatives of order k, for vectors $h_1, \ldots, h_k \in \mathbb{R}^n$, we recursively define for $k \geq 1$*

$$D^k F(x)[h_1, \ldots, h_k] :=$$
$$\lim_{\epsilon \to 0} \frac{D^{k-1} F(x + \epsilon h_k)[h_1, \ldots, h_{k-1}] - D^{k-1} F(x)[h_1, \ldots, h_{k-1}]}{\epsilon},$$

and we define $D^k F(x) := F(x)$ for $k = 0$, where we have assumed that $x \in \operatorname{int}(\mathcal{K})$. For $x \notin \operatorname{int}(\mathcal{K})$, we let $F(x) = \infty$. We call such an F a regular self-concordant barrier if it satisfies the following conditions:

1. *(Convex, Smooth): F is convex and thrice continuously differentiable on* $\text{int}(\mathcal{K})$*.*

2. *(Barrier): For every sequence of points* $\{x_k\} \in \text{int}(\mathcal{K})$ *converging to a point* $x \notin \text{int}(\mathcal{K})$*, we have* $\lim_{k\to\infty} F(x_k) = \infty$*.*

3. *(Differential Inequalities) For all* $h \in \mathbb{R}^n$ *and all* $x \in \text{int}(\mathcal{K})$*, the following inequalities hold:*

   (a) $F(x)$ *is "self-concordant," i.e.,*

   $$D^3F(x)[h,h,h] \le 2(D^2F(x)[h,h])^{\frac{3}{2}}.$$

   (b) For some $\nu \in \mathbb{R}$*,*

   $$|DF(x)[h]|^2 \le \nu D^2F(x)[h,h].$$

   The smallest such ν *is the "self-concordance parameter" of the barrier.*

Self-concordant barriers are general in the sense that for every convex body $\mathcal{K} \subset \mathbb{R}^n$, there exists a self-concordant barrier with self-concordance parameter $\nu = (1 + o(1))n$ (Nesterov and Nemirovsky, 1994; Bubeck and Eldan, 2014), though this barrier may not be easy to compute. Additionally, if F is a self-concordant barrier for $\mathcal{K}$ and A is a non-singular affine transformation, then $F_A(x) := F(A^{-1}(x))$ is a self-concordant barrier for $A(\mathcal{K})$. We shall refer to a general self-concordant barrier as F_s in the sequel. We now describe some concrete examples of self-concordant barriers given in (Narayanan, 2016).

Logarithmic barrier of a polytope

Given a polytope $\mathcal{P} = \left\{x \in \mathbb{R}^n \mid a_i^\top x \le b_i,\ i = 1, \ldots, m\right\}$, the logarithmic barrier is the real-valued function

$$F_\ell(x) = -\sum_{i=1}^{m} \log(b_i - a_i^\top x). \tag{2.1}$$

The self-concordance parameter associated with F_ℓ is $\nu_\ell = m$.

Hyperbolic Barriers

The vector space of polynomials has connections to convex analysis (Güler, 1997). We summarize the results here and an example. A polynomial $p : \mathbb{C}^n \to \mathbb{C}$ is said to be *homogenous* of degree q if

$$p(tx) = t^q p(x)$$

for all $x \in \mathbb{C}^n$ and all $t \in \mathbb{C}$. We may define hyperbolic homogenous polynomials as follows.

Definition 2.1.11 (Hyperbolic polynomial)**.** *A homogeneous polynomial p of degree q is said to be hyperbolic with respect to a direction $d \in \mathbb{R}^n$ if and only if $p(d) \neq 0$ and there exists a constant $t_0 \in \mathbb{R}$ such that $p(x + itd) \neq 0$ if $x \in \mathbb{R}^n$ and $t < t_0$.*

Associated with such polynomials is a cone of hyperbolicity.

Definition 2.1.12 (Hyperbolic cone)**.** *Let p be a homogeneous hyperbolic polynomial of degree q with respect to direction d. The set*

$$\left\{x \in \mathbb{R}^n \;\middle|\; t \mapsto p(x + td) \text{ has all negative roots}\right\}$$

is called the cone of hyperbolicity of p.

It may be shown (Güler, 1997) that such cones are convex. For each polynomial (and the convex cone it defines) we may construct a barrier function as follows:

$$F_h(x) = -\log p(x).$$

For concrete examples, consider the polynomial $p(X) = \det X$, where $X \in S^m$. Its hyperbolicity cone is the cone of $m \times m$ symmetric positive definite matrices. In this case, $F_h(x) = -\log \det x$ is a hyperbolic barrier with self-concordance parameter $\nu_h = m$. Additionally, let $A_1, \ldots, A_k$ be non-singular affine transformations, and consider

$$\mathcal{K} = \bigcap_{i=1}^{k} A_i(\mathcal{B}^n),$$

where $\mathcal{B}^n$ is the unit Euclidean ball in $\mathbb{R}^n$. $\mathcal{K}$ is an intersection of k ellipsoids, and thus is convex. The function

$$F_h(x) = -\sum_{i=1}^{k} \log(1 - \|A_i(x)\|^2)$$

is a hyperbolic barrier for $\mathcal{K}$ with self-concordance parameter $\nu_h = k$.

2.2 Dikin Ellipsoids

Suppose a convex body $\mathcal{K} \subset \mathbb{R}^n$ is equipped with a self-concordant barrier F. Dikin ellipsoids are ellipsoids which adapt to the local geometry of $\mathcal{K}$ via the Hessian of F at x. Dikin ellipsoids are defined as follows.

Definition 2.2.1 (Dikin ellipsoid)**.** *Given any point $x \in \mathcal{K}$ and a self-concordant barrier F for $\mathcal{K}$, the Dikin ellipsoid of radius r is*

$$\mathcal{D}_x(r) := \left\{ y \in \mathbb{R}^n \;\middle|\; D^2F(x)[x-y, x-y] \leq r^2 \right\}. \tag{2.2}$$

If $r = 1$, we let $\mathcal{D}_x := \mathcal{D}_x(r)$.

We now enumerate some properties of Dikin ellipsoids. Firstly, the Dikin ellipsoid locally defines the norm

$$\|h\|_x^2 = D^2F(x)[h,h], \tag{2.3}$$

where $x \in \operatorname{int}\mathcal{K}$ and $h \in \mathbb{R}^n$. Secondly, Nesterov and Nemirovsky (1994) showed that

$$\mathcal{D}_x \subset \mathcal{K}$$

for any $x \in \operatorname{int}\mathcal{K}$. Thirdly, Dikin ellipsoids are affine-invariant in that $\mathcal{D}_x(r)$ is the Dikin ellipsoid of radius r about a point $x \in \mathcal{K}$ and $T(\cdot)$ is a non-singular affine transformation, then the Dikin ellipsoid of radius r centered at the point $T(x)$ is $T(\mathcal{D}_x(r))$, provided the new barrier that is used is $H(y) := F(T^{-1}(y))$ (Narayanan, 2016).

2.2.1 The Dikin Ellipsoid for a Polytope

In chapter 5, we use some properties of the Dikin ellipsoid with regards to the logarithmic barrier defined in equation (2.1). We illustrate some of properties of the Dikin ellipsoid for this special case here, noting that they hold more generally, as shown in Chapters 2 and 3 of (Nesterov and Nemirovsky, 1994). Our illustration does not make explicit use of the theory of self-concordance, and is similar to that of Vishnoi (2014) with some modifications.

Given a polytope $\mathcal{P}$ as specified in definition 2.1.7, define the *slacks* of each inequality defining $\mathcal{P}$ to be

$$s_i(x) := b_i - a_i^\top x.$$

The logarithmic barrier thus can be written as

$$F(x) = \sum_{i=1}^{m} \log(s_i(x)).$$

The gradient of this barrier is

$$\nabla F(x) = \sum_{i=1}^{m} \frac{a_i}{s_i(x)},$$

and the Hessian is

$$H(x) := \nabla^2 F(x) = \sum_{i=1}^{m} \frac{a_i a_i^\top}{s_i(x)^2}.$$

Letting $S_x := \operatorname{diag}(s_1(x), \ldots, s_m(x))$, we may equivalently write $H(x) = A^\top S_x^{-2} A$.

We first show that the Dikin ellipsoid of radius 1 is contained in $\mathcal{P}$.

Proposition 2.2.2. *For every $x \in \operatorname{int} \mathcal{P}$, we have $\mathcal{D}_x \subset \mathcal{P}$.*

Proof. Let $y \in \mathcal{D}_x$. Then $(y-x)^\top H(x)(y-x) \leq 1$. Equivalently,

$$\sum_{i=1}^{m} \frac{|a_i^\top (y-x)|^2}{|s_i(x)|^2} \leq 1,$$

so $|a_i^\top (y-x)|^2 / s_i(x)^2 \leq 1$ for all $i \in \{1, \ldots, m\}$. Equivalently, for each i, we have

$$\left(\frac{s_i(x) - s_i(y)}{s_i(x)} \right)^2 \leq 1.$$

This implies that $0 \leq s_i(y)/s_i(x) \leq 2$ for each i where $s_i(x) > 0$ since $x \in \operatorname{int} \mathcal{P}$. Thus $s_i(y) \geq 0$, and $y \in \mathcal{P}$. $\square$

The *analytic center* of a polytope is the interior point which minimizes the logarithmic barrier. Denote this point x_{ac}, and note the following result, which states that the Dikin ellipsoid centered at x_{ac} inflated by a factor of m contains the polytope.

Proposition 2.2.3. *Given the analytic center of $\mathcal{P}$ is x_{ac}, we have*

$$\mathcal{P} \subset \mathcal{D}_{x_{ac}}(m).$$

Proof. Given any $x \in \mathcal{P}$, we seek to show that $(x - x_{ac})^\top H(x_{ac})(x - x_{ac}) \leq m^2$. We have

$$\begin{aligned}
(x - x_{ac})^\top H(x_{ac})(x - x_{ac}) &= (x - x_{ac})^\top \left(\sum_{i=1}^m \frac{a_i a_i^\top}{s_i(x_{ac})^2} \right) (x - x_{ac}) \\
&= \sum_{i=1}^m \frac{|a_i^\top (x - x_{ac})|^2}{|s_i(x_{ac})|^2} \\
&= \sum_{i=1}^m \frac{|s_i(x_{ac}) - s_i(x)|^2}{|s_i(x_{ac})|^2} \\
&= m - 2 \sum_{i=1}^m \frac{s_i(x)}{s_i(x_{ac})} + \sum_{i=1}^m \frac{|s_i(x)|^2}{|s_i(x_{ac})|^2}.
\end{aligned}$$

To analyze this sum, let $f(x) = \sum_{i=1}^m \frac{s_i(x)}{s_i(x_{ac})}$. Clearly $f(x_{ac}) = m$. The gradient of f as a function of x is

$$\nabla f(x) = \frac{a_i}{s_i(x_{ac})} = \nabla F(x_{ac}),$$

and x_{ac} minimizes $F(x)$, so $\nabla f(x_{ac}) = 0$. Thus $f(x) = m$ for all $x \in \mathcal{P}$. Consequently,

$$\begin{aligned}
(x - x_{ac})^\top H(x_{ac})(x - x_{ac}) &= -m + \sum_{i=1}^m \left(\frac{s_i(x)}{s_i(x_{ac})} \right)^2 \\
&\leq -m + \left(\sum_{i=1}^m \frac{s_i(x)}{s_i(x_{ac})} \right)^2 \\
&\leq -m + m^2 \\
&\leq m^2,
\end{aligned}$$

where the first inequality followed since each term in the sum is non-negative. Thus $x \in \mathcal{D}_{x_{ac}}(m)$. $\square$

Finally, we note that m may be replaced with $\sqrt{m}$ in the previous proposition in the case of a symmetric polytope.

Proposition 2.2.4. *Given the analytic center of a symmetric polytope $\mathcal{P}$ is x_{ac}, we have*

$$\mathcal{P} \subset \mathcal{D}_{x_{ac}}(\sqrt{m}).$$

Proof. We may shift $\mathcal{P}$ such that $x_{ac} = 0$. Since $0 \in \operatorname{int} \mathcal{P}$, we must have $b_i \neq 0$ for each i. Thus we may rescale each a_i by dividing by b_i. Thus no generality is lost in assuming that the polytope is defined by

$$\mathcal{P} = \left\{x \mid a_i^\top x \leq 1,\ i = 1, \ldots, m\right\},$$

and by symmetry for each inequality $a_i^\top x \leq 1$, there is a corresponding inequality $a_j^\top x \geq -1$. The Dikin ellipsoid centered at $x_{ac} = 0$ is

$$\mathcal{D}_0 = \left\{x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{m} |a_i^\top x_i|^2 \leq m\right\}.$$

Given any $x \in \partial\mathcal{D}_0$, we have $\sum_{i=1}^{m} |a_i^\top x_i|^2 = m$, so there must exist at least one i such that $|a_i^\top x|^2 \geq 1$. Thus either $a_i^\top x \geq 1$ or $a_i^\top x \leq -1$, whence $x \notin \operatorname{int} \mathcal{P}$. $\square$

To summarize, Dikin ellipsoids thus provide a good approximation of a polytope $\mathcal{P}$ in the sense that

$$\mathcal{D}_{x_{ac}} \subset \mathcal{P} \subset \mathcal{D}_{x_{ac}}(m)$$

where we may instead use the radius $\sqrt{m}$ if $\mathcal{P}$ is symmetric. In general $\mathcal{P}$ must be specified by at least $n+1$ inequality constraints. The ellipsoid defined in the next section gives us a better approximation with n rather than m.

2.3 John's Maximal Volume Ellipsoid Theorem

For a convex body $\mathcal{K} \subset \mathbb{R}^n$, Fritz John considered finding the ellipsoid of maximal volume $\mathcal{E}$ such that $\mathcal{E} \subset \mathcal{K}$. John (1948) showed that each convex body contains a unique maximum volume ellipsoid $\mathcal{E}$, and characterized how to find it. John proved the following theorem.

Theorem 2.3.1 (John's Maximal Volume Ellipsoid Theorem)**.** *Each convex body $\mathcal{K} \subset \mathbb{R}^n$ contains a unique ellipsoid of maximal volume. The ellipsoid is $\mathcal{B}$ if and only if the following conditions are satisfied: $\mathcal{B} \subset \mathcal{K}$, and for some $m \geq n$ there are Euclidean unit vectors $\{u_i\}_{i=1}^m$ on the boundary of $\mathcal{K}$ and positive constants $\{c_i\}_{i=1}^m$ satisfying,*

$$\sum_{i=1}^{m} c_i u_i = 0, \tag{2.4}$$

$$\sum_{i=1}^{m} c_i u_i u_i^\top = I_n, \tag{2.5}$$

where I_n denotes the identity matrix in $\mathbb{R}^{n\times n}$.

Note that condition (2.5) is sometimes written equivalently as

$$\langle x,\, y\rangle = \sum_{i=1}^{m} c_i \langle u_i,\, x\rangle \langle u_i,\, y\rangle$$

for all $x, y \in \mathbb{R}^n$. Using the cyclic invariance of the trace and that the $\{u_i\}$ are unit vectors, condition (2.5) implies that

$$\sum_{i=1}^{n} c_i = n, \tag{2.6}$$

a property we employ in the analysis of the random walk we develop in chapter 5.

We now enumerate some properties from Ball (1992) which provide additional insight into the geometric properties of John's ellipsoids. Note that condition (2.4) implies that all the contact points do not lie in one in one half-space of the unit ball, and this condition is redundant in the symmetric case, since for every contact point u_i, its reflection about the origin $-u_i$ is also a contact point. Condition (2.5) guarantees such contact points do not lie close to a proper subspace. Furthermore, it may be shown that there are at most $n(n+3)/2$ contact points for general $\mathcal{K}$, and $n(n+1)/2$ non-redundant contact points if $\mathcal{K}$ is origin-symmetric (Gruber, 1988). At each u_i, the supporting hyperplane to $\mathcal{K}$ is unique and orthogonal to u_i, since this is the case for the unit ball. Thus considering the polytope resulting from such supporting hyperplanes, $\mathcal{P} = \left\{x \in \mathbb{R}^n \;\middle|\; \langle u_i,\, x\rangle \leq 1,\, i = 1, \ldots, m\right\}$, the convex set $\mathcal{K}$ obeys the sandwiching $\mathcal{B} \subset \mathcal{K} \subset \mathcal{P}$. By Cauchy-Schwarz, for any $x \in \mathcal{P}$, we have

$$-|x| \leq \langle u_i, x \rangle \leq 1.$$

Since the weights $\{c_i\}$ are positive, it follows by employing conditions (2.4), (2.5), and (2.6) that

$$\begin{aligned} 0 &\leq \sum_i c_i(1 - \langle u_i, x \rangle)(|x| + \langle u_i, x \rangle) \\ &= |x| \sum_i c_i + (1 - |x|) \left\langle \sum_i c_i u_i, x \right\rangle - \sum_i c_i \langle u_i, x \rangle^2 \\ &= n|x| - |x|^2, \end{aligned}$$

from which it follows that $|x| \leq n$. If the convex body is origin-symmetric, then by substituting $-u_i$ for u_i, for any $x \in \mathcal{P}$, we have

$$|\langle u_i, x \rangle| \leq 1.$$

It follows that

$$|x|^2 = \sum_{i=1}^m c_i \langle u_i, x \rangle^2 \leq \sum_{i=1}^m c_i = n,$$

so $|x| \leq \sqrt{n}$. The following corollary of John's theorem results.

Corollary 2.3.2. *After an appropriate affine transformation $A(\cdot)$, any convex body $\mathcal{K} \subset \mathbb{R}^n$ satisfies*

$$\mathcal{B} \subset A(\mathcal{K}) \subset n\mathcal{B}.$$

If $A(\mathcal{K})$ is origin-symmetric, then the containment is

$$\mathcal{B} \subset A(\mathcal{K}) \subset \sqrt{n}\mathcal{B}. \tag{2.7}$$

The containments are tight, as indicated in by taking $\mathcal{K}$ to be the unit cube in the symmetric case, and to be the standard regular simplex in the non-symmetric case (Ball, 1997).

For completeness with regards to the Markov chain we develop in chapter 5, we provide a proof of John's theorem for a general convex body similar to the proof of Ball (1997). We refer the reader to Appendix A for the proof.

2.4 Cutting Plane Methods

In this section we review cutting plane methods for finding a point in a convex body, and also discuss extending the feasibility method to minimization.

2.4.1 The Feasibility Problem

The feasibility problem in cutting-plane methods are as follows: given a convex body $\mathcal{K} \subset \mathbb{R}^n$, find a point $x \in \mathcal{K}$. To this end, cutting-plane methods assume that one has access to a separation oracle for $\mathcal{K}$. I.e., given any $y \in \mathcal{K}$ and a test point x, the separation oracle gives a vector w_x such that

$$(y - x)^\top w_x \leq 0.$$

Thus we are performing localization of $\mathcal{K}$ in the sense that we are certifying that $\mathcal{K}$ is contained in a half-space,

$$\mathcal{K} \subset \left\{ y \in \mathbb{R}^n \mid (y - x)^\top w_x \leq 0 \right\}.$$

Generally, a cutting-plane scheme to find such a point $x \in \mathcal{K}$ is given in algorithm 2.4.1.

Given $\mathcal{S}_0 \supset \mathcal{K}$, a point $x_0 \in \text{int}(\mathcal{S}_0)$, and the number of iterations T. For $t \geq 0$:

1. Call the separation oracle with x_t.

   (a) If $t < T$ and $x_t \in \mathcal{K}$, then terminate the loop, returning the feasible point $x_t \in \mathcal{K}$.

   (b) Otherwise, use the vector $w_t := w_{x_t}$ returned from the separation oracle to localize $\mathcal{K}$, letting

   $$\mathcal{S}'_{t+1} = \left\{ y \in \mathbb{R}^n \mid (y - x_t)^\top w_t \leq 0 \right\} \cap \mathcal{S}_t,$$

   so $\mathcal{K} \subset \mathcal{S}'_{t+1}$. Embed the intermediate localizer $\mathcal{S}'_{t+1}$ into a convex body $\mathcal{S}_{t+1}$, and find a point $x_{t+1} \in \text{int}(\mathcal{S}_{t+1})$. Increment t and loop.

   (c) If $t = T$ and $x_t \notin \mathcal{K}$, terminate the loop with no point found, declaring that $\lambda_n(\mathcal{K}) \leq \lambda_n(\mathcal{S}_T)$, where $\lambda_n(\cdot)$ denotes the n-dimensional volume.

Algorithm 2.4.1: Cutting Plane Method

There are many different methods which follow this general technique. For example, consider the following methods.

- *Ellipsoid Method*: Choose $\mathcal{S}_0$ to be an ellipsoid containing $\mathcal{K}$ centered at some known point x_0 for some known positive definite matrix Σ_0:

  $$\mathcal{S}_0 = \left\{ y \in \mathbb{R}^n \mid (y - x_0)^\top \Sigma_0^{-1} (y - x_0) \leq 1 \right\}.$$

  $\mathcal{S}'_{t+1}$ is chosen to be the half-ellipsoid which contains $\mathcal{K}$,

  $$\mathcal{S}'_{t+1} = \mathcal{S}_t \cap \left\{ y \in \mathbb{R}^n \mid w_t^\top (y - x_t) \leq 0 \right\}.$$

  $\mathcal{S}'_{t+1}$ is chosen to be the unique minimum volume volume ellipsoid containing $\mathcal{S}'_{t+1}$, given as (see, e.g., Bubeck (2015))

  $$\mathcal{S}_{t+1} = \left\{ y \in \mathbb{R}^n \mid (y - x_{t+1})^\top \Sigma_{t+1}^{-1} (y - x_{t+1}) \leq 1 \right\},$$

where

$$x_{t+1} = x_t - \frac{1}{n+1} \frac{\Sigma_t w_t}{\sqrt{w_t^\top \Sigma_t w_t}},$$

$$\Sigma_{t+1} = \frac{n^2}{n^2-1} \left(\Sigma_t - \frac{2}{n+1} \Sigma_t \tilde{w}_t \tilde{w}_t^\top \Sigma_t \right).$$

Note that it may be shown that $\lambda_n(\mathcal{S}_{t+1}) \leq \exp\left(-\frac{1}{2n}\right) \lambda_n(\mathcal{S}_t)$.

- *Center of Gravity Method*: Compute the center of gravity of $\mathcal{S}_t$,

$$x_t = \frac{1}{\lambda_n(\mathcal{S}_t)} \int_{x \in \mathcal{S}_t} x \, dx,$$

and query the separation oracle to get w_t. Let

$$\mathcal{S}_{t+1} = \mathcal{S}'_{t+1} = \mathcal{S}_t \cap \left\{ y \in \mathbb{R}^n \mid (y - x_t)^\top w_t \leq 0 \right\}.$$

Note that finding the center of gravity is a challenging problem unto itself which we revisit in chapter 4, so this is mainly used as a proof of concept. Additionally, Grunbaum's theorem (see, e.g., Bubeck (2015)) implies that $\lambda_n(\mathcal{S}_{t+1}) \leq \left(1 - \frac{1}{e}\right) \lambda_n(\mathcal{S}_t)$.

2.4.2 From Feasibility to Minimization

To modify the cutting-plane method to perform minimization, the steps are essentially the same, but with one modification. The minimization problem is essentially a feasibility problem in which we instead seek a point $\hat{x}$ in the set $\mathcal{K} \cap \{x \mid f(x) - f(x^\star) \leq \gamma\}$, where $\gamma > 0$ is an error tolerance and $x^\star$ is any minimizer of f on $\mathcal{K}$. If we find a point $y \in \mathcal{K}$, we instead use the oracle specified by any subgradient $w_y \in \partial f(y)$ to localize an optimal solution. If $0 \in \partial f(y)$, then y is an optimal point, and we are done. Otherwise, we use the hyperplane $\left\{x \mid w_y^\top (x - y) \leq 0\right\}$ within which the set $\{x \mid f(x) \leq f(y)\}$ is contained, and proceed. If an optimal $x^\star$ was not found in T iterations, we find an approximate solution as follows. Letting $T' \subset \{0, 1, 2, \ldots, T\}$ denote the steps for which an $x_t \in \mathcal{K}$ was found, after T iterations we return

$$\hat{x}_T \in \operatorname*{argmin}_{x_s,\, s \in T'} f(x_s). \tag{2.8}$$

Note that $f(x^\star)$ is not known, so we cannot directly evaluate whether any approximate solution $\hat{x}_T \in \mathcal{K}$ satisfies the error tolerance. However, given information on the geometry of $\mathcal{K}$ and on the objective function f we may choose T to guarantee that this is the case. To that end, we follow the general discussion of cutting plane methods of Nemirovski (1995). Let $x \in \mathcal{K}$, and given $\epsilon \in (0,1)$, let

$$\begin{aligned}\mathcal{K}_\epsilon(x) &:= x + \epsilon(\mathcal{K} - x)\\ &= \{(1-\epsilon)x + \epsilon z \mid z \in \mathcal{K}\}.\end{aligned}$$

Now let $f_\epsilon^\star := \inf_{x\in\mathcal{K}} f_\epsilon(x)$ where $f_\epsilon(x) := \sup_{y\in\mathcal{K}_\epsilon(x)} f(y)$. Nemirovski (1995) defines an ϵ-solution to be any $x \in \mathcal{K}$ such that

$$f(x) \le f_\epsilon^\star,$$

and provides the following theorem.

Theorem 2.4.1. *Assume that after T steps the method has not terminated with an optimal solution. Then given that $T' \neq \emptyset$, any solution $\hat{x}_T$ of equation (2.8) is an ϵ-solution for any $\epsilon \in (0,1)$ such that*

$$\epsilon^n > \frac{\lambda_n(\mathcal{S}_T)}{\lambda_n(\mathcal{K})}. \tag{2.9}$$

If the function f is convex and continuous on K, then any ϵ-solution x satisfies

$$f(x) - f(x^\star) \le \epsilon\left(\sup_{x\in\mathcal{K}} f(x) - f(x^\star)\right) \tag{2.10}$$

for any optimal $x^\star \in \mathcal{K}$. If the function f is Lipschitz continuous on $\mathcal{K}$ with respect to some norm $|\cdot|$ with Lipschitz constant L_f, then any ϵ-solution x satisfies

$$f(x) - f(x^\star) \le \mathrm{diam}(\mathcal{K}) L_f \epsilon \tag{2.11}$$

for any optimal $x^\star \in \mathcal{K}$, where $\mathrm{diam}(\mathcal{K})$ *is the diameter with respect to the same norm.*

Proof. Given $x \in \mathcal{K}$ and $\epsilon \in (0,1)$, let $y \in \mathcal{K}_\epsilon(x)$. Then $y = (1-\epsilon)x + \epsilon z$ for some $z \in \mathcal{K}$. By the continuity and convexity of f, we have

$$\begin{aligned}\sup_{y\in\mathcal{K}_\epsilon(x)} f(y) &= \sup_{z\in\mathcal{K}} f((1-\epsilon)x+\epsilon z) \\ &\leq f(x) + \epsilon\left(\sup_{z\in\mathcal{K}} f(z) - f(x)\right).\end{aligned}$$

Thus

$$f_\epsilon^\star - \inf_{x\in\mathcal{K}} f(x) \leq \epsilon\left(\sup_{z\in\mathcal{K}} f(z) - \inf_{x\in\mathcal{K}} f(x)\right),$$

which proves (2.10). Furthermore if f is Lipschitz with respect to a norm $|\cdot|$, we have

$$\begin{aligned}f_\epsilon^\star - \inf_{x\in\mathcal{K}} f(x) &\leq \epsilon \sup_{z,x\in\mathcal{K}} \left(f(z) - f(x)\right) \\ &\leq \epsilon\, L_f \sup_{z,x\in\mathcal{K}} |z-x|,\end{aligned}$$

which proves (2.11). To prove (2.9), since $\lambda_n(\mathcal{S}_T) < \epsilon^n \lambda_n(\mathcal{K}) = \lambda_n(\mathcal{K}_\epsilon(x))$, by monotonicity we must have that $\mathcal{K}_\epsilon(x) \not\subset \mathcal{S}_T$. Thus there must exist a $y \in \mathcal{K}_\epsilon(x) \setminus \mathcal{S}_T$. It follows that for any $\hat{x}_T$ of (2.8) satisfies

$$f(\hat{x}_T) < f(y) \leq \sup_{y\in\mathcal{K}_\epsilon(x)} f(y) = f_\epsilon(x),$$

since any such y must have been removed by the subgradient oracle for some $t \in T'$. Since $x \in \mathcal{K}$ was arbitrary, it follows that $f(\hat{x}_T) \leq f_\epsilon^\star = \inf_{x\in\mathcal{K}} f_\epsilon(x)$, or that $\hat{x}_T$ is an ϵ-solution. □

Thus to satisfy a general error tolerance $\gamma > 0$, one need only to know that either f is bounded on $\mathcal{K}$ or to know the Lipschitz constant associated with f on $\mathcal{K}$ and the diameter of $\mathcal{K}$ with respect to some norm $|\cdot|$.

2.4.3 Vaidya's Algorithm

In Chapter 5 and Chapter 6, we make use of Vaidya's algorithm to minimize a convex function over a convex set. We state this algorithm here for the feasibility case, noting we may employ the modification of section 2.4.2 to perform minimization.

Vaidya's algorithm is a cutting-plane method which seeks a feasible point in an arbitrary convex set,

$$\mathcal{X} \subset \mathcal{S}_0 := \left\{ x \in \mathbb{R}^n \;\middle|\; \|x\|_\infty \leq \rho \right\},$$

where we note that the implementation of Anstreicher (1997) assumes $\rho = 1$. The set $\mathcal{X}$ is specified by a separation oracle: given a point $y \in \mathbb{R}^n$, the oracle either certifies that $y \in \mathcal{X}$, or returns a separating hyperplane between y and $\mathcal{X}$ (i.e., a vector w such that $\mathcal{X} \subset \left\{ x \;\middle|\; w \cdot (x - y) \leq 0 \right\}$). The algorithm initializes with a polytope $\mathcal{S}_0 \supset \mathcal{K}$ and an interior point x_0 of $\mathcal{S}_0$, and maintains a polytope $\mathcal{S}_t \supset \mathcal{X}$ and an interior point x_t of $\mathcal{S}_t$ at each iteration t, where $\mathcal{S}_t$ is defined via the separation oracle. At each iteration t, a constraint is either added or deleted depending on certain criteria, and Anstreicher (1997) shows that the polytope $\mathcal{S}_t$ is specified by no more than $200n$ constraints throughout the algorithm. After $T = O(n(L + \log \rho))$ calls to the separation oracle, we have

$$\lambda_n(\mathcal{S}_T) < \lambda_n(2^{-L}\mathcal{B}^n),$$

where $\lambda_n(\cdot)$ denotes the n-dimensional Lebesgue measure and $L = \Omega(\log n)$ is a user-specified constant. Thus, the algorithm certifies that if no feasible point is found within T iterations, the volume of $\mathcal{X}$ is less than that of an n-dimensional ball of radius 2^{-L}. We remark that the value of T in our case (where $\rho \neq 1$ in general) is easily determined via an argument along the lines of lemma 3.1 in (Anstreicher, 1997), and is given as

$$T \geq \frac{n \left[1.4L + 2\log n + 2\log(1 + 1/\epsilon) + 0.5 \log\left(\frac{1+\tau}{1-\epsilon}\right) + 2\log(\rho) - \log(2) \right]}{\Delta V}, \tag{2.12}$$

where $\epsilon = 0.005$ and $\tau = .007$ are parameters of the algorithm, and $\Delta V = 0.00037$. The algorithm uses a total of

$$O(n(L + \log \rho)\kappa + n^{\omega+1}(L + \log \rho)) \tag{2.13}$$

operations, where κ is the cost of evaluating the separation oracle.

Chapter 3

MARKOV CHAINS ON GENERAL STATE SPACES

This chapter provides an introduction to Markov chains on general state spaces which provides the foundation for proving that the Markov chains with transition measures defined using the ellipsoids introduced in Chapter 2 are rapidly mixing in terms of the dimension.

3.1 Definition

We first formally define a Markov chain for general state spaces and discuss some of their properties. To that end, consider defining a probability space. A convex body $\mathcal{K} \subset \mathbb{R}^n$ will define a state space, and let $\mathcal{A}$ be the σ-algebra generated by the set of all Lebesgue-measurable subsets of $\mathcal{K}$. Additionally, for any $u \in \mathcal{K}$ and any $A \in \mathcal{A}$, we define the one-step probability measure $P_u(A)$ as the probability of being in the set A after taking one step from u, or equivalently,

$$P_u(A) = \int_{v \in A} dP_u(v).$$

Finally, given an initial probability measure π_0 on $\mathcal{K}$, we let $\pi_0(A)$ represent the probability of starting in the set $A \in \mathcal{A}$. With this setup, we may now define a discrete-time, homogeneous *Markov chain* on general state spaces.

Definition 3.1.1 (Markov chain)**.** *A Markov chain is a sequence of points $X_0, X_1, X_2, \ldots$ such that $P(X_0 \in A) = \pi_0(A)$ and*

$$P(X_{t+1} \in B | X_t = x_t, \ldots, X_0 = x_0) = P(X_{t+1} \in B | X_t = x_t) = P_{x_t}(B)$$

for any $A, B \in \mathcal{A}$ and any $x_s \in \mathcal{K}$ for $s \in \{0, 1, \ldots, t\}$. More generally, a Markov chain is the triplet $\{\mathcal{K}, \mathcal{A}, P_x\}$, where the transition measure P_x is specified for all $x \in \mathrm{int}(\mathcal{K})$, along with an initial measure π_0.

Thus, for a Markov chain, the probability of being in any given set $A \in \mathcal{A}$ after one step depends only on starting position of the step, and not on any previous positions. Note that some authors (Lovász and Simonovits, 1993) use the terminology *Markov scheme* to refer to the case where the initial probability measure π_0 for the starting point X_0 is unspecified, and a Markov chain to refer to the triplet $\{\mathcal{K}, \mathcal{A}, P_x\}$ along with the initial measure. We avoid this distinction in the sequel, and refer to either case as a Markov chain depending on the context.

Before proceeding, it is illustrative to consider some common examples of Markov chains on general state spaces. Consider random walks on convex bodies $\mathcal{K} \subset \mathbb{R}^n$, where the state is the current location in the body. Here are two such walks which assume only a membership oracle, and their associated transition measures:

- *Ball Walk*: P_x is uniform on $\mathcal{B}^n(x, \delta)$, the n-dimensional Euclidean ball of radius δ centered at x. The ball walk picks a random point y from ball of radius δ centered at x. If $y \in \mathcal{K}$, the step is accepted. Otherwise, the step is rejected.

- *Hit and Run*: P_x is uniform on random chord through $\mathcal{K}$. To generate the next step given the current location $x \in \mathcal{K}$, hit and run picks a random direction which defines a line ℓ through $x \in \mathcal{K}$. The next step y is chosen uniformly on $\ell \cap \mathcal{K}$.

3.2 Stationarity, Reversibility, and Laziness

One natural question that arises is if a Markov chain reaches a steady state distribution, known as the *stationary* measure. We formally define this probability measure as follows.

Definition 3.2.1 (Stationary distribution)**.** *A distribution π is called stationary if for all $A \in \mathcal{A}$, we have*

$$\pi(A) = \int_{u \in \mathcal{K}} P_u(A) d\pi(u). \tag{3.1}$$

In other words, a probability measure π is stationary if the probability of being in any set $A \in \mathcal{A}$ is the same after one step. Note that this distribution is also referred to as the *invariant* measure given this property.

The notion of the stationary distribution is related to reversibility.

Definition 3.2.2 (Reversible)**.** *A Markov chain is reversible with respect to the distribution* π *if for all* $u, v \in \mathcal{K}$*, we have*

$$dP_u(v)d\pi(u) = dP_v(u)d\pi(v). \tag{3.2}$$

Given any $A, B \in \mathcal{A}$*, it follows that*

$$\int_{u \in A} P_u(B)d\pi(u) = \int_{v \in B} P_v(A)d\pi(v).$$

Equation (3.2) is sometimes referred to as *detailed balance.* Reversibility is a useful property because of the following fact.

Proposition 3.2.3 (Reversibility and Stationary Distribution)**.** *If a Markov chain is reversible with respect to a distribution* π*, then* π *is stationary for the chain.*

Proof. For any $A \in \mathcal{A}$, we have

$$\begin{aligned} \int_{u \in \mathcal{K}} P_u(A)d\pi(u) &= \int_{v \in A} P_v(\mathcal{K})d\pi(v) \\ &= \int_{v \in A} d\pi(v) \\ &= \pi(A), \end{aligned}$$

where the first equality follows from reversibility. Thus, $\pi(A)$ is stationary by definition 3.2.1. $\square$

The Markov chains we employ in chapters 4 and 5 will be *lazy* chains, which we define as follows.

Definition 3.2.4 (Laziness)**.** *A Markov chain with transition measure* P_u *may be made "lazy" if with probability 1/2, we do not make any transition from the current location* $u \in \mathcal{K}$*, and instead do nothing. This implies that the transition measure of a lazy chain is*

$$Q_u(A) = \frac{1}{2}P_u(A) + \frac{1}{2}I_A(u) \tag{3.3}$$

for any $A \in \mathcal{A}$*, where* $I_A(u)$ *is the indicator for the set* A*.*

It is useful to note a stationary measure is preserved under making the chain lazy.

Proposition 3.2.5. *Given a Markov chain with a stationary distribution π, the lazy version of the chain also has the stationary distribution π.*

Proof. Let Q_u denote the transition measure of the lazy version of the chain at a point $u \in \mathcal{K}$ as in equation (3.3). Then

$$\begin{aligned}\int_{u\in\mathcal{K}} Q_u(A)d\pi(u) &= \frac{1}{2}\int_{u\in\mathcal{K}} P_u(A)d\pi(u) + \frac{1}{2}\int_{u\in\mathcal{K}} I_A(u)d\pi(u)\\ &= \frac{1}{2}\pi(A) + \frac{1}{2}\pi(A)\\ &= \pi(A).\end{aligned}$$

□

3.3 The Metropolis Filter

The Metropolis filter (Roberts et al., 2004) is means by which we may guarantee that a Markov chain reaches the stationary measure is π. Given the current point x and the transition measure P_x, we seek to generate the next point in the chain. We then propose a point y sampled from P_x. Let $q = \frac{d\pi}{d\lambda}$ and $p_x = \frac{dP_x}{d\lambda}$ be Radon-Nikodym derivatives of π and P_x, respectively, with respect to some dominating measure λ. Similarly let $p_y = \frac{dP_y}{d\lambda}$. The Metropolis filter is as follows: we independently let the next point be y with probability

$$\alpha_x(y) = \min\left(1,\ \frac{q(y)p_y(x)}{q(x)p_x(y)}\right).$$

This produces a new Markov chain with transition density $\widetilde{p}_x(y) = p_x(y)\alpha_x(y)$ which is reversible with respect to π. For $y \neq x$, we have

$$\begin{aligned}d\pi(x)\, d\widetilde{P}_x(y) &= [q(x)\, d\lambda(x)]\, [p_x(y)\alpha_x(y)\, d\lambda(y)]\\ &= \min[q(x)p_x(y),\ q(y)p_y(x)]\, d\lambda(x)\, d\lambda(y)\end{aligned}$$

This expression is the same if we swap x and y, whence (3.2) holds.

3.4 Convergence and Conductance

The notion of stationarity is closely related to the notion of flow between sets in the state space. Another way to look at the one-step probability distribution is to consider the *ergodic flow*.

Definition 3.4.1 (Ergodic flow)**.** *The ergodic flow $\Phi(A)$ of a set $A \in \mathcal{A}$ is the probability of transitioning from A to $\mathcal{K} \setminus A$ under the stationary measure π, i.e.,*

$$\Phi(A) = \int_{u \in A} P_u(\mathcal{K} \setminus A) d\pi(u).$$

The following theorem formally links the notion of stationarity and ergodic flow.

Proposition 3.4.2. *A distribution π is stationary if and only if $\Phi(A) = \Phi(\mathcal{K} \setminus A)$.*

Proof. Considering the difference, we have

$$\begin{aligned}
\Phi(A) - \Phi(\mathcal{K} \setminus A) &= \int_{u \in A} P_u(\mathcal{K} \setminus A) d\pi(u) - \int_{u \notin A} P_u(A) d\pi(u) \\
&= \int_{u \in A} (1 - P_u(A))\, d\pi(u) - \int_{u \notin A} P_u(A) d\pi(u) \\
&= \pi(A) - \int_{u \in \mathcal{K}} P_u(A) d\pi(u).
\end{aligned}$$

The last quantity is the probability of staying in A after one step, and by definition 3.2.1 is equal to $\pi(A)$ if and only if π is stationary. $\square$

This proposition is stating the intuitive result that for stationarity, the ergodic flow out of any set $A \in \mathcal{A}$ should be equal to the flow into that set, which is consistent with the notion of reversibility.

Given the existence and uniqueness of a stationary distribution, a natural question that rises is how quickly such a chain converges to this distribution. Intuitively, a fastly converging chain should have high ergodic flow out of any set $A \in \mathcal{A}$ with respect to the stationary distribution. The definition of conductance formalizes this notion.

Definition 3.4.3 (Conductance)**.** *The conductance of a Markov chain is defined as*

$$\phi = \inf_{A \in \mathcal{A} \colon \pi(A) \leq \frac{1}{2}} \frac{\Phi(A)}{\pi(A)},$$

where π is the stationary distribution of the chain.

Note that the ergodic flow $\Phi(A)$ is the probability of transitioning from A to $\mathcal{K} \setminus A$ under the stationary distribution, and $\pi(A)$ is the probability of being in the set A under the stationary distribution. Thus conductance represents the worst-case scenario of the probability of transitioning out of a set A adjusted for the probability that the chain is in set A under the stationary distribution.

Fastly mixing Markov chains should intuitively avoid staying in any set A for too long (i.e., avoid a "bottleneck"), so a high value of ϕ informs us that our chain mixes rapidly. We use the approach of Lovász and Simonovits (1993) of lower-bounding the conductance of the chain to prove mixing times. Recall the total variation distance between two measures P_1, P_2 on a measurable space $(\mathcal{K}, \mathcal{A})$ is

$$d_{TV}(P_1, P_2) = \sup_{A \in \mathcal{A}} |P_1(A) - P_2(A)|.$$

Note that $|P_1(A) - P_2(A)| = |P_1(\mathcal{K} \setminus A) - P_2(\mathcal{K} \setminus A)|$, so if the supremum is attained on any $A \in \mathcal{A}$, then it is attained on $\mathcal{K} \setminus A \in \mathcal{A}$ as well. If P_1 and P_2 are both absolutely continuous with respect to a dominating measure μ and thus have densities $p_1 := \frac{dP_1}{d\mu}$ and $p_2 := \frac{dP_2}{d\mu}$, respectively, the total variation distance may also be written as

$$\begin{aligned} d_{TV}(P_1, P_2) &= \frac{1}{2} \int \big| p_1 - p_2 \big| d\mu \\ &= 1 - \int \min(p_1, p_2)\, d\mu \\ &= 1 - \int_{S_1} \left[\min\left(1, \frac{p_2}{p_1}\right) \right] p_1 \, d\mu \\ &= 1 - \mathbb{E}_{P_1} \left[\min\left(1, \frac{p_2}{p_1}\right) \right], \end{aligned} \tag{3.4}$$

where $S_1 = \big\{ x \;\big|\; p_1(x) > 0 \big\}$. Recall that (3.4) does not depend on the choice of dominating measure μ but rather that the densities are correctly specified with respect to the dominating measure. Additionally, note that the equality is attained on $\{x \mid p_1(x) \geq p_2(x)\}$ almost everywhere with respect to μ (or alternatively on its complement). The following relationship between conductance and the total variation distance to the stationary measure was proven in (Lovász and Simonovits, 1993).

Theorem 3.4.4 (Lovász and Simonovits). *Let π_0 be the initial distribution for a lazy, reversible Markov chain with conductance ϕ and stationary measure π, and let π_t denote the distribution after t steps. Let π_0 be an M-warm start for π, i.e., we have $M := \sup_{A \in \mathcal{A}} \frac{\pi_0(A)}{\pi(A)} = O(1)$. Then*

$$d_{TV}(\pi_t, \pi) \leq \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

As a consequence of this theorem, we have the following bound on the mixing time.

Corollary 3.4.5. *Given $\epsilon > 0$, $M := \sup_{A \in \mathcal{A}} \frac{\pi_0(A)}{\pi(A)} = O(1)$, and $t \geq \tau(\epsilon) := \lceil \frac{2}{\phi^2} \log(\sqrt{M}/\epsilon) \rceil$, we have $d_{TV}(\pi_t, \pi) \leq \epsilon$. Thus the Markov chain mixes in $\widetilde{O}(\phi^{-2})$ steps from a warm start.*

To find mixing times, it then suffices to lower-bound the conductance ϕ.

Occasionally, as in the case of the geometric walk known as the *ball walk* (in which steps are generated from uniformly distributed samples on the Euclidean ball of a certain radius), it is useful to define a local version of conductance, as follows.

Definition 3.4.6 (Local Conductance). *The local conductance of a point $u \in \mathcal{K}$ for a Markov chain is defined as*

$$\ell(u) = 1 - P_u(\{u\}),$$

i.e., the probability of leaving the point $u \in \mathcal{K}$.

We mention this definition here to note an advantage of the random walks we employ in Chapter 4 and develop in Chapter 5 as compared to the ball walk. Namely, points near corners of some convex bodies may be such that the probability of making a transition is quite low, and thus we must make some adjustments to our chain to avoid this problem of not transitioning. For example, in the hypercube case with the ball walk, the probability of staying within the body decays exponentially with dimension at a corner, and thus the local conductance at a corner is quite small. In a strict sense, then, random walks using the uniform distribution on a Euclidean ball as a transition measure are not rapidly mixing, since an exponential number of steps in terms of the dimension will be required to leave the corner. To circumvent this problem, the ball walk is often instead employed in the smoothed body $\widetilde{\mathcal{K}} = \mathcal{K} + \alpha\mathcal{B}^n$ (Vempala, 2005). Alternatively, in analyzing the ball walk, points with small local conductance are ignored (Vempala, 2005). For hit and run, the algorithm is known to mix fast from any interior point of $\mathcal{K}$ (Lovász and Vempala, 2006a), but the mixing time depends on the ratio R^2/r^2 where it assumed $r\mathcal{B}^n \subset \mathcal{K} \subset R\mathcal{B}^n$, as well as $\log(R/d)$ where d is the distance of the starting point from the boundary.

Affine-invariant random walks are random walks for which the mixing time is unaffected by affine transformations of $\mathcal{K}$. The affine-invariant random walks we employ in Chapter 4 and design in Chapter 5 do not suffer from poor local conductance, since the ellipsoids they use to make steps adapt to the local geometry, and no modifications to the body $\mathcal{K}$ or ignoring portions of the body in the analysis will need to be made. Additionally, the mixing times will not depend on the ratio R/r. However, it is unknown whether these random walks mix rapidly from any given interior point.

3.5 Conditional Expectation as a Bounded Linear Operator

In this section we discuss conditional expectation as a bounded linear operator on Hilbert spaces. Necessary background information from functional analysis is provided in Appendix B. Markov chains and their stationary measure $\pi(\cdot)$ define a Hilbert space $L^2(\pi)$ with inner product

$$\langle f,\, g\rangle_{L^2(\pi)} := \int_{u\in\mathcal{K}} f(u)g(u)\, d\pi(u),$$

which induces the norm

$$\|f\|_{L^2(\pi)} := \langle f,\, f\rangle_{L^2(\pi)}^{1/2} = (\mathbb{E}_\pi[|f|^2])^{1/2}.$$

More generally, Markov chains define a Banach space $L^q(\pi)$ with $1 \leq q < \infty$ with the norm

$$\|f\|_{L^q(\pi)} := (\mathbb{E}_\pi[|f|^q])^{1/q},$$

or in the $q = \infty$ case, with the norm

$$\begin{aligned}\|f\|_{L^\infty(\pi)} &= \operatorname*{ess\,sup}_{x \in \mathcal{K}} |f(x)| \\ &:= \inf \{c > 0 \mid \pi \{x \in \mathcal{K} \mid |f(x)| > c\} = 0\}.\end{aligned}$$

The transition measure P_u may be viewed as a linear operator P on $L^q(\pi)$. The operator acts on measures on the left, creating a measure $\mu\mathsf{P}$. That is, for every $A \in \mathcal{A}$, we have

$$\mu\mathsf{P}(A) := \int_{x \in \mathcal{K}} P_x(A)\, d\mu(x),$$

and condition (3.1) may equivalently be written as $\pi\mathsf{P} = \pi$. Additionally, this operator acts on functions on the right. Note that

$$(\mathsf{P}f)(u) := \mathbb{E}_{P_u}[f] = \int_{v \in \mathcal{K}} f(v)\, dP_u(v),$$

i.e., the expected value of $f(v_{t+1})$ given that $v_t = u$. Note that it is straightforward to show that this operator maps $L^q(\pi)$ to $L^q(\pi)$ for any q such that $1 \leq q < \infty$, since

$$\begin{aligned}|(\mathsf{P}f)(u)|^q &= \left| \int_{v \in \mathcal{K}} f(v)\, dP_u(v) \right|^q \\ &\leq \int_{v \in \mathcal{K}} |f(v)|^q \; dP_u(v),\end{aligned}$$

by Jensen's inequality, and thus

$$\begin{aligned}\|\mathsf{P}f\|_q^q &= \int_{u \in \mathcal{K}} |(\mathsf{P}f)(u)|^q \; d\pi(u) \\ &\leq \int_{u \in \mathcal{K}} \left(\int_{v \in \mathcal{K}} |f(v)|^q \, dP_u(v) \right) d\pi(u) \\ &= \int_{v \in \mathcal{K}} |f(v)|^q d\pi(v) \\ &= \|f\|_q^q,\end{aligned}$$

where we applied Fubini's theorem and $\pi\mathsf{P} = \pi$ in the third line. Thus the operator P acts as a contraction in that the P is bounded and the operator norm satisfies

$$\|\mathsf{P}\|_{L^q(\pi)} := \sup_{f \in L^q(\pi):\, \neq 0 \ a.e.-\pi} \frac{\|\mathsf{P}f\|_{L^q(\pi)}}{\|f\|_{L^q(\pi)}} \leq 1.$$

Equality is achieved with $f = 1$ almost everywhere with respect to π, so $\|\mathsf{P}\|_{L^q(\pi)} = 1$. Similar statements apply for the $q = \infty$ case.

We will be concerned with the $q = 2$ case. Note that in this case the conditional expectation defines a self-adjoint linear operator.

Proposition 3.5.1. *If a Markov-chain is reversible with respect to* π*, then* $\mathsf{P} : L^2(\pi) \to L^2(\pi)$ *is a self-adjoint linear operator.*

Proof. For any $f, g \in L^2(\pi)$, by reversibility we have

$$\begin{aligned}
\langle \mathsf{P}f,\, g \rangle_{L^2(\pi)} &= \int_x \left(\int_y f(y)\, dP_x(y) \right) g(x)\, d\pi(x) \\
&= \int_y f(y) \left(\int_x g(x)\, dP_y(x) \right) d\pi(y) \\
&= \langle f,\, \mathsf{P}g \rangle_{L^2(\pi)}.
\end{aligned}$$

□

3.6 Ergodicity, the Spectral Gap, and Mixing Times

Before reading this section, we again encourage the reader to review some results from functional analysis provided in Appendix B. We consider only lazy reversible chains, though mention that the theory extends to the non-lazy non-reversible case as well (Roberts and Rosenthal, 1997; Kontoyiannis and Meyn, 2012). For our chains, we let P_x^t to denote the distribution of X_t conditioned on $X_0 = x$, and the measure of our chain after t steps is $\pi_0 \mathsf{P}^t$ where π_0 is the measure of X_0. Roberts et al. (2004) define ergodicity in terms of the total variation distance as follows.

Definition 3.6.1 (Ergodicity). *A discrete-time homogenous Markov chain with invariant distribution π, state space $\mathcal{K}$, and transition measure $P_x(\cdot)$ is said to be uniformly ergodic if there exists $\rho < 1, M < \infty$ such that*

$$d_{TV}(P_x^t, \pi) = M\rho^t, \quad t \in \mathbb{N}.$$

A weaker condition is geometric ergodicity in which the constant M depends on the initial state $x \in \mathcal{K}$. A chain is said to be geometrically ergodic if, given $x \in \mathcal{K}$, there exists $M(x) < \infty$ such that

$$d_{TV}(P_x^t, \pi) = M(x)\rho^t, \quad t \in \mathbb{N}.$$

Note that given a warm start and a lower bound on the conductance ϕ for a lazy, reversible Markov chain on a general state space, theorem 3.4.4 implies that the Markov chain is geometrically ergodic.

We next define the spectrum of the Markov chain as determined by the conditional expectation linear operator P, and its spectral gap.

Definition 3.6.2 (Spectrum). *Given a Markov chain $\{\mathcal{K}, \mathcal{A}, P_u\}$ and the linear operator P defined as in section 3.5, the spectrum of the chain is*

$$S_2 := \left\{\lambda \in \mathbb{C} \setminus \{0\} \mid (\mathsf{P} - \lambda\mathsf{I})^{-1} \textit{ does not exist as a bounded linear operator on } L^2(\pi)\right\},$$

where I is the identity map on $L^2(\pi)$.

We additionally define the spectral gap as follows.

Definition 3.6.3 (Spectral Gap). *The spectral gap of Markov chain is*

$$\gamma := \begin{cases} \sup\{|\gamma| \mid \gamma \in S_2, \gamma \neq 1\}, & \textit{if eigenvalue 1 has multiplicity 1,} \\ 0, & \textit{otherwise.} \end{cases}$$

If the chain is reversible, then by theorem B.7, the spectrum lies on the real line and is contained in the closed interval $[-1, 1]$. The lazy version of the chain (with transition measure $Q_u(A) = \frac{1}{2}P_u + \frac{1}{2}I_A(u)$, $A \in \mathcal{A}$) then satisfies

$$\begin{aligned}\langle f, \mathsf{Q}f\rangle_{L^2(\pi)} &= \frac{1}{2}\langle f, \mathsf{P}f\rangle_{L^2(\pi)} + \frac{1}{2}\langle f, \mathsf{I}f\rangle_{L^2(\pi)} \\ &\geq -\frac{1}{2}\|f\|_{L^2(\pi)} + \frac{1}{2}\|f\|_{L^2(\pi)} \\ &= 0,\end{aligned}$$

so the lazy chain's spectrum lies in the closed interval $[0, 1]$. That geometrically ergodicity is equivalent to admitting a non-zero spectral gap is shown in (Roberts and Rosenthal, 1997) and (Kontoyiannis and Meyn, 2012). Given theorem 3.4.4, we thus may presume the Markov chains in this dissertation admit a spectral gap.

Finally, we remark that one contribution of Paulin (2015) is relating the spectral gap to the mixing time for ergodic chains. We state this theorem as follows, and refer the reader to (Paulin, 2015) for a proof.

Theorem 3.6.4 (Relationship between Mixing Times and Spectral Gap)**.** *For a lazy, reversible, ergodic Markov chain, we have*

$$\gamma \geq \frac{1}{1 + \tau(\epsilon)/\log(1/\epsilon)},$$

where $0 \leq \epsilon < 1/2$, and $\tau(\epsilon)$ is as defined in corollary 3.4.5. For an arbitrary initial distribution π_0 which is absolutely continuous with respect to π, we have

$$d_{TV}(\pi_0\mathsf{P}^t, \pi) \leq \frac{1}{2}(1-\gamma)^t\sqrt{N_{\pi_0} - 1},$$

where we define

$$N_{\pi_0} := \mathbb{E}_\pi\left[\left(\frac{d\pi_0}{d\pi}\right)^2\right] = \int_{x\in\mathcal{K}} \frac{d\pi_0}{d\pi}(x)\, d\pi_0(x). \tag{3.5}$$

Chapter 4

ROUNDING A CONVEX BODY SPECIFIED BY A SELF-CONCORDANT BARRIER

4.1 Introduction

The recent $\widetilde{O}(n^3)$ algorithm of Cousins and Vempala (2015) for volume computation presumes the convex body K is suitably rounded in the sense that

$$\mathcal{B}^n \subset \mathcal{K} \subset O(\sqrt{n})\mathcal{B}^n, \tag{4.1}$$

but omits an algorithm to perform such a rounding step. Although for simple convex bodies one may be able to find a suitable affine transformation such that the body is suitably rounded by inspection, this in general will not be the case, necessitating the need for rounding algorithms. Rounding algorithms have been analyzed in (Kannan et al., 1997) and (Lovász and Vempala, 2006b), but have complexities $\widetilde{O}(n^5)$ and $\widetilde{O}(n^4)$, respectively, dominating the volume estimation. Furthermore, these algorithms require multiple phases, either interlaced with the volume computation algorithm (Kannan et al., 1997), or requiring several phases of sampling and incrementally rounding as a preprocessing step (Lovász and Vempala, 2006b), which adds additional complication. Due to the affine-invariant nature of the Dikin walk, this interweaving of rounding and sampling phases will be unnecessary, and just one phase of sampling is required. Given a warm start, our algorithm requires $\widetilde{O}(n^3\nu)$ samples, where $\mathcal{K}$ is described by a self-concordant barrier F with complexity parameter ν.

4.2 Isotropic Position

Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex body. Let π denote the uniform measure on $\mathcal{K}$, and let X be a random sample from π. Denote the centroid of $\mathcal{K}$ as $\mu = \mathbb{E}_\pi[X]$ and the covariance (or inertia) matrix $\Sigma = \mathbb{E}_\pi[(X-\mu)(X-\mu)^\top]$. Such a convex body is said to be in *isotropic position* if $\mu = 0$ and $\Sigma = I$. Equivalently, if $\mu = 0$, the convex body is in isotropic position if for every $u \in S^{n-1}$ (where S^{n-1} denotes the unit sphere in $\mathbb{R}^n$), we have $\mathbb{E}_\pi[|u^\top X|^2] = 1$.

The connection between isotropic positioning and the assumption in (4.1) are as follows. Kannan et al. (1995) showed that for $\mathcal{K}$ in isotropic position, we have

$$\sqrt{\frac{n+1}{n}}\,\mathcal{B}^n \subset \mathcal{K} \subset \sqrt{n(n+1)}\,\mathcal{B}^n,$$

so $\mathcal{B}^n \subset \mathcal{K}$. Note also that any uniform random sample Y from isotropic $\mathcal{K}$ satisfies

$$\begin{aligned}\mathbb{E}|Y|^2 &= \mathbb{E}[\mathrm{tr}(YY^\top)] \\ &= \mathrm{tr}(\mathbb{E}[YY^\top]) \\ &= n.\end{aligned}$$

It follows by Markov's inequality that for any Y chosen uniformly at random from an isotropic $\mathcal{K} \subset \mathbb{R}^n$, we have

$$\mathbb{P}(|Y| \geq s\sqrt{n}) \leq \frac{1}{s^2},$$

so all but a fraction of $\delta > 0$ of $\mathcal{K}$ is contained within a Euclidean ball of radius $\sqrt{n/\delta}$. In fact this may be strengthened for large enough s by lemma 5.17 in (Lovász and Vempala, 2007), which states that

$$\mathbb{P}(|Y| \geq s\sqrt{n}) \leq e^{-s+1}.$$

In either case, if we ignore a small portion of $\mathcal{K}$, assumption (4.1) holds.

Note that any convex body with a non-empty interior may be transformed into an isotropic body $\widetilde{\mathcal{K}}$ by an affine transformation,

$$\begin{aligned}\widetilde{\mathcal{K}} &= \left\{y \in \mathbb{R}^n \;\middle|\; \Sigma^{1/2}v + \mu \in \mathcal{K}\right\} \\ &= \left\{\Sigma^{-1/2}(x-\mu) \;\middle|\; x \in \mathcal{K}\right\}.\end{aligned}$$

Finding such an affine transformation necessitates estimating the mean μ and the covariance matrix Σ. As these cannot be estimated with infinite precision, the best one can hope for is approximate isotropic positioning. A convex body $\mathcal{K}$ is said to be in *φ-nearly isotropic position* for $\varphi \in (0,1)$ if

$$1 - \varphi \leq \mathbb{E}_\pi[XX^\top] \leq 1 + \varphi, \tag{4.2}$$

Note that (4.2) is equivalent to the statement that

$$\|\Sigma + \mu\mu^\top - I\| \leq \varphi,$$

or that the eigenvalues of $\Sigma + \mu\mu^\top$ are bounded between $(1-\varphi)$ and $(1+\varphi)$.

Several researchers (Rudelson, 1999; Adamczak et al., 2010) have addressed a related question: given an isotropic log-concave distribution on $\mathbb{R}^n$ (which includes the uniform distribution on a convex body), how many independent points X_i are required such that for some given $\varphi > 0$, we have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i X_i^\top - I \right\| \leq \varphi \tag{4.3}$$

with overwhelming probability, where $\|M\|$ denotes the spectral norm a matrix M. Note that for a positive semidefinite matrix M we have $\|M\| = \sup_{u \in S^{n-1}} u^\top A u$, whence equation (4.3) is equivalent to

$$1 - \varphi \leq \frac{1}{N} \sum_{i=1}^{N} |X_i^\top u|^2 \leq 1 + \varphi \quad \text{for all} \quad u \in S^{n-1}.$$

This implies (4.2) for the empirical measure given $\mu = 0$ is known. While the results of (Rudelson, 1999) and (Adamczak et al., 2010) among others show that $N = \widetilde{O}(n)$ uniform samples suffice, they are presumed to be independent, which is not easily obtained for a general convex body $\mathcal{K} \subset \mathbb{R}^n$.

The typical approach instead is to acquire samples from the convex body $\mathcal{K} \subset \mathbb{R}^n$ using geometric random walks such as the ball walk (Kannan et al., 1997; Lovász and Vempala, 2006b). Unfortunately, these approaches suffer from an assumption on $\mathcal{K}$ being rounded such that $r\mathcal{B}^n \subset \mathcal{K} \subset R\mathcal{B}^n$, where the ratio R/r is not too large. However, the assumption for volume computation of (4.1) was $r = 1$ and $R = \sqrt{n}$. Thus, multi-phase Monte Carlo algorithms are used such that either volume computation steps are interwoven with rounding steps (Kannan et al., 1997), or many sampling and rounding phases are performed initially as a preprocessing step (Lovász and Vempala, 2006b). These approaches require $\widetilde{O}(n^5)$ and $\widetilde{O}(n^4)$ membership oracle calls, respectively. We instead employ an affine-invariant random walk, the Dikin walk (Narayanan, 2016), which will require only one sampling phase. While our analysis is specific to that of the Dikin walk, it is easily modified to any geometrically ergodic affine-invariant random walk.

4.3 Sampling with the Dikin Walk

In this section we describe the Dikin walk on a convex body $\mathcal{K}$ specified by a self-concordant barrier. We begin with the description of the body.

4.3.1 Defining the Convex Set

We assume that $\mathcal{K} := \mathcal{K}_\ell \cap \mathcal{K}_h \cap \mathcal{K}_s$, where $\mathcal{K}_\ell$ is a polytope with m faces accompanied by a logarithmic barrier F_ℓ, $\mathcal{K}_h$ is a convex set accompanied by a hyperbolic barrier F_h with self-concordance parameter ν_h, and $\mathcal{K}_s$ is a convex set accompanied by a self-concordant barrier F_s whose self-concordance parameter ν_s. We assume that $\mathcal{K}$ is bounded and contains the origin, though each of the convex sets comprising the intersection defining $\mathcal{K}$ may be unbounded.

Definition 4.3.1 (General self-concordant barrier)**.** *Given F_ℓ, F_h, F_s as above, we define the self-concordant barrier function*

$$F := F_\ell + nF_h + n^2 F_s,$$

and define

$$\nu := m + n\nu_h + (n\nu_s)^2$$

to be the complexity parameter of F (this differs from its self-concordance parameter, which is bounded above by $m + n\nu_h + n^2\nu_s$).

Given the current step $x \in \mathrm{int}(\mathcal{K})$, generate the next step y as follows.

1. Toss a fair coin. If the result is heads, let $y = x$.

2. Otherwise,

   (a) Sample z from the density g_x.

      i. If $z \notin \mathcal{K}$, let $y = x$.

      ii. If $z \in \mathcal{K}$, let

$$y = \begin{cases} z, & \text{with probability } \min\left(1, \frac{g_z(x)}{g_x(z)}\right) \\ x & \text{else.} \end{cases}$$

Algorithm 4.3.1: Dikin Walk Step

4.3.2 Dikin Ellipsoids and the Dikin Walk

We refer the reader to section 2.2 to review Dikin ellipsoids and their properties. Namely, the Dikin ellipsoid $\mathcal{D}_x(r)$ is given by (2.2) and the local norm it induces is given by (2.3). Using the local norm, one may then define a Gaussian proposal distribution to use in the walk we will define. For $x \in \mathrm{int}\,\mathcal{K}$, let g_x denote the Gaussian density function given by

$$g_x(y) := \left(\frac{n}{\pi r^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n\|y - x\|_x^2}{r^2} + V(x)\right),$$

where

$$V(x) = \left(\frac{1}{2}\right) \log\det D^2 F(x).$$

In other words, $g_x(y)$ is the density of a Gaussian distribution centered at x with covariance matrix $\frac{r^2}{n}(D^2F(x))^{-1}$. We may now define the Dikin walk, provided in algorithm 4.3.1.

4.3.3 Mixing Time of the Dikin Walk

Narayanan (2016) analyzed the convergence of this walk to the uniform distribution on a convex body $\mathcal{K} \subset \mathbb{R}^n$, and proved the following theorem.

Theorem 4.3.2 (Mixing Time of the Dikin Walk)**.** *Let $\mathcal{K} \subset \mathbb{R}^n$ be an n-dimensional convex set accompanied by a self-concordant barrier function F as defined in section 4.3.1 with complexity parameter ν. Let $x_0 = x \in \operatorname{int} \mathcal{K}$ be a "s-central" starting point, where s is defined such that $s \geq \frac{|p-x|}{|q-x|}$ for any chord $\overline{pq}$ in $\mathcal{K}$ through x. If τ_1 is the time of the first non-trivial move of the Markov chain, and π_t is the measure of the chain after t steps, then the number of steps after τ_1 before $d_{TV}(\pi_t, \pi) \leq \epsilon$ is $O(n\nu(n \log(\nu s) + \log(\frac{1}{\epsilon})))$. The waiting time τ_1 of the first non-trivial move has a geometric distribution whose mean is bounded above by a universal constant. If $x_0 = x \in \operatorname{int} \mathcal{K}$ is generated via a warm start, then the number of steps needed such that $d_{TV}(\pi_t, \pi) \leq \epsilon$ is $O(n\nu \log \frac{1}{\epsilon})$.*

Given theorem 3.6.4, the spectral gap of the Dikin walk from a warm start immediately follows.

Corollary 4.3.3 (Dikin Walk Spectral Gap)**.** *The spectral gap γ of the Dikin walk given a warm start is $\Omega(\frac{1}{n\nu})$.*

We will use the Dikin walk as a means to bring a convex body into approximate isotropic position. Before we analyze this method, however, we derive some moment bounds for isotropic random vectors.

4.4 Moment Bounds for Log-Concave Densities in Isotropic Position

Lovász and Vempala (2007) address the general case of random variables Y with densities with respect to the Lebesgue measure which are *log-concave*. Assume λ denotes the Lebesgue measure in $\mathbb{R}^n$ and the Radon-Nikodym derivative (i.e., density) is $f = \frac{d\pi}{d\lambda}$. Log-concave densities are defined as densities $f : \mathbb{R}^n \to \mathbb{R}_+$ satisfying

$$f(\alpha x + (1-\alpha)y) \geq f(x)^{\alpha} f(y)^{1-\alpha}$$

for every $x, y \in \mathbb{R}^n$ and all $\alpha \in [0,1]$. Equivalently, the support of f is convex and $\log f$ is concave. Trivially the uniform density on a convex body $\mathcal{K}$ meets this assumption.

We will require bounds on $\mathbb{E}|Y|^k$ for $k = 1$ and $k = 4$, as well as a bound on $|Y|$. We employ two useful lemmas from (Lovász and Vempala, 2007) to this end. By theorem 5.22 of (Lovász and Vempala, 2007), for any Y with a log-concave density, we have

$$(\mathbb{E}|Y|)^k \leq \mathbb{E}|Y|^k \leq (2k\,\mathbb{E}|Y|)^k, \tag{4.4}$$

where first inequality is just Jensen's inequality and the second inequality is a generalized Khinchine inequality. Secondly, lemma 5.18 from (Lovász and Vempala, 2007) states that for any log-concave density corresponding to a body in isotropic position, we have

$$|Y| \leq n + 1. \tag{4.5}$$

almost everywhere. Using (4.4), given isotropic $\mathcal{K}$, have the expected norm bound,

$$\frac{\sqrt{n}}{16} \leq \mathbb{E}(|Y|) \leq \sqrt{n},$$

which implies by using (4.4) again that

$$\frac{n^2}{16^4} \leq \mathbb{E}(|Y|^4) \leq 8^4 n^2. \tag{4.6}$$

4.5 Rounding the Body

Assume that the mean and covariance of the uniform distribution π on $\mathcal{K}$ are μ and Σ, respectively. Given $X \sim \pi$, we seek a vector θ and a square invertible matrix A such that $Y = A(X - \theta)$ is an approximately isotropic random vector. Clearly, $\mathbb{E}[Y] = A(\mu - \theta)$ and $\mathrm{Cov}(Y) = A\Sigma A^\top$, so

$$\begin{aligned}\mathbb{E}[YY^\top] &= \mathrm{Cov}(Y) + \mathbb{E}[Y]\mathbb{E}[Y]^\top \\ &= A(\Sigma + (\mu - \theta)(\mu - \theta)^\top)A^\top.\end{aligned}$$

For approximate isotropic position, we require

$$\|A(\Sigma + (\mu - \theta)(\mu - \theta)^\top)A^\top - I\| \leq \varphi. \tag{4.7}$$

which follows from the triangle inequality if we can show that

$$\|A(\mu-\theta)(\mu-\theta)^\top A^\top\| \le \varphi_1 \tag{4.8}$$

and

$$\|A\Sigma A^\top - I\| \le \varphi_2, \tag{4.9}$$

where φ_1, φ_2 satisfy $\varphi_1 + \varphi_2 \le \varphi$.

To choose A and θ, we sample from the body. We presume that we have a sample $\{X_1, X_2, \ldots, X_T\}$ of points taken from the a Dikin walk on $\mathcal{K} \subset \mathbb{R}^n$ where $d_{TV}(\pi_1, \pi) \le \epsilon$. We then form the usual estimates,

$$\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t, \tag{4.10}$$

and

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})^\top. \tag{4.11}$$

Note that exact isotropic positioning with $\varphi = 0$ in (4.7) is achieved via $\theta = \mu$ and $A = \Sigma^{-1/2}$. Thus, the natural choices for θ and A given samples are $\bar{X}$ and $\hat{\Sigma}^{-1/2}$, respectively.

4.5.1 Analysis Using Markov Chain Concentration Inequalities

We first presume that we have drawn (dependent) samples starting from the uniform distribution on a body $\mathcal{K}$ in isotropic position. We then extend the analysis to the general case in which our first sample is nearly uniform and $\mathcal{K}$ is not isotropic. The following Bernstein inequality for ergodic Markov chains provides the means by which we prove near-isotropic positioning.

Theorem 4.5.1 (Paulin (2015))**.** *Let $Z_1, \ldots, Z_T$ be samples taken from a stationary, lazy, reversible Markov chain with spectral gap γ and stationary measure π. Consider any measurable $f \in L^2(\pi)$ be such that*

$$|f(Z) - \mathbb{E}_\pi(f(Z))| \le C,$$

and define $V_f := \mathrm{Var}_\pi(f)$. *Let* $S := \sum_{t=1}^{\top} f(Z_t)$. *Then*

$$\mathbb{P}_\pi(|S - \mathbb{E}_\pi(S)| \geq a) \leq 2\exp\left(-\frac{\gamma a^2}{4TV_f + 10aC}\right). \tag{4.12}$$

To bound the probability in (4.12) by $\eta \in (0,1)$, let $a = \varphi T$. Then inverting (4.12), we see that

$$T = \left\lceil \frac{(4V_f + 10C\varphi)\log(2/\eta)}{\varphi^2\gamma} \right\rceil$$

samples are required to bound (4.12) by η. Note that the spectral gap γ of the Dikin walk provided by corollary 4.3.3. Given any function f, it remains to bound C and V_f to bound the number of samples T such that $\frac{1}{T}(S - \mathbb{E}_\pi(S))$ is between $-\varphi$ and φ with probability $1-\eta$.

Now consider bounding the affine-invariant quantities

$$\|\Sigma^{-1/2}(\bar{X} - \mu)(\bar{X} - \mu)^\top \Sigma^{-1/2}\| \tag{4.13}$$

and

$$\|\Sigma^{-1/2}(\hat{\Sigma} - \Sigma)\Sigma^{-1/2}\|. \tag{4.14}$$

Note that $\hat{\Sigma} = \hat{\Sigma}_0 - \bar{X}\bar{X}^\top$ where $\hat{\Sigma}_0 = \frac{1}{T}\sum_{t=1}^{\top} X_i X_i^\top$. For $\mathcal{K}$ in isotropic position, (4.13) reduces to

$$\|\bar{X}\bar{X}^\top\| = |\bar{X}|^2$$

and (4.14) reduces to

$$\begin{aligned}\|\hat{\Sigma} - I\| &= \|(\hat{\Sigma} - \hat{\Sigma}_0) + (\hat{\Sigma}_0 - I)\| \\ &\leq \| - \bar{X}\bar{X}^\top\| + \|\hat{\Sigma}_0 - I\| \\ &= |\bar{X}|^2 + \|\hat{\Sigma}_0 - I\|. \end{aligned} \tag{4.15}$$

We consider each term in (4.15) separately, applying (4.12) to each.

Considering the mean, note the following lemma.

Lemma 4.5.2 (Mean Bound)**.** *Let $X_1, \ldots, X_{T_1}$ be sampled from a stationary, lazy, reversible Markov chain with spectral gap γ on an isotropic convex body $\mathcal{K}$. Then $|\bar{X}| \leq \varphi_1$ with probability at least $1 - \eta_1$ for*

$$T_1 \geq \left\lceil \frac{[(15/4)n + 10(n+1)\varphi_1] \log(2/\eta_1)}{\varphi_1^2 \gamma} \right\rceil. \tag{4.16}$$

Proof. We have

$$\begin{aligned} \frac{1}{T}|S - \mathbb{E}_\pi(S)| &= \left| \frac{1}{T} \sum_{t=1}^{T} X_t \right| \\ &\leq \frac{1}{T} \sum_{t=1}^{T} |X_t|. \end{aligned}$$

Thus, let $Y \sim \pi$ where π denotes the uniform distribution on an isotropic body $\mathcal{K}$, and let $f(Y) = |Y|$. Applying (4.5), we have

$$\begin{aligned} |f(Y) - \mathbb{E}_\pi[f(Y)]| &\leq |Y| \\ &\leq n + 1, \end{aligned}$$

so we make take $C = n + 1$. Applying (4.4), we have

$$\begin{aligned} V_f &= \mathbb{E}_\pi(|Y|^2) - (\mathbb{E}_\pi|Y|)^2 \\ &\leq n - (1/16)\mathbb{E}_\pi(|Y|^2) \\ &= (15/16)n. \end{aligned}$$

Thus, the lower bound for T_1 in (4.16) follows. ☐

Considering the variance, note the following lemma.

Lemma 4.5.3 (Covariance Bound)**.** *Let $X_1, \ldots, X_{T_2}$ be sampled from a stationary, lazy, reversible Markov chain with spectral gap γ on an isotropic convex body $\mathcal{K}$. Then $|\hat{\Sigma}_0 - I| \leq \varphi_2$ with probability at least $1 - \eta_2$ for*

$$T_2 \geq \left\lceil \frac{(4(8^4 - 1)n^2 + 10(n^2 + n - 1)\varphi_2) \log(2/\eta_2)}{\varphi_2^2 \gamma} \right\rceil \tag{4.17}$$

Proof. Noting that

$$\begin{aligned}\sup_{u\in S^{n-1}} u^\top(\hat{\Sigma}_0 - I)u &= \sup_{u\in S^{n-1}} u^\top\left[\frac{1}{T}\sum_{t=1}^{\top}\left(X_tX_t^\top - I\right)\right]u \\ &\leq \frac{1}{T}\sum_{i=1}^{\top}\sup_{u\in S^{n-1}} u^\top(X_tX_t^\top - I)u.\end{aligned}$$

we may apply (4.12) with

$$\begin{aligned}f(Y) &= \sup_{u\in S^{n-1}} u^\top(YY^\top - I)u \\ &= |Y|^2 - 1.\end{aligned}$$

To find a bound C, inequality (4.5) implies that

$$\begin{aligned}|f(Y) - \mathbb{E}_\pi[f(Y)]| &= |Y|^2 - n \\ &\leq (n+1)^2 - n,\end{aligned}$$

so we make take $C = n^2 + n - 1$. To bound V_f, using (4.6), we have

$$\begin{aligned}V_f &= \mathbb{E}_\pi f(Y)^2 - (\mathbb{E}_\pi f(Y))^2 \\ &= \mathbb{E}_\pi[(|Y|^2 - 1)^2] - (n-1)^2 \\ &= \mathbb{E}_\pi[|Y|^4] - 2\mathbb{E}_\pi[|Y|^2] - n^2 + 2n \\ &\leq (8^4 - 1)n^2.\end{aligned}$$

Thus, the lower bound for T_2 in (4.17) follows. $\square$

Since $T_2 > T_1$ and $\gamma \geq \frac{c}{n\nu}$ for some universal constant $c > 0$, we conclude that with $T = T_2 = \widetilde{O}(n^3\nu)$ Dikin walk samples from the stationary measure on an isotropic body, we have

$$\|\hat{\Sigma} - I\| \leq \varphi$$

with probability at least $1 - \eta$, where $\varphi = \varphi_1^2 + \varphi_2$ and $\eta = 2\eta_2$.

It remains to show approximate isotropic positioning for the general case in which the body $\mathcal{K}$ is not isotropic, and the samples are taken from a distribution which is only approximately stationary. We prove (4.7) using lemmas 4.5.2 and 4.5.3 as follows.

Theorem 4.5.4 (Near Isotropic Position)**.** *Let*

$$\widetilde{\mathcal{K}} = \left\{ \hat{\Sigma}^{-1/2}(x - \bar{X}) \;\middle|\; x \in \mathcal{K} \right\}.$$

If $T \geq T_2$, then $\widetilde{\mathcal{K}}$ is nearly isotropic in that given any Y uniformly drawn from $\widetilde{\mathcal{K}}$, we have

$$\|\mathbb{E}[YY^\top] - I\| \leq \frac{\varphi_1^2}{(1-\varphi_2)^2} + \frac{\varphi_2}{1-\varphi_2}. \tag{4.18}$$

with probability at least $1 - \eta$.

Proof. Lemma 4.5.3 provides a bound

$$\|MM^\top - I\| \leq \varphi_2$$

where $M = \Sigma^{-1/2}\hat{\Sigma}^{1/2}$. We require a bound of the form of (4.9), i.e., we need to bound $\|(MM^\top)^{-1} - I\|$. Given lemma 4.5.3, the eigenvalues of $(MM^\top)^{-1}$ are contained in the interval $[(1+\varphi_2)^{-1}, (1-\varphi_2)^{-1}]$. Thus

$$\|(MM^\top)^{-1} - I\| \leq \frac{\varphi_2}{1-\varphi_2}.$$

Similarly, lemma 4.5.2 provides a bound

$$\|\Sigma^{-1/2}(\bar{X} - \mu)(\bar{X} - \mu)^\top \Sigma^{-1/2}\| \leq \varphi_1^2,$$

but we require a bound of the form (4.8). Since the spectral norm is sub-multiplicative, we have

$$\begin{aligned}
\|\hat{\Sigma}^{-1/2}(\bar{X} - \mu)(\bar{X} - \mu)^\top \hat{\Sigma}^{-1/2}\| &= \|M^{-1}\Sigma^{-1/2}(\bar{X} - \mu)(\bar{X} - \mu)^\top \Sigma^{-1/2}(M^\top)^{-1}\| \\
&\leq \varphi_1^2 \|M^{-1}\|^2 \\
&\leq \frac{\varphi_1^2}{(1-\varphi_2)^2}.
\end{aligned}$$

Combining the two bounds yields (4.18). □

Finally, note that we have assumed throughout that we are sampling precisely from the uniform measure. In practice, after an initial "burn-in" phase, the first sample in the estimators (4.10) and (4.11) will be drawn from a measure that is approximately uniform. We state the following theorem from Paulin (2015) which shows that we pay a penalty of just ϵ for taking our first sample a point X_1 such that $d_{TV}(\pi_1, \pi) \leq \epsilon$.

Theorem 4.5.5 (Bounds for Non-stationary Chains)**.** *Let $X_1, \dots, X_t$ be a time homogenous Markov chain with state space $\mathcal{K}$ and stationary distribution π. Suppose that $g(X_1, \dots, X_t)$ is a real-valued measurable function, and X_1 is distributed according to some distribution q. Then*

$$\mathbb{P}_q(g(X_1, \dots, X_t) \geq a) \leq N_q^{1/2} \left[\mathbb{P}_q(g(X_1, \dots, X_t) \geq a)\right]^{1/2},$$

where N_q was defined in (3.5). If we "burn" the first t_0 observations, we have

$$\mathbb{P}_q(g(X_{t_0+1}, \dots, X_{t_0+t}) \geq a) \leq \mathbb{P}_\pi(g(X_{t_0+1}, \dots, X_{t_0+t}) \geq a) + d_{TV}(q\mathsf{P}^{t_0}, \pi).$$

4.5.2 Complexity of Rounding

Note that an approximately uniform random sample may be generated in $\widetilde{O}(n\nu)$ steps of the Dikin Walk given a warm start (which may be generated in $\widetilde{O}(n^2\nu)$ steps after the first move from an s-central point). Since our algorithm required $\widetilde{O}(n^3\nu)$ samples, in either the case of a warm start or after the first move from an s-central point, a total of $\widetilde{O}(n^3\nu)$ steps are required. If the complexity parameter of the Dikin walk is $\nu = n$, then this matches the number of membership oracle calls of Lovász and Vempala (2006b). However, each sample of the Dikin walk requires additionally the computation of the Hessian of the self-concordant barrier which specifies the body of the random walk. Denoting this cost ζ, the complexity of the algorithm is thus $\widetilde{O}(\zeta n^3 \nu)$.

We state ζ for two special cases. For the case of a polytope described by m inequality constraints, we have self-concordance parameter m. Let $s_i(x) := 1 - a_i^\top x$ denote the slack in the i-th constraint for $i = 1, \dots, m$. Given the current location of the random walk is $x \in \mathcal{K}$, the Hessian of the logarithmic barrier is $H(x) = A^\top S_x^{-2} A$, where $S_x := \operatorname{diag}(s_1(x), \dots, s_m(x))$. Letting $A_x := A^\top S_x^{-1}$, the Hessian is equivalently $H(x) = A_x^\top A_x$, where $A_x^\top = [a_1/s_1, \ldots, a_m/s_m]$. By padding $A_x^\top$ with zeros as necessary and then partitioning the resultant matrix into less than $(m+n)/n$ submatrices of dimension $n \times n$, we may compute the Hessian in $\zeta = O(mn^{\omega-1})$ arithmetic operations. In the case of a semidefinite constraint of rank r, the number of arithmetic steps required for computing the Hessian of the log det barrier is $\zeta = O(n^2r^2 + nr^\omega)$ (Nesterov and Nemirovsky, 1994).

4.6 Simulation

Consider the following polytope defined by $n+1$ inequality constraints,

$$\Delta_c^n := \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i \le 1 \text{ and } x_i \ge 0 \text{ for } i = 1, \ldots, n \right\}.$$

Δ_c^n is the simplex in $\mathbb{R}^n$ corresponding to a corner of a unit cube located in the non-negative orthant. To generate a uniform random sample in this polytope, we may use the Dirichlet distribution. The Dirichlet distribution is parameterized by a vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{n+1})$ and is supported on the n-standard simplex in $\mathbb{R}^{n+1}$, i.e.,

$$\Delta^n := \left\{ x \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=1}^{n+1} x_i = 1 \text{ and } x_i \ge 0,\; i = 1, 2, \ldots, n+1 \right\}.$$

The Dirichlet distribution's density function is

$$f(x \mid \alpha) = \frac{1}{B(\alpha)} \left(\prod_{i=1}^{n+1} x_i^{\alpha_i - 1} \right) I_{\Delta^n}(x),$$

where $I_S(x) = 1$ if $x \in S$ and 0 otherwise, and

$$B(\alpha) := \prod_{i=1}^{n+1} \frac{\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{n+1} \alpha_i)}$$

is the multivariate beta function. The mean of the Dirichlet distribution is

$$\mathbb{E}X = \frac{\alpha}{\alpha_0}$$

where $\alpha_0 := \sum_{i=1}^{n+1} \alpha_i$, and the covariance is

$$\mathrm{Cov}(X_i,\, X_j) = \begin{cases} \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0+1)} & i = j, \\ \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)} & i \ne j. \end{cases}$$

For the special case corresponding to the uniform distribution on Δ^n, we have $\alpha = 1_{n+1} := [1, \ldots, 1]^\top$, and $\alpha_0 = (n+1)$. The mean is

$$\mathbb{E}X = \left(\frac{1}{n+1}\right) 1_{n+1}.$$

and the covariance matrix is

$$\mathrm{Cov}(X) = \left(\frac{1}{(n+1)^2(n+2)}\right) \left[(n+1)I_{n+1} - 1_{n+1}1_{n+1}^\top\right].$$

To generate a uniform random sample on Δ_c^n, it suffices to sample X from a Dirichlet distribution with $\alpha = 1_{n+1}$, and keep only the first n coordinates. Similarly, the mean is the first n coordinates of $\mathbb{E}X$, and the covariance is the $n \times n$ submatrix of $\mathrm{Cov}(X)$ corresponding to the first n coordinates. Letting $c = 1/((n+1)^2(n+2))$, it may be verified that the covariance matrix corresponding to the uniform distribution on Δ_c^n has eigenvalues $c(n+1)$ of multiplicity n and c of multiplicity 1.

We may further transform this simplex via a linear transformation. Let the covariance matrix of Δ_c^n be Σ, and its eigenvalue decomposition be $U\Lambda U^\top$. Let $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where $\lambda_1 = c$. To transform the simplex such that we have a condition number κ, consider the matrix $D = \mathrm{diag}(\sqrt{(n+1)/\kappa}, 1, \ldots, 1)$. Then we may form the new covariance matrix, by letting $\tilde{\Lambda} = D\Lambda D$, and thus,

$$\begin{aligned}
\tilde{\Sigma} &= U\tilde{\Lambda}U^\top \\
&= UD\Lambda DU^\top \\
&= UD(U^\top U)\Lambda(U^\top U)DU^\top \\
&= (UDU^\top)\Sigma(UDU^\top) \\
&\equiv L\Sigma L^\top,
\end{aligned}$$

where $L = UDU^\top$. Thus our linear transformation is $y = L(x)$, where $L = UDU^\top$.

We simulated the Dikin walk in the Python programming language on the simplex Δ_c^n and transformed versions thereof. We chose the radius to be 0.35, which empirically led to approximately 30 to 40 percent of the samples being rejected by the Metropolis filter.

We chose the transformation L such that the condition number of the covariance matrix corresponding to the transformed body was $\kappa = (n+1)^k$ for $k = 1, 2, 3$, or 4, where the $k = 1$ case corresponds to no transformation. We set the dimension to be $n = 10, 20$, or 40. For each combination of k and n, we took an initial (stationary) random sample using the Dirichlet distribution, and then simulated a total of 10^6 steps of the Dikin walk. We then formed the mean estimator (4.10) and the covariance estimator (4.11) by taking consecutive windows of size T of the 10^6 steps of the walk. We choose the window size $T = \lceil n^\alpha \rceil$ for $\alpha \in \{1.5, 1.75, 2, \ldots\}$, where the maximum α is chosen such that we have at least consecutive 10 windows of size T for each value of n.

For mean estimator $\bar{X}$, we considered the error $\epsilon := |\Sigma^{-1/2}(\bar{X} - \mu)|^2$. For the covariance estimator $\hat{\Sigma}$, we let $M = \Sigma^{-1/2}\hat{\Sigma}^{1/2}$, and found the smallest value C such that

$$\frac{1}{C} \leq \lambda(MM^\top) \leq C,$$

where $\lambda(S)$ denotes the eigenvalues of a symmetric matrix S. Note that this value of C bounds the spectrum of $(MM^\top)^{-1}$ as well. Plots showing these estimators are shown in figures 4.1 through 4.6. The median error is indicated with the solid line, and the bands indicate the middle 90% of all simulations, where we interpolated if the 0.05 and 0.95 quantiles were not precisely defined. Given the results of lemmas 4.5.2 and 4.5.3, we expect both of these errors to relatively controlled with $\widetilde{O}(n^3\nu) = \widetilde{O}(n^4)$ samples. This is indeed the case. Additionally, the affine-invariance is readily apparent in that for a given n, each plot exhibits similar behavior irrespective of the value of k.

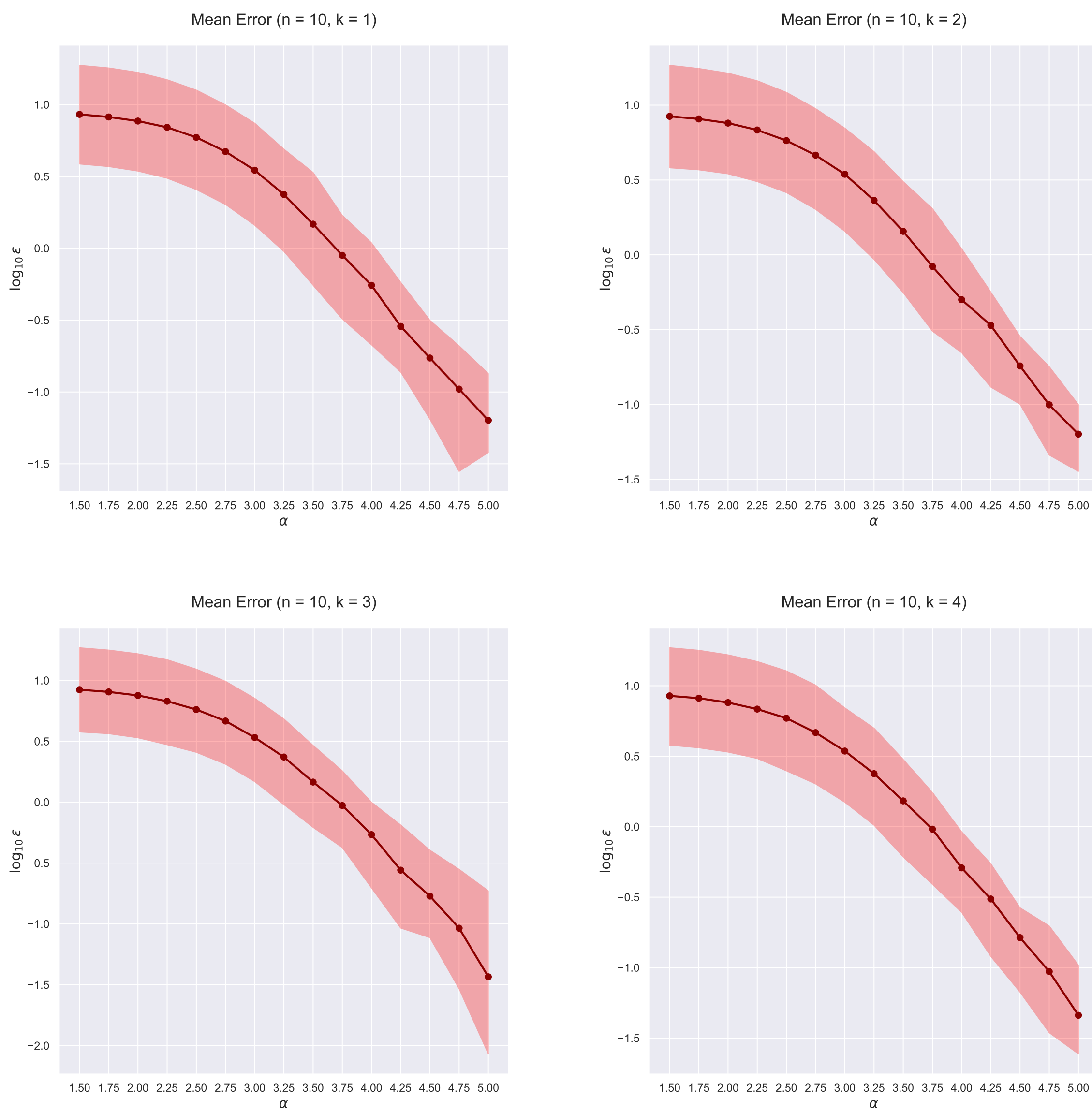


Figure 4.1: Mean errors for $n = 10$.

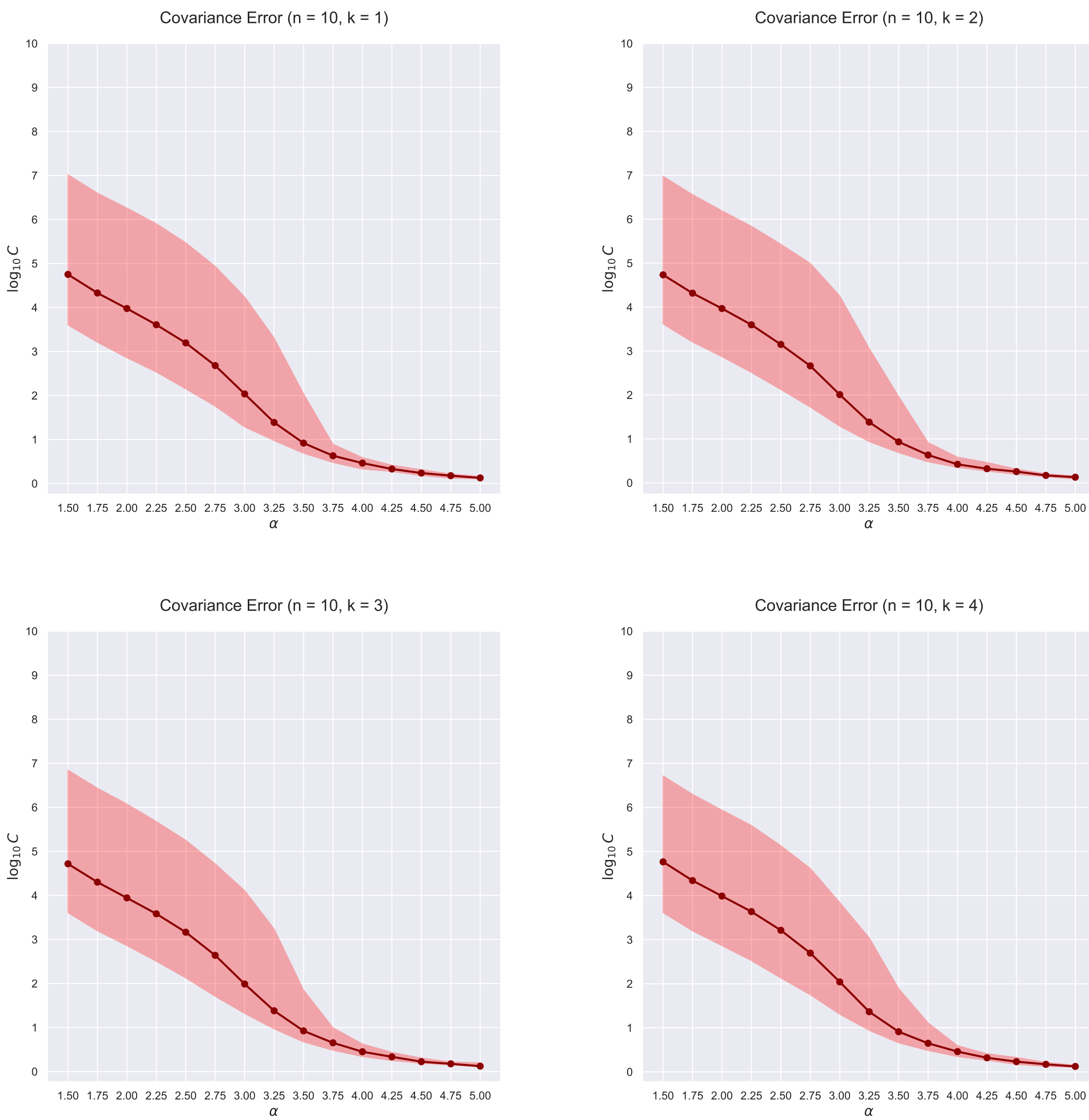


Figure 4.2: Covariance errors for $n = 10$.

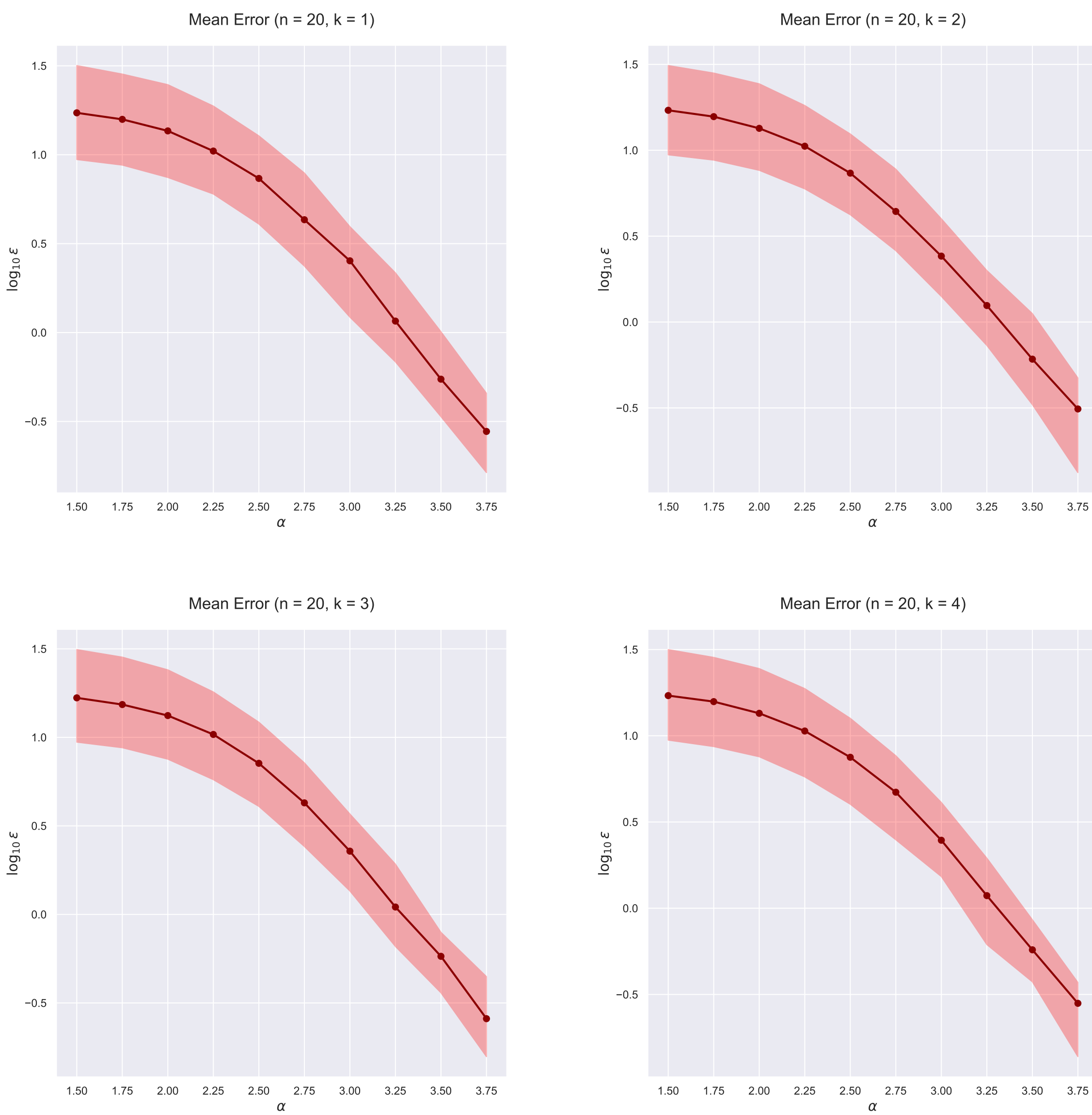


Figure 4.3: Mean errors for $n = 20$.

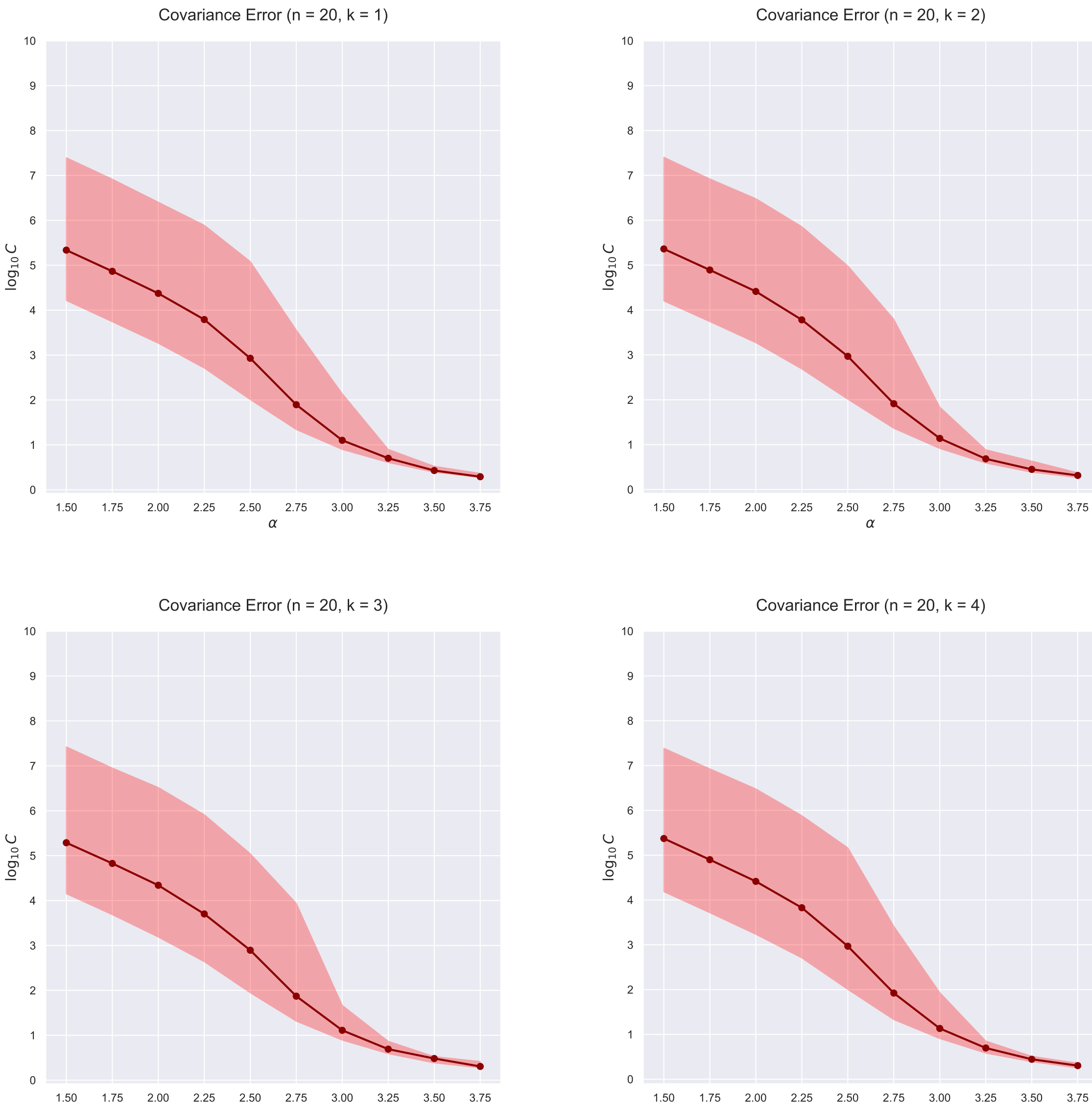


Figure 4.4: Covariance errors for $n = 20$.

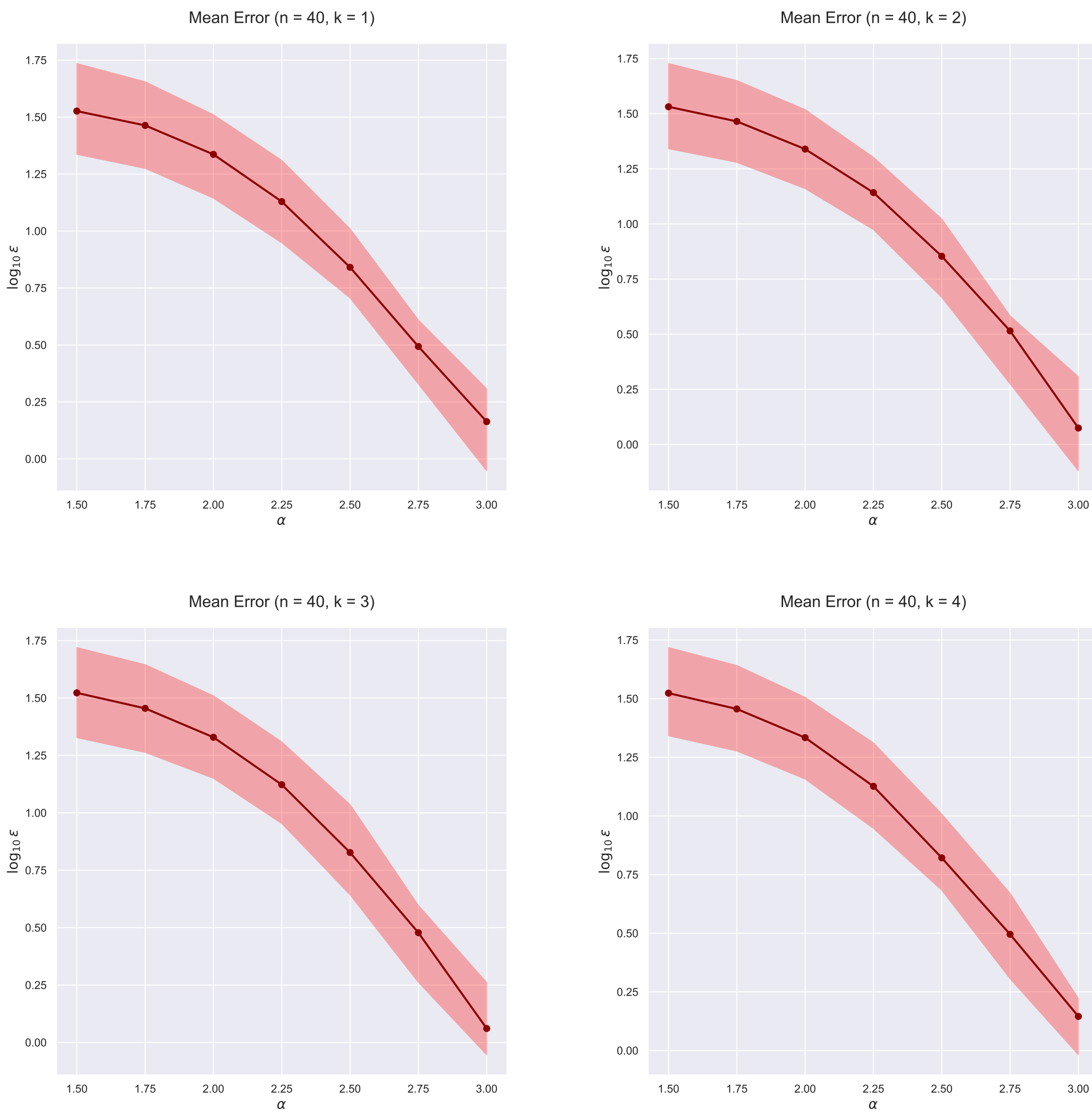


Figure 4.5: Mean errors for $n = 40$.

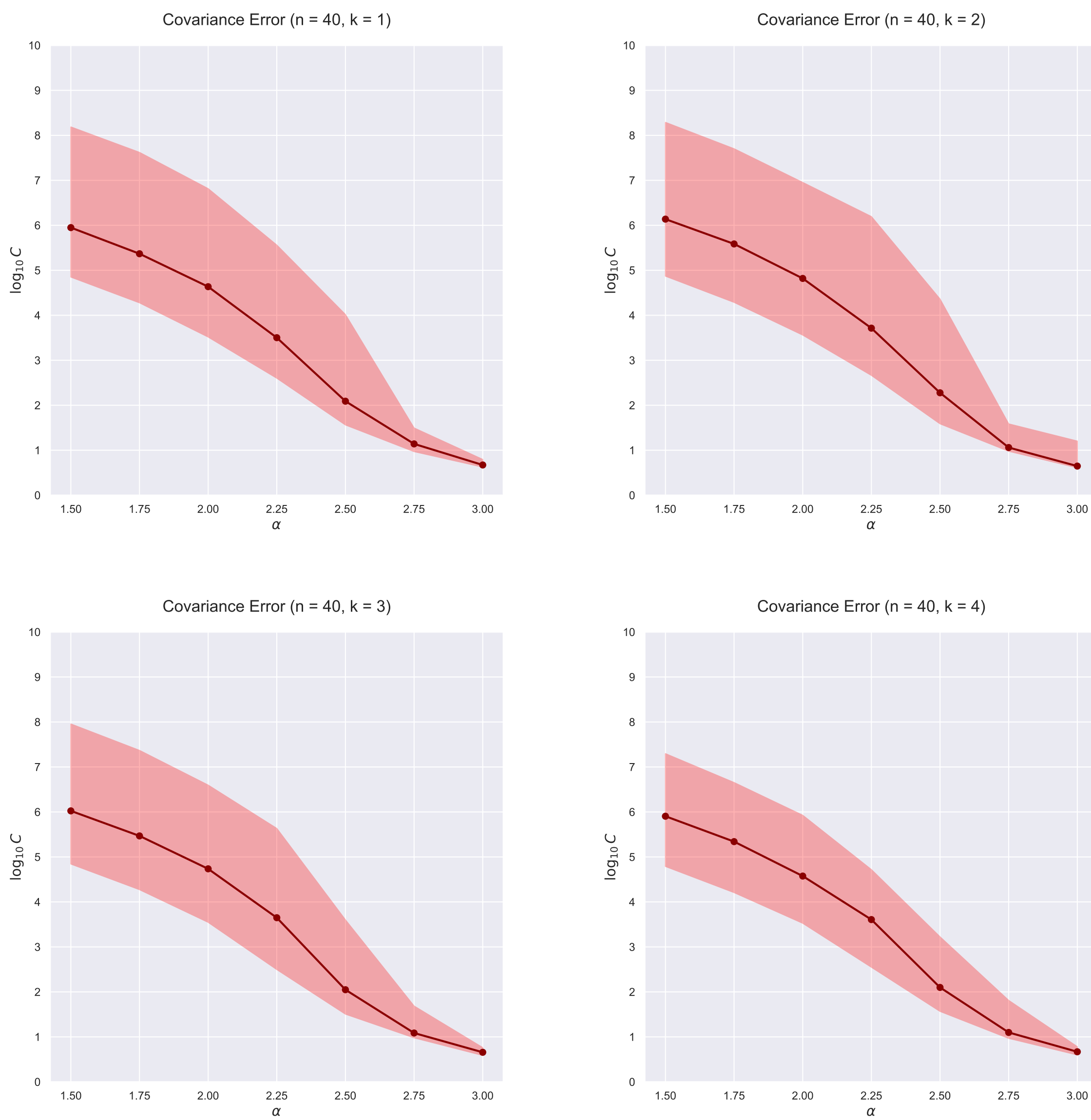


Figure 4.6: Covariance errors for $n = 40$.

4.7 Discussion

We have provided an algorithm to round a convex body with complexity $\widetilde{O}(\kappa n^3 \nu)$. The benefit of using this random walk is that it generalizes to convex bodies specified by a self-concordant barrier, and avoids the interleaving of sampling and rounding phases of other algorithms. Additionally, while our algorithm's complexity is derived in terms of the Dikin walk, our analysis was general in that it applies to any affine-invariant random walk which admits a spectral gap.

We have also supported our algorithm with simulations illustrating the number of samples required to bring polytopes into an approximately isotropic position, matching roughly our theoretical bounds. We believe with further analysis we may be able to reduce the power of n in the number of samples required, and this may necessitate developing tools akin to the matrix concentration inequalities of Tropp (2012) for the case of general state-space Markov chains. We leave this problem open to future work.

Chapter 5

AN AFFINE-INVARIANT RANDOM WALK USING JOHN'S MAXIMAL VOLUME ELLIPSOID

5.1 Introduction

Drawing random samples from a convex body in $\mathcal{K} \subset \mathbb{R}^n$ is an important problem for volume computation and optimization which has generated a large body of research. Usually $\mathcal{K}$ is specified by a membership oracle which certifies whether or not a test point $x \in \mathbb{R}^n$ is contained in $\mathcal{K}$. Given such an oracle, geometric random walks are then used to explore $\mathcal{K}$ such that after a sufficient number of steps, the walk has "mixed" in the sense that the current point is suitably close to a point uniformly drawn from $\mathcal{K}$ in terms of statistical distance. To use such walks, an assumption that $\mathcal{B}(r) \subset \mathcal{K} \subset \mathcal{B}(R)$ is often made, where $\mathcal{B}(r)$ represents the Euclidean ball of radius $r > 0$. One common example of such geometric walks is the Ball Walk, which generates the next point by uniformly randomly sampling from a ball of radius $\delta \leq r/\sqrt{n}$ centered at the current point, and mixes in $\widetilde{O}(n(R^2/\delta^2))$ steps from a warm start (Kannan et al., 1997). Another is Hit and Run, where the next point is chosen uniformly at random from a random chord in $\mathcal{K}$ which intersects the current point. Hit and Run mixes in $O(n^3(R^2/r^2)\log(R/(d\epsilon)))$ where the starting point is a distance d from the boundary and ϵ is the desired distance to stationarity (Lovász and Vempala, 2006a). Both Ball Walk and Hit and Run thus depend on the rounding of the convex body in question, i.e., the term R/r, which may be arbitrarily large in terms of n. An affine transformation is required as a preprocessing step to bring the convex body into a position for which R/r is suitably small. Such rounding transformations depend on drawing samples to estimate the covariance matrix of the body, and thus the complexity of rounding dwarfs the complexity of sampling. As such, random walks which circumvent the problem of rounding are desirable.

One recent algorithm which avoids rounding the body is known as projected Langevin Monte Carlo (Bubeck et al., 2018), and uses a discretization of Langevin diffusion with a projection of each step onto the body if necessary. This algorithm mixes in $\widetilde{O}(n^7)$ steps from any starting point, but the complexity of each iteration depends additionally on the projection which may be costly for general bodies. Affine-invariant walks (i.e., geometric walks whose mixing time is invariant to such affine transformations) are another class of random walks which avoid the problem of rounding. One such random walk is known as Dikin Walk (Kannan and Narayanan, 2012), which uses uniform sampling from Dikin ellipsoids to make steps. Given a polytope with m inequality constraints, the Dikin Walk mixes in $\widetilde{O}(mn)$ steps from a warm start. This random walk was extended to general convex bodies equipped with a ν-self-concordant barrier by Narayanan (2016), and mixes in $\widetilde{O}(n^3\nu^2)$ steps from a warm start. For the case of a polytope, this implies that the Dikin walk equipped with the Lee-Sidford (LS) barrier (Lee and Sidford, 2013) mixes in $\widetilde{O}(n^5)$ steps from a warm start, though at each step one must additionally compute the LS barrier which requires $O(nnz(A) + n^2)$ arithmetic operations, where $nnz(A)$ is the number of non-zeros in the matrix A which defines the polytope.

This chapter introduces another affine-invariant random walk akin to Dikin Walk which uses uniform sampling from John's ellipsoids to make steps, and show that this walk mixes in $\widetilde{O}(n^7)$ steps from a warm start. The logarithmic factors in this mixing time depend only on the warm start and error parameters (and not on the dimension or the number of inequality constraints defining, say, a polytope). The type of convex body $\mathcal{K}$ is not specified (i.e., need not be a polytope) in our analysis of the mixing time, but one must have access to the John's ellipsoid of the current symmetrization of the convex body. For the special case of a polytope, we describe an algorithm to compute the John's ellipsoid of the current symmetrization in $\widetilde{O}(mn^{\omega+1} + n^{2\omega+2})$ iterations, noting that the Dikin Walk also has computational complexity for each step of the walk which is linear in m. Thus, since the mixing time does not depend on m, our algorithm may be suitable in cases in which $m \gg n$ and n is not prohibitively large.

This chapter is organized as follows. In section 5.2 we recall John's theorem which characterizes the maximum volume ellipsoids contained in a convex body, describe the convex program to find such an ellipsoid in the case of a polytope, and additionally describe our algorithm for John's Walk. In section 5.3, we analyze the mixing time of our algorithm. In section 5.4, we describe an algorithm to compute suitably approximate John's ellipsoids for the special case of a polytope, and show that the mixing time is maintained with using these approximate ellipsoids.

5.2 John's Walk

In this section, we describe John's maximum volume ellipsoid for a convex body $\mathcal{K} \subset \mathbb{R}^n$, and describe a geometric random walk using such ellipsoids. We refer the reader to section 2.3 to review John's theorem and its consequences for a general convex body $\mathcal{K} \subset \mathbb{R}^n$. We discuss John's maximal volume ellipsoid for the special case of a polytope next.

5.2.1 John's Ellipsoids Via Convex Programming

For general convex bodies, finding John's maximal volume inscribed ellipsoid is hard to compute. However for polytopes, methods exist which calculate the maximal volume ellipsoid up to a user-specified tolerance parameter. Now let $\mathcal{P} = \left\{x \in \mathbb{R}^n \;\middle|\; Ax \leq 1\right\}$ denote a polytope in $\mathbb{R}^n$, where $A \in \mathbb{R}^{m \times n}$. Note that we may parameterize any ellipsoid centered at x_c as

$$\mathcal{E} = \left\{Ey + x_c \;\middle|\; |y| \leq 1\right\},$$

or equivalently by

$$\mathcal{E} = \left\{x \;\middle|\; (x - x_c)^\top E^{-2}(x - x_c) \leq 1\right\},$$

where $E \in S^n_{++}$. For a general convex body $\mathcal{K} \subset \mathbb{R}^n$, since the volume of $\mathcal{E}$ is proportional to $\det E$, the convex optimization problem to be solved is as follows (Vandenberghe et al., 1998):

$$\begin{aligned} &\min_{E \in S^n_{++},\, x_c \in \mathbb{R}^n} && -\log\det E \\ &\text{subject to} && \sup_{|y|\le 1} I_{\mathcal{K}}(Ey + x_c) \le 0, \end{aligned}$$

where $I_{\mathcal{K}}(x) = \infty$ if $x \notin \mathcal{K}$, $I_{\mathcal{K}}(x) = 0$ if $x \in \mathcal{K}$, and we note that $\log\det(\cdot)$ is concave on S^n_{++}. In the special case of a polytope $\mathcal{P} = \left\{x \in \mathbb{R}^n \;\middle|\; Ax \le 1\right\}$, letting $\{a_i\}_{i=1}^m$ denote the rows of A, note that $\sup_{|y|\le 1} \langle a_i,\, Ey + x_c\rangle \le 1$ if and only if $|Ea_i| + \langle a_i,\, x_c\rangle \le 1$. Thus, the maximum volume ellipsoid is found as the solution of the optimization problem,

$$\begin{aligned} &\min_{E \in S^n_{++},\, x_c \in \mathbb{R}^n} && -\log\det E \\ &\text{s.t.} && |Ea_i| + \langle a_i,\, x_c\rangle \le 1, \quad i = 1, \dots, m. \end{aligned} \tag{5.1}$$

We address the issue of computing approximate John's ellipsoids for polytopes using convex programming in section 5.4.

5.2.2 The John's Walk Algorithm

We state the algorithm for a general convex body $\mathcal{K}$. At a given point $x \in \mathcal{K}$, let the symmetrization of $x \in \mathcal{K}$ be

$$\mathcal{K}^s_x := \mathcal{K} \cap \left\{2x - y \;\middle|\; y \in \mathcal{K}\right\},$$

and let $\mathcal{E}_x = \left\{E_x u + x \;\middle|\; |u| \le 1\right\}$ denote the John's ellipsoid of $\mathcal{K}^s_x$. Similarly, let the rescaled John's ellipsoid be $\mathcal{E}_x(r) = \left\{r(E_x u) + x \;\middle|\; |u| \le 1\right\}$, where the radius $r > 0$ will be specified in section 5.3. Assume $0 = x_0 \in \operatorname{int}(\mathcal{K})$, and we have computed $\mathcal{E}_{x_0}$. To generate a sample x_i given x_{i-1}, we use algorithm 5.2.1, where $\lambda(\cdot)$ denotes the Lebesgue measure on $\mathbb{R}^n$:

Algorithm 5.2.1 is a Metropolis-Hastings geometric random walk which uses the uniform measure $Q_x(\cdot)$ on the dilated John's ellipsoid $\mathcal{E}_x(r)$ as the proposal distribution. Tossing a fair coin ensures the transition probability kernel defined by the algorithm is positive definite, which is known as making the walk lazy. Lazy random walks have the same stationary distribution as the original walk at the cost of a constant increase in mixing time (we will

Given $x \in \mathcal{K}_x^s$, $r > 0$, and $\mathcal{E}_x$, generate the next step y as follows:

1. Toss a fair coin. If the result is heads, let $y = x$.

2. If the result is tails:

   (a) Draw a uniformly distributed random point z from $\mathcal{E}_x(r)$.

   (b) Compute $\mathcal{E}_z$ using $\mathcal{K}_z^s$.

   (c) If $x \notin \mathcal{E}_z(r)$, let $y = x$. Otherwise, let

$$y = \begin{cases} z, & \text{with probability } \min\left(1, \frac{\lambda(\mathcal{E}_x(r))}{\lambda(\mathcal{E}_z(r))}\right) = \min\left(1, \frac{\det E_x}{\det E_z}\right), \\ x, & \text{else.} \end{cases}$$

Algorithm 5.2.1: John's Walk Step

analyze the non-lazy walk, noting that the mixing time is not affected in terms of complexity as a function of m and n). The rejection of any sample z such that $x \notin \mathcal{E}_z(r)$ is necessary to ensure the random walk is reversible.

The uniform measure on the John's ellipsoid $\mathcal{E}_x(r)$ is absolutely continuous with respect to the Lebesgue measure λ, and thus the Radon-Nikodym derivative (i.e., density) for the proposal distribution is

$$\begin{aligned} q_x(y) &:= \frac{dQ_x}{d\lambda}(y) \\ &= \left(\frac{1}{\lambda(\mathcal{E}_x(r))}\right) \cdot 1_{\{y \in \mathcal{E}_x(r)\}}. \end{aligned} \tag{5.2}$$

The acceptance probability corresponding to the uniform stationary measure in the Metropolis filter is

$$\alpha_x(y) = \min\left[1, \frac{\lambda(\mathcal{E}_x(r))}{\lambda(\mathcal{E}_y(r))}\right].$$

By the Lebesgue decomposition, the transition probability measure $P_x(\cdot)$ of the non-lazy version of algorithm 5.2.1 is absolutely continuous with respect to the measure

$$\mu := \lambda + \delta_x, \tag{5.3}$$

where $\delta_x(\cdot)$ is the Dirac measure at x corresponding to a rejected move. The transition density is thus

$$\begin{aligned} p_x(y) &:= \frac{dP_x}{d\mu}(y) \\ &= \alpha_x(y) q_x(y) 1_{\{y \neq x,\, x \in \mathcal{E}_y(r)\}} + \rho(x) 1_{\{y = x\}} \\ &= \min\left[\frac{1}{\lambda(\mathcal{E}_x(r))}, \frac{1}{\lambda(\mathcal{E}_y(r))}\right] 1_{\{y \neq x,\, x \in \mathcal{E}_y(r),\, y \in \mathcal{E}_x(r)\}} + \rho(x) 1_{\{y = x\}}, \end{aligned}$$

where $1_{\{\cdot\}}$ is the indicator function and the rejection probability is denoted $\rho(x)$. We next analyze the mixing time of the walk.

5.3 Analysis of Mixing Time

In what follows we let a discrete-time, homogeneous Markov chain be the triple $\{\mathcal{K}, \mathcal{A}, P_x\}$ along with a distribution π_0 for the starting point, where the sample space is the convex body $\mathcal{K} \subset \mathbb{R}^n$, the measurable sets on $\mathcal{K}$ are denoted by $\mathcal{A}$, and P_x denotes the uniform measure on the scaled John's ellipsoid of radius of $\mathcal{K}_x^s$.

5.3.1 Conductance and Mixing Times

We use the approach from Lovász and Simonovits (1993) of lower-bounding the conductance of the chain to prove mixing times. We refer the reader to section 3.4 to review the relationship between convergence and conductance.

5.3.2 Isoperimetry

The typical means by which one finds lower bounds on the conductance is via isoperimetric inequalites. We first restate the cross-ratio used in isoperimetric inequality we will employ.

Definition 5.3.1 (Cross-Ratio)**.** *Let $x, y \in \mathcal{K}$, and let p, q be the end points of a chord in $\mathcal{K}$ passing through x, y where the cross-ratio is defined to be*

$$\sigma(x, y) = \frac{|x - y||p - q|}{|p - x||y - q|},$$

where $|\cdot|$ denotes the Euclidean norm.

Additionally, for any $S_1, S_2 \subset \mathcal{K}$, let

$$\sigma(S_1, S_2) = \inf_{x \in S_1, y \in S_2} \sigma(x, y).$$

Lovász (1999) proved an isoperimetric inequality involving the cross-ratio from which the conductance ϕ may be lower-bounded for the special case of the uniform distribution on a convex body $\mathcal{K} \subset \mathbb{R}^n$. It was extended to log-concave measures by Lovász and Vempala (2007) for which the uniform measure on convex body is a special case. We state the latter result as follows.

Theorem 5.3.2 (Lovász and Vempala)**.** *For any log-concave measure $\pi(\cdot)$ supported $\mathcal{K}$ and a partition of $\mathcal{K}$ into measurable subsets S_1, S_2, and $S_3 = \mathcal{K} \setminus (S_1 \cup S_2)$, we have*

$$\pi(S_3) \geq \sigma(S_1, S_2)\pi(S_1)\pi(S_2).$$

5.3.3 Mixing of John's Walk

The key step in proving conductance lower bounds is to show that if two points are close in geometric distance, then they are close in statistical distance. Note that given John's ellipsoid $\mathcal{E}_x = \left\{E_x u + x \;\middle|\; |u| \leq 1\right\}$, a local norm is induced via

$$\|y - x\|_x^2 = (y - x)^\top E_x^{-2}(y - x).$$

We first relate this local norm to the cross-ratio as follows.

Theorem 5.3.3. *Let $\|\cdot\|_x$ denote the norm induced by the John's ellipsoid of $\mathcal{K}_x^s$. Then*

$$\sigma(x, y) \geq \frac{1}{\sqrt{n}}\|y - x\|_x.$$

Proof. Noting that the cross-ratio is invariant to affine transformations, without loss of generality we may assume by a suitable affine transformation that the John's ellipsoid of $\mathcal{K}_x^s$ is the unit ball, and thus $\|y-x\|_x = |y-x|$. Let p, x, y, q denote successive points on a chord through $\mathcal{K}_x^s$. Then

$$\begin{aligned}
\sigma(x,y) &= \frac{|x-y||p-q|}{|p-x||y-q|} \\
&\geq \frac{|x-y|\left(|p-x|+|y-q|\right)}{|p-x||y-q|} \\
&\geq \max\left(\frac{|x-y|}{|y-q|}, \frac{|x-y|}{|p-x|}\right) \\
&\geq \frac{|x-y|}{\sqrt{n}},
\end{aligned}$$

where the last inequality follows from the containment in equation (2.7). □

Before bounding the statistical distance between P_x and P_y given a bound on the geometric distance between x and y, we first state some useful lemmas regarding the ellipsoids $\mathcal{E}_x$ and $\mathcal{E}_y$. The next lemma is a generalization of the Cauchy-Schwarz inequality to semidefinite matrices.

Lemma 5.3.4 (Semidefinite Cauchy-Schwarz)**.** *Let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and let $A_1, \ldots, A_m \in \mathbb{R}^{r\times n}$. Then*

$$\left(\sum_{i=1}^m \alpha_i A_i\right)\left(\sum_{i=1}^m \alpha_i A_i\right)^\top \preceq \left(\sum_i \alpha_i^2\right)\left(\sum_{i=1}^m A_i A_i^\top\right),$$

where $A \preceq B$ signifies that $B - A$ is positive semidefinite.

Proof. The proof is as in lemma 3.11 in (Kannan and Narayanan, 2012), which we repeat here for convenience. For all i and j,

$$(\alpha_j A_i - \alpha_i A_j)(\alpha_j A_i - \alpha_i A_j)^\top \succeq 0.$$

Thus

$$
\begin{aligned}
0 \preceq \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_j A_i - \alpha_i A_j)(\alpha_j A_i - \alpha_i A_j)^\top \\
= \frac{1}{2} \sum_{i=1}^{m} \Bigg[\left(\sum_{i=1}^{m} \alpha_j^2 \right) A_i A_i^\top - \alpha_i A_i \sum_{j=1}^{m} (\alpha_j A_j^\top) \\
- \left(\sum_{j=1}^{m} (\alpha_j A_j) \right) (\alpha_i A_i^\top) + \alpha_i^2 \sum_{i=1}^{m} A_j A_j^\top \Bigg] \\
= \left(\sum_i \alpha_i^2 \right) \left(\sum_{i=1}^{m} A_i A_i^\top \right) - \left(\sum_{i=1}^{m} \alpha_i A_i \right) \left(\sum_{i=1}^{m} \alpha_i A_i \right)^\top .
\end{aligned}
$$

□

Now we study how the volume and aspect ratio of the John's ellipsoid changes from a move from x to y. If the John's ellipsoid centered at $x = 0$ is the unit ball, and we make a move to y, the matrix E_y which induces $\mathcal{E}_y$ satisfies the solution of

$$
\begin{aligned}
\min_{E \in S_{++}^n} \quad & -\log \det E \\
\text{s.t.} \quad & |Eu_i| + \langle u_i, y \rangle \leq 1, \quad i = 1, \dots, m,
\end{aligned}
\tag{5.4}
$$

where $m = n(n+1)/2$ and the u_i's are the contact points of $\mathcal{K}$ (i.e, those not induced by the symmetrization). Using the constraint of (5.4), theorem 2.3.1, and lemma 5.3.4, we deduce an upper bound on $\det E_y$ as follows.

Lemma 5.3.5. *Let $c > 0$ be some universal constant, let $r = cn^{-5/2}$, and presume y is chosen from a ball of radius r such that $\|y - x\|_x = |y - x| \leq r$. Then*

$$\det E_y \leq 1 + cn^{-2}.$$

Proof. By (5.4), E_y is the maximum of $\log \det E$ of under the constraints $|Eu_i| \leq 1 - u_i^\top y$. Since the weights c_i corresponding to the John's ellipsoid $\mathcal{K}_x^s$ are positive, the constraint implies that

$$\sum_i c_i u_i^\top E_y^2 u_i \leq \sum_i c_i (1 - u_i^\top y)^2.$$

By (2.5), (2.6), and using the linearity and cyclic invariance of the trace, we have

$$\begin{aligned}\operatorname{tr}(E_y^2) &\leq \sum_i c_i - 2\sum_i c_i u_i^\top y + \sum_i c_i (u_i^\top y)^2 \\ &= n - 2\left(\sum_i c_i u_i\right)^\top y + y^\top \left(\sum_i c_i u_i u_i^\top\right) y \\ &= n - 2\left(\sum_i c_i u_i\right)^\top y + |y|^2.\end{aligned}$$

Considering $|\sum_i c_i u_i|$ to bound the middle term, we may employ lemma 5.3.4. Letting $\alpha_i = \sqrt{c_i}$ and $A_i = \sqrt{c_i} u_i$, we have

$$\left(\sum_i c_i u_i\right)\left(\sum_i c_i u_i\right)^\top \preceq \left(\sum_i c_i\right)\left(\sum_i c_i u_i u_i^\top\right)$$

Noting the right side is equal to nI_n, it follows that

$$\left|\sum_i c_i u_i\right| \leq \sqrt{n}.$$

Therefore, if y is chosen from a ball of radius $cn^{-5/2}$, by Cauchy-Schwarz we conclude that

$$\operatorname{tr}(E_y^2) \leq n + cn^{-2},$$

where we have absorbed the $|y|^2 \leq c^2 n^{-5}$ term into the constant. Now letting the eigenvalues of E_y be denoted $d_i > 0$, we have by the arithmetic-geometric mean inequality,

$$\begin{aligned}(\det E_y)^{2/n} &= \left(\prod_{i=1}^n d_i^2\right)^{1/n} \\ &\leq \frac{1}{n}\sum_{i=1}^n d_i^2 \\ &\leq \frac{1}{n}\left(n + cn^{-2}\right),\end{aligned}$$

Thus,

$$\begin{aligned}
\det E_y &\leq (1+cn^{-3})^{n/2} \\
&\leq (1+cn^{-3})^{n} \\
&= 1 + cn^{-2} + \sum_{k=2}^{n} \binom{n}{k} (cn^{-3})^k \\
&\leq 1 + cn^{-2} + \sum_{k=2}^{n} \left(\frac{ne}{k}\right)^k (cn^{-3})^k \\
&\leq 1 + cn^{-2} + \sum_{k=2}^{\infty} (ecn^{-2})^k \\
&= 1 + cn^{-2} + (ecn^{-2})^2 \left(\frac{1}{1-ecn^{-2}}\right) \\
&= 1 + cn^{-2}\left(1 + \frac{ce^2}{n^2 - ce}\right).
\end{aligned}$$

Absorbing terms into the constant, the claim holds.

□

We deduce a lower-bound on $\det E_y$ by considering a positive definite matrix of the form $E = \beta(I - \alpha y y^\top)$ that is feasible for (5.4). Note that such a matrix has eigenvalue $\beta(1-\alpha|y|^2)$ of multiplicity 1 corresponding to unit eigenvector $y/|y|$, and eigenvalues β of multiplicity $n-1$ corresponding to any unit vector z which is orthogonal to y.

Lemma 5.3.6. *The matrix $\beta(I_n - \alpha y y^\top)$ is feasible for (5.4) with $\beta = \left(1 - \frac{|y|}{\sqrt{n}}\right)$ and $\alpha = \frac{2\sqrt{n}}{|y|}$.*

Proof. We divide the contact points u_i which do not arise from the symmetrization into two sets: $A = \left\{ i \;\middle|\; \langle u_i, y\rangle \leq \frac{|y|}{\sqrt{n}} \right\}$ and $B = \left\{ i \;\middle|\; \langle u_i, y\rangle > \frac{|y|}{\sqrt{n}} \right\}$. If $i \in B$, we have

$$1 - \langle u_i, y\rangle > 1 - \frac{|y|}{\sqrt{n}},$$

and noting that $E \succ 0$ and $0 \prec (I_n - \alpha y y^\top) \prec I_n$ for $\beta > 0$ and $0 < \alpha < |y|^2$,

$$|Eu_i| \leq \beta |I_n - \alpha y y^\top| \leq \beta.$$

Thus it suffices to choose $\beta = (1 - \frac{|y|}{\sqrt{n}})$ for feasible E if $i \in B$. If $i \in A$, we have

$$\begin{aligned}
|Eu_i|^2 &= \beta^2 u_i^\top (I - \alpha y y^\top)^2 u_i \\
&\leq u_i^\top (I - \alpha y y^\top)^2 u_i \\
&= 1 - \alpha \langle u_i, y \rangle^2 \\
&\leq \left(1 - \frac{\alpha}{2} \langle u_i, y \rangle\right)^2 \\
&\leq \left(1 - \frac{\alpha |y|}{2\sqrt{n}}\right)^2,
\end{aligned}$$

where the second inequality follows from $I - \alpha y y^\top \succ (I - \alpha y y^\top)^2$. Thus the choice of $\alpha = \frac{2\sqrt{n}}{|y|}$ guarantees E is feasible. □

Lemma 5.3.7. *Let c be some universal constant, let $r = cn^{-5/2}$, and presume y is chosen from a ball of radius r such that $\|y - x\|_x = |y - x| \leq r$. Then*

$$\det E_y \geq 1 - cn^{-2}.$$

Proof. Considering the matrix E as provided by lemma 5.3.6, E_y satisfies

$$\begin{aligned}
\det E_y &\geq \det E \\
&= \beta^n (1 - \alpha |y|^2) \\
&= \left(1 - \frac{|y|}{\sqrt{n}}\right)^n (1 - 2\sqrt{n}|y|).
\end{aligned}$$

Considering $(1 - \frac{|y|}{\sqrt{n}})^n$, noting that $|y| \leq cn^{-5/2}$, we have

$$\begin{aligned}
\log \left(1 - \frac{|y|}{\sqrt{n}}\right)^n &\geq n \log(1 - cn^{-3}) \\
&= -n \left(\sum_{k=1}^{\infty} \frac{(cn^{-3})^k}{k}\right) \\
&\geq -\sum_{k=1}^{\infty} \frac{(cn^{-2})^k}{k} \\
&= \log(1 - cn^{-2}).
\end{aligned}$$

Thus

$$\begin{aligned}\det E_y &\geq (1 - cn^{-2})(1 - 2cn^{-2}) \\ &\geq (1 - 3cn^{-2}),\end{aligned}$$

and the claim follows by absorbing 3 into the constant. □

Lemmas 5.3.5 and 5.3.7 establish that for some universal constant $c > 0$ and $|y - x| \leq cn^{-5/2}$, the volume ratio of $\mathcal{E}_y$ and $\mathcal{B}_x$ satisfies

$$1 - cn^{-2} \leq \frac{\lambda(\mathcal{E}_y)}{\lambda(\mathcal{B}_x)} \leq 1 + cn^{-2}, \tag{5.5}$$

This does not necessarily indicate that the John's ellipsoid $\mathcal{E}_y$ does not lie close to some proper subspace of $\mathbb{R}^n$, a property we require so rejection does not occur too frequently. The following lemma guarantees that this is indeed not the case.

Lemma 5.3.8. *Let $d_1 \geq d_2 \geq \ldots \geq d_n > 0$ denote the eigenvalues of E_y. Then the minimum eigenvalue d_n satisfies*

$$d_n \geq 1 - cn^{-1}$$

for large enough n.

Proof. Presume the eigenvalues are ordered such that $d_1 \geq \ldots \geq d_n$. We have

$$\det E_y = \prod_{i=1}^{n} d_i \geq 1 - cn^{-2}.$$

By the power mean inequality and $\operatorname{tr}(E_y^2) \leq n + c/n^2$, it follows that

$$\begin{aligned}\frac{1}{n}\operatorname{tr}(E_y) &\leq \sqrt{\frac{1}{n}\operatorname{tr}(E_y^2)} \\ &= \sqrt{1 + cn^{-3}} \\ &\leq 1 + cn^{-3},\end{aligned}$$

so $\operatorname{tr}(E_y) \leq n + cn^{-2}$. By the arithmetic-geometric mean inequality, we thus have

$$\begin{aligned}
(1 - cn^{-2})/d_n &\leq \prod_{i<n} d_i \\
&\leq \left(\frac{\sum_{i<n} d_i}{n-1}\right)^{n-1} \\
&\leq \left(\frac{n + cn^{-2} - d_n}{n-1}\right)^{n-1} \\
&= \left(1 + \frac{1 + cn^{-2} - d_n}{n-1}\right)^{n-1}.
\end{aligned}$$

Since $(1 + z/m)^m \leq \exp(z)$ for $z \geq 0$ and integer $m \geq 1$,

$$(1 - c/n^2) \leq \exp(1 + cn^{-2} - d_n)d_n,$$

or equivalently,

$$-\left[(1 - d_n) + \log(1 - (1 - d_n))\right] \leq cn^{-2} - \log(1 - cn^{-2}).$$

The claim is true if $d_n \geq 1$, so we may assume that $d_n < 1$. Expanding terms, we have

$$\begin{aligned}
-(1 - d_n) - \log(1 + (1 - d_n)) &= \sum_{k=2}^{\infty} \frac{(1 - d_n)^k}{k} \\
&\geq \frac{(1 - d_n)^2}{2},
\end{aligned}$$

and

$$\begin{aligned}
cn^{-2} - \log(1 - cn^{-2}) &= 2cn^{-2} + \sum_{k=2}^{\infty} \frac{(cn^{-2})^k}{k} \\
&\leq \left(2cn^{-2} + (cn^{-2})^2 \left(\frac{1}{1 - cn^{-2}}\right)\right) \\
&= cn^{-2}\left(2 + \frac{c^2}{n^4 - cn^2}\right)
\end{aligned}$$

Thus by absorbing terms into the constant, $(1 - d_n)^2/2 \leq cn^{-2}$, which implies that $d_n \geq 1 - (\sqrt{2c})n^{-1}$. Absorbing terms again, the claim holds. □

Now to derive a lower bound on the conductance for John's walk, we first must bound the statistical distance between two points given a bound on their geometric distance with respect to the local norm. Again without loss of generality in what follows we may assume $x = 0$ and the John's ellipsoid centered at x is the unit ball $\mathcal{B}_x$ (otherwise perform an affine transformation such that this is the case). Let $x, y \in \mathcal{K}$ represent any two points in the body such that $\|y - x\|_x = |y - x| \le r$, where $r \in (0, 1)$ is a constant to be specified in terms of the dimension n. Let P_x and P_y denote the one-step transition probability measures defined at x and y, respectively. Let the uniform probability measures defined by the rescaled John's ellipsoids $\mathcal{E}_x(r)$ and $\mathcal{E}_y(r)$ be denoted Q_x and Q_y, respectively. We seek to bound

$$d_{TV}(P_x, P_y) \le d_{TV}(P_x, Q_x) + d_{TV}(Q_x, Q_y) + d_{TV}(Q_y, P_y) \tag{5.6}$$

by choosing r such that the right side of (5.6) is $1 - \Omega(1)$.

To bound $d_{TV}(Q_x, Q_y)$ in (5.6), letting $\widetilde{Q}_y$ denote the probability measure corresponding to the uniform distribution on a ball of radius r centered at y, we may alternatively bound

$$d_{TV}(Q_x, Q_y) \le d_{TV}(Q_x, \widetilde{Q}_y) + d_{TV}(\widetilde{Q}_y, Q_y). \tag{5.7}$$

We bound each term in (5.7) separately. To bound $d_{TV}(Q_x, \widetilde{Q}_y)$, note that by our assumption that $\mathcal{E}_x = \mathcal{B}_x$ (the unit ball at x), the corresponding densities with respect to the dominating Lebesgue measure λ are

$$q_x(z) = \left(\frac{1}{\lambda(\mathcal{B}_x(r))}\right) \cdot 1_{\{z \in \mathcal{B}_x(r)\}}$$

and

$$\widetilde{q}_y(z) = \left(\frac{1}{\lambda(\mathcal{B}_y(r))}\right) \cdot 1_{\{z \in \mathcal{B}_y(r)\}}.$$

Thus using (3.4) and noting $\lambda(\mathcal{B}_x(r)) = \lambda(\mathcal{B}_y(r))$, we have

$$\begin{aligned} d_{TV}(Q_x, \widetilde{Q}_y) &= 1 - \int_{\mathcal{B}_x(r) \cap \mathcal{B}_y(r)} q_x(z)\, d\lambda(z) \\ &= 1 - \frac{\lambda\left(\mathcal{B}_x(r) \cap \mathcal{B}_y(r)\right)}{\lambda(\mathcal{B}_x(r))} \\ &= 1 - \frac{\lambda\left(\mathcal{B}_x \cap \mathcal{B}_y\right)}{\lambda(\mathcal{B}_x)}. \end{aligned} \tag{5.8}$$

The Lebesgue measure of $\mathcal{B}_x \cap \mathcal{B}_y$ is equal to twice the volume of a spherical cap. The following lemma regarding the volume of a hyperspherical cap from Lovász and Simonovits (1993) is useful.

Lemma 5.3.9. *Let $\mathcal{B}_x \subset \mathbb{R}^n$ be the Euclidean ball of unit radius centered at x. Let $\mathcal{H} \subset \mathbb{R}^n$ define a halfspace at a distance of at least t from x (so x is not contained in the halfspace). Then for $t \leq \frac{1}{\sqrt{n}}$, we have*

$$\lambda(\mathcal{H} \cap \mathcal{B}_x) \geq \frac{1}{2}\left(1 - t\sqrt{n}\right)\lambda(\mathcal{B}_x).$$

The following lemma results trivially from lemma 5.3.9 and (5.8).

Lemma 5.3.10. *Let $t \leq 1$. If $\|y - x\|_x = |y - x| \leq \frac{rt}{\sqrt{n}}$, then*

$$d_{TV}(Q_x, \widetilde{Q}_y) \leq t.$$

To bound $d_{TV}(\widetilde{Q}_y, Q_y)$, note that we are bounding the total variation distance between a density supported on a ball and a density supported on an ellipsoid with the same center. The following lemma provides the bound.

Lemma 5.3.11. *If $\|y - x\|_x \leq r = cn^{-5/2}$, the total variation distance between $\widetilde{Q}_y$ and Q_y satisfies*

$$d_{TV}(\widetilde{Q}_y, Q_y) \leq 1/4.$$

Proof. Note that by (3.4), we have

$$\begin{aligned}
d_{TV}(\widetilde{Q}_y, Q_y) &= 1 - \mathbb{E}_{\widetilde{Q}_y}\left[\min\left(1, \frac{\lambda(\mathcal{B}_y(r))}{\lambda(\mathcal{E}_y(r))}\right)\right] \\
&= 1 - \min\left[1, \frac{\lambda(\mathcal{B}_y)}{\lambda(\mathcal{E}_y)}\right]\mathbb{P}_{\widetilde{Q}_y}(Z \in \mathcal{E}_y(r)) \\
&\leq 1 - \min\left[1, \frac{\lambda(\mathcal{B}_y)}{\lambda(\mathcal{E}_y)}\right]\mathbb{P}_{\widetilde{Q}_y}(Z \in \mathcal{E}_y(r)|B)\mathbb{P}_{\widetilde{Q}_y}(Z \in B),
\end{aligned}$$

where B denotes the event in which $Z \in (1 - \frac{c}{n}) \cdot \mathcal{B}_y(r)$. By lemma 5.3.8, it follows that $\mathbb{P}_{\widetilde{Q}_y}(Z \in \mathcal{E}_y(r)|B) = 1$ since the smallest eigenvalue of E_y is at least $1 - cn^{-1}$. Additionally by (5.5),

$$\begin{aligned}
\min\left[1, \frac{\lambda(\mathcal{B}_y)}{\lambda(\mathcal{E}_y)}\right] &\geq \frac{1}{1 + cn^{-2}} \\
&\geq \exp(-cn^{-2}).
\end{aligned}$$

Now noting that $(1-\frac{x}{2}) \geq e^{-x}$ for $x \in [0,1]$, we have $\mathbb{P}_{\widetilde{Q}_y}(Z \in B) = \left(1-\frac{c}{n}\right)^n \geq e^{-2c}$, and

$$d_{TV}(\widetilde{Q}_y, Q_y) \leq 1 - \exp\left(-c(2+n^{-2})\right).$$

The claim holds choosing c large enough relative to n. $\square$

To bound $d_{TV}(P_x, Q_x)$, we provide the following lemma.

Lemma 5.3.12. *If $\|y-x\|_x \leq r = cn^{-5/2}$, the total variation distance between P_x and Q_x satisfies*

$$d_{TV}(P_x, Q_x) \leq 1/4.$$

Proof. With some abuse of notation with regards to (5.2), temporarily let the density of Q_x with respect to the dominating measure μ as defined by (5.3) be

$$q_x(y) := \frac{dQ}{d\mu}(y) = \left(\frac{1}{\lambda(\mathcal{E}_x(r))}\right) \cdot 1_{\{y \in \mathcal{E}_x(r),\, y \neq x\}}.$$

Then since $q_x(x) = 0$ and $p_x(y) \leq q_x(y)$ for $y \neq x$, by (3.4) we have

$$\begin{aligned}
d_{TV}(P_x, Q_x) &= 1 - \int_{\left\{y \,\middle|\, q_x(y) \geq p_x(y)\right\}} \min\left[q_x(y),\, p_x(y)\right] d\mu(y) \\
&= 1 - \int_{\mathcal{K}\setminus\{x\}} \min\left[q_x(y),\, p_x(y)\right] d\lambda(y) \\
&= 1 - \int_{\mathcal{E}_x(r)\cap\setminus\{x\}} \min\left[\frac{1}{\lambda(\mathcal{E}_x(r))},\, \frac{1}{\lambda(\mathcal{E}_y(r))}\right] \cdot 1_{\{x \in \mathcal{E}_y(r)\}}\, d\lambda(y) \qquad , \\
&= 1 - \int_{\mathcal{E}_x(r)} \left[\min\left(1,\, \frac{\lambda(\mathcal{E}_x(r))}{\lambda(\mathcal{E}_y(r))}\right)\right] \cdot 1_{\{x \in \mathcal{E}_y(r)\}} \cdot \left(\frac{1}{\lambda(\mathcal{E}_x(r))}\right) d\lambda(y) \\
&= 1 - \mathbb{E}_{Q_x}\left[\min\left(1,\, \frac{\lambda(\mathcal{E}_x)}{\lambda(\mathcal{E}_Y)}\right) \cdot 1_{\{Y \in A\}}\right].
\end{aligned}$$

where we let A denote the "accept" event in which $x \in \mathcal{E}_Y(r)$. Since $\mathcal{E}_x = \mathcal{B}_x$, as in the proof to lemma 5.3.11, we have for all $Y \in A$

$$\min\left[1,\, \frac{\lambda(\mathcal{E}_x)}{\lambda(\mathcal{E}_Y)}\right] \geq \frac{1}{1+cn^{-2}},$$

Therefore,

$$\begin{aligned} d_{TV}(P_x, Q_x) &\leq 1 - \left(\frac{1}{1+cn^{-2}}\right)\mathbb{P}_{Q_x}(Y \in A) \\ &\leq 1 - \left(\frac{1}{1+cn^{-2}}\right)\mathbb{P}_{Q_x}(Y \in A | Y \in B)\mathbb{P}_{Q_x}(Y \in B), \end{aligned}$$

where B is the event in which $Y \in (1 - \frac{c}{n}) \cdot \mathcal{E}_x(r) = r(1 - \frac{c}{n}) \cdot \mathcal{B}_x$. Again by lemma 5.3.8, $\mathbb{P}_{Q_x}(Y \in A | Y \in B) = 1$. The remainder of the proof is as in lemma 5.3.11. $\square$

Note that by a similar argument, $d_{TV}(Q_y, P_y) \leq 1/4$ for some universal $c > 0$ as well. Combining this with lemmas 5.3.10, 5.3.11, and 5.3.12, the following theorem results.

Theorem 5.3.13. *If $\|y - x\|_x \leq \frac{rt}{\sqrt{n}} = ctn^{-3}$ for some universal constant $c > 0$ and some $t \leq 1$, the total variation distance between P_x and P_y satisfies*

$$d_{TV}(P_x, P_y) \leq 3/4 + t = 1 - \epsilon.$$

In particular, we may choose $t = 1/8$ so $\epsilon = 1/8$.

We finally arrive at a lower bound on the conductance for John's Walk using theorems 5.3.2, 5.3.3, and 5.3.13. The proof of the next result is similar to corollary 10 and theorem 11 in (Lovász, 1999).

Theorem 5.3.14 (Conductance Lower Bound)**.** *Consider the partition $\mathcal{K} = S_1 \cup S_2$ where $S_1, S_2 \in \mathcal{A}$, and let π be the uniform measure on $\mathcal{K}$, i.e.,*

$$\pi(A) = \frac{\lambda(A)}{\lambda(\mathcal{K})} \quad \text{for all } A \in \mathcal{A}.$$

Then for large enough n and $t = 1/8$, we have

$$\int_{S_1} P_x(S_2) d\pi(x) \geq \left(\frac{c}{512 n^{7/2}}\right) \min(\pi(S_1), \pi(S_2)),$$

so $\phi = \Omega(n^{-7/2})$.

Proof. Note that the Radon-Nikodym derivative of P_x with respect to the Lebesgue measure λ is is well-defined for all $y \in \mathcal{K} \setminus \{x\}$, and is given as

$$\frac{dP_x}{d\lambda}(y) = \min\left[\frac{1}{\lambda(\mathcal{E}_x(r))}, \frac{1}{\lambda(\mathcal{E}_y(r))}\right] 1_{\{x \in \mathcal{E}_y(r),\, y \in \mathcal{E}_x(r)\}}.$$

Let

$$\rho(x) := \frac{d\pi}{d\lambda}(x) = \frac{1}{\lambda(\mathcal{K})} \cdot 1_{\{x \in \mathcal{K}\}}$$

be the density for π. Then for any $x, y \in \mathcal{K}$ such that $y \neq x$, we have

$$\rho(x)\frac{dP_x}{d\lambda}(y) = \rho(y)\frac{dP_y}{d\lambda}(x),$$

from which it follows that π is the stationary measure for the chain.

Now consider points far inside S_1 that are unlikely to cross over to S_2. Letting $t = 1/8$ so $\epsilon = 1/8$ as in theorem 5.3.13, we define

$$S_1' := S_1 \cap \left\{x \mid \rho(x)P_x(S_2) < \frac{\epsilon}{2\lambda(\mathcal{K})}\right\}.$$

Similarly, let

$$S_2' := S_2 \cap \left\{y \mid \rho(x)P_y(S_1) < \frac{\epsilon}{2\lambda(\mathcal{K})}\right\}.$$

Since $\rho(x)P_x(S_2) \geq \epsilon/(2\lambda(\mathcal{K}))$ for $x \in S_1 \setminus S_1'$, we have

$$\begin{aligned}\int_{S_1} P_x(S_2)d\pi(x) &\geq \int_{S_1 \setminus S_1'} \rho(x)P_x(S_2)d\lambda(x) \\ &\geq \frac{\epsilon\lambda(S_1 \setminus S_1')}{2\lambda(\mathcal{K})} \\ &= (\epsilon/2)\pi(S_1 \setminus S_1').\end{aligned}$$

Similarly for $y \in S_2 \setminus S_2'$, we have

$$\int_{S_2} P_y(S_1)d\pi(x) \geq (\epsilon/2)\pi(S_2 \setminus S_2').$$

By the reversibility of the chain, we have

$$\int_{S_1} P_x(S_2)d\pi(x) = \int_{S_2} P_y(S_1)d\pi(y), \tag{5.9}$$

so it follows that

$$\begin{aligned}\int_{S_1} P_x(S_2)d\pi(x) &= \frac{1}{2}\int_{S_1} P_x(S_2)d\pi(x) + \frac{1}{2}\int_{S_2} P_x(S_1)d\pi(x)\\ &\geq \frac{\epsilon}{4}\left(\pi(S_1 \setminus S_1') + \pi(S_2 \setminus S_2')\right)\\ &= (\epsilon/4)\pi(\mathcal{K} \setminus (S_1' \cup S_2')).\end{aligned}$$

Now let $\delta = ctn^{-3}$. Assuming that $\pi(S_1') \leq (1-\delta)\pi(S_1)$, we have $\pi(S_1 \setminus S_1') = \pi(S_1) - \pi(S_1') \geq \delta\pi(S_1)$, and thus

$$\begin{aligned}\int_{S_1} P_x(S_2)d\pi(x) \geq \epsilon\delta\pi(S_1) &\geq (\epsilon\delta/2)\min(\pi(S_1), \pi(S_2))\\ &= \left(\frac{c}{128n^3}\right)\min(\pi(S_1), \pi(S_2)),\end{aligned}$$

which proves the claim. Similarly if $\pi(S_2') \leq (1-\delta)\pi(S_2)$, the claim is proved again using (5.9). Thus assume that $\pi(S_1') > (1-\delta)\pi(S_1)$ and $\pi(S_2') > (1-\delta)\pi(S_2)$. By theorem (5.3.2), we have

$$\pi(\mathcal{K} \setminus (S_1' \cup S_2')) \geq \sigma(S_1', S_2')\pi(S_1')\pi(S_2').$$

Now given $x \in S_1'$ and $y \in S_2'$, the total variation between P_x and P_y satisfies

$$\begin{aligned}d_{TV}(P_x, P_y) &\geq P_x(S_1) - P_y(S_1)\\ &= 1 - P_x(S_2) - P_y(S_1)\\ &\geq 1 - P_x(S_2') - P_y(S_1')\\ &> 1 - \epsilon.\end{aligned}$$

By theorem 5.3.13, it follows that $\|y - x\|_x > \delta$. Then by theorem 5.3.3, it follows that

$$\sigma(S_1', S_2') \geq n^{-1/2}\|y - x\|_x > \delta n^{-1/2}.$$

Finally, we deduce that

$$\begin{aligned}\int_{S_1} P_x(S_2)d\pi(x) &\geq (\epsilon\delta n^{-1/2}/4)\pi(S_1')\pi(S_2')\\ &\geq \left(\frac{\epsilon\delta(1-\delta)^2}{4\sqrt{n}}\right)\pi(S_1)\pi(S_2)\\ &\geq \left(\frac{\epsilon\delta(1-\delta)^2}{8\sqrt{n}}\right)\min(\pi(S_1),\pi(S_2))\\ &= \left(\frac{c(1-\delta)^2}{512n^{7/2}}\right)\min(\pi(S_1),\pi(S_2)).\end{aligned}$$

The claim follows by absorbing terms into the constant for large enough n. □

5.4 Approximate John's Ellipsoids

We describe a cutting plane algorithm for computing approximate John's ellipsoids for the current symmetrization of a polytope $\mathcal{P} = \left\{z \in \mathbb{R}^n \;\middle|\; Az \leq 1\right\}$, and show the mixing time is preserved for such ellipsoids. While algorithms such as that of (Anstreicher, 2002) requires $\widetilde{O}(m^{3.5})$ arithmetic operations to find an approximate John's ellipsoid, our algorithm retains linear complexity in m and thus is appropriate for the $m \gg n$ case. To find an approximate John's ellipsoid while keeping the computational complexity linear in m, we may use the volumetric cutting plane method of (Vaidya, 1996), and remark that its complexity has recently been improved in the algorithm by Lee et al. (2015). For an efficient implementation of Vaidya's algorithm, we refer the reader to (Anstreicher, 1997), and make use of a slightly modified version of that algorithm here.

We refer the reader to review Vaidya's algorithm in section 2.4.3, and note that we will be applying the method with here with dimension $d := (n(n-1))/2$. Given that the current location of the random walk is some x in the interior of $\mathcal{P}$, the current symmetrization of the polytope is

$$\begin{aligned}\mathcal{P}_x^s &= \mathcal{P} \cap \left\{2x - y \;\middle|\; y \in \mathcal{P}\right\}\\ &= \left\{z \;\middle|\; Az \leq 1\right\} \cap \left\{2x - y \;\middle|\; Ay \leq 1\right\}\\ &= x + \left\{y \in \mathbb{R}^n \;\middle|\; \|Ay\|_\infty \leq 1 - Ax\right\}.\end{aligned}$$

For any x in the interior of $\mathcal{P}$, we have $0 < 1 - Ax$. With $b = 1 - Ax$ and denoting the rows of A as $\{a_i^\top\}_{i=1}^m$, let

$$\tilde{a}_i^\top = \begin{cases} a_i^\top / b_i & i = 1, \ldots, m, \\ -a_{i-m}^\top / b_{i-m} & i = m+1, \ldots, 2m. \end{cases}$$

This defines a matrix $\tilde{A} \in \mathbb{R}^{2m \times n}$ which may be computed in $O(mn)$ arithmetic operations. We thus seek the maximum volume ellipsoid centered at the origin contained in

$$\widetilde{\mathcal{P}} = \left\{ y \in \mathbb{R}^n \;\middle|\; \tilde{A} y \leq 1 \right\},$$

and we note that the analytic center of this polytope is the origin given the symmetrization.

To keep ρ and L suitably small, it will be useful to make a further linear transformation to $\widetilde{\mathcal{P}}$ before using Vaidya's algorithm. We review some properties of Dikin ellipsoids which will enable us to do so. For any z in the interior of $\widetilde{\mathcal{P}}$, the Dikin ellipsoid centered at z of radius r is defined by

$$\mathcal{D}_z(r) = \left\{ y \in \mathbb{R}^n \;\middle|\; (y - z)^\top H(z)(y - z) \leq r^2 \right\},$$

where $H(z)$ is the Hessian of the log-barrier at z, i.e.,

$$H(z) = \sum_{i=1}^{2m} \frac{\tilde{a}_i \tilde{a}_i^\top}{(1 - \tilde{a}_i^\top z)^2} = \tilde{A}^\top S_z^{-2} \tilde{A},$$

where $s_i(z) = 1 - \tilde{a}_i^\top z$ and $S_z = \mathrm{diag}(s_i(z))$. Dikin ellipsoids are affine invariants in that if $\mathcal{D}_z(r)$ is the Dikin ellipsoid for the polytope $\widetilde{\mathcal{P}}$ centered at some z in the interior of $\widetilde{\mathcal{P}}$ and $T(\cdot)$ is an affine transformation, the Dikin ellipsoid of radius r centered at the point $T(z)$ for the polytope $T(\widetilde{\mathcal{P}})$ is $T(D_z(r))$.

Let $\mathcal{D}_z$ denote the Dikin ellipsoid of radius 1 centered at z. We have

$$\mathcal{D}_z \subset \widetilde{\mathcal{P}},$$
$$\mathcal{D}_0 \subset \widetilde{\mathcal{P}} \subset \sqrt{2m}\mathcal{D}_0,$$

for any z in the interior of $\widetilde{\mathcal{P}}$, and $z = 0$, respectively. Here $\mathcal{D}_0 = \left\{y \;\middle|\; y^\top H y \leq 1\right\}$ where $H := H(0) = \tilde{A}^\top \tilde{A}$. Letting $T(y) = H^{1/2} y$, we thus have $T(\mathcal{D}_0) = \mathcal{B}$. The resultant polytope is

$$T(\widetilde{\mathcal{P}}) = \left\{v \in \mathbb{R}^n \;\middle|\; \tilde{A} H^{-1/2} v \leq b\right\}.$$

The complexity of the transformation T is as follows. Computing $\tilde{A} S_0^{-1}$ requires $O(mn)$ arithmetic operations. By padding $\tilde{A}$ with zeros as necessary and then partitioning the resultant matrix into less than $(2m+n)/n$ submatrices of dimension $n \times n$, we may compute $H = \tilde{A}^\top \tilde{A}$ in $O(mn^{\omega-1})$ arithmetic operations. Computing $\tilde{A} H^{-1/2}$ given $H^{-1/2}$ similarly requires $O(mn^{\omega-1})$ arithmetic operations, and computing $H^{-1/2}$ costs $O(n^\omega)$ arithmetic operations (Demmel et al., 2007). Thus the linear transformation T costs $O(mn^{\omega-1})$ arithmetic operations since $m > n$.

With some abuse of notation, now let $\mathcal{P} = \left\{y \;\middle|\; Ay \leq 1\right\}$ where $A \in \mathbb{R}^{2m \times n}$ denote the centrally symmetric polytope for which $\mathcal{D}_0 = \mathcal{B}$, noting we can again rescale the rows of A in $O(mn)$ arithmetic operations to ensure $b = 1$. To find the maximum volume inscribed ellipsoid centered at $y = 0$, we may find the solution of the following optimization problem:

$$\begin{aligned} \min \quad & -\log\det X \\ \text{s.t.} \quad & \left\langle X,\, a_i a_i^\top \right\rangle \leq 1, \quad i = 1, \ldots, 2m, \\ & X \succeq I_n/n. \end{aligned} \tag{5.10}$$

Here $\langle A,\, B\rangle = \operatorname{tr}(A^\top B)$ is the inner product which induces the Frobenius norm, and the maximum volume ellipsoid in this parameterization is given by $\mathcal{E}_0 = \left\{y \;\middle|\; y^\top E^{-1} y \leq 1\right\}$ if $E \succ 0$ is optimal. Since we are optimizing over symmetric matrices, the cutting plane method to solve (5.10) has dimension $d = (n+1)n/2$. The constraint $X \succeq I_n/n$ arises from the nesting

$$\mathcal{D}_0 \subset \mathcal{P} \subset \sqrt{n}\mathcal{E}_0.$$

Finally, note that we may initialize the cutting-plane algorithm using I_n as a feasible point.

To specify the separation oracles, we let $\mathcal{X}$ denote the feasible region of (5.10). The objective function is $f_0(X) = -\log\det X$, which has gradient $\nabla f_0(X) = -X^{-1}$. The region

$\mathcal{X}$ is defined by the intersection of the sublevel sets of convex functions $f_i(X) = \langle X, a_i a_i^\top \rangle - 1$, which have gradients $\nabla f_i(X) = a_i a_i^\top$, and the positive definite cone $X \succeq I_n/n$. To specify a separating hyperplane for the constraints $X \succeq I_n/n$, by the eigenvalue decomposition we have $X = UDU^\top$ where $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. If $d_i < 1/n$, then let $v_i = Ue_i$, where e_i is i^{th} standard basis vector in $\mathbb{R}^n$. Then the matrix $v_i v_i^\top$ defines the separation oracle, since

$$\langle v_i v_i^\top, X \rangle = v_i^\top X v_i = e_i^\top D e_i = d_i < 1/n.$$

and for any matrix $Z \succeq I_n/n$, we have

$$\langle v_i v_i^\top, Z \rangle = v_i^\top Z v_i \geq 1/n.$$

Thus with $C = -v_i v_i^\top$ and $b = 1/n$, we have $\langle C, X \rangle > b$ and

$$\mathcal{X} \subset \left\{ Z \in \mathbb{R}^{n \times n} \;\middle|\; \langle C, Z \rangle \leq b \right\}.$$

To check that $\langle X, a_i a_i^\top \rangle \leq 1$ for all i, note that again we may compute $\operatorname{diag}(AXA^\top) - 1$ in $O(mn^{\omega-1})$ arithmetic operations, and check the $2m$ constraints. Evaluating $f_0(X) = -\log\det X$, $\nabla f_0(X) = -X^{-1}$, and computing the eigenvalue decomposition each require $O(n^\omega)$ arithmetic operations (Demmel et al., 2007). Thus the net cost of evaluating the separation oracle is $\kappa = O(mn^{\omega-1})$ since $m > n$.

It remains to set the parameters ρ and L, which determine the requisite number of iterations T in equation (2.12) such that we still have mixing with the approximate ellipsoids defined by an ϵ-solution of (5.10). Considering ρ, given that $\mathcal{P} \subset \sqrt{2m}\mathcal{D}_0$, it follows that any feasible X satisfies $\|X\|_2 \leq 2m$. Letting

$$\|X\|_{\max} = \max_{j \geq i} |X_{ij}|,$$

one may easily verify (Golub and Van Loan, 2012) that $\|X\|_{\max} \leq \|X\|_2$. Thus we may take $\rho = 2m$.

To determine L, note that in section 5.4.1 we assume that solving (5.10) yields a positive definite matrix $\hat{E}$ such that $-\log\det \hat{E} + \log\det E \leq \epsilon := 2n^{-10}$, where E is the optimal solution. By theorem 2.4.1, we must guarantee that

$$\frac{\lambda_d(\mathcal{S}_T)}{\lambda_d(\mathcal{X})} < \epsilon^d.$$

We may do so by upper-bounding $\lambda_d(\mathcal{S}_T)$ and lower-bounding $\lambda_d(\mathcal{X})$. The algorithm guarantees that

$$\lambda_d(\mathcal{S}_T) < 2^{-dL}\lambda_d(\mathcal{B}^d). \tag{5.11}$$

To provide a lower bound for $\lambda_d(\mathcal{X})$, note that since I_n is feasible, we must have $|a_i| \leq 1$. Additionally the point $(1/2)I_n$ is feasible for $n \geq 2$. Now consider any positive semidefinite X such that

$$\|X - (1/2)I_n\|_F \leq 1/4,$$

where $\|\cdot\|_F$ denotes the Frobenius norm in $\mathbb{R}^{n\times n}$. Since $\|X\|_F \leq \|X\|_2$, it follows that $\|X - (1/2)I_n\|_2 \leq 1/4$. By the triangle inequality,

$$\|X\|_2 \leq \|X - (1/2)I_n\|_2 + \|(1/2)I_n\|_2 \leq 3/4.$$

Since $|a_i| \leq 1$, we have

$$\left\langle X,\, a_i a_i^\top \right\rangle = a_i^\top X a_i \leq \|X\|_2 \leq 3/4,$$

so X is feasible. Thus $\mathcal{X}$ contains a Frobenius norm ball of radius $r = 1/4$, and we deduce that

$$\lambda_d(\mathcal{X}) > \frac{\lambda_d(\mathcal{B}^d)}{4^d}. \tag{5.12}$$

Thus by equations (5.11) and (5.12), it suffices to choose L such that

$$4^d 2^{-dL} < \epsilon^d,$$

or equivalently,

$$\begin{aligned} L &> \log_2\left(\frac{4}{\epsilon}\right) \\ &= \log_2\left(2n^{10}\right) \\ &= 1 + 10\log_2(n). \end{aligned}$$

Noting equations (2.12) and (2.13), we thus find a suitably approximate John's ellipsoid in $\widetilde{O}(mn^{\omega+1} + n^{2\omega+2})$ operations.

5.4.1 Mixing with Approximate Ellipsoids

It remains to show that sampling from the approximate ellipsoids is sufficient to retain the mixing time derived in section 5.3. We may again assume without loss of generality that the John's ellipsoid $\mathcal{E}_x$ of the current symmetrization $\mathcal{K}_x^s$ is the unit ball $\mathcal{B}_x$ centered at $x = 0$, and we sample uniformly at random from $\mathcal{B}_x(r)$ where $r = cn^{-5/2}$. Let $\hat{E}_y$ denote the square root of the matrix generated by (5.10), and E_y denote the square root of the optimal matrix. Thus necessarily we have $\det \hat{E}_y \leq \det E_y$. Using Vaidya's algorithm, we may find $\hat{E}_y$ such that $-\log\det \hat{E}_y + \log\det E_y \leq n^{-10}$. Equivalently, we have

$$\frac{\det \hat{E}_y}{\det E_y} \geq e^{-n^{-10}} \geq 1 - n^{-10}.$$

Thus by lemmas 5.3.5 and 5.3.7, it follows that for some universal constant $c > 0$, we have

$$1 - cn^{-2} \leq \frac{\lambda(\hat{\mathcal{E}}_y)}{\lambda(\mathcal{B}_x)} \leq 1 + cn^{-2},$$

where $\hat{\mathcal{E}}_y$ is the ellipsoid induced by $\hat{E}_y$.

To show that the aspect ratio of $\hat{\mathcal{E}}_y$ is well-behaved, note that $\log(\det E_y / \det \hat{E}_y)$ is invariant to an affine transformation applied to both $\mathcal{E}_y$ and $\hat{\mathcal{E}}_y$. We thus may presume that $\mathcal{E}_y = \mathcal{B}$. Since $E_y = I_n$ is the solution to (5.1) with center $x_c = 0$ and $\hat{E}_y$ is feasible, it follows from the constraints that

$$\mathrm{tr}(\hat{E}_y^2) = \mathrm{tr}\left(\left(\hat{E}_y^2\right)\left(\sum_i c_i u_i u_i^\top\right)\right) \leq \sum_i c_i = n.$$

Lemma 5.3.8 then follows for $\hat{E}_y$. The remaining steps in proving a lower bound on the conductance of the chain using the approximate ellipsoids are precisely as before.

5.5 Discussion

We have shown that John's Walk mixes in $\widetilde{O}(n^7)$ steps from a warm start. We again remark that our mixing analysis did not presume any structure on $\mathcal{K}$ beyond being a convex body, and as such there is likely considerable room for improvement by assuming more structure on $\mathcal{K}$. There is also a bottleneck in our analysis stemming from the rejection step, since we are forced to control the aspect ratio of the ellipsoid after a move such that rejection does not occur too frequently. Additionally, isoperimetric inequalities which depend a notion of average (rather than minimum) distance between sets such as that of Lovász and Vempala (2006a) may be of use in reducing the mixing time.

We also have described an $\widetilde{O}(mn^{\omega+1}+n^{2\omega+2})$ algorithm for generating approximate John's ellipsoids for the special case of a polytope such that this mixing time is preserved. For the case $m \gg n$ in which n is large, dependence on $n^{2\omega+2}$ may be prohibitive. In the preparation of this manuscript, we had pondered the question of whether the recent approximate John's ellipsoids of Lee (2016) may be employed for the special case of a polytope, which may be formed in $\widetilde{O}(mn^{\omega-1})$ arithmetic operations. We learned this question has been recently answered in the affirmative in Chen et al. (2017). Their algorithm reaches a total variation distance of ϵ from the uniform measure in $O\left(n^{2.5}\log^4\left(\frac{2m}{n}\right)\log\left(\frac{M}{\epsilon}\right)\right)$ steps from an M-warm start. We note that while our mixing time is substantially worse in terms of the polynomial order in n, our mixing time did not depend on m. Additionally, their algorithm had a cost of $O(mn^2\log(m))$ for each step. Our algorithm required worse per step complexity in n to form the approximate ellipsoids while we retaining the $m\log m$ dependence.

Chapter 6

REGRESSION FOR DIFFERENTIABLE FUNCTIONS WITH LIPSCHITZ GRADIENTS IN MULTIPLE DIMENSIONS

One means of fitting functions to high-dimensional data is by providing smoothness constraints. Recently, the following smooth function approximation problem was proposed by Herbert-Voss, Hirn, and McCollum (2017): given a finite set $E \subset \mathbb{R}^d$ and a function $f : E \to \mathbb{R}$, interpolate the given information with a function $\widehat{f} \in \dot{C}^{1,1}(\mathbb{R}^d)$ (the class of first-order differentiable functions with Lipschitz gradients) such that $\widehat{f}(a) = f(a)$ for all $a \in E$, and the value of $\mathrm{Lip}(\nabla \widehat{f})$ is minimal. An algorithm is provided that constructs such an approximating function $\widehat{f}$ and estimates the optimal Lipschitz constant $\mathrm{Lip}(\nabla \widehat{f})$ in the noiseless setting.

In this chapter we address statistical aspects of reconstructing the approximating function $\widehat{f}$ from a closely-related class $C^{1,1}(\mathbb{R}^d)$ given samples from noisy data. We observe independent and identically distributed samples $y(a) = f(a) + \xi(a)$ for $a \in E$, where $\xi(a)$ is a noise term and the set $E \subset \mathbb{R}^d$ is fixed and known. We state uniform bounds relating the empirical risk and true risk over the class $\mathcal{F}_{\widetilde{M}} = \left\{ f \in C^{1,1}(\mathbb{R}^d) \mid \mathrm{Lip}(\nabla f) \leq \widetilde{M} \right\}$, where the quantity $\widetilde{M}$ grows with the number of samples at a rate governed by the metric entropy of the class $C^{1,1}(\mathbb{R}^d)$. Finally, we provide an implementation using Vaidya's algorithm, supporting our results via numerical experiments on simulated data.

6.1 Introduction

Regression tasks are prevalent throughout statistical learning theory and machine learning. Given n samples in $E \subset \mathbb{R}^d$ and corresponding values $\mathcal{Y} = \{y(a)\}_{a \in E} \subset \mathbb{R}$, a regression function $f : \mathbb{R}^d \to \mathbb{R}$ learns a model for the data $(E, \mathcal{Y})$ that best generalizes to new points

$x \notin E$. Absent any prior information on x, the best regression function $\widehat{f}$, as measured by the squared loss, is obtained by minimizing the ℓ^2 empirical risk over a specified function class $\mathcal{F}$,

$$\widehat{f} = \arg\inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{a \in E} |f(a) - y(a)|^2,$$

subject to a regularization penalty. If $\mathcal{F}$ is equipped with a norm or semi-norm $\|\cdot\|_{\mathcal{F}}$, then the regularized risk can take either the form

$$\widehat{f} = \arg\inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{a \in E} |f(a) - y(a)|^2 + \lambda \cdot \Omega(\|f\|_{\mathcal{F}}), \tag{6.1}$$

or

$$\widehat{f} = \arg\inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{a \in E} |f(a) - y(a)|^2 \quad \text{subject to} \quad \|f\|_{\mathcal{F}} \leq M, \tag{6.2}$$

where λ and M are hyper-parameters, and $\Omega : [0, \infty) \to \mathbb{R}$ is a monotonically increasing function.

In either case, the quality of $\widehat{f}$ is primarily determined by the functional class $\mathcal{F}$. Recently, numerous state of the art empirical results have been obtained by using neural networks, which generate a functional class $\mathcal{F}$ through the architecture of the network. The class $\mathcal{F}$ is also often taken as the span of a suitably defined dictionary of functions, or a reproducing kernel Hilbert space (RKHS) with an appropriate kernel. For example, when $\Omega(\|f\|_{\mathcal{F}}) = \frac{1}{2}\|f\|_{\mathcal{F}}^2$ and $\mathcal{F}$ is an RKHS, equation (6.1) leads to the popular kernel ridge regression scheme, which has a closed form solution that is simple to compute.

When $\mathcal{F} = \operatorname{span}(\{\phi_k\}_k)$, the smoothness of $\widehat{f}$ is determined by the dictionary $\{\phi_k\}_k$, or if $\mathcal{F}$ is a reproducing kernel Hilbert space, the regularity of $\widehat{f}$ is determined by the kernel. An alternate approach that does not require choice of dictionary or kernel is to specify the smoothness of $\widehat{f}$ directly, by taking $\mathcal{F} = \dot{C}^m(\mathbb{R}^d)$ or $\mathcal{F} = \dot{C}^{m-1,1}(\mathbb{R}^d)$. (The former class contains the functions continuously differentiable up to order m. The class of $\dot{C}^{m-1,1}(\mathbb{R}^d)$ functions is similar, but it consists of the functions that are differentiable up to order $m-1$, with the highest-order derivatives having a finite Lipschitz constant.) However, the computational complexity of minimizing the regularized risk over these spaces is generally prohibitive. An exception is the space $\dot{C}^{0,1}(\mathbb{R}^d)$, which consists of functions f with finite Lipschitz constant, and for which several regression algorithms exist (von Luxburg and Bousquet, 2004; Beliakov, 2006; Gottlieb et al., 2013; Kyng et al., 2015).

In recent work, Herbert-Voss, Hirn, and McCollum (2017) provide an efficient algorithm for computing the interpolant $\widehat{f} \in \dot{C}^{1,1}(\mathbb{R}^d)$ that, given noiseless data $(E, \mathcal{Y})$, minimizes the Lipschitz constant of the gradient. In this paper we extend the methods of Herbert-Voss et al. (2017) to regularized risk optimizations of the form (6.2). In particular, we consider the noisy scenario in which the function to be reconstructed is not measured precisely on a finite subset, but instead is measured with some uncertainty.

An outline of this chapter is as follows. In Section 6.1.1, we introduce the function interpolation problem considered by Herbert-Voss et al. (2017), and summarize the solution in the noiseless case in Section 6.1.2. Next, we consider the setting where the function is measured under uncertainty, and derive uniform sample complexity bounds on our estimator in Section 6.2.1. The resulting optimization problem can be solved using an algorithm due to Vaidya (1996); we provide details on computing the solution to the regularized risk in Sections 6.2.2 and 6.2.4. We implement the estimator and present reconstruction results on simulated data examples in Section 6.3, supporting our theoretical contributions, and close with a discussion.

6.1.1 Noiseless Function Interpolation Problem

Here we summarize the function approximation problem considered by Herbert-Voss et al. (2017). First, recall that the Lipschitz constant of an arbitrary function $g : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathrm{Lip}(g) := \sup_{x,y \in \mathbb{R}^d,\, x \neq y} \frac{|g(x) - g(y)|}{|x - y|},$$

where $|\cdot|$ denotes the standard Euclidean norm, and we note that in the sequel we use this definition on domains other than $\mathbb{R}^d$. Second, denote the gradient of such an arbitrary g as $\nabla g = (\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_d})$. Finally, let $\dot{C}^{m-1,1}(\mathbb{R}^d)$ be the class of $(m-1)$-times continuously differentiable functions whose derivatives have a finite Lipschitz constant. In the function approximation problem, we are given a finite set of points $E \subset \mathbb{R}^d$ such that $|E| = n$, and a function $f : E \to \mathbb{R}$ specified on E. The $\dot{C}^{1,1}(\mathbb{R}^d)$ function approximation problem as stated by Herbert-Voss et al. (2017) is to compute an interpolating function $\widehat{f} : \mathbb{R}^d \to \mathbb{R}$ that minimizes

$$\|f\|_{\dot{C}^{1,1}(\mathbb{R}^d)} := \inf \left\{ \mathrm{Lip}(\nabla \widetilde{f}) \mid \widetilde{f}(a) = f(a) \text{ for all } a \in E \right\}.$$

The question of whether one can even reconstruct such an interpolating function was answered by Whitney (1934). Presume that we also have access to the gradients of f on E, and denote them $\{D_a f\}_{a \in E}$. In the case of $\dot{C}^{1,1}(\mathbb{R}^d)$, the polynomials are defined by the specified function and gradient information:

$$P_a(x) := f(a) + D_a f \cdot (x - a), \quad a \in E, \ x \in \mathbb{R}^d.$$

Letting $\mathcal{P}$ denote the space of first-order polynomials (i.e., affine functions), the map $P : \mathbb{R}^d \to \mathcal{P}, a \mapsto P_a$ is known as a *1-field*. For any $f \in \dot{C}^{1,1}(\mathbb{R}^d)$, the first order Taylor expansions of f are elements of $\mathcal{P}$, and are known as *jets* (Fefferman and Klartag, 2009), defined as:

$$J_a f(x) := f(a) + \nabla f(a) \cdot (x - a), \quad a, x \in \mathbb{R}^d.$$

Whitney's Extension Theorem for $\dot{C}^{1,1}(\mathbb{R}^d)$ may be stated as follows:

Theorem 6.1.1 (Whitney's Extension Theorem for $\dot{C}^{1,1}(\mathbb{R}^d)$)**.** *Let $E \subset \mathbb{R}^d$ be closed and let $P : E \to \mathcal{P}$ be a* 1*-field with domain E. If there exists a constant $M < \infty$ such that*

(W_0) $|P_a(a) - P_b(a)| \leq M|a - b|^2$ *for all $a, b \in E$, and*

(W_1) $|\frac{\partial P_a}{\partial x_i}(a) - \frac{\partial P_b}{\partial x_i}(a)| \leq M|a - b|$ *for all $a, b \in E$, and $i \in \{1, \ldots, d\}$,*

then there exists an extension $\widehat{f} \in \dot{C}^{1,1}(\mathbb{R}^d)$ such that $J_a \widehat{f} = P_a$ for all $a \in E$.

Given a finite set E as in the function interpolation problem, these conditions are automatically satisfied. However, this theorem does not provide a solution for the minimal Lipschitz constant of ∇f. Le Gruyer (2009) provides a solution to both problems, which we discuss next.

6.1.2 Minimal Value of $\mathrm{Lip}(\nabla \widehat{f})$

Herbert-Voss et al. (2017) define the following norm for when the first-order polynomials are known

$$\|P\|_{\dot{C}^{1,1}(E)} := \inf \left\{ \mathrm{Lip}(\nabla \widetilde{f}) \mid J_a \widetilde{f} = P_a \text{ for all } a \in E \right\}, \tag{6.3}$$

and similarly define

$$\|f\|_{\dot{C}^{1,1}(E)} := \inf \left\{ \mathrm{Lip}(\nabla \widetilde{f}) \mid \widetilde{f}(a) = f(a) \text{ for all } a \in E \right\}, \tag{6.4}$$

when the gradients $\{D_a f\}_{a \in E}$ are unknown, where in both cases the infimum is taken over functions $\widetilde{f} : E \to \mathbb{R}$.

Presuming we are given the 1-field $P : E \to \mathcal{P}$, $a \mapsto P_a$, Le Gruyer (2009) defines the functional Γ^1 as follows.

$$\Gamma^1(P; E) = 2 \sup_{x \in \mathbb{R}^d} \left(\max_{a,b \in E,\, a \neq b} \frac{P_a(x) - P_b(x)}{|a - x|^2 + |b - x|^2} \right). \tag{6.5}$$

Given only functions $f : E \to \mathbb{R}$, Le Gruyer (2009) also defines the functional Γ^1 in terms of f as

$$\Gamma^1(f; E) = \inf \left\{ \Gamma^1(P; E) \mid P_a(a) = f(a) \text{ for all } a \in E \right\},$$

The following theorem is proven by Le Gruyer (2009), which shows that (6.5) and its equivalent formulation in (6.6) provides a solution for (6.3).

Theorem 6.1.2 (Le Gruyer)**.** *Given a set $E \subset \mathbb{R}^d$ and a 1-field $P : E \to \mathcal{P}$,*

$$\Gamma^1(P; E) = \|P\|_{\dot{C}^{1,1}(E)}.$$

An equivalent formulation of (6.5) which is amenable to implementation is as follows. Consider the following functionals mapping $E \times E \to [0, \infty]$:

$$\begin{aligned} A(P; a, b) &= \frac{(P_a(a) - P_b(a)) + (P_a(b) - P_b(b))}{|a-b|^2} \\ &= \frac{2(f(a) - f(b)) + (D_a f - D_b f) \cdot (b-a)}{|a-b|^2}, \\ B(P; a, b) &= \frac{|\nabla P_a(a) - \nabla P_b(a)|}{|a-b|} \\ &= \frac{|D_a f - D_b f|}{|a-b|}. \end{aligned}$$

Proposition 2.2 of (Le Gruyer, 2009) states that

$$\Gamma^1(P; E) = \max_{a \neq b \in E} \sqrt{A(P; a, b)^2 + B(P; a, b)^2} + |A(P; a, b)|, \tag{6.6}$$

whence a naive implementation allows $\Gamma^1(P; E)$ to be found in $O(n^2)$ computations. Inspired by Fefferman and Klartag (2009), Herbert-Voss et al. (2017) also construct algorithms which will solve for the order of magnitude of $\|P\|_{\dot{C}^{1,1}(E)}$ in $O(n \log n)$ time, but we omit the details here. Additionally, as a consequence of the proof of Proposition 2.2 of (Le Gruyer, 2009), equation (6.5) may alternatively be written as

$$\Gamma^1(P; E) = 2 \max_{a, b \in E:\, a \neq b} \sup_{x \in \bar{B}^d\left(\frac{a+b}{2}, \frac{|a-b|}{2}\right)} \frac{P_a(x) - P_b(x)}{|a-x|^2 + |b-x|^2}, \tag{6.7}$$

where $\bar{B}^d(z, r)$ denotes the closed d-dimensional Euclidean ball centered at z with radius r.

Recall that the gradients $\{D_a f\}_{a \in E}$ are typically not known in applications. As a corollary, we have the following convex optimization problem for finding (6.4), and the minimizing 1-field provides the gradients $\{D_a f\}_{a \in E}$.

Corollary 6.1.3. *Given a set $E \subset \mathbb{R}^d$ and a function $f : E \to \mathbb{R}$,*

$$\Gamma^1(f; E) = \|f\|_{\dot{C}^{1,1}(E)}.$$

Recall that $P_a(x) = f(a) + D_a f \cdot (x - a)$. The set $E \subset \mathbb{R}^d$ and the values $\{f(a)\}_{a \in E}$ are fixed, so the optimization problem is to solve for the gradients $\{D_a f\}_{a \in E}$ that minimize $\Gamma^1(P; E)$.

6.2 Regression

In statistical applications where $f(a)$ is observed with uncertainty, one often assumes that we observe $\{y(a)\}_{a\in E}$, where $y(a) = f(a) + \xi(a)$, and $\xi(a)$ is assumed to be independent and identically distributed Gaussian noise for each $a \in E$. Since both the function values and the gradients $\{f(a), D_a f\}_{a\in E}$ are unknown, we minimize an empirical squared error loss over the $k := (d+1)n$ variables defining the 1-field. Given a bound on the $\dot{C}^{1,1}(E)$ seminorm of the unknown 1-field, regression entails solving an optimization problem of the form

$$\begin{aligned} \min_{P} \quad & \frac{1}{n}\sum_{a\in E}(y(a) - P_a(a))^2 \\ \text{s.t.} \quad & \|P\|_{\dot{C}^{1,1}(E)} \leq M. \end{aligned} \tag{6.8}$$

This is a convex optimization problem: the objective function of the empirical squared error loss in (6.8) is convex, as is the constraint set since it is a ball specified by a seminorm. This section proceeds as follows: we begin by analyzing the sample complexity of the function class. These risk bounds establish almost sure convergence of the empirical risk minimizer, and guides the choice of M. Given M, we next appeal to Vaidya's algorithm to solve the resulting optimization problem (6.8). We then apply the efficient algorithm of (Herbert-Voss et al., 2017) to compute the optimal interpolating function.

6.2.1 Sample Complexity and Empirical Risk Minimization

In this section we summarize the results from (Gustafson et al., 2018) of coauthors Mohammed and Xu. The constant $M > 0$ will be chosen via sample complexity arguments. To this end, Mohammed and Xu derived uniform risk bounds for classes of continuous functions $f : \mathcal{B}^d \to \mathbb{R}$, where $\mathcal{B}^d$ denotes the unit Euclidean ball in $\mathbb{R}^d$. The function classes of interest are defined in terms of $C^{1,1}$-norm balls as

$$\mathcal{F}_{\widetilde{M}} = \left\{ f \mid \|f\|_{C^{1,1}(\mathcal{B}^d)} \leq \widetilde{M} \right\},$$

where we are using the norm

$$\|f\|_{C^{1,1}(\mathcal{B}^d)} := \max\left\{ \sup_{x\in\mathcal{B}^d} |f|\,,\ \sup_{x\in\mathcal{B}^d} |\nabla f|\,,\ \mathrm{Lip}(\nabla f) \right\}. \tag{6.9}$$

We note that in order to derive the uniform risk bounds for the function classes $\mathcal{F}_{\widetilde{M}}$, we require such classes be compact, which necessitates the choice of the $C^{1,1}(\mathcal{B}^d)$ norm in equation (6.9) as opposed to the $\dot{C}^{1,1}(\mathcal{B}^d)$ seminorm.

With some abuse of notation, in this section we let f^* denote the underlying function from which we observe noisy samples. We observe an i.i.d. sample

$$S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

drawn from a probability distribution $\mathcal{P} = S_X \times S_{Y|X}$ supported on $X \times Y \subset \mathcal{B}^d \times \mathbb{R}$ under the assumption

$$S_{Y|X} \sim \mathcal{N}(f^*(X), \sigma^2), \quad \text{where } \|f^*\|_{C^{1,1}(\mathcal{B}^d)} = M^*.$$

Since we are in the regression setting, we use squared error loss

$$L(f(x), y) = (f(x) - y)^2 .$$

The true risk is defined as the expectation of L over $\mathcal{P}$,

$$R(f) = \mathbb{E}_{\mathcal{P}}\left[L(f(x), y)\right],$$

and the empirical risk is the expectation over S_S, the empirical distribution on the sample S,

$$\widehat{R}(f) = \mathbb{E}_{S_S}\left[L(f(x), y)\right].$$

In order for the empirical risk minimization procedure to converge to a minimizer of the true risk, we need to bound

$$\sup_{f\in\mathcal{F}_{\widetilde{M}}} \left|\widehat{R}(f) - R(f)\right|$$

with high probability. The most natural way to do so is by expanding the risk and appealing to entropy methods (i.e., covering number bounds) and standard concentration results. We refer the reader to (Gustafson et al., 2018) for the details, but state the main result regarding the empirical risk as follows.

Theorem 6.2.1. *Suppose we set* $\widetilde{M} := n^{1/(2\widetilde{d})}$*, where* $\widetilde{d} := \max\{d, 5\}$*, and let* $\mathcal{F}_{\widetilde{M}}$ *be the class of functions with* $C^{1,1}$*-norm bounded above by* $\widetilde{M}$*.*

(i) For $0 < \delta < 1$*,*

$$\mathbb{P}\left[\sup_{f \in \mathcal{F}_{\widetilde{M}}} \left| R(f) - \widehat{R}(f) \right| < \varepsilon \right] > (1-\delta)(1 - e^{-n^{2/\max\{d,5\}}/2\sigma^2}),$$

where ε *is a monotonically-decreasing function of* n *for large enough* n *and* $\lim_{n\to\infty} \varepsilon = 0$*.*

(ii)

$$\sup_{f \in \mathcal{F}_{\widetilde{M}}} \left| R(f) - \widehat{R}(f) \right| \xrightarrow{a.s.} 0.$$

Additionally, the following theorem of Mohammed and Xu from (Gustafson et al., 2018) establishes the almost sure convergence of the empirical risk minimizer.

Theorem 6.2.2. *Let* $X \sim P_X$*, where* P_X *has density* p *on* $\mathcal{B}^d$ *such that* $0 < c \leq \inf_x p$ *for some constant* c*. Let* f^* *be the true regression function in that observations follow* $S_{Y|X} \sim \mathcal{N}(f^*(X), \sigma^2)$*. Suppose we set* $\widetilde{M} := n^{1/(2\widetilde{d})}$*, where* $\widetilde{d} := \max\{d, 5\}$*, and let* $\widehat{f} \in \mathcal{F}_{\widetilde{M}}$ *be the empirical risk minimizer.*

(i) For $0 < \delta < 1$*,*

$$\mathbb{P}\left[\sup_{x \in \mathcal{B}^d} \left| \widehat{f} - f^* \right| < \beta \right] > (1-\delta)(1 - e^{-n^{2/\max\{d,5\}}/2\sigma^2}),$$

where β *is a monotonically-decreasing function of* n *for large enough* n *and* $\lim_{n\to\infty} \beta = 0$*.*

(ii)

$$\sup_{x \in \mathcal{B}^d} \left| \widehat{f} - f^* \right| \xrightarrow{a.s.} 0.$$

6.2.2 Vaidya's Algorithm

Given the value of M, it remains to solve the optimization problem (6.8). This is a convex program over a set that is not a polytope, and can be solved using interior point methods with a barrier function constructed for the $\dot{C}^{1,1}(E)$ seminorm constraint. However, such methods would have no guarantees on the convergence time without deriving properties of the barrier such as self-concordance (Nesterov and Nemirovsky, 1994), a topic we leave open for future study. Instead, we detail a solution by way of Vaidya's algorithm (Vaidya, 1996), making use of a slight modification of an efficient implementation provided by Anstreicher (1997). We refer the reader to section 6.2.2 for the details of the algorithm, and equation 2.12 and theorem 2.4.1 for the number of iterations required, noting again that in this application we are optimizing over $k = (d+1)n$ variables.

Before finding the requisite number of iterations T, let us first derive the separation oracles, starting with the oracle for $K_1(M) := \left\{ P \mid \|P\|_{\dot{C}^{1,1}(E)} \leq M \right\}$. Presume at the current step of the algorithm, we have a set of function values and gradients $\{f(a), D_a f\}_{a \in E}$. which generate a candidate 1-field P. By Theorem 6.1.2 and equation (6.6), we may find the $a, b \in E, a \neq b$ such that $\Gamma^1(P; E) = \|P\|_{\dot{C}^{1,1}(E)}$ in $n(n-1)/2$ operations. Thus, to determine whether the 1-field at the current step is contained in the constraint set, we simply check if $\|P\|_{\dot{C}^{1,1}(E)} \leq M$. Otherwise, we must return a separating hyperplane in the space of 1-fields. Let $a^\star b^\star \in E$, $a^\star \neq b^\star$ be any elements of E that solve (6.7), with $a^\star$ denoting the first element of E in the numerator of (6.7). Specifying the separating hyperplane requires finding the $x \in \mathbb{R}^d$ that solves (6.7), that is, $x \in \mathbb{R}^d$ such that

$$\Gamma^1(P; E) = 2 \sup_{x \in \bar{B}^d\left(\frac{a^\star + b^\star}{2}, \frac{|a^\star - b^\star|}{2}\right)} \frac{P_{a^\star}(x) - P_{b^\star}(x)}{|a^\star - x|^2 + |b^\star - x|^2}. \tag{6.10}$$

Equation (6.10) is a nonlinear fractional program, and is equivalent to minimizing the ratio

$$R(x) := \frac{N(x)}{D(x)}, \tag{6.11}$$

where $N(x) = |a^\star - x|^2 + |b^\star - x|^2$ and $D(x) = 2(P_{a^\star}(x) - P_{b^\star}(x)) > 0$. Here, we additionally know that the minimizer of (6.11) attains the optimal value $1/\Gamma^1(P; E)$ due to equation (6.6).

Jagannathan (1966) and Dinkelbach (1967) showed that for $N(x)$ continuous, $D(x) > 0$ continuous, the solution to $\min_{x\in\mathcal{X}} R(x)$ over a compact subset $\mathcal{X} \subset \mathbb{R}^d$ is $z \in \mathcal{X}$ if and only if $z \in \mathcal{X}$ is also an optimal solution for

$$\min_{x\in\mathcal{X}} N(x) - R(z)D(x).$$

Plugging in the optimal value $R(y) = 1/\Gamma^1(P;E)$ yields the minimization

$$\min_{x} |a^\star - x|^2 + |b^\star - x|^2 - \left(\frac{2(P_{a^\star}(x) - P_{b^\star}(x))}{\Gamma^1(P;E)}\right).$$

Thus finding the x which solves (6.10) amounts to minimizing a convex quadratic in x. The solution is

$$z = \left(\frac{a^\star + b^\star}{2}\right) + \left(\frac{D_{a^\star} f - D_{b^\star} f}{2\Gamma^1(P;E)}\right).$$

The separation oracle for feasibility is thus specified as follows. For a candidate 1-field P, if $\|P\|_{\dot{C}^{1,1}(E)} \leq M$, then certify that P is a feasible 1-field. Otherwise, separate all other 1-fields $\widetilde{P} = \left\{\widetilde{f}(a), \widetilde{D_a f}\right\}_{a\in E}$ from P via

$$\left\{\widetilde{P} \;\middle|\; \frac{2(\widetilde{P}_{a^\star}(z) - \widetilde{P}_{b^\star}(z))}{|a^\star - z|^2 + |b^\star - z|^2} \leq \Gamma^1(P;E)\right\},$$

or equivalently

$$(\widetilde{f}(a^\star) + \widetilde{D_{a^\star} f} \cdot (z - a^\star)) - (\widetilde{f}(b^\star) + \widetilde{D_{b^\star} f} \cdot (z - b^\star)) \leq \Gamma^1(P;E) \left(|a^\star - z|^2 + |b^\star - z|^2\right).$$

The separation oracle for $K_1(M)$ is thus equivalent to using the vector

$$w_{a^\star,b^\star} = \begin{pmatrix} 1 \\ z - a^\star \\ -1 \\ -(z - b^\star) \end{pmatrix} \in \mathbb{R}^{2(d+1)}$$

and the scalar $u_{a^\star,b^\star} = \Gamma^1(P;E)\left(|a^\star - z|^2 + |b^\star - z|^2\right)$ to define the hyperplane

$$\left\{v \in \mathbb{R}^{2(d+1)} \mid w_{a^\star,b^\star} \cdot v \leq u_{a^\star,b^\star}\right\}. \tag{6.12}$$

Appropriately padding $w_{a^\star,b^\star}$ and v with zeros over the remaining possible choices of a, b thus defines a k-dimensional separating hyperplane, and this separating hyperplane is constructed in $O(n^2 + d)$ operations.

Note that the objective function in (6.8) is equivalent to

$$g(P) = \frac{1}{n}|y - f|^2.$$

To construct the separation oracle for $K_2(\gamma) := \{P \mid g(P) - g(P^\star) \leq \gamma\}$, taking the gradient with respect to f, we have $w = \frac{2}{n}(f - y)$. Thus, given a current feasible field P with function values f, the separating hyperplane is specified as

$$\left\{\widetilde{f} \in \mathbb{R}^n \mid w \cdot \widetilde{f} \leq u\right\}, \tag{6.13}$$

where $u = w \cdot f$. Suitably concatenating w and $\widetilde{f}$ with zeros to form vectors in $\mathbb{R}^k$ thus specifies the separation oracle for $K_2(\gamma)$, and requires $O(n)$ operations to evaluate.

Now, to find the requisite number of iterations T to find a feasible point in $K_1(M) \cap K_2(\gamma)$, we sandwich the set $K_1(M)$ with Euclidean balls. The next result characterizes the Euclidean ball inside $K_1(M)$.

Lemma 6.2.3. *Assume $|a - b| \geq r > 0$ for all $a, b \in E$, $a \neq b$. Then*

$$\rho_1 \mathcal{B}^k \subset K_1(M),$$

where

$$\rho_1 = \left(\frac{r^2 M}{8(1 + r)}\right)\sqrt{n}.$$

Proof. Let $P = \{f(a), D_a f\}_{a \in E}$ be any 1-field, and assume that

$$|P_a| = \sqrt{|f(a)|^2 + |D_a f|^2} \leq \rho$$

for all $a \in E$, so P is represented by a vector in $(\rho\sqrt{n})\mathcal{B}^k$. Note that the numerator of (6.7) may be written as

$$P_a(x) - P_b(x) = (f(a) - f(b)) + \frac{1}{2}(D_a f + D_b f) \cdot (b - a) + (D_a f - D_b f) \cdot \left(x - \frac{a + b}{2}\right).$$

Thus,

$$|P_a(x) - P_b(x)| \leq 2\rho + \rho|b-a| + 2\rho\left|x - \frac{a+b}{2}\right|$$

$$\leq 2\rho(1 + |b-a|),$$

where we have used the fact that $x \in \bar{B}^d\left(\frac{a+b}{2}, \frac{|a-b|}{2}\right)$ in (6.7). The denominator is minimized at $x = \frac{a+b}{2}$ with minimal value $|a-b|^2/2$. Thus

$$\Gamma^1(P;E) \leq \frac{8\rho(1+|b-a|)}{|b-a|^2}$$

$$\leq 8\rho(r^{-1} + r^{-2}).$$

It follows that $\|P\|_{\dot{C}^{1,1}(E)} \leq M$ if $\rho \leq \frac{Mr^2}{8(1+r)}$. $\square$

We now derive a bounding ball for $K_1(M)$ which indicates the value of ρ to use in Vaidya's algorithm.

Lemma 6.2.4. *Assume E is an $\varepsilon-$net of $\mathcal{B}^d$, the unit ball in $\mathbb{R}^d$, where $\varepsilon < 1/10$. Suppose for all distinct $a, b \in E$, $|a-b| > r$. Then*

$$K_1(M) \subset \rho_2\mathcal{B}^k,$$

where

$$\rho_2 = \sqrt{n}\left[\frac{|y|^2}{n} + 4\left(\frac{10|y|}{r} + \frac{5M}{2}\right)^2\right]^{1/2}.$$

Proof. Let $\prod_f$ denote the projection of any subspace of $\mathbb{R}^k$ onto the n-dimensional subspace corresponding to $D_a f = 0$ for all $a \in E$. If $y \in \prod_f K_1(M)$, then an optimal solution is given by $\{y(a), 0\}_{a\in E}$. Thus we may presume that $y \notin \prod_f K_1(M)$, whence

$$|f| \leq |y|.$$

To bound $|D_a f|$ given $a \in E$, note that by equation (6.7) we have

$$\frac{2|P_a(x) - P_b(x)|}{|a-x|^2 + |b-x|^2} \leq M$$

for any $b \in E \setminus \{a\}$ and $x \in \mathcal{B}^d\left(\frac{a+b}{2}, \frac{|a-b|}{2}\right)$. Choosing $x = b$ implies that

$$\begin{aligned}|D_a f \cdot (b-a)| &\leq |f(a) - f(b)| + \frac{M}{2}|a-b|^2 \\ &\leq 2|y| + \frac{M}{2}|a-b|^2.\end{aligned}$$

By the conditions of the lemma, given a, there exist $b \in E \setminus \{a\}$ such that

$$|D_a f \cdot (b-a)| \geq \frac{|D_a f||(b_1 - a)|}{10}.$$

It follows that

$$|D_a f| \leq \frac{20|y|}{r} + 5M.$$

Thus

$$\begin{aligned}|P|^2 &= |f|^2 + \sum_{a \in E} |D_a f|^2 \\ &\leq |y|^2 + 4n\left(\frac{10|y|}{r} + \frac{5M}{2}\right)^2.\end{aligned}$$

□

We arrive at the number of iterations required such that $g(P) - g(P^\star) \leq \gamma$.

Theorem 6.2.5. *Let $\gamma > 0$ be an error tolerance parameter, let $P^\star$ be any optimal solution to (6.8), and assume $0 < r \leq |a-b| \leq R$ for all $a, b \in E$. Applying Vaidya's algorithm for minimization as in (2.8) using the separation oracles specified in (6.12) and (6.13) yields an approximate solution $\widehat{P}_T$ to (6.8) such that*

$$g(\widehat{P}_T) - g(P^\star) \leq \gamma$$

where we choose

$$L \geq \log_2\left(\frac{4|y|^2}{n\gamma\rho_1}\right),$$

with ρ_1 as stated in lemma 6.2.3 and T is given in equation (2.12) using $\rho = \rho_2$ from lemma 6.2.4.

Proof. Recall that if $y \in \prod_f K_1(M)$, then an optimal 1-field is returned without calling Vaidya's algorithm via $\{y(a), 0\}_{a \in E}$. Thus we may presume that $|f| \leq |y|$ on $\prod_f K_1(M)$. Thus

$$g(P) = \frac{1}{n}|y - f|^2 \leq \frac{4|y|^2}{n}$$

for any $P \in K_1(M)$. Thus we set $\varepsilon = \frac{n\gamma}{4|y|^2}$, and apply theorem 2.4.1. From lemma 6.2.3, we have that $\lambda_k(K_1(M)) \geq \rho_1^k \lambda_k(\mathcal{B}^k)$. Since $\lambda_k(\mathcal{S}_T) < 2^{-kL}\lambda_k(\mathcal{B}^k)$, it suffices to choose L such that

$$2^{-kL}\rho_1^{-k} \leq \left(\frac{n\gamma}{4|y|^2}\right)^k,$$

from which the statement results. □

6.2.3 *Constraining the $\dot{C}^{1,1}$-Seminorm Rather Than the $C^{1,1}$-Norm*

In the previous section, we showed how to use Vaidya's algorithm to solve the optimization problem (6.8) that is central to this chapter. Before we use the output 1-field to actually construct the interpolant, we need to show that the risk bounds derived in Section 6.2.1 apply to our optimization scheme. Here we briefly summarize work of coauthor Narayanan from (Gustafson et al., 2018) which relates the risk bounds to our optimization scheme.

The only potential conflict is that our solution to (6.8) involves constraining the $\dot{C}^{1,1}$-seminorm of functions defined on a discrete set $E \subset \mathbb{R}^d$; however, the risk bounds given in Section 6.2.1 are based on the overall $C^{1,1}$-norm of functions defined on the unit ball $\mathcal{B}^d \subset \mathbb{R}^d$ (the overall norm is the maximum of the $\dot{C}^{1,1}$-seminorm, the Euclidean norm of the gradient, and the absolute value of the function values). Narayanan showed that as long as the sample size is large enough, the $\dot{C}^{1,1}$-seminorm $\|\cdot\|_{\dot{C}^{1,1}(E)}$ determines the $C^{1,1}$-norm $\|\cdot\|_{C^{1,1}(\mathcal{B}^d)}$ in our setup with high probability.

Recall that for a finite set of points $E \in \mathbb{R}^d$ and a function $f : E \to \mathbb{R}$, norms and seminorms of f are defined in terms of their analogues for continuous-domain extensions of

f. Specifically,

$$
\begin{aligned}
\|f\|_{\dot{C}^{1,1}(E)} &:= \inf\left\{\mathrm{Lip}(\nabla\widetilde{f}) \mid \widetilde{f}(a) = f(a) \text{ for all } a \in E\right\}, \text{ where } \widetilde{f} \in \dot{C}^{1,1}\left(\mathcal{B}^d\right), \\
\|f\|_{C^0(E)} &:= \inf\left\{\sup_{x\in\mathcal{B}^d} |\widetilde{f}(x)| \mid \widetilde{f}(a) = f(a) \text{ for all } a \in E\right\}, \text{ where } \widetilde{f} \in C^0\left(\mathcal{B}^d\right), \\
\|f\|_{C^1(E)} &:= \inf\left\{\sup_{x\in\mathcal{B}^d} \|\nabla\widetilde{f}(x)\| \mid \widetilde{f}(a) = f(a) \text{ for all } a \in E\right\}, \text{ where } \widetilde{f} \in C^1\left(\mathcal{B}^d\right), \text{ and} \\
\|f\|_{C^{1,1}(E)} &:= \max\left\{\|f\|_{C^0(E)}, \|f\|_{C^1(E)}, \|f\|_{\dot{C}^{1,1}(E)}\right\}.
\end{aligned}
$$

The main results are Theorem 6.2.6 and Theorem 6.2.7. Let f^* be the true function that appears in the generative process. Let $\|f^*\|_{C^{1,1}(\mathcal{B}^d)} \le M^*$. Let a set X containing n random points be chosen i.i.d from $\mathcal{P}$, which we assume has a density $\rho(x)$ with respect to the Lebesgue measure on $\mathcal{B}^d$ and a minimum density ρ_{min}. Let $y_i = f^*(x_i) + \xi_i$, where $x_i \in X_0$ and ξ_i is a Gaussian with mean 0 and variance σ^2 that is independent of all the other ξ_j. Set $\widetilde{M} := n^{1/(2\widetilde{d})}$, where $\widetilde{d} := \max\{d, 5\}$. We will denote in this section by C_d constants depending only on d. Suppose we project y onto the set of all functions f such that $\|f\|_{\dot{C}^{1,1}(X)} \le \widetilde{M}$. Theorem 6.2.6 states that for a large-enough sample, the $C^0(X)$- and $C^1(X)$-norms of the projection are less than $\widetilde{M}/2$. Furthermore, theorem 6.2.7 states that the extension of this projection to the unit ball has $C^{1,1}(\mathcal{B}^d)$-norm no more than $\widetilde{M}$. This is enough to show that the sample complexity results are compatible with the construction of the interpolant in Sections 6.2.2 and 6.2.4.

Theorem 6.2.6. *Let $K \subseteq L^2(X)$ be the closed convex set of all functions f such that*

$$\|f\|_{\dot{C}^{1,1}(X)} \le \widetilde{M}.$$

Let h be the projection of y onto K with respect to the Hilbert space $L^2(X)$. Then when n is sufficiently large, with probability at least $1 - \exp(-n^{1/100})$,

$$\max\left(\|h\|_{C^0(X)}, \|h\|_{C^1(X)}\right) < \widetilde{M}/2.$$

Theorem 6.2.7. *Let* $\|\widetilde{h}\|_{\dot{C}^{1,1}(X)} \leq 1$ *and* $\max(\|\widetilde{h}\|_{C^1(X)}, \|\widetilde{h}\|_{C^0(X)}) \leq 1/2$*. Then, with probability at least* $1 - \exp(-n^{1/100})$ *any minimal* $\dot{C}^{1,1}(\mathcal{B}^d)$*-norm extension* f *of* $\widetilde{h}$ *to the unit ball satisfies*

$$\max\left(\|f\|_{C^1(\mathcal{B}^d)}, \|f\|_{C^0(\mathcal{B}^d)}\right) \leq 1.$$

6.2.4 Wells' Construction for $\widehat{f}$ *Given* $\mathrm{Lip}(\nabla \widehat{f})$

Given $M = \mathrm{Lip}(\nabla \widehat{f})$ and the estimates $\widehat{f}(a)$ and $\widehat{D_a f}$ for all $a \in E$, it remains to construct the interpolant $\widehat{f} \in \dot{C}^{1,1}(\mathbb{R}^d)$. We may now apply solution methods from the noiseless function interpolation problem. We summarize the solution provided by Wells et al. (1973) here.

Wells' construction takes $E \subset \mathbb{R}^d$, the 1-field $P : E \to \mathcal{P}$ consisting of function values $\{f(a)\}_{a \in E}$ and gradients $\{D_a f\}_{a \in E}$, and a value $M = \mathrm{Lip}(\nabla \widehat{f})$ as inputs. A necessary condition for Wells' construction to hold is that

$$f(b) \leq f(a) + \frac{1}{2}(D_a f + D_b f) \cdot (b - a) + \frac{M}{4}|b - a|^2 - \frac{1}{4M}|D_a f - D_b f|^2, \quad \forall a, b \in E, \tag{6.14}$$

for which the optimal objective function value and gradients returned by the methods in Sections 6.1.2 and 6.2.2 satisfy.

For all $a \in E$, Wells defines the shifted points

$$\widetilde{a} = a - \frac{D_a f}{M},$$

and associates a type of distance function for any $x \in \mathbb{R}^d$ to that point,

$$d_a(x) = f(a) - \frac{1}{2M}|D_a f|^2 + \frac{M}{4}|x - \widetilde{a}|^2.$$

Using the shifted points, every subset $S \subset E$ is associated with several new sets:

$$\begin{aligned}
\widetilde{S} &= \{\widetilde{a} \mid a \in S\}, \\
S_H &= \text{the smallest affine space containing } \widetilde{S}, \\
\widehat{S} &= \text{the convex hull of } \widetilde{S}, \\
S_E &= \left\{x \in \mathbb{R}^d \mid d_a(x) = d_b(x) \text{ for all } a, b \in S\right\}, \\
S_* &= \left\{x \in \mathbb{R}^d \mid d_a(x) = d_b(x) \le d_c(x) \text{ for all } a, b \in S,\ c \in E\right\}, \\
S_C &= S_H \cap S_E.
\end{aligned}$$

Note that $S_H \perp S_E$, so S_C is a singleton. Wells next defines the collection of subsets

$$\mathcal{K} = \left\{S \subset E \mid \exists x \in S_* \text{ such that } d_S(x) < d_{E \setminus S}(x)\right\},$$

and a new collection of sets $\{T_S\}_{S \in \mathcal{K}}$, where

$$T_S = \frac{1}{2}(\widehat{S} + S_*) = \left\{\frac{1}{2}(y+z) \mid y \in \widehat{S},\ z \in S_*\right\}, \quad S \in \mathcal{K}. \tag{6.15}$$

The collection $\{T_S\}_{S \in \mathcal{K}}$ form a partition of $\mathbb{R}^d$ in the sense that overlapping sets have Lebesgue measure 0. On each set T_S, Wells defines a function $\widehat{f}_S : T_S \to \mathbb{R}$ which is a local piece of the interpolating function $\widehat{f}$:

$$\widehat{f}_S(x) = d_S(S_C) + \frac{M}{2}\operatorname{dist}(x, S_H)^2 - \frac{M}{2}\operatorname{dist}(x, S_E)^2, \quad x \in T_S,\ S \in \mathcal{K}, \tag{6.16}$$

where as usual for sets $A, B \subset \mathbb{R}^d$, we have $\operatorname{dist}(A, B) = \inf_{x \in A,\ y \in B} |x - y|$. The final function $\widehat{f} : \mathbb{R}^d \to \mathbb{R}$ is then defined using (6.15) and (6.16):

$$\widehat{f}(x) = \widehat{f}_S(x), \quad \text{if } x \in T_S. \tag{6.17}$$

The gradient of $\widehat{f}_S$ is

$$\nabla \widehat{f}_S(x) = \frac{M}{2}(z - y), \quad \text{where } x = \frac{1}{2}(y+z),\ y \in \widehat{S}, z \in S_*.$$

The function $\widehat{f} : \mathbb{R}^d \to \mathbb{R}$ of (6.17) satisfies the following:

Theorem 6.2.8 (Wells' Construction)**.** *Given a finite set $E \subset \mathbb{R}^d$, a 1-field $P : E \to \mathcal{P}$, and a constant M satisfying (6.14), the function $\widehat{f} : \mathbb{R}^d \to \mathbb{R}$ defined by (6.16) is in $\dot{C}^{1,1}(\mathbb{R}^d)$ and satisfies*

1. $J_a \widehat{f} = P_a$ *for all* $a \in E$.

2. $\mathrm{Lip}(\nabla \widehat{f}) = M$.

(Herbert-Voss et al., 2017) provides efficient algorithms to implement Wells' construction. We refer the reader to (Herbert-Voss et al., 2017) for the details, but state briefly the computational cost of their methods. Let $m = |\mathcal{K}|$, and note that in the worst case $m = O(n^{\lceil d/2 \rceil})$. A pessimistic bound on the storage as well as the number of computations for the one time work is $O(m^2)$. The query work is then also bounded by $O(m^2)$, however using more efficient querying algorithms to find the set $T_S \in \mathcal{K}$ to which a given $x \in \mathbb{R}^d$ belongs can lessen the work significantly. Using a tree structure, for example, will require $O(\log m) = O(\log n)$ work per query if the tree is balanced, but this need not be the case in general.

6.3 Simulation

In this section we describe work with coauthor Matthew Hirn to illustrate the algorithm. We numerically compute the empirical risk minimizer $\widehat{f}$ over the functional class,

$$\mathcal{F}_M = \{f \mid \|f\|_{C^{1,1}(\mathcal{B}^d)} \leq M\}.$$

To do so, we solve the optimization problem (6.8) for a 1-field P over the finite set E. This can be done efficiently with the algorithm described in Section 6.2.2, or using any constrained, convex optimization algorithm (such as interior point methods). The 1-field P is extended to a function $\widehat{f}$ in $\mathcal{F}_M$, using the algorithm described by Herbert-Voss et al. (2017).

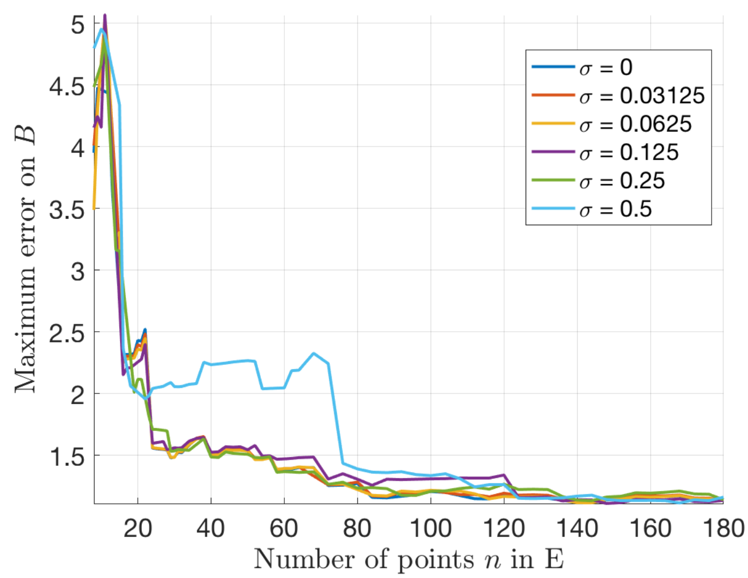
The underlying function $f : \mathbb{R}^2 \to \mathbb{R}$ is taken as:

$$\forall\, x = (x_1, x_2) \in \mathbb{R}^2, \quad f(x_1, x_2) = \begin{cases} \cos(\pi x_1)\sin(\pi x_2)\exp\left(-\frac{1}{1-|x|^2}\right), & |x| < 1, \\ 0, & |x| \geq 1, \end{cases}$$

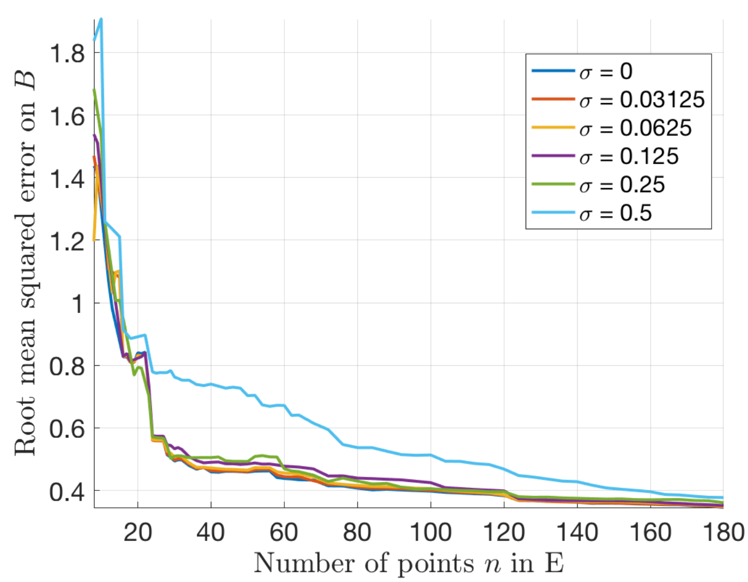
which is supported in the unit ball $\mathcal{B}^2$. The training points E are sampled uniformly from $\mathcal{B}^2$, and noisy function values $y(a) = f(a) + \xi(a)$ for $a \in E$ are recorded, where $\xi(a)$ is i.i.d. Gaussian white noise with standard deviation σ. To approximate the generalization error between $\widehat{f}$ and f on $\mathcal{B}^2$, the unit cube in which $\mathcal{B}^2$ is inscribed is sampled on a grid containing 2^{14} points (2^7 along each axis). As such, all errors over $\mathcal{B}^2$ described below are numerically approximated on the intersection of this grid with $\mathcal{B}^2$.

The error $\sup_{x\in\mathcal{B}^2} |f(x) - \widehat{f}(x)|$ between f and the empirical risk minimizer is plotted in Figure 6.1a as a function of n, for various values of the noise standard deviation σ. The value of M grows with n according to $M = O(1/n^{10})$, as in the proof of Theorem 6.2.1. Figure 6.1b plots the generalization error for the quadratic loss, i.e., $\left(\int_{\mathcal{B}^2} |f(x) - \widehat{f}(x)|^2\, dx\right)^{1/2}$. Both figures show that the generalization error generally decreases as n (and correspondingly M) increase.

Figure 6.2 gives a qualitative assessment of the empirical risk minimizer $\widehat{f}$. In each figure the original function is shown in the upper left plot. In figure 6.2a, the subsequent plots (in left to right order, and then proceeding to the second row) are for the noiseless ($\sigma = 0$) setting for $N = 8, 23, 38, 84, 180$ samples. Figure 6.2b shows the noisy scenario for $N = 84$ samples with $\sigma = 2^{-j}$, $j = 5, 4, \ldots, 1$ (in the analogous order). As expected, the empirical risk minimizer $\widehat{f}$ visually appears to better match the underlying function f as n increases. Figure 6.2b fixes $n = 84$ and plots $\widehat{f}$ for increasing values of the noise standard deviation. While for large noise ($\sigma = 0.5$) the empirical risk minimizer $\widehat{f}$ deviates noticeably from f, for lower noise values ($\sigma \leq 0.25$), $\widehat{f}$ is a stable approximation of f.

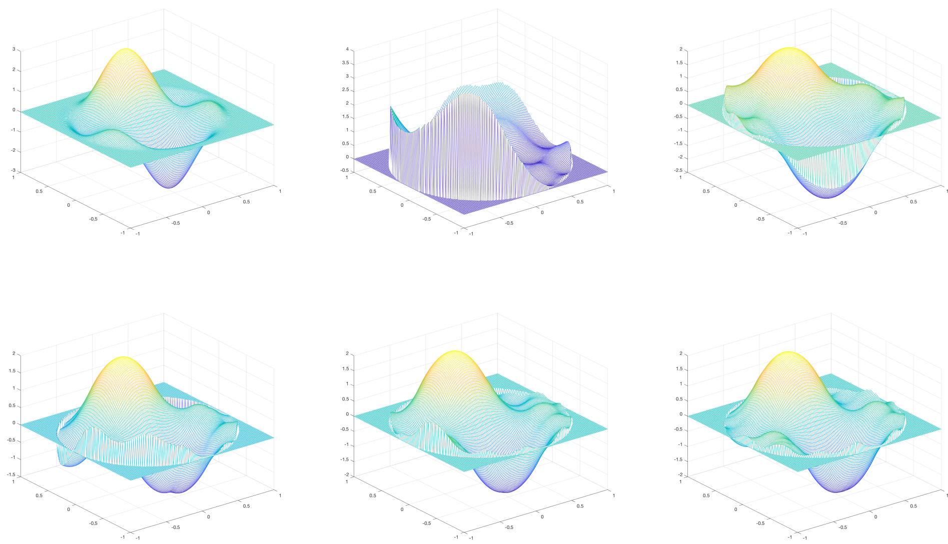


(a) Maximum error

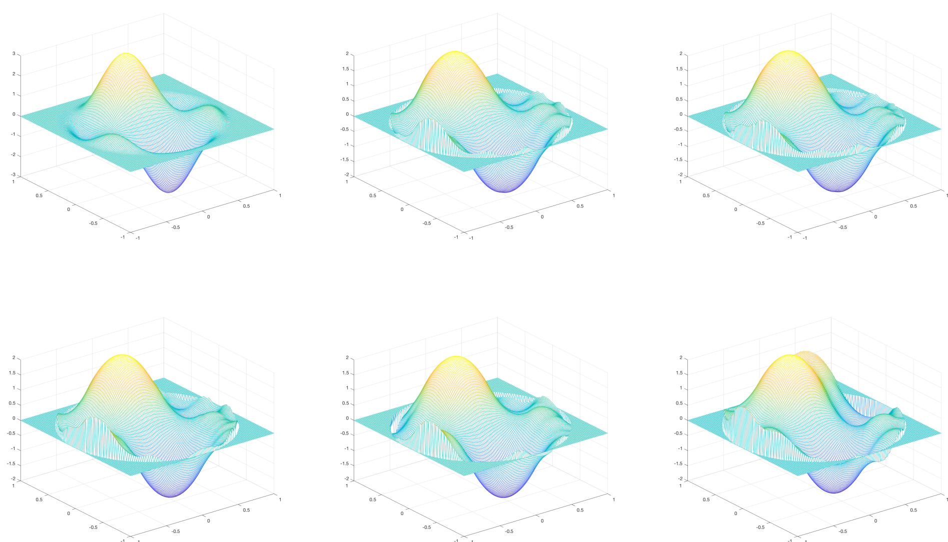


(b) Root mean square error

Figure 6.1: Generalization error as a function of $|E| = n$



(a) Noiseless recovery



(b) Noisy recovery

Figure 6.2: Interpolants $\widehat{f}$ in the noisy and noiseless scenarios

6.4 Discussion

In this chapter, we extened the function interpolation problem considered by Herbert-Voss et al. (2017) to the regression setting, where function values $f(a)$ are observed with uncertainty over finite $a \in E$. We imposed smoothness on the approximating function by considering regression solutions in the class of $C^{1,1}(\mathbb{R}^d)$ functions. Minimizing the risk over this function class is computationally tractable optimization problem, requiring $O((d+1)^2n^2)$ calls to a separation oracle using Vaidya's algorithm. We presented a separating hyperplane that requires $O(n^2)$ operations, and given the output of Vaidya's algorithm, reconstruct the interpolant using efficient implementations of Wells' construction proposed by Herbert-Voss et al. (2017).

We stated uniform bounds relating the empirical risk of the regression solution to the true risk using empirical processes methods. The covering number of the class of $C^{1,1}(\mathbb{R}^d)$ functions is known and can be used to derive the covering number of Lipschitz loss classes. Our loss class is unbounded, but by conditioning on a suitable bound that increases with n, we obtain high probability bounds. As a consequence of the uniform risk bounds, almost sure convergence of the empirical risk minimizer is also guaranteed. These theoretical contributions were supported by numerical results via simulation.

Chapter 7

REGRESSION FOR HIGHER-ORDER DIFFERENTIABLE FUNCTIONS IN MULTIPLE DIMENSIONS

In this chapter we describe an algorithm to generalize the $\dot{C}^{1,1}(E)$ regression of section 6.2 to $\dot{C}^m(E)$ regression. We discuss a Calderón-Zygmund decomposition of a local section of $\mathbb{R}^d$ that was proposed by (Fefferman, 2011), prove some basic properties regarding this decomposition, and implement it in $\mathbb{R}^2$ to illustrate these properties. We then also show how to use this decomposition to solve $\dot{C}^m(E)$ regression. As a step to this end, we prove Whitney's extension theorem using this decomposition in a manner similar to (Fefferman et al., 2016). We leave the implementation of the algorithm for $\dot{C}^m(E)$ regression and the derivation of uniform risk bounds for future work.

7.1 The Function and Jet Interpolation Problems

As in Chapter 6, we presume that we observe a function f on a finite set $E \subset \mathbb{R}^d$, where $N = |E| \geq 2$. The *function interpolation problem* is to extend our observations of f on E to $\mathbb{R}^d$ by forming an interpolant $F : \mathbb{R}^d \to \mathbb{R}$ such that for all $a \in E$, we have $f(a) = F(a)$. Further, we will require that F satisfies certain smoothness constraints. For any non-negative integer m, let $C^m(\mathbb{R}^d)$ denote the Banach space of real-valued m-times continuously differentiable functions F on $\mathbb{R}^d$ for which the norm

$$\|F\|_{C^m(\mathbb{R}^d)} = \sup_{x \in \mathbb{R}^d} \max_{|\alpha| \leq m} |\partial^\alpha F(x)| \tag{7.1}$$

is finite. Here α is a multi-index, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where each α_i is a non-negative integer. With some abuse of notation (with regards to the cardinality of a set E denoted by

$|E|$), we use the notational shorthand as described in the following operations,

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n, \qquad \text{(sum of components)}$$

$$\alpha! = \alpha_1!\alpha_2!\ldots\alpha_n!, \qquad \text{(factorial)}$$

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \qquad \text{(power)}$$

$$\partial^\alpha = \frac{\partial_1^{\alpha_1}\partial_2^{\alpha_2}, \ldots \partial_n^{\alpha_n}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}. \qquad \text{(high-order partial derivative)}$$

Note also that $\partial^0 F(x)$ is defined to be $F(x)$.

The function interpolation problem is to find an interpolant $F \in C^m(\mathbb{R}^d)$ for which the norm $\|F\|_{C^m(\mathbb{R}^d)}$ is as small as possible such that $F(a) = f(a)$ for all $a \in E$. Whereas in chapter 6 using theorem 6.1.2 we were able to construct an interpolant which had minimal $C^{1,1}(\mathbb{R}^d)$ norm, in what follows we will be satisfied to find an $F \in C^m(\mathbb{R}^d)$ which has a minimal norm up to a multiplicative constant, i.e.,

$$\|F\|_{C^m(\mathbb{R}^d)} \leq C\|\widetilde{F}\|_{C^m(\mathbb{R}^d)},$$

where $\widetilde{F}$ is any minimal interpolating function and C depends only on m and d.

To solve the function interpolation problem, we again instead first consider the *jet interpolation problem*. Recall Taylor's theorem (Folland, 2005), which states that if $F \in C^m(S)$ on a convex set $S \subset \mathbb{R}^d$, then for any $x, y \in S$

$$F(y) = \sum_{|\alpha| \leq m-1} \frac{\partial^\alpha F(x)}{\alpha!}(y-x)^\alpha + R_{x,m-1}(y-x).$$

The *jet* of F at x is the Taylor polynomial of F at x of order $m-1$,

$$J_x(F)(y) = \sum_{|\alpha| \leq m-1} \frac{\partial^\alpha F(x)}{\alpha!}(y-x)^\alpha.$$

Taylor's theorem states that the jet $J_x(F)$ is the only polynomial of degree $m-1$ which agrees with $F(x)$ to order k at $y-x$. If $|\partial^\alpha F(x)| \leq M$ for $x \in S$ and $|\alpha| = m$, the remainder satisfies

$$|R_{x,m-1}(y-x)| = \frac{M}{m!}\|y-x\|_\infty^m,$$

where $\|h\|_\infty = \sum_{i=1}^d |h_i|$. With Taylor's theorem in mind, let $\mathcal{P}$ denote the vector space of all real-valued polynomials in $\mathbb{R}^d$ of degree at most m. Clearly $J_x(F) \in \mathcal{P}$.

If F is a real-valued function on a cube Q, we write $F \in C^m(Q)$ to denote that F and its derivatives up to order m extend continuously to the closure of the cube Q, and define the norm,

$$\|F\|_{C^m(Q)} = \sup_{x\in Q} \max_{|\alpha|\leq m} |\partial^\alpha F(x)|,$$

which coincides with (7.1) restricted the cube Q. If x belongs to the boundary of Q, we still write $J_x(F)$ to denote the Taylor polynomial of degree $m-1$ of F at x even though F is not defined on a full neighborhood of x. For a finite set $E \subset \mathbb{R}^n$ such that $|E| \geq 2$, a *Whitney field* on E is a family of polynomials

$$P = \{P_a\}_{a\in E}$$

such that $P_a \in \mathcal{P}$ for all $a \in E$. Considering the vector space of all Whitney fields on E, we may define the seminorm

$$\|P\|_{\dot{C}^m(E)} = \max_{a,b\in E,\, a\neq b,\, |\alpha|\leq m} \frac{|\partial^\alpha (P_a - P_b)(a)|}{|a-b|^{m-|\alpha|}}, \tag{7.2}$$

The jet interpolation problem is then as follows: given a function f and its jets P defined on E, find an interpolant $F : \mathbb{R}^d \to \mathbb{R}$ and its jets such that $J_a(F) = P_a$ for all $a \in E$ and $\|F\|_{C^m(\mathbb{R}^d)}$ is minimal. Again we will be satisfied to solve this problem up to a multiplicative constant depending only on m and d.

7.2 Whitney's Extension Theorem

The question of the existence such a function F was solved by Whitney (1934). We will provide a similar proof in the subsequent section, but first we discuss Calderón-Zygmund (CZ) decompositions.

7.2.1 A Calderón-Zygmund Decomposition and Its Properties

A closed, axis-aligned *cube* in $\mathbb{R}^d$ is a set

$$Q = I_1 \times I_2 \times \ldots \times I_d$$

where the I_j are closed intervals in $\mathbb{R}$ with common lengths $|I_1| = |I_2| = \ldots = |I_d|$. Let the sidelength $\delta_Q = |I_j|$ denote this common length. With some abuse of notation, for all $a > 0$, we let aQ denote the cube with the same center as Q but with sidelength $\delta_{aQ} = a\delta_Q$. A *dyadic cube* is any cube in $Q \in \mathbb{R}^d$ such that for each $j \in \{1, 2, \ldots, d\}$, we have $I_j = [2^k \cdot i_j, 2^k \cdot (i_j+1)]$ for integers k, i_j. Thus a dyadic cube Q has sidelength $\delta_Q = 2^k$ for some integer k, and this cube is contained in precisely one parent cube $\widetilde{Q}$ with sidelength $\delta_{\widetilde{Q}} = 2\delta_Q = 2^{k+1}$.

Now presume that $E \subset Q_0$ for some dyadic cube Q_0, and given any $a, b \in E$, we have $|a-b| \geq \eta\delta_{Q_0}$ for some $\eta > 0$. A *Calderón-Zygmund (CZ) decomposition* is a finite collection subcubes of Q_0 such that the subcubes satisfy the following properties:

- *Pairwise Disjoint Interiors*: If $Q_1, Q_2 \in \mathcal{CZ}$, then $\operatorname{int}(Q_1) \cap \operatorname{int}(Q_2) = \emptyset$.
- *Finite Covering*: $Q_0 = \cup_{Q \in \mathcal{CZ}} Q$ and $|\mathcal{CZ}| < \infty$.
- *Good Geometry*: If $Q_1, Q_2 \in \mathcal{CZ}$, then $\frac{1}{2}\delta_{Q_1} \leq \delta_{Q_2} \leq 2\delta_{Q_1}$.

The following decomposition of Q_0 is proposed by Fefferman (2011):

$$\mathcal{CZ} = \left\{ \text{dyadic } Q \subset Q_0 \;\middle|\; Q \text{ maximal and } |3Q \cap E| \leq 1 \right\}. \tag{7.3}$$

A dyadic subcube $Q \subset Q_0$ is said to be *maximal* if it is the largest Q satisfying the condition $|3Q \cap E| \leq 1$. Equivalently, if $\widetilde{Q}$ is the parent cube of Q, then $|3\widetilde{Q} \cap E| \geq 2$. To verify that $\mathcal{CZ}$ given in (7.3) is in fact a CZ decomposition, note that since the cubes dyadic and maximal, any two cubes in the decomposition have pairwise disjoint interiors. Also the collection $\mathcal{CZ}$ of cubes clearly covers Q_0 and the collection is finite since $N = |E| < \infty$, and for all $a, b \in E$ such that $a \neq b$, we have $|a - b| > \eta\delta_{Q_0}$ for some $\eta > 0$. To show the good geometry property, we require the following lemma.

Lemma 7.2.1. *Suppose $Q_1, Q_2 \subset \mathbb{R}^d$ are any closed cubes such that $Q_1 \cap Q_2 \neq \emptyset$ and $\delta_{Q_1} \leq \frac{1}{2}\delta_{Q_2}$. Then $3Q_1 \subset 3Q_2$.*

Proof. Let $d_\infty : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be defined as

$$d_\infty(a, b) = \max\left\{|a_1 - b_1|, |a_2 - b_2|, \ldots, |a_d - b_d|\right\}.$$

Let $a \in 3Q_1$, and suppose $Q_1 \cap Q_2 \neq \emptyset$. Then

$$d_\infty(a, Q_2) = \min_{b \in Q_2} d_\infty(a, b) \leq 2\delta_{Q_1} \leq \delta_{Q_2}.$$

Since $3Q_2 = \left\{a \in \mathbb{R}^d \;\middle|\; d_\infty(a, Q_2) \leq \delta_{Q_2}\right\}$, it follows that $b \in 3Q_2$. $\square$

With lemma 7.2.1, we may show the good geometry property and that $\mathcal{CZ}$ is a properly defined CZ decomposition.

Theorem 7.2.2. *The decomposition $\mathcal{CZ}$ in (7.3) is a CZ-decomposition.*

Proof. It remains to show the good geometry property using lemma 7.2.1. Assume that $Q_1, Q_2 \in \mathcal{CZ}$ are such that $Q_1 \cap Q_2 \neq \emptyset$, and assume without loss of generality that $\delta_{Q_1} \leq \delta_{Q_2}$. For any Q_1 such that $\delta_{Q_1} \leq \frac{1}{4}\delta_{Q_2}$, by lemma 7.2.1 we have $3Q_1 \subset 3Q_2$, whence

$$|3Q_1 \cap E| \leq |3Q_2 \cap E| \leq 1.$$

Now let $\widetilde{Q}_1$ be the parent cube of any such Q_1, so $\delta_{\widetilde{Q}_1} = 2\delta_{Q_1} \leq \frac{1}{2}\delta_{Q_2}$. Then again by lemma 7.2.1 we have $3\widetilde{Q}_1 \subset 3Q_2$, and thus

$$|3\widetilde{Q}_1 \cap E| \leq |3Q_2 \cap E| \leq 1,$$

whence Q_1 was not maximal. Continuing taking parent cubes of Q_1 in this fashion if necessary, we see that a maximal dyadic cube Q_1 such that $\delta_{Q_1} \leq \delta_{Q_2}$ must satisfy $\delta_{Q_1} \geq \frac{1}{2}\delta_{Q_2}$. The symmetric assumption that $\delta_{Q_1} \geq \delta_{Q_2}$ leads to the good geometry property. $\square$

Finally, it is important to note the following two properties about the cubes in the CZ decomposition. Firstly, there are a finite number of cubes that intersect any given cube.

Theorem 7.2.3. *Given $Q \in \mathcal{CZ}$, there are at most 4^d cubes $Q' \in \mathcal{CZ}$ such that $Q \cap Q' \neq \emptyset$.*

Proof. Given any $Q \in \mathcal{CZ}$, by the good geometry property, any cube $Q' \in \mathcal{CZ}$ such that $Q \cap Q' \neq \emptyset$ must have sidelength at least $\frac{1}{2}\delta_Q$. Thus given that Q' has the minimal sidelength, we have $Q' \subset 2Q$. The cube $2Q$ contains less than 4^d cubes of with sidelength $\frac{1}{2}\delta_Q$, so at most 4^d cubes Q' touch Q. □

Secondly, if we dilate any cube in the decomposition slightly, any point in Q_0 is still contained in a finite number of cubes.

Theorem 7.2.4. *Let $0 < \epsilon < 1$. Any point $y \in Q_0$ is contained in at most 4^d cubes of the form $Q^* = (1+\epsilon)Q$, where $Q \in \mathcal{CZ}$.*

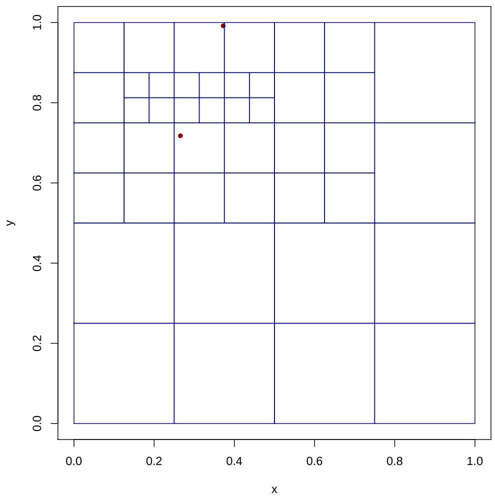
Proof. For any $y \in Q_0$, there exists a cube $Q_y \in \mathcal{CZ}$ such that $y \in Q_y$. Fix any cube $Q \in \mathcal{CZ}$. We claim that Q^* intersects Q_y only if $Q \cap Q_y \neq \emptyset$, whence at most 4^d cubes Q^* may contain y in light of the previous theorem. By the good geometry property, since any cube $Q' \in \mathcal{CZ}$ that intersects Q has sidelength $\delta_{Q'} \geq \frac{1}{2}\delta_Q$, and Q^* has sidelength $\delta_{Q^*} < 2\delta_Q$, it follows that

$$Q^* \subset \bigcup \left\{ Q' \in \mathcal{CZ} \mid Q' \cap Q \neq \emptyset \right\}.$$

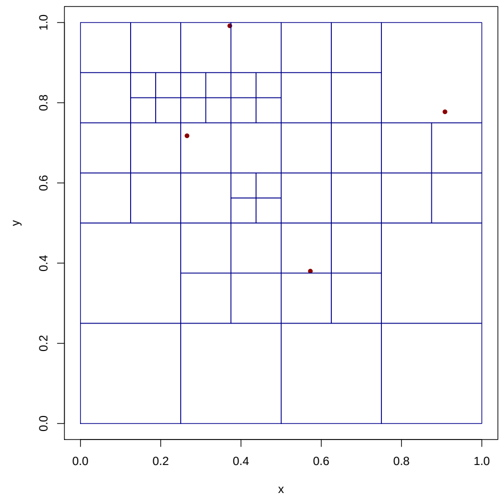
Thus if $y \in Q^*$, we have $Q \cap Q_y \neq \emptyset$. □

7.2.2 Illustration of the $\mathcal{CZ}$ Decomposition

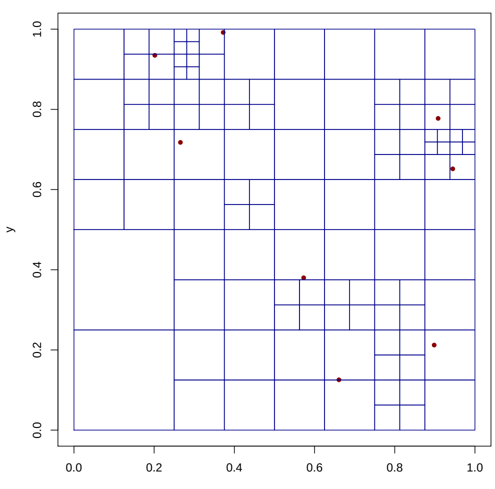
Here we illustrate the $\mathcal{CZ}$ decomposition of a set $E \subset Q_0 := [0,1]^2$ of equation (7.3) for various $n = |E|$. These decompositions were formed by taking one uniform random sample of E' from $[0,1]^2$ of size 16, and letting E be the first n elements of E' for $n = 2, 4, 8, 16$. We note that the good geometry property is easily verified visually, and that $|E \cap 3Q| \leq 1$ for each $Q \in \mathcal{CZ}$. We also see that $\delta_Q \leq \frac{1}{4}$ in all cases. Finally note that as $\eta = \min_{x,y \in E} |x - y|$ decreases, the cardinality of the $\mathcal{CZ}$ decomposition increases. The cubes in this decomposition will be crucial to solving the jet interpolation problem.



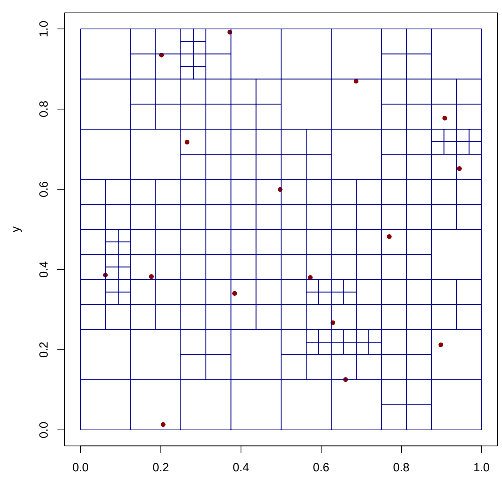
(a) $n = |E| = 2$.



(b) $n = |E| = 4$.



(c) $n = |E| = 8$.



(d) $n = |E| = 16$.

Figure 7.1: CZ Decompositions for $d = 2$ and $n = 2, 4, 8, 16$.

7.2.3 Proof of Whitney's Extension Theorem

To prove Whitney's extension theorem in general for a finite set $E \subset \mathbb{R}^d$ such that $|E| \geq 2$, consider the CZ decomposition of equation (7.3). Let

$$\Gamma(a, M) = \left\{ P \in \mathcal{P} \;\middle|\; \text{there exists } F \in C^m(\mathbb{R}^d) \text{ with } \|F\|_{C^m(\mathbb{R}^d)} \leq M \text{ and } J_a(F) = P \right\}.$$

Whitney's theorem is as follows.

Theorem 7.2.5. *Let $\{P_a\}_{a \in E}$ be a Whitney field on a finite set E, and let $M > 0$. Given*

$$P_a \in \Gamma(a, M) \text{ for each } a \in E,$$
$$|\partial^\alpha (P_b - P_a)(b)| \leq M|b - a|^{m-|\alpha|} \text{ for all } a, b \in E,\ |\alpha| \leq m - 1,$$

there exists $F \in C^m(\mathbb{R}^d)$ such that $\|F\|_{C^m(\mathbb{R}^d)} \leq CM$ and $J_a(F) = P_a$ for all $a \in E$.

Proof. Note in what follows the constants $C > 0$ depend only on m and d, and may change line to line as necessary. Consider the collection of cubes $\mathcal{CZ}$ given in (7.3). Without loss of generality we may presume $Q_0 = [0, 1]^d$ by translating and rescaling as necessary. Note that since we presume $n = |E| \geq 2$, it follows that $\delta_Q \leq \frac{1}{4}$ for any $Q \in \mathcal{CZ}$. We classify the cubes in $Q \in \mathcal{CZ}$ into two types,

(i) $E \cap 3Q \neq \emptyset$,

(ii) $E \cap 3Q = \emptyset$.

If $Q \in \mathcal{CZ}$ is type (i), then $E \cap 3Q = \{a_Q\}$. By hypothesis, since $P_{a_Q} \in \Gamma(a_Q, M)$, there exists $F_Q \in C^m(\mathbb{R}^d)$ such that

$$\|F_Q\|_{C^m(\mathbb{R}^d)} \leq M \text{ and } J_{a_Q}(F_Q) = P_{a_Q}. \tag{7.4}$$

We choose such an F_Q as in (7.4). If $Q \in \mathcal{CZ}$ is type (ii), then Q's parent cube $\widetilde{Q}$ satisfies $|E \cap 3\widetilde{Q}| \geq 2$. We may pick any of the $a_Q \in E \cap 3\widetilde{Q}$. By hypothesis, since $P_{a_Q} \in \Gamma(a_Q, M)$, again there exists $F_Q \in C^m(\mathbb{R}^d)$ satisfying (7.4). Thus F_Q has been defined for both types of cubes for any cube $Q \in \mathcal{CZ}$, and by hypothesis we have

$$J_a(F_Q) = P_a \text{ for all } a \in E \cap 3Q, \tag{7.5}$$

noting that for type (ii) cubes (7.5) holds vacuously since $E \cap 3Q = \emptyset$.

Now let $Q, Q' \in \mathcal{CZ}$, and consider their dilations $(1+\epsilon)Q, (1+\epsilon')Q'$ for $\epsilon, \epsilon' \in (0,1)$. Suppose that $(1+\epsilon)Q \cap (1+\epsilon')Q' \neq \emptyset$. We will show that

$$|\partial^\alpha(F_Q - F_{Q'})| \leq CM\delta_Q^{m-|\alpha|} \text{ on } (1+\epsilon)Q \cap (1+\epsilon')Q' \text{ for } |\alpha| \leq m. \tag{7.6}$$

Let $\widetilde{Q}, \widetilde{Q}'$ denote the parent of a cubes $Q, Q' \in \mathcal{CZ}$, respectively. Since each cube is either of type (i) or type (ii), we have $a_Q \in E \cap 3\widetilde{Q}$ and $a_{Q'} \in E \cap 3\widetilde{Q}'$ where $(1+\epsilon)Q \cap (1+\epsilon')Q' \neq \emptyset$. By good geometry and theorem 7.2.4, we have that

$$|a_Q - a_{Q'}| \leq C\delta_Q.$$

Additionally by good geometry and theorem 7.2.4, we have that

$$|x - a_Q|, |x - a_{Q'}| \leq C\delta_Q \text{ for all } x \in (1+\epsilon)Q \cap (1+\epsilon')Q'.$$

Thus by equation (7.4) it follows that

$$|\partial^\alpha(F_Q - P_{a_Q})(x)| \leq CM|x - a_Q| \leq CM\delta_Q^{m-|\alpha|} \tag{7.7}$$

and that

$$|\partial^\alpha(F_{Q'} - P_{a_{Q'}})(x)| \leq CM|x - a_{Q'}| \leq CM\delta_Q^{m-|\alpha|} \tag{7.8}$$

for all $x \in (1+\epsilon)Q \cap (1+\epsilon')Q'$ and for $|\alpha| \leq m$. By the theorem's hypothesis, equations (7.7) and (7.8), and a straightforward application of the triangle inequality, the bound in (7.6) follows directly.

Next we introduce a partition of unity,

$$1 = \sum_{Q \in \mathcal{CZ}} \phi_Q, \tag{7.9}$$

where each $\phi_Q \in C^m(\mathbb{R}^d)$, and each ϕ_Q is supported in $(1+\epsilon)Q$, whence

$$|\partial^\alpha \phi_Q| \leq C\delta_Q^{-|\alpha|} \text{ for } |\alpha| \leq m. \tag{7.10}$$

We finally define our extension operator as

$$F = \sum_{Q \in \mathcal{CZ}} \phi_Q F_Q, \tag{7.11}$$

where we note we have defined $F_Q \in C^m(\mathbb{R}^d)$ for each type of cube in $\mathcal{CZ}$, and $\mathcal{CZ}$ is finite. Thus $F \in C^m(\mathbb{R}^d)$. For any y, let Q_y be the unique cube in $\mathcal{CZ}$ which contains y. Then for $|\alpha| \leq m$, we have by the product rule and equation (7.9) that

$$\partial^\alpha F(y) = \partial^\alpha F_{Q_y}(y) + \sum_{Q \in CZ} \partial^\alpha(\phi_Q \cdot (F_Q - F_{Q_y}))(y),$$

where the sum is finite due to theorem 7.2.4. Note additionally that given the support of ϕ_Q, any $Q \in \mathcal{CZ}$ enters the sum only if $y \in (1+\epsilon)Q$. Since $\delta_Q \leq \frac{1}{4} < 1$, it follows from (7.6) and (7.10) that

$$|\partial^\alpha(\phi_Q \cdot (F_Q - F_{Q_y}))(y)| \leq CM\delta_Q^{m-|\alpha|} \leq CM \text{ for } |\alpha| \leq m.$$

Additionally we have that $|\partial^\alpha F_{Q_y}(y)| \leq CM$ for $|\alpha| \leq m$ by (7.4). We have shown that $F \in C^m(\mathbb{R}^d)$ and $\|F\|_{C^m(\mathbb{R}^d)} \leq CM$.

Now to prove that $J_a(F) = P_a$ for all $a \in E$, we need to define multiplication of jets. Given any $x \in \mathbb{R}^d$, we may define a multiplication operator $\odot_x$ such that

$$P \odot_x Q = J_x(PQ) \text{ for } P, Q \in \mathcal{P}.$$

Note that such an operation is equivalent to polynomial multiplication in which terms of order $m+1$ or greater are truncated. Given $Q \in \mathcal{CZ}$, and $a \in E$, we have $J_a(F_Q) = P_a$ by the good geometry property. Since the support of ϕ_Q is contained in $(1+\epsilon)Q$, it follows that for each $Q \in \mathcal{CZ}$ we have

$$J_x(\phi_Q F_Q) = J_a(\phi_Q) \odot_a P_a$$

It therefore follows by equation (7.9) that

$$J_a(F) = \sum_{Q \in \mathcal{CZ}} J_a(\phi_Q F_Q) = \left(\sum_{Q \in \mathcal{CZ}} J_a(\phi_Q) \right) \odot_a P_a = P_a.$$

Thus $F \in C^m(\mathbb{R}^d)$ is such that $\|F\|_{C^m(\mathbb{R}^d)} \leq CM$ and $J_a(F) = P_a$ for all $a \in E$. $\square$

7.3 Function Interpolation and Regression

7.3.1 Interpolation by Linear Programming

Theorem 7.2.5 provides the existence of an interpolant $F \in C^m(\mathbb{R}^d)$ which agrees with any interpolant of minimal $C^m(\mathbb{R}^d)$ norm up to a multiplicative constant, but it does not provide a means by which to find such an interpolant. Furthermore, in applications we are not given the Whitney field $\{P_a\}_{a \in E}$ beyond the $\alpha = 0$ term, since we observe only a function value $f(a)$ for all $a \in E$. However, we may solve the following optimization problem. Given a non-negative integer m, we solve

$$\begin{aligned} \min_{z \in \mathbb{R}^k, M \in \mathbb{R}} \quad & M \\ \text{s.t.} \quad & \|P\|_{\dot{C}^m(E)} \leq M \\ & P_a(a) = f(a) \quad \text{for all } a \in E. \end{aligned} \tag{7.12}$$

Recall that $\|P\|_{\dot{C}(E)}$ was defined in (7.2). In the optimization problem (7.12), we note that $z \in \mathbb{R}^k$ represents the coefficients of the polynomials defined by the Whitney field $\{P_a\}_{a \in E}$. The number of monomials in the multinomial expansion of degree $|\alpha|$ over d variables is

$$\binom{d + |\alpha| - 1}{|\alpha| - 1}.$$

Thus, it follows that the dimension of z is

$$k = |E| \left(\sum_{|\alpha| \leq m} \binom{d + |\alpha| - 1}{|\alpha| - 1} \right).$$

This optimization problem has a linear objective function and is specified by linear equality and inequality constraints in the variable $z \in \mathbb{R}^k$, and hence is a linear program. Given its solution, we form the extension F using the CZ decomposition and the partition of unity as discussed in the previous section. Namely, (7.11) yields the interpolant $F \in C^m(\mathbb{R}^d)$. Though this linear program will not be tractable for large m or d, for applications we typically will have $d = 2$ or $d = 3$, and small m will yield a parsimonious model of observed data.

7.3.2 Regression by Quadratic Programming

It remains to describe the function regression problem akin to section 6.2. We again leave this topic for future work, but describe what remains to be done to extend the methods of chapter 6 to the m-times differentiable case. Again, we observe noisy samples $y(a) = f(a) + \xi(a)$ for all $a \in E$, and the constant M will be chosen via sample complexity arguments as in section 6.2.1. Given the value of M, one obtains the quadratic program

$$\begin{aligned} \min_{z \in \mathbb{R}^k} \quad & \frac{1}{n} \sum_{a \in E} (y(a) - P_a(a))^2 \\ \text{s.t.} \quad & \|P\|_{\dot{C}^m(E)} \leq M \end{aligned}$$

where k is as in the linear program, and the extension F is again constructed via (7.11).

7.3.3 Constructing a Partition of Unity

Finally, it is worth mentioning how to construct the partition of unity (7.9). We state the approach shown in (Tu, 2010). It suffices to use a $C^\infty(\mathbb{R})$ function to construct *bump functions* in $\mathbb{R}$, and then use the Cartesian product of such functions to form a bump function on $\mathbb{R}^d$. It may be shown by induction that the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(t) = \begin{cases} e^{-1/t} & t > 0, \\ 0 & t \leq 0 \end{cases}$$

is a $C^\infty(\mathbb{R})$ function. Using this function, we construct a *step function*

$$\begin{aligned} g(t) &= \frac{f(t)}{f(t)+f(1-t)} \\ &= \begin{cases} 0 & t \le 0, \\ \frac{1}{1+\frac{e^{1/t}}{e^{1/(1-t)}}} & 0 < t < 1, \\ 1 & t \ge 1. \end{cases} \end{aligned}$$

It may be shown that $g \in C^\infty(\mathbb{R})$ and is strictly increasing on $(0,1)$. We may form a step function which is increasing on an interval (a^2, b^2) by letting

$$h(t) = g\left(\frac{t-a^2}{b^2-a^2}\right).$$

Letting $h_s(t) = h(t^2)$, we then have a symmetric function $h_s \in C^\infty(\mathbb{R})$. Finally, we let

$$\rho(t) = 1 - h_s(t) = 1 - g\left(\frac{t^2-a^2}{b^2-a^2}\right),$$

we have formed a $C^\infty(\mathbb{R})$ bump function centered at 0 which is 1 on $[-a,a]$ and is supported on $[-b,b]$. Translating by considering $\rho(t-s)$ yields a bump function centered at s, and the a and b may be modified as necessary. Finally, a bump function $\rho \in C^\infty(\mathbb{R}^d)$ may be formed via

$$\varphi(x) = \prod_{i=1}^{d} \rho(x_i),$$

Since each $x \in Q_0$ is contained by only finitely many cubes $Q \in \mathcal{CZ}$, the partition of unity in (7.9) may be formed via normalizing the bump functions φ_Q.

BIBLIOGRAPHY

Radosław Adamczak, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *Journal of the American Mathematical Society*, 23(2):535–561, 2010.

Kurt M Anstreicher. On Vaidya's volumetric cutting plane method for convex programming. *Mathematics of Operations Research*, 22(1):63–89, 1997.

Kurt M Anstreicher. Improved complexity for maximum volume inscribed ellipsoids. *SIAM Journal on Optimization*, 13(2):309–320, 2002.

Keith Ball. Ellipsoids of maximal volume in convex bodies. *Geometriae Dedicata*, 41(2): 241–250, 1992.

Keith Ball. An elementary introduction to modern convex geometry. *Flavors of geometry*, 31:1–58, 1997.

Gleb Beliakov. Interpolation of Lipschitz functions. *Journal of Computational and Applied Mathematics*, 196:20–44, 2006.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Sébastien Bubeck. Convex optimization: algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. *arXiv preprint arXiv:1412.1587*, 2014.

Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. Sampling from a log-concave distribution with projected langevin monte carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.

Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. Fast mcmc sampling algorithms on polytopes. *arXiv preprint arXiv:1710.08165*, 2017.

Benjamin Cousins and Santosh Vempala. Bypassing KLS: Gaussian cooling and an $O^*(n^3)$ volume algorithm. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 539–548. ACM, 2015.

Alexander Munro Davie and Andrew James Stothers. Improved bound for complexity of matrix multiplication. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, 143(02):351–369, 2013.

James Demmel, Ioana Dumitriu, and Olga Holtz. Fast linear algebra is stable. *Numerische Mathematik*, 108(1):59–91, 2007.

Werner Dinkelbach. On nonlinear fractional programming. *Management Science*, 13(7): 492–498, 1967.

Charles Fefferman. Interpolation by linear programming I. *Discrete and Continuous Dynamical Systems*, 30(2):477–492, 2011.

Charles Fefferman and Bo'az Klartag. Fitting a C^m-smooth function to data I. *Annals of mathematics*, pages 315–346, 2009.

Charles Fefferman, Arie Israel, and Garving K Luli. Interpolation of data by smooth nonnegative functions. *arXiv preprint arXiv:1603.02330*, 2016.

Gerald Folland. Higher-order derivatives and taylor's formula in several variables, 2005.

Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. Efficient regression in metric spaces via approximate Lipschitz extension. In Edwin Hancock and Marcello Pelillo, editors, *Similarity-Based Pattern Recognition, SIMBAD 2013*, volume 7953 of *Lecture Notes in Computer Science*, pages 43–58. Springer, Berlin, Heidelberg, 2013.

Peter M Gruber. Minimal ellipsoids and their duals. *Rendiconti del Circolo Matematico di Palermo*, 37(1):35–64, 1988.

Osman Güler. Hyperbolic polynomials and interior point methods for convex programming. *Mathematics of Operations Research*, 22(2):350–377, 1997.

Adam Gustafson, Matthew Hirn, Kitty Mohammed, Hariharan Narayanan, and Jason Xu. Structural risk minimization for $C^{1,1}(\mathbb{R}^d)$ regression. *arXiv preprint arXiv:1803.10884*, 2018.

Ariel Herbert-Voss, Matthew J Hirn, and Frederick McCollum. Computing minimal interpolants in $C^{1,1}(\mathbb{R}^d)$. *Revista Matemática Iberoamericana*, 33(1):29–66, 2017.

John K Hunter and Bruno Nachtergaele. *Applied analysis*. World Scientific, 2001.

R Jagannathan. On some properties of programming problems in parametric form pertaining to fractional programming. *Management Science*, 12(7):609–615, 1966.

F. John. Extremum Problems with Inequalities as Subsidiary Conditions. In K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, editors, *Studies and Essays: Courant Anniversary Volume*, pages 187–204. Wiley-Interscience, New York, 1948.

Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3-4):541–559, 1995.

Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random structures and algorithms*, 11(1):1–50, 1997.

Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37 (1):1–20, 2012.

Ioannis Kontoyiannis and Sean P Meyn. Geometric ergodicity and the spectral gap of non-reversible markov chains. *Probability Theory and Related Fields*, 154(1-2):327–339, 2012.

Rasmus Kyng, Anup Rao, Sushant Sachdeva, and Daniel A. Spielman. Algorithms for Lipschitz learning on graphs. *JMLR: Workshop and Conference Proceedings*, 40:1–34, 2015.

François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, pages 296–303. ACM, 2014.

Erwan Le Gruyer. Minimal lipschitz extensions to differentiable functions defined on a Hilbert space. *Geometric and Functional Analysis*, 19(4):1101–1118, 2009.

Yin Tat Lee. *Faster algorithms for convex and combinatorial optimization.* PhD thesis, Massachusetts Institute of Technology, 2016.

Yin Tat Lee and Aaron Sidford. Path finding I: Solving linear programs with $\widetilde{O}(\sqrt{rank})$ linear system solves. *arXiv preprint arXiv:1312.6677*, 2013.

Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1049–1065. IEEE, 2015.

László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures & Algorithms*, 4(4):359–412, 1993.

László Lovász and Santosh Vempala. Hit-and-run from a corner. *SIAM Journal on Computing*, 35(4):985–1005, 2006a.

László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006b.

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

Hariharan Narayanan. Randomized interior point methods for sampling and optimization. *The Annals of Applied Probability*, 26(1):597–641, 2016.

Arkadi Nemirovski. Information-based complexity of convex programming. *Lecture notes, Technion*, 1995.

Yu Nesterov and Arkadi Nemirovsky. Interior-point polynomial methods in convex programming, vol. 13 of. *Studies in Applied Mathematics*, 1994.

Daniel Paulin. Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability*, 20, 2015.

Gareth O Roberts and Jeffrey S Rosenthal. Geometric ergodicity and hybrid markov chains. *Electron. Comm. Probab*, 2(2):13–25, 1997.

Gareth O Roberts, Jeffrey S Rosenthal, et al. General state space markov chains and mcmc algorithms. *Probability surveys*, 1:20–71, 2004.

Mark Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.

Walter Rudin. Functional analysis. international series in pure and applied mathematics, 1991.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

L.W. Tu. *An Introduction to Manifolds*. Springer New York, 2010.

Pravin M Vaidya. A new algorithm for minimizing convex functions over convex sets. *Mathematical Programming*, 73(3):291–341, 1996.

Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.

Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

Nisheeth Vishnoi. A mini-course on convex optimization, 2014.

Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with Lipschitz functions. *Journal of Machine Learning Research*, 5:669–695, 2004.

John C Wells et al. Differentiable functions on Banach spaces with Lipschitz derivatives. *Journal of Differential Geometry*, 8(1):135–152, 1973.

Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.

Virginia Vassilevska Williams. Breaking the Coppersmith-Winograd barrier. 2011.

Appendix A

PROOF OF JOHN'S MAXIMAL VOLUME ELLIPSOID THEOREM

In this appendix we prove John's maximal volume ellipsoid theorem providing a proof similar to that of Ball (1997). We first enumerate some properties on the polars of convex sets which will be necessary to establish the proof. In particular we will need to derive the polar of an ellipsoid.

A.1 Polars of Sets

In this section we review some properties of the polar of a set. The definition of the polar of a set is as follows.

Definition A.1 (Polar of a Set)**.** *Let $S \subset \mathbb{R}^n$. Then the polar of S is defined to be*

$$S^\circ = \{y \in \mathbb{R}^n : \langle y,\, x\rangle \leq 1 \textit{ for all } x \in S\}\,.$$

The polar of a set is convex even if the original set is not.

Proposition A.2. *S° is convex.*

Proof. Let $y_1, y_2 \in S^\circ$, and $\alpha_1 + \alpha_2 = 1$ where $\alpha_1, \alpha_2 \geq 0$. Then given any $x \in S$, we have

$$\langle \alpha_1 y_1 + \alpha_2 y_2,\, x\rangle = \alpha_1 \langle y_1,\, x\rangle + \alpha_2 \langle y_2,\, x\rangle \leq \alpha_1 + \alpha_2 = 1,$$

so $(\alpha_1 y_1 + \alpha_2 y_2) \in S^\circ$. □

The polar of a ball is given the ball defined with respect to the dual norm.

Proposition A.3. *Let* $\mathcal{B} := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$*, where* $\|\cdot\|$ *is a norm. Then* $\mathcal{B}^\circ = \mathcal{B}^\star := \{y \in \mathbb{R}^n \mid \|y\|_\star \leq 1\}$*, where* $\|\cdot\|_\star$ *is the dual norm defined by*

$$\|y\|_\star = \sup_{x \in \mathbb{R}^n \colon \|x\| \leq 1} \langle x,\, y\rangle\,.$$

Proof. Note that it is straightforward to verify that the dual norm is a norm, and by its definition the generalized Cauchy-Schwarz (GCS) inequality

$$\langle x,\, y\rangle \leq \|x\|\|y\|_\star$$

holds. If $y \in \mathcal{B}^\circ$, we have $\langle y,\, x\rangle \leq 1$ for all $x \in \mathcal{B}$. Picking x such that $\|x\| = 1$, it follows that $\langle y,\, x\rangle \leq \|x\|$. By GCS, this implies $\|y\|_\star \leq 1$, and thus $y \in \mathcal{B}^\star$. Conversely, if $y \in \mathcal{B}^\star$, then for all $x \in \mathcal{B}$, we have $\langle y,\, x\rangle \leq \|x\|\|y\|_\star = 1$ by GCS, so $y \in \mathcal{B}^\circ$. □

Lemma A.4. *Let* $\mathcal{E}$ *be an ellipsoid in* $\mathbb{R}^n$*, defined as*

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n \mid \sum_{j=1}^{n} \frac{\langle x, e_j\rangle^2}{\alpha_j^2} \leq 1 \right\},$$

where $\{e_i\}_{i=1}^n$ *is an orthonormal basis for* $\mathbb{R}^n$ *and* $\alpha_i > 0$ *for* $i = 1, \ldots, n$*. Then*

$$\mathcal{E}^\circ = \left\{ y \in \mathbb{R}^n \mid \sum_{j=1}^{n} \alpha_j^2 \langle y, e_j\rangle^2 \leq 1 \right\}.$$

Proof. Since the situation is invariant to rotations, without loss of generality we may assume the ellipsoid $\mathcal{E}$ is axis-aligned, i.e.,

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^{n} \frac{x_i^2}{\alpha_i^2} \leq 1 \right\}.$$

Note that the ball

$$\mathcal{B}_D := \{x \in \mathbb{R}^n \mid \|x\|_D \leq 1\}$$

is equivalent to

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n \mid \|x\|_D^2 \leq 1 \right\},$$

where the norm $\|\cdot\|_D$ is defined by

$$\|x\|_D := (x'Dx)^{1/2} = \|D^{1/2}x\|,$$

with $D = \operatorname{diag}\left(\frac{1}{\alpha_1^2}, \ldots, \frac{1}{\alpha_n^2}\right)$. Applying proposition A.3, we have $\mathcal{E}^\circ = \mathcal{B}_D^\star$. It remains to find the dual norm which defines this ball. We have

$$\begin{aligned}
\|y\|_{D\star} &:= \sup_{x:\|x\|_D \le 1} \langle x,\, y\rangle \\
&= \sup_{x:\|D^{1/2}x\| \le 1} \langle x,\, y\rangle \\
&= \sup_{u:\|u\| \le 1} \left\langle D^{-1/2}u,\, y\right\rangle \\
&= \sup_{u:\|u\| \le 1} \langle u,\, v\rangle\,,
\end{aligned}$$

where $v = D^{-1/2}y$. By the Cauchy-Schwarz inequality, the supremum is $\|u\|\|v\|$ achieved by $u = cv$ where c is chosen such that $\|u\| \le 1$. Thus the supremum is $1 \cdot \|v\| = \|v\| = \|D^{-1/2}y\|$, and

$$\begin{aligned}
\mathcal{B}_D^\star &= \{y \in \mathbb{R}^n \mid \|y\|_{D\star} \le 1\} \\
&= \left\{y \in \mathbb{R}^n \mid \|y\|_{D\star}^2 \le 1\right\} \\
&= \mathcal{E}^\circ,
\end{aligned}$$

where $\mathcal{E}^\circ$ is as defined in the statement of the lemma. □

Lemma A.4 will be used in the proof of John's theorem regarding the contact points of the maximum volume ellipsoid. We conclude this aside into the polar of a set by noting that the polar of the polar is the original set when that set is convex. Thus the polar of the polar ellipsoid is the original ellipsoid.

Proposition A.5. *If C is a closed convex set with $0 \in \operatorname{int} C$, then $(C^\circ)^\circ = C$.*

Proof. Assume $x \in C$. Then $\langle y,\, x\rangle \le 1$ for all $y \in C^\circ$ by definition. However,

$$(C^\circ)^\circ = \{z \in \mathbb{R}^n \mid \langle z,\, y\rangle \le 1 \text{ for all } y \in C^\circ\}\,,$$

so $x \in (C^\circ)^\circ$. Conversely, assume that $z \in (C^\circ)^\circ$, so $\langle z,\, y\rangle \le 1$ for all $y \in C^\circ$. If $z \notin C$, then by theorem 2.1.3, there exists a hyperplane $\{y \in \mathbb{R}^n \mid \langle w,\, y\rangle = 1\}$ such that $\langle w,\, z\rangle > 1$ and $\langle w,\, x\rangle \le 1$ for all $x \in C$. This implies that $w \in C^\circ$, so $\langle z,\, w\rangle \le 1$, a contradiction. □

A.2 Proof of John's Theorem

With lemma A.4, we now prove theorem 2.3.1.

A.2.1 Sufficiency

The easier direction is to prove that if such a set of weights and contact points exist, then $\mathcal{B}_2^n$ is the unique ellipsoid of maximal volume, which we prove first. Let us assume that such a set of contact points and weights exist. Let

$$\mathcal{E} = \left\{ x : \sum_{i=1}^{n} \frac{\langle x, e_i\rangle^2}{\alpha_i^2} \le 1 \right\}$$

be an ellipsoid in $\mathcal{K}$ for some orthonormal basis $\{e_i\}_{i=1}^n$ and positive α_i. We want to show that

$$\prod_{i=1}^{n} \alpha_i \le 1,$$

from which it follows that the product equals 1 if and only if $\alpha_i = 1$ for all i, or equivalently that $\mathcal{E} = \mathcal{B}_2^n$.

Since $\mathcal{E} \subset \mathcal{K}$, for each i the hyperplane $\{x : \langle x, u_i\rangle = 1\}$ does not cut $\mathcal{E}$. This implies that each u_i belongs to the polar ellipsoid,

$$\mathcal{E}^\circ = \left\{ y : \sum_{j=1}^{n} \alpha_j^2 \langle y, e_j\rangle^2 \le 1 \right\}.$$

Thus, for each i, we have that

$$\sum_{j=1}^{n} \alpha_j^2 \langle u_i, e_j\rangle^2 \le 1.$$

Since

$$n = \mathrm{tr}(I_n) = \mathrm{tr}\left(\sum_{i=1}^{m} c_i \,\langle u_i,\, u_i\rangle \right) = \sum_{i=1}^{m} c_i \mathrm{tr}(\langle u_i,\, u_i\rangle) = \sum_{i=1}^{m} c_i,$$

it follows that

$$\sum_{i=1}^{m} c_i \sum_{j=1}^{n} \alpha_j^2 \langle u_i, e_j \rangle^2 \leq \sum_{i=1}^{m} c_i = n.$$

By condition (2.5), it follows that the left side of the above inequality is $\sum_j \alpha_j^2$ since e_j is a unit vector. Thus, the inequality is equivalent to

$$\frac{1}{n} \sum_{j=1}^{n} \alpha_j^2 \leq 1.$$

By the arithmetic-geometric mean inequality, we have

$$\left(\prod_{j=1}^{n} \alpha_j^2 \right)^{1/n} \leq \frac{1}{n} \sum_{j=1}^{n} \alpha_j^2 \leq 1,$$

which proves that $\mathcal{E} = \mathcal{B}_2^n$.

A.2.2 Necessity

We want to show that if $\mathcal{B}_2^n$ is the maximal volume ellipsoid contained in $\mathcal{K}$, then there exist contact points $\{u_i\}_{i=1}^m$ such that for each i, we have $u_i \in \mathcal{K} \cap \mathcal{B}_2^n$, and there exists $c_i > 0$ where

$$\sum_{i=1}^{m} \left(\frac{c_i}{n} \right) (u_i u_i^\top) = \frac{I_n}{n}.$$

If this is possible, we know by taking the trace that

$$\sum_{i=1}^{m} \frac{c_i}{n} = 1,$$

so we want to show that the matrix I_n/n can be written as a convex combination of a finite number of matrices of the form $uu^\top$. Equivalently, we need to show that I_n/n belongs to the convex hull of the set of all such rank one matrices,

$$T = \left\{ uu^\top \mid u \text{ is a contact point} \right\}.$$

We will prove this by showing a contradiction that if $I_n/n \notin T$, then we can perturb the unit ball slightly to get a new ellipsoid in $\mathcal{K}$ of larger volume than the unit ball.

Assume $I_n/n \notin T$. There exists a separating hyperplane in the spaces of matrices that separates I_n/n from T. Thus there is a linear functional $\phi(\cdot)$ on the spaces of matrices such that for each contact point u, we have

$$\phi\left(\frac{I_n}{n}\right) < \phi(uu^\top).$$

By letting the space of matrices in $\mathbb{R}^{n\times n}$ be represented by a vector in $\mathbb{R}^{n^2}$, ϕ may be represented by an $n \times n$ matrix H so that for any matrix A,

$$\phi(A) = \sum_{jk} H_{jk}A_{jk}.$$

Since the matrices $uu^\top$ and I_n/n are symmetric, we may assume that H is as well. Note that the matrices $uu^\top$ and I_n/n both have trace of 1, so the situation remains unchanged if we add a constant to each diagonal entry of H. Thus, we may assume that $\mathrm{tr}(H) = 0$, but this implies that $\phi(I_n) = 0$. Subsequently, unless the $\{u\}$ are the canonical basis of $\mathbb{R}^n$, we have found a symmetric matrix H with zero trace such that

$$\sum_{jk} H_{jk}(uu^\top)_{jk} > 0.$$

For each vector u, the left side of the above inequality is equivalent to $u'Hu$. For $\delta > 0$, the set

$$\mathcal{E}_\delta = \{x \in \mathbb{R}^n \mid x'(I_n + \delta H)x \leq 1\}$$

is an ellipsoid, and this set tends to $\mathcal{B}_2^n$ as δ tends to 0. If u is a contact point, then

$$u'(I_n + \delta H)u = 1 + \delta u'Hu > 1,$$

since $u'Hu > 0$. Since the boundary of $\mathcal{K}$ is compact and the map $x \mapsto x'Hx$ is continuous, $\mathcal{E}_\delta$ will not contain any other point of the boundary of $\mathcal{K}$ as long as δ is sufficiently small. For such δ, the ellipsoid $\mathcal{E}_\delta$ is strictly inside $\mathcal{K}$ and some slightly expanded ellipsoid is inside of $\mathcal{K}$.

It remains to prove that the volume of $\mathcal{E}_\delta$ is at least that of $\mathcal{B}_2^n$, i.e., $v_n(\mathcal{E}_\delta) \geq v_n(\mathcal{B}_2^n)$. Let λ_j denote the eigenvalues of the symmetric matrix $I_n + \delta H$. Then

$$v_n(\mathcal{E}_\delta) = \frac{v_n(\mathcal{B}_2^n)}{\prod_{j=1}^n \lambda_j},$$

so $v_n(\mathcal{E}_\delta) \geq v_n(\mathcal{B}_2^n)$ if $\prod_{i=1}^n \lambda_j \leq 1$. But $\mathrm{tr}(I_n + \delta H) = \sum_{j=1}^n \lambda_j = \mathrm{tr}(I_n) + \delta \mathrm{tr}(H) = n$ since $\mathrm{tr}(H) = 0$. Thus we have

$$1 = \frac{1}{n} \sum_{j=1} \lambda_j \geq \left(\prod_{j=1}^{n} \lambda_j \right)^{1/n},$$

where the last inequality follows from the arithmetic-geometric mean inequality, whence $\prod_{j=1}^n \lambda_j \leq 1$ as necessary.

Appendix B

THE SPECTRUM OF BOUNDED LINEAR OPERATORS

We recall the following definitions and theorems from functional analysis, and refer the reader to (Hunter and Nachtergaele, 2001) and (Rudin, 1991) for details. A real-valued *normed linear space* is a metric space with the metric derived its norm, i.e., $d(x, y) = \|x - y\|$. If a normed linear space is complete, it is known as a *Banach space*. If the norm is induced by an inner product, i.e., $\|x - y\| = \sqrt{\langle x,\, y\rangle}$, then the Banach space is known as a *Hilbert space*.

An *operator* between real-valued Banach spaces X, Y is a map $\mathsf{A} : X \to Y$. The operator A is said to be *linear* if for any $x_1, x_2 \in X$ and any $\alpha_1, \alpha_2 \in \mathbb{R}$, we have

$$\mathsf{A}(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 \mathsf{A}(x_1) + \alpha_2 \mathsf{A}(x_2).$$

A linear operator A is said to be *bounded* if there exists a constant M such that

$$|\mathsf{A}(x)| \leq M\|x\| \quad \text{for all } x \in X.$$

The *norm* of a bounded linear operator is defined as

$$\|\mathsf{A}\| \equiv \sup_{x \in X:\, \|x\| \neq 0} \frac{\|\mathsf{A}(x)\|}{\|x\|}.$$

An important theorem is which allows us to further characterize bounded linear operators is the following.

Theorem B.1 (Riesz Representation Theorem)**.** *If* A *is a bounded linear operator on a separable Hilbert space* H*, then for every* $x \in H$*, there exists a* $y \in H$ *such that*

$$\mathsf{A}(x) = \langle x,\, y\rangle$$

Moreover, the operator norm satisfies $\|\mathsf{A}\| = \|y\|$.

The *adjoint* of a bounded linear operator on a Hilbert space is defined as in the following theorem, which results from the Riesz representation theorem.

Theorem B.2 (Existence and Uniqueness of the Adjoint Operator)**.** *If* A *is a bounded linear operator on a Hilbert space* H*, then there is a unique operator* A^* *on* H *such that*

$$\langle x,\, \mathsf{A}^* y\rangle = \langle \mathsf{A}x,\, y\rangle \qquad \text{for all } x, y \in H.$$

Furthermore, A^* *is linear and bounded,* $\|\mathsf{A}^*\| = \|\mathsf{A}\|$*, and* $(\mathsf{A}^*)^* = \mathsf{A}$*. The operator* A^* *is known as the adjoint of* A*.*

In the case that $\mathsf{A} = \mathsf{A}^*$, we say that A is a *self-adjoint operator*.

We now characterize the *resolvent* and the *spectrum* of bounded linear operators on a Hilbert space.

Definition B.3 (Resolvent and Spectrum)**.** *The resolvent set of an a bounded linear operator* A *on a Hilbert space* H*, denoted by* $\rho(\mathsf{A})$*, is the set of complex numbers* λ *such that* $(\mathsf{A} - \lambda \mathsf{I}) : H \mapsto H$ *is a bijection, where* I *denotes the identity operator. The spectrum of* A*, denoted by* $\sigma(\mathsf{A})$*, is the complement of the resolvent set in* $\mathbb{C}$*, i.e.,* $\sigma(\mathsf{A}) = \mathbb{C} \setminus \rho(\mathsf{A})$*.*

The *open mapping theorem* states that a bounded linear operator A which is surjective is an open map. Thus the resolvent set is open, and the spectrum is closed. When $\lambda \in \rho(\mathsf{A})$, both $(\mathsf{A} - \lambda\mathsf{I})$ and $(\mathsf{A} - \lambda\mathsf{I})^{-1}$ are bijective bounded linear operators. Thus the spectrum may $\sigma(\mathsf{A})$ may also be characterized as the set of $\lambda \in \mathbb{C}$ such that $(\mathsf{A} - \lambda\mathsf{I})^{-1}$ does not exist as a bounded linear operator.

We may relate the resolvent set and the spectrum to the operator norm via the following theorem.

Theorem B.4. *If* A *is a bounded linear operator on a Hilbert space, then the resolvent set* $\rho(\mathsf{A})$ *is an open subset of* $\mathbb{C}$ *that contains the "exterior disk"* $\{\lambda \in \mathbb{C} \mid |\lambda| > \|\mathsf{A}\|\}$*. Equivalently, the spectrum* $\sigma(\mathsf{A})$ *is a closed subset of* $\mathbb{C}$ *which is contained in* $\{\lambda \in \mathbb{C} \mid |\lambda| \leq \|\mathsf{A}\|\}$*.*

The *spectral radius* of a bounded linear operator is defined as follows.

Definition B.5 (Spectral Radius)**.** *The spectral radius of a bounded linear operator* A *is defined as*

$$r(\mathsf{A}) \equiv \sup \left\{ |\lambda| \;\middle|\; \lambda \in \sigma(\mathsf{A}) \right\}.$$

Theorem B.4 may be refined as follows.

Theorem B.6 (Gelfand's Formula)**.** *If* A *is a bounded linear operator, then*

$$r(\mathsf{A}) = \lim_{n\to\infty} \|\mathsf{A}^n\|^{1/n}.$$

If A is self-adjoint, then $r(\mathsf{A}) = \|\mathsf{A}\|$.

Finally, we have the following result which characterizes the spectrum of bounded self-adjoint operators and is analogous to the finite-dimensional case.

Theorem B.7. *If* A *is a bounded, self-adjoint operator on a Hilbert space, then the spectrum of* A *is real-valued and is contained in the closed interval* $[-\|\mathsf{A}\|, \|\mathsf{A}\|]$.

VITA

Adam Marc Gustafson was born in Phoenix, Arizona, and was raised there as well as in Springfield, Illinois. He majored in electrical engineering for bachelors studies at Tufts University, where he graduated summa cum laude, and masters studies at University of Illinois. His interest in probability and statistics arose from his enjoyment of his first random processes course, upon which he decided to pursue a doctoral degree in statistics at University of Washington. In his spare time, he enjoys playing classical and jazz guitar, reading about world affairs and politics, alpine skiing, and practicing yoga.