Rates of Convergence in the Central Limit Theorem for Markov Chains

Your Name Goes Here, Ph.D.
University of Connecticut, 2025

ABSTRACT

Put your abstract here. Try to use as little symbols as possible. This abstract will be typeset for web publishing and various databases (not by you), and symbols don't transfer well.

Rates of Convergence in the Central Limit Theorem for Markov Chains

Your Name Goes Here

M.Sc. Mathematics, Universiteit Antwerpen, Antwerp, Belgium, 1992
M.Sc. Theoretical Physics, Université Libre de Bruxelles, Brussels, Belgium, 1993

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2025

APPROVAL PAGE

Doctor of Philosophy Dissertation

Rates of Convergence in the Central Limit Theorem for Markov Chains

Presented by
Your Name Goes Here, M.Sc. Math., M.Sc. Th.Phys.

Major Advisor ______________________________
Richard Bass

Associate Advisor ______________________________
Evarist Giné-Masdéu

Associate Advisor ______________________________
Yung-Sze Choi

University of Connecticut
2025

ACKNOWLEDGEMENTS

At this point, I would like to thank the many people who - directly or indirectly - have contributed to this dissertation. ....

TABLE OF CONTENTS

NOTATION

(This section should be for reference only and most of these symbols should be defined in the text)
First recall the conventional multiindex notation. Let $\mathbb{N}_0$ denote the nonnegative integers. A multiindex $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}_0^n$ is an n-tuple of nonnegative integers. The degree of a multiindex is $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. Multivariate exponentiation is defined as follows. For $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

Here we will use the convention $0^0 = 1$ so that for example $(x, y, z)^{(0,1,0)} = y$. We will use $e_i \in \mathbb{R}^n$ to represent the canonical unit vector

$$e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$

which can also be interpreted as a multiindex so for example

$$(x, y, z)^{2e_1+e_3} = x^2 z$$

The multinomial formula gives a convenient expansion for multinomial powers. Let $k \in \mathbb{N}_0$. Then

$$(b_1 + b_2 + \cdots + b_n)^k = \sum_{|\alpha|=k} \binom{k}{\alpha} b^\alpha$$

where the multinomial coefficients are defined

$$\binom{k}{\alpha} = \frac{k!}{\alpha_1!\alpha_2!\cdots\alpha_n!}.$$

Note that the multinomial expansion sums over all multiindeces $\alpha \in \mathbb{N}_0^n$ of degree k.
We denote the standard euclidean inner product

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

where $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ and the Euclidean norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

We may slightly abuse notation by using the inner product and norm in different dimensions, even in the same equation. This can be excused if one views each euclidean space $\mathbb{R}^n$ as embedded in the sequence space $\ell^2(\mathbb{R}^{\mathbb{N}})$ in which the inner product and norm are equivalent.

When discussing hyperplanes in $\mathbb{R}^n$ we index them by a unit normal vector, $\omega \in S^{n-1}$, and distance from origin $-\infty < p < \infty$, and we write for example the implicit hyperplane equation $\langle x, \omega \rangle = p$. Note that $\langle x, -\omega \rangle = -p$ represents the same hyperplane. It is perhaps more correct, in later defining the Radon and Gaussian Radon transforms, to identify these indexes and define the transforms over a projective space. We omit this discussion as it is not relevant within the scope of this work.

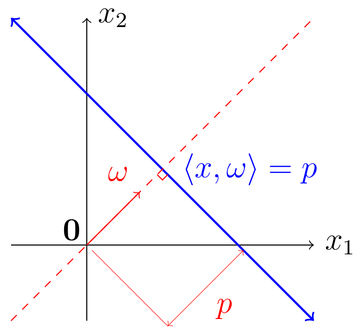

Figure 0.0.1: Hyperplane equation

Unless otherwise indicated all measures are Euclidean measures, that is the Borel measure assocciated with the standard Euclidean metric on a given space. In particular the Euclidean measure on a hyperplane of $\mathbb{R}^n$ is equivalent to the Euclidean measure on $\mathbb{R}^{n-1}$. We chose for convenience to denote measures by their associated variable such as dx, dp, dz, and others. It should be stated that this notation, while uniform, is context dependent. For example in the integrals

$$\int_{\mathbb{R}^n} dx \qquad \text{and} \qquad \int_{\langle x, \omega \rangle = p} dx,$$

the measure dx is to be understood as the n-dimensional and $(n-1)$-dimensional Euclidean measure respectively.

Introduction

Introduction

0.1 Shape Reconstruction

Here is a sort of to-do list for whatever this document is:

- To write:
  - ~~Document. Transfer into the thesis template.~~
  - Document. Fix bibliography
  - ~~Everywhere. Remove the term classical.~~
  - ~~Intro. Rephrase stuff about numerical “in theory”.~~
  - Intro. Add formulas to intro.
  - ~~Intro. BE MORE PRECISE~~
  - Moments. Prove: Bounded regions are uniquely determined by moments? Maybe fits in RT/GRT since it can be formulated in terms of RT.
  - GF/M. Define Hermite polynomials, uni- and multivariate.
  - ~~GF/M. Example: Integral of Gaussian, and even multivariable moments.~~
  - RT/GRT. Work on notation for Gaussian measure. In particular w and ω are too similar. ~~Also the Euclidean norm is inconsistent when going from~~ $\cancel{\mathbb{R}^n}$ ~~to~~ $\cancel{\mathbb{R}^{n-1}}$~~. We think the notation is understandble, but it might be more precise to disambiguate norms in different dimensions? Possible that they can all be dexribed in terms of a seminorm on some space in which each~~ $\cancel{\mathbb{R}^n}$ ~~is embedded?~~
  - RT/GRT. Search for references on G slice theorem and G projection moments etc
  - RT/GTT. Expand background, historical context, etc…
  - ~~RT/GRT. Swap sufficient working condition with Fubini condition in slice theorems.~~
  - Implementation. Output some figures from Mathematica. For example sample points, moment tables, and reconstructions.

- Example. Try to prove the conjecture (RT Hermite)

- To research:
  - Everywhere (ongoing). Provide examples for $n = 2, 3$ of as many results as we can.
  - GRT. Domains of defintion, i.e. $GRT : L^2 \to L^2$ and such.
  - GRT. Integration by parts formula for GRT of $\partial f / \partial x_i$ (found dR/dp and thus dGR/dp)
  - GRT. Prove the conjecture on GRT of Hermite polynomials.
  - Mathematica. Implement Gaussian shape reconstruction method, try some example shapes. Clearly lay out the choices made: Series of expanding images, or compactified $\mathbb{R}^n$? How to complexify points?

The shape reconstruction method: In a 2005 paper, Annie Cuyt et. al. [?Cuyt05] proposed a method for shape reconstruction from moments, via Pade approximants to a multidimensional integral transform. Given a set of multivariate moments of some region A in $\mathbb{R}^n$, the method produces a pixel image approximating A.
Suppose $A \subset \mathbb{R}^n$, and let $f(x)$ be its indicator function. We may assume f is measurable and has bounded support, and thus finite moments. The moment sequence gives Taylor coefficients in a neighborhood of zero for a certain holomorphic function with integral representation

$$g(y) = \int_{\mathbb{R}^n} \frac{f(x)}{1 + \langle x, y \rangle} \, dx \approx \sum_{\alpha \in \mathbb{N}_0^n} \binom{|\alpha|}{\alpha} c_\alpha (-y)^\alpha.$$

Thus the function $g(y)$ may be approximated to a certain degree of accuracy depending on the number of available moments.
At the same time g can be approximated by a multivariate quadrature formula

$$\int_{\mathbb{R}^n} \frac{f(x)}{1 + \langle x, y \rangle} \, dx = \sum_i \frac{1}{1 + \langle x_i, y \rangle} f(x_i) = \sum_i w(x_i, y) f(x_i)$$

Where the nodes x_i lie, for example, on a cubic lattice. We can then sample a Pade approximant to g at some sufficinetly large group of points $y = y_j$ forming a linear system of equations, from which we solve for $f(x_i)$. If all goes well we have approximations of $f(x_i)$ on a node lattice, which can be turned into a pixel image of A.
For now we will gloss over the numerical discussion of quadrature and linear systems, taking for granted that such methods are applicable in at least some simple cases. Computational tests (to be included in, say, section 10) further support the validity of the method. Our focus will be on demonstrating that one can approximate g by rational Padé approximants constructed from moments. To this end we show that, when restricted to one dimensional subspaces, g is equivalent to the Stieltjes transform

of the Radon transform of f at a fixed projection angle.

$$g(z\omega) = \int_{-\infty}^{\infty} \frac{R_f(\omega, p)}{1 + zp} dp$$

where $\omega \in S^{n-1}, z \in \mathbb{R}$. Thus we are able to leverage the more well understood properties of these univariate transforms.

In particular, the "projection moments" of $R_f(\omega, p)$ in any fixed direction ω, may be computed from the multivariate moments of f. Further, it can be shown that Padé approximants to a moment sequence converge to the Stieltjes transform, and thus to g. Returning to shape reconstruction, it then follows that one can approximate g at various sample points $y_j = z_j \omega_j$ via Padé approximants on linear subspaces at a selection of angles ω, as is needed to form our linear system.

Now let us briefly discuss our proposed modification to the shape reconstruction method, the theoretical complications and computational considerations. Essentially what we propose is to recreate the method with a Gaussian measure applied, thus allowing for convergence on a larger class of regions A. In particular, the proposed method would apply to unbounded regions, for which we cannot guarentee finite moments, convergence of the function $g(y)$ (let alone the explicit series expansion in terms of moments), or convergence of the Radon transform.

Here we will briefly discuss the potential challenges that this proposed method presents, which we address more carefully in sections 5 through 8. First and foremost instead of the standard moments (with respect to the Lebesgue measure), we are now given a Gaussian moment sequence — that is, moments with respect to the standard Gaussian measure on $\mathbb{R}^n$. On the other end, the reconstructed region may be recovered as normal by simply inverting the Gaussian weight.

From a theoretical standpoint, proving the validity of the proposed method presents a few challenges. Firstly, even a quick glance at the classical literature tells us that moment problems on bounded domains are substantially better behaved than their unbounded counterparts. Whereas all moment problems on a bounded interval are determinate, we now have the potential to run in to indeterminate problems. We show that we can guarentee determinacy in a Gaussian-weighted L^2 completion of the space of polynomials. This of course includes our primary use case: indicator functions.

Secondly, we must address the effect of the proposed modifications on the Radon transform. We will define the Gaussian Radon transform, which unlike the Radon transform exists for unbounded regions A.

The Stieltjes transform on unbounded domains is significantly more complicated than the original case. From the integral representation

$$g(y) = \int_{\mathbb{R}^n} \frac{f(x)}{1 + \langle x, y \rangle} \, dx,$$

one can start to see a problem: If $f(x)$ has unbounded support $g(y)$ has the potential to not converge for any $y \in \mathbb{R}^n$. In particular $g(y)$ may not exist in a neighborhood of 0, meaning our series expansion and thus the explicit conection to moments, may be broken. We need now to consider g as a function on some non-real domain. Here we show Pade approximants can be computed from moments which still approximate g on, for example, a half space. The convergence of these approximants is tied to the question determinacy of the moment problem, which we resolve as described above. Thus we simply take sample points avoiding the problematic real space and the method is validated.

Chapter 1

Multivariate moment problems and the Radon transform

1.1 Classical and multivariate moment problems

Let $f(x)$ be a measurable function on $\mathbb{R}$. For $k \in \mathbb{N}_0$, define the kth moment of f as

$$c_k = \int_{-\infty}^{\infty} f(x) x^k \, dx.$$

The sequence $(c_k)_{k \in \mathbb{N}_0}$ is called a moment sequence or a moment problem, and the function f is called a solution to the moment problem. Loosely speaking a moment problem poses the question: Under given constraints (e.g. domain, continuity, etc…), to what extent can one determine the solution f from its moments?
The classical study of moment problems is divided into three cases depending on the domain:

(i) Markov (or Haussdorf) moment problems for bounded domains (i.e. the unit interval),

(ii) Stieltjes moment problems for one-sided unbounded domains (i.e. the positive real line), and

(iii) Hamburger moment problems for bi-infinite domains (i.e. the real line).

There are two natural questions one can ask about a moment problem:

(i) Solvability: Does a solution f exist possessing the given moments?

(ii) Determinacy: Is a solution unique? If not, what can be said about the set of solutions?

In the classical cases (Markov, Stieltjes, Hamburger) these questions have been more or less resolved. Precise conditions for solvable and determinate moment problems have been found, and the nature of solution sets to indeterminate moment problems are well understood. For instance, all solvable Markov moment problems are unique.

Proposition 1.1.1. Let μ, ν be Borel measures on the unit interval $[0,1]$ with equal moments:

$$\int_{[0,1]} x^k d\mu = c_k = \int_{[0,1]} x^k d\nu$$

for all $k \in \mathbb{N}_0$. Then $\mu = \nu$.

However many classes of generalized moment problems remain in active study.
In particular let us discuss multivariate moment problems. Let $f(x)$ be a measureable function now on $\mathbb{R}^n$. For $\alpha \in \mathbb{N}_0^n$, define the αth multivariate moment of f as

$$c_\alpha = \int_{\mathbb{R}^n} f(x) x^\alpha \, dx.$$

The moment sequence $(c_\alpha)_{\alpha \in \mathbb{N}_0^n}$ is now multiindexed. No precise general conditions are known for the solvability or determinacy of multivariate moment problems. However some sufficient conditions have been discovered by leveraging classical moment problem theory [1].

Example 1.1.2. Let $A = [0,1]^n \subseteq \mathbb{R}^n$ be the unit cube and f its characteristic function. Then f has multivariable moments

$$\begin{aligned}\int_{\mathbb{R}^n} f(x) x^\alpha \, dx &= \int_0^1 x_1^{\alpha_1} \, dx_1 \int_0^1 x_2^{\alpha_2} \, dx_2 \cdots \int_0^1 x_n^{\alpha_n} \, dx_n \\ &= \frac{1}{(\alpha_1 + 1)(\alpha_2 + 1) \cdots (\alpha_n + 1)}\end{aligned}$$

Similarly for any rectanglular region $A = \prod_{i=1}^n [a_i, b_i]$

$$\int_A x^\alpha \, dx = \prod_{i=1}^n \int_{a_i}^{b_i} x_i^{\alpha_i} = \prod_{i=1}^n \frac{b_i^{\alpha_i+1} - a_i^{\alpha_i+1}}{\alpha_i + 1}$$

We claim that multivariate moment problems on bounded domains are determinate, provided a solution exists. Without loss of generality:

Claim 1.1.3. If μ is a Radon measure on the unit cube $A = [0,1]^n$, with moments

$$c_\alpha = \int_A x^\alpha d\mu, \qquad \alpha \in \mathbb{N}_0^n$$

then the moment sequence $(c_\alpha)_{\alpha \in \mathbb{N}_0^n}$ is determinate.

1.2 Gaussian functions and measures

Definition 1.2.1. Let $w(p) : \mathbb{R} \to \mathbb{R}$ be the standard Gaussian density on $\mathbb{R}$,

$$w(p) = \frac{1}{\sqrt{2\pi}} e^{\frac{-p^2}{2}}$$

and $w_n(x) : \mathbb{R}^n \to \mathbb{R}$ the standard Gaussian density on $\mathbb{R}^n$,

$$w_n(x) = (2\pi)^{-\frac{n}{2}} e^{\frac{-\|x\|^2}{2}}.$$

Note that $w = w_1$.

A couple of formula: First the Gaussian integral,

$$\int_{-\infty}^{\infty} w(p)\, dp = 1.$$

This can be proven in a number of clever ways.
Second: The derivative

$$w'(p) = -pw(p).$$

Third: The Gaussian function dominates polynomials in the sense that

$$\lim_{p \to -\infty} P(p)w(p) = 0 = \lim_{p \to \infty} P(p)w(p)$$

for any polynomial P.

Definition 1.2.2. The *Gaussian moments* of a function $f : \mathbb{R}^n \to \mathbb{R}$ are

$$c_\alpha^G = \int_{\mathbb{R}^n} f(x) x^\alpha w(x)\, dx$$

for $\alpha \in \mathbb{N}_0^n$.

Example 1.2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be the constant function $f(p) = 1$. The Gaussian moments of f, i.e. the moments of $w(p)$, are

$$c_k^G = \begin{cases} (k-1)!!, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

We have already seen $c_0 = 1$. Furthermore it is not hard to see that for odd k,

$$c_k^G = \int_{-\infty}^{\infty} p^k w(p)\, dp = 0$$

since $p^k w(p)$ is an odd function.
Now for $k \geq 2$ even, we derive a recurrence by integrating by parts

$$\begin{aligned} c_k^G &= \int_{-\infty}^{\infty} p^k w(p)\, dp \\ &= -\int_{-\infty}^{\infty} p^{k-1} \left(-pw(p)\right)\, dp \\ &= -\left[p^{k-1} w(p)\right]_{-\infty}^{\infty} + (k-1) \int_{-\infty}^{\infty} p^{k-2} w(p)\, dp \\ &= (k-1) c_{k-2}^G. \end{aligned}$$

Thus by induction,

$$c_k^G = (k-1)(k-3)\cdots c_0 = (k-1)!!.$$

A relationship between the univariate and multivariate Gaussian densities can be seen as follows: Let $x, y \in \mathbb{R}^n$, $n \geq 2$. Suppose $y = p\omega$ where $\omega \in S^1$ and $p \in \mathbb{R}$, so that $p\omega$ is the orthogonal projection of x onto the span of y. Then the Pythagorean relation,

$$\|x\|^2 = \|x - p\omega\|^2 + \|p\omega\|^2 = \|x - p\omega\|^2 + p^2$$

implies that

$$\begin{aligned} w_n(x) &= (2\pi)^{-\frac{n}{2}} e^{\frac{-\|x\|^2}{2}} \\ &= (2\pi)^{-\frac{n-1}{2}} e^{\frac{-\|x-p\omega\|^2}{2}} (2\pi)^{-\frac{1}{2}} e^{\frac{-p^2}{2}} \\ &= w_{n-1}(x - p\omega) w(p) \end{aligned}$$

The equation

$$w_n(x) = w_{n-1}(x - p\omega)w(p) \tag{1.2.1}$$

is in some respect the defining property of the Gaussian measure described below. Indeed if $x = (x_1, x_2, \ldots, x_n)$, by repeated application of (1.2.1) one can write the decomposition

$$w_n(x) = \prod_{k=1}^{n} w(x_k).$$

Thus the $w_n(x)$ is the product of n copies of w. The standard Gaussian measure γ^n is thus a measure whose cardinal projections are standard Gaussian measures.

Definition 1.2.4. Let γ^n be a the Borel measure on $\mathbb{R}^n$ with density w_n,

$$\int_{\mathbb{R}^n} f(x) d\gamma^n = \int_{\mathbb{R}^n} f(x) w_n(x)\, dx.$$

We call γ^n the *standard Gaussian measure*. More generally, for mean $a \in \mathbb{R}^n$, and variance $\sigma > 0$, the following γ^n_{a,σ^2} are *Gaussian measures*,

$$\begin{aligned} \int_{\mathbb{R}^n} f(x) d\gamma^n_{a,\sigma^2} &= \frac{(2\pi)^{-\frac{n}{2}}}{\sigma} \int_{\mathbb{R}^n} f(x) e^{-\frac{\|x-a\|^2}{2\sigma^2}}\, dx \\ &= \int_{\mathbb{R}^n} f(\sigma x + a) d\gamma^n \end{aligned}$$

Example 1.2.5. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the constant function $f(x) = 1$. The Gaussian moments of f are the multivariate moments of $w_n(x)$, and can be written as

$$\begin{aligned} c_\alpha^G &= \int_{\mathbb{R}^n} x^\alpha w_n(x)\, dx \\ &= \prod_{\ell=1}^{n} \int_{-\infty}^{\infty} x_\ell^{\alpha_\ell} w(x_\ell)\, dx \\ &= \prod_{\ell=1}^{n} c_{\alpha_\ell}^G \end{aligned}$$

where $\{c_k\}_{k\in\mathbb{N}}$ are the univariate moments of $w(p)$. Thus $c_\alpha^G = 0$ if any α_i is odd. Otherwise, if every α_i is even then

$$c_\alpha^G = \prod_{\ell=1}^{n} (\alpha_\ell - 1)!!$$

Definition 1.2.6. The *Hermite polynomials* can be defined by the Rodriguez formula

$$H_k(p) = (-1)^k \frac{w^{(k)}(p)}{w(p)}$$

note that H_k is a polynomial of degree k.

The Hermite polynomials have the following properties:

$$\int_{-\infty}^{\infty} H_k(p) H_\ell(p) w(p)\, dp = \begin{cases} \sqrt{2\pi} k!, & k = \ell \\ 0, & k \neq \ell \end{cases}$$

The Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}, w)$.

1.3 The Radon and Gaussian Radon transforms

Let f be a multivariable function on the n-dimensional Euclidean space $\mathbb{R}^n$. We imagine taking "slices" of f by restricting it to a $(n-1)$-dimensional hyperplane Λ. These hyperplanes form the domain of the Radon Transform. More precisely, the Radon Transform associates each slice with a corresponding integral

$$R_f(\Lambda) = \int_\Lambda f(x)\, dx,$$

which can be thought of as a $(n-1)$-dimensional measurement of f. To be more precise we parametrize the collection of hyperplanes Λ by normal vector, $\omega \in S^{n-1}$, and (signed) distance from the origin $-\infty < p < \infty$. Indeed any hyperplane can be described in the form $\Lambda = \{x \in \mathbb{R}^n : \langle x, \omega \rangle = p\}$. As mentioned, there is a slight inconsistency in these definitions where a hyperplane Λ can be indexed by both $\langle \omega, x \rangle = p$ and $\langle -\omega, x \rangle = -p$. This difference is inconsequential for our purposes so we choose the latter definition for clarity.

The Radon transform [?Helg65] (RT) is a thing that I will discuss the history of, with references, in this paragraph. The transform gets its name from Johann Radon, whose first defined the tranform in the form below in 1917, although a similar transform was introduced by Paul Funk in 1911 [?????]. It is interesting to note that the primary application of the RT in medical imaging (CT scans) was not invented for another half century. I can only hope that in 2076 my dissertation will serve as an absorbant coffee coaster for a sleep deprived student.

Definition 1.3.1. Let f be a nonnegative measurable function on $\mathbb{R}^n$. The *Radon transform* $R_f : S^{n-1} \times \mathbb{R} \to \mathbb{R}$ of f is a function which, given a unit vector $\omega \in \mathbb{R}^n$ and $-\infty < p < \infty$, is defined as

$$R_f(\omega, p) = \int\limits_{\langle x,\omega\rangle = p} f(x)\, dx,$$

provided the integral converges.

Example 1.3.2. For computation it is convenient to write the Radon Transform with an explicit isometric parameterization $x(t)$ of the hyperplane $\langle x, \omega\rangle = p$. In particular we note that there exists a map $x : \mathbb{R}^{n-1} \to \mathbb{R}^n$ such that $x(0) = p\omega$ and

$$R_f(\omega, p) = \int_{\mathbb{R}^{n-1}} f(x(t))dt.$$

For reference let's specify an hyperplane parameterization for the $n = 2$ case. In $\mathbb{R}^2$ often we identify ω with the angle $0 \le \theta < 2\pi$ such that $\omega = (\cos\theta, \sin\theta)$. We define $x(t)$ by

$$x(t) = (t\sin\theta + p\cos\theta, -t\cos\theta + p\sin\theta), \qquad -\infty < t < \infty.$$

It is not difficult to show that $x(0) = p\omega$, $\langle x(t), \omega\rangle = p$, and most importantly

$$R_f(\omega, p) = \int_{-\infty}^{\infty} f(t\sin\theta + p\cos\theta, -t\cos\theta + p\sin\theta)\, dt$$

Example 1.3.3. Let $B(r) = \{x : |x| \le r\} \subseteq \mathbb{R}^n$ be the ball of radius r.

$$R_{B(r)}(\omega, p) = \begin{cases} V_{n-1}\left(\sqrt{r^2 - p^2}\right), & |p| \le r \\ 0 & |p| > r \end{cases}$$

where $V_n(r)$ is the volume of a hypersphere of radius r,

$$V_n(r) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n.$$

Note that as a bounded domain, we can guarentee $B(r)$ is integrable on all hyperplanes. In fact the RT of any subset of $B(r)$ is sharply bounded by $R_{B(r)} \le V_{n-1}(r)$. Also note that $B(r)$ is rotation invariant, hence $R_{B(r)}(\omega, p)$ is independent of ω.

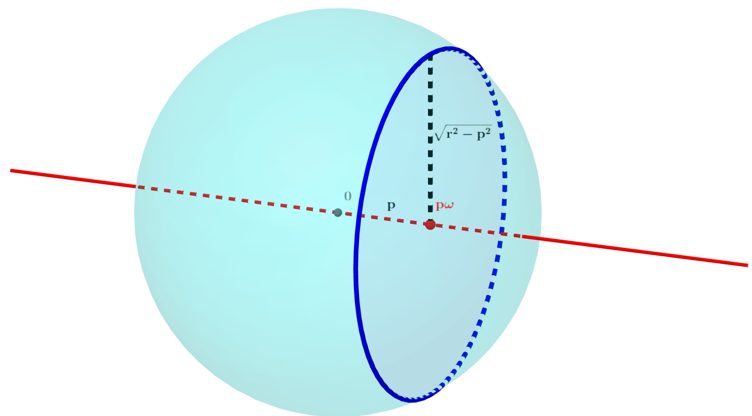

Figure 1.3.1: The RT of a ball

We can use this formula to determine the RT of an anulus. Let $A(r_1, r_2) = \{x \in \mathbb{R}^n : r_1 \leq \|x\| \leq r_2\}$. Then

$$\begin{aligned} R_{A(r_1,r_2)}(\omega, p) &= R_{B(r_2)}(\omega, p) - R_{B(r_1)}(\omega, p) \\ &= \begin{cases} V_{n-1}(\sqrt{r_2^2 - p^2}) - V_{n-1}(\sqrt{r_1^2 - p^2}), & |p| < r_1 \\ V_{n-1}(\sqrt{r_2^2 - p^2}), & r_1 \leq |p| \leq r_2 \\ 0, & |p| \geq r_2 \end{cases} \end{aligned}$$

Now consider an example of an unbounded region.

Example 1.3.4. Let $S = \{(x, y) : |y| \leq \frac{1}{2}\} \subseteq \mathbb{R}^2$ be a strip centered on the x-axis with width 1. Clearly if $\theta = \pi/2$ or $3\pi/2$ then

$$R_S(\theta, p) = \begin{cases} \infty, & |p| \leq \frac{1}{2} \\ 0, & |p| > \frac{1}{2} \end{cases}.$$

Otherwise,

$$R_S(\theta, p) = \sec\theta$$

Note because S is unbounded, that R_S is not only unbounded, but even divergent in some cases.

Our main addition to previous work will be the use of a modified RT, the Gaussian Radon transform (GRT). This transform is very similar to the RT, but the inclusion of a Gaussian density $w_{n-1}(x)$ in the integral allows for convergence on a larger class of functions f. This includes for example unbounded regions. In a broader context the Gaussian Radon transform also has the advantages of generalizng to infite dimensional

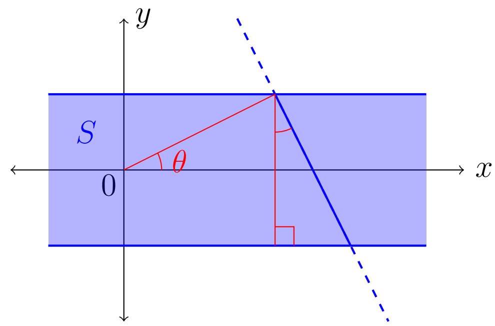

Figure 1.3.2: RT of a strip

Hilbert spaces [?Seng14] (on which the Lebesgue measure is not defined), as well as having a natural probabilistic interpretation.

Definition 1.3.5. The *Gaussian Radon transform* $GR_f : S^{n-1} \times \mathbb{R} \to \mathbb{R}$ of f is defined similarly to the RT. Given $\omega \in S^{n-1}$ and $-\infty < p < \infty$, the GRT is

$$GR_f(\omega, p) = \int\limits_{\langle x,\omega\rangle = p} f(x) w_{n-1}(x - p\omega)\, dx.$$

provided the integral converges. Note the Gaussian density

$$w_{n-1}(x - p\omega) = (2\pi)^{-(n-1)/2} e^{-\|x-p\omega\|^2/2}$$

is centered on the point of the hyperplane closest to the origin.

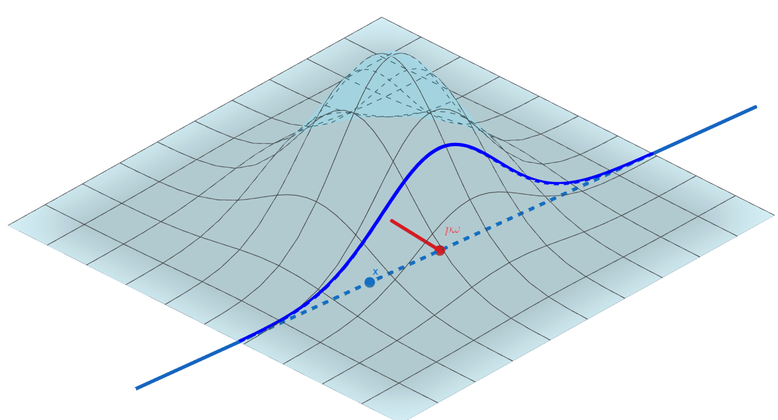

Figure 1.3.3: The GRT

Remark 1.3.6. It may be helpful to understand the GRT as a simple modification of the RT with respect to a Gaussian measure on $\mathbb{R}^n$. From the relation

$$\int\limits_{\langle x,\omega\rangle=p} f(x)w_n(x)\,dx = \int\limits_{\langle x,\omega\rangle=p} f(x)w_{n-1}(x)\,dx\; w(p),$$

we can express the GRT of f in terms of the RT of the function $g(x) = f(x)w_n(x)$:

$$R_g(\omega,p) = GR_f(\omega,p)w(p), \qquad g(x) := f(x)w_n(x). \tag{1.3.1}$$

The relation above provides decent intuition for the GRT, and is also a useful tool proving some basic properties of the transform.

Example 1.3.7. If $x(t) : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}^n$ is a parametrization of $\langle x,\omega\rangle = p$ as described above, then

$$GR_f(\omega,p) = \int_{\mathbb{R}^{n-1}} f(x(t))w_{n-1}(t)dt$$

In particular for $f : \mathbb{R}^2 \to \mathbb{R}$

$$GR_f(\omega,p) = \int_{-\infty}^{\infty} f(t\sin\theta + p\cos\theta, -t\cos\theta + p\sin\theta)w(t)\,dt$$

where $\omega = (\cos\theta, \sin\theta)$.

Example 1.3.8. The Gaussian Radon transform is bounded for any measurable region $A \subseteq \mathbb{R}^n$. Indeed

$$GR_A(\omega,p) \le GR_1(\omega,p) = \int_{\mathbb{R}^{n-1}} w_{n-1}(t)dtw(p) = w(p)$$

Now imagine sweeping a hyperplanar "slice" across $\mathbb{R}^n$. As a function of p, $R(\omega,p)$ can be seen as a projection of f onto the linear subspace spanned by ω. It is not surprising that integrating this projection over $-\infty < p < \infty$ we get the same result as the n-fold integral of f over $\mathbb{R}^n$.

$$\int_{-\infty}^{\infty} R(\omega,p)\,dp = \int_{\mathbb{R}^n} f(x)\,dx$$

The so called "slice theorem" further generalizes this observation.

Proposition 1.3.9 (Slice Theorem)**.** If $f \in L^1(\mathbb{R}^n)$ and $F(p) \in L^\infty(\mathbb{R})$ then

$$\int_{-\infty}^{\infty} R_f(\omega,p)F(p)dp = \int_{\mathbb{R}^n} f(x)F(\langle x,\omega\rangle)dx. \tag{1.3.2}$$

Proof. Inserting the definition of the RT, the left side is

$$\int_{-\infty}^{\infty}\int\limits_{\langle x,\omega\rangle=p} f(x)\,dx\;F(p)\,dp = \int_{-\infty}^{\infty}\int\limits_{\langle x,\omega\rangle=p} f(x)F(\langle x,\omega\rangle)\,dxdp.$$

Up to a rigid transformation (under which the Euclidean measures are invariant) this is essentially an itterated integral over $\mathbb{R}$ and $\mathbb{R}^{n-1}$. Thus given the integrability requirement, Fubini's theorem applies and the slice theorem is proved. $\square$

Remark 1.3.10. The sufficient conditions for the slice theorem above can loosened significantly. We can take for example the straightforward Fubini condition

$$\int_{\mathbb{R}^n} |f(x)F(\langle x,\omega\rangle)| < \infty$$

which is necessary, but not particularly verifiable. We may also use the condition f has bounded support and $F \in L^1_{loc}(\mathbb{R})$.

If $F(p) = e^{-ip}$ and $f(x)$ is such that $\int_{-\infty}^{\infty} R_f(\omega,p)dp < \infty$ then (1.3.2) becomes the well known Fourier slice theorem

$$\int_{-\infty}^{\infty} R_f(\omega,p)e^{-ip}\,dp = \int_{\mathbb{R}^n} f(x)e^{-i\langle x,\omega\rangle}\,dx,$$

which is often articulated as saying that the 1-dimensional Fourier transform of the Radon transform is the n-dimenstional Fourier transform of f.
An early and natural question in the study of the RT is that of inversion. Radon himself derived the "Radon inversion formula" [?Rado17] [?Rado86], which is often proved via the above Fourier slice theorem. The groundbreaking inversion formula is the basis for what, in application, called "filtered backpropogation".
If one is interested in inverting the RT then a prerequisite consern is of course: Is the transform injective? The answer clearly depends on what space we draw the function f from. Radon [?Rado17] [?Rado86] provides a set of sufficient regularity conditions such that the RT is invertible. Other similar results followed [?????]. On the other hand counterexamples have been constructed by, for example, Lawrence Zalcman [?Zalc82], of continuous and nontrivial functions for which the RT is identically zero.
By way of the relation (1.3.1) we can prove an analagous slice theorem for the GRT.

Theorem 1.3.11 (Gaussian Slice Theorem)**.** If $f \in L^1(\mathbb{R}^n, w_n)$ and $F \in L^\infty(\mathbb{R}, w)$ then

$$\int_{-\infty}^{\infty} GR_f(\omega,p)F(p)w(p)\,dp = \int_{\mathbb{R}^n} f(x)F(\langle x,\omega\rangle)w_n(x)\,dx. \tag{1.3.3}$$

Remark 1.3.12. Again the more general Fubini condition

$$\int_{\mathbb{R}^n} |f(x)F(\langle x,\omega\rangle)w_n(x)|dx < \infty$$

may be used. Further, whatever conditions on f and F are sufficient for convergence in (1.3.2) are then sufficient conditions to be checked of $f(x)w_n(x)$ and $F(p)w(p)$. I'll need to check these conditions more carefully.

Proof. From (1.3.1)

$$\int_{-\infty}^{\infty} GR_f(\omega,p)F(p)w(p)\ dp = \int_{-\infty}^{\infty} R_g(\omega,p)F(p)\ dp$$

where $g(x) = f(x)e^{-\|x\|^2/2}$. Then applying the slice theorem:

$$\int_{-\infty}^{\infty} R_g(\omega,p)F(p)\ dp = \int_{\mathbb{R}^n} f(x)F(\langle x,\omega\rangle)w_n(x)\ dx,$$

completing the proof. □

As an applications of the slice theorems we first show that, fixing ω, the kth projection moment at ω can be written as a weighted sum of the degree k multivariate moments of f.

Proposition 1.3.13. Let $c_\alpha(\omega) = \int_{-\infty}^{\infty} R_f(\omega,p)p^k dp$ be the projection moments of f at a fixed ω, and c_α the multivariate moments of f. Then

$$c_k(\omega) = \sum_{|\alpha|=k} \binom{k}{\alpha}\omega^\alpha c_\alpha$$

where $\binom{k}{\alpha} = \frac{k!}{\alpha_1!\alpha_2!\cdots\alpha_n!}$ are multinomial coefficients.

Proof. By the slice theorem (1.3.2) with $F(p) = p^k$,

$$\int_{-\infty}^{\infty} R_f(\omega,p)p^k\ dp = \int_{\mathbb{R}^n} f(x)\langle x,\omega\rangle^k\ dx.$$

Now $\langle x,\omega\rangle^k = (x_1\omega_1 + \cdots + x_n\omega_n)^k$ has the multinomial expansion

$$\langle x,\omega\rangle^k = \sum_{|\alpha|=k} \binom{k}{\alpha} x^\alpha\omega^\alpha.$$

Thus after a bit of rearranging we get

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)\langle x,\omega\rangle^k\ dx &= \int_{\mathbb{R}^n} f(x)\sum_{|\alpha|=k}\binom{k}{\alpha}x^\alpha\omega^\alpha\ dx\\ &= \sum_{|\alpha|=k}\binom{k}{\alpha}\omega^\alpha\int_{\mathbb{R}^n} f(x)x^\alpha\ dx,\end{aligned}$$

where the integrands are precisely the kth degree multivariate moments of f. □

Example 1.3.14. Let $e_1, e_2, \ldots, e_n \in S^{n-1}$ be the standard unit vectors

$$e_i = (0,\ldots,0,1,0,\ldots,0)$$

Then $\langle x,e_i\rangle = x_i$ is the natural projection of $\mathbb{R}^n$ onto the e_i axis. The standard projection moments can be calculated as follows

$$\begin{aligned} c_k(e_i) &= \sum_{|\alpha|=k} \binom{k}{\alpha} e_i^\alpha c_\alpha \\ &= c_{ke_i} \end{aligned}$$

since $e_i^\alpha = 0$ unless $\alpha = ke_i$.

Proposition 1.3.15. If $A \subseteq \mathbb{R}^n$ is a measurable and bounded region, then the multivariate moment problem is determinate.

Similarly, moments of the GRT (Gaussian projection moments) can be expressed in terms of multivariate gaussian moments.

Proposition 1.3.16. Let $c_k^G(\omega) = \int_{-\infty}^{\infty} GR_f(\omega, p)p^k w(p)dp$ be the Gaussian weighted moments of the GRT of f at a fixed ω. Let $c_\alpha^G = \int_{\mathbb{R}^n} f(x)w_n(x)x^\alpha dx$ be the Gaussian weighted multivariate moments of f. Then

$$c_k^G(\omega) = \sum_{|\alpha|=k} \binom{k}{\alpha} \omega^\alpha c_\alpha^G.$$

Proof. The proof follows as it did for the RT. This time we apply the GRT slice theorem (1.3.3) with $F(p) = p^k$,

$$\int_{-\infty}^{\infty} GR_f(\omega, p)p^k w(p)\, dp = \int_{\mathbb{R}^n} f(x)\langle x, \omega\rangle^k w_n(x)\, dx$$

Again we use the multinomial expansion of $\langle x, \omega\rangle^n$ and rearange:

$$\int_{\mathbb{R}^n} f(x)\langle x, \omega\rangle^k w_n(x)\, dx = \sum_{|\alpha|=k} \binom{k}{\alpha} \omega^\alpha \int_{\mathbb{R}^n} f(x) w_n(x) x^\alpha dx.$$

Thus

$$c^G(\omega) = \sum_{|\alpha|=k} \binom{k}{\alpha} \omega^\alpha c_\alpha^G.$$

□

Chapter 2

Shape reconstruction

2.1 The Markov and Hamburger transforms

Let μ be a Borel measure with infinite support and finite moments $c_k = \int_{-\infty}^{\infty} p^k d\mu$.

Definition 2.1.1. The function on $\mathbb{C}$ defined by

$$g(z) = \int_{\infty}^{\infty} \frac{d\mu}{1+zp}$$

is called the Markov (Hamburger resp.) transform of μ when μ has bounded (unbounded resp.) support.

Let us consider the domain of definition for $g(z)$. Since p is a real number, the denominator $1+zp$ can only vanish if Im $z=0$ In fact:

Proposition 2.1.2. The function $g(z)$ is holomorphic on the upper half plane Im $z>0$.

Proof. Let γ be a closed piecewise C^1 curve in the upper half plane.

$$\begin{aligned}
|zg(z)| &\leq \int_{-\infty}^{\infty} \frac{d\mu}{|\frac{1}{z}+p|} \\
&\leq \int_{-\infty}^{\infty} \frac{d\mu}{|\text{Im } \frac{1}{z}|} \\
&= \frac{c_0}{|\text{Im } \frac{1}{z}|}.
\end{aligned}$$

Thus

$$|g(z)| \leq \frac{c_0}{|z\text{Im } \frac{1}{z}|} = c_0 \left| \frac{z}{\text{Im } z} \right|.$$

Note that as a compact subset of the upper half plane γ must have positive distance from the real line. As a consequence $g(z)$ is bounded on γ, and by Fubini's theorem

$$\begin{aligned}
\oint_\gamma g(z)dz &= \oint_\gamma \int_{-\infty}^{\infty} \frac{d\mu}{1+zp}\, dz \\
&= \int_{-\infty}^{\infty} \oint_\gamma \frac{dz}{1+zp}\, d\mu \\
&= \int_{-\infty}^{\infty} 0\, d\mu = 0.
\end{aligned}$$

By Morera's theorem $g(z)$ is holomorphic on the upper half plane. $\square$

Remark 2.1.3. Similarly one can show $g(z)$ is holomorphic on the lower half plane $\operatorname{Im} z < 0$. Moreover $g(z)$ commutes with conjugation:

$$g(\bar{z}) = \int_{\infty}^{\infty} \frac{d\mu}{1+\bar{z}p} = \overline{\int_{\infty}^{\infty} \frac{d\mu}{1+zp}} = \overline{g(z)}$$

since μ is a real measure.

When z is a real number $g(z)$ may not converge. In particular $g(z)$ does not converge when $z = -1/p$ for some p in the support of μ. Thus if μ has bounded support g converges in a neighborhood of 0. On the other hand if μ has unbounded support $g(0)$ is not defined. However — as we will see — even in the Hamburger transform case, a formal series expansion at $z = 0$ holds valuable information.
By a simple geometric series expansion,

$$\frac{1}{1+zp} = \sum_{k=0}^{\infty} (-z)^k p^k$$

when $|zp| < 1$. This suggests the connection between $g(z)$ and the moment sequence $(c_k)_{k\in\mathbb{N}_0}$. Indeed, at least formally,

$$\begin{aligned}
\int_{\infty}^{\infty} \frac{d\mu}{1+zp} &= \int_{\infty}^{\infty} \sum_{k=0}^{\infty} (-z)^k p^k \, d\mu \\
&= \sum_{k=0}^{\infty} (-z)^k \int_{\infty}^{\infty} p^k \, d\mu \\
&= \sum_{k=0}^{\infty} c_k (-z)^k.
\end{aligned}$$

In the Markov case we have a positive radius of convergence. But even in the Hamburger case, there is a sense in which this assymptotic expansion holds "non-tangentially". Non-tangential convergence at $z = 0$ is, as the name implies, convergence along curves not tangent to the real line at 0. More precisely, one defines a "wedge domain"

$$\Gamma_\delta = \{\delta \leq \arg z \leq \pi - \delta\}$$

in which curves approach 0 with an angle at least δ from the real line.

Definition 2.1.4. (Non-Tangential Limit) We say $g(z) \to z_0$ non-tangentially at $z = 0$ if, for any $\delta > 0$, the limit

$$\lim_{z\to 0} g(z) = z_0$$

holds for z in the Γ_δ.

While $g(z)$ does not generally converge at $z = 0$, in a non-tangential sense the afformentioned assymptotic expansion holds:

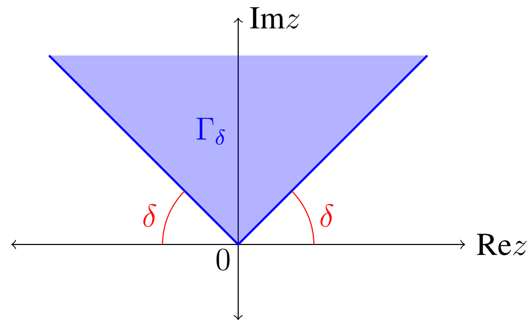

Figure 2.1.1: A wedge domain

Proposition 2.1.5. (i) If μ has bounded support

$$g(z) = \sum_{k=0}^{\infty} c_k(-z)^k$$

in a neighboorhood of $z = 0$.

(ii) For all $n \in \mathbb{N}_0$

$$g(z) = \sum_{k=0}^{n-1} c_k(-z)^k + z^n h_n(z) \tag{2.1.1}$$

where h_n is such that $\lim_{z\to 0} h_n(z) = 0$ non-tangentially.

Proof. We first note that

$$\sum_{k=0}^{n} c_k(-z)^k = \sum_{k=0}^{n} \int_{-\infty}^{\infty} (-zp)^k d\mu = \int_{-\infty}^{\infty} \frac{1-(-zp)^{n+1}}{1+zp} d\mu$$

and so

$$h_n(z) = \frac{1}{z^n}\left\{g(z) - \sum_{k=0}^{n} c_k(-z)^k\right\} = \int_{-\infty}^{\infty} \frac{zp^{n+1}}{1+zp} d\mu.$$

It remains to determine if and in what sense this trailing term $h_n(z)$ vanishes at the origin.

As expected in the case when μ has bounded support, $h_n(z)$ exists in a neighborhood of 0 and vanishes as $z \to 0$ in this neighboorhood. Thus the assymptotic expansion 2.1.1 is the Taylor expansion: The Markov transform is analytic at the origin.

When μ has unbounded support and $h_n(z)$ may not be well defined on the real line, so we must weaken our result to non-tangential convergence. If $z \in \Gamma_\delta$ then we have the following inequalities for any $p \in \mathbb{R}$,

$$|p - z| \geq |p| \sin\delta \quad \text{and} \quad |p - z| \geq |p| \sin\delta.$$

We will show that

$$\lim_{z\to 0} h_{2n}(z) = 0, \quad z \in \Gamma_\delta$$

from which the corresponding limit for odd orders will immediately follow from the relation

$$h_{2n-1}(z) = z h_{2n}(z) + c_{2n} z^{2n}.$$

Here for the sake of convenience we define $w = -\frac{1}{z}$, following a more classical approach. Note that $|w| = \frac{1}{|z|}$ and if z is in the wedge Γ_δ then so is w. Now

$$\begin{aligned}
|h_{2n}(z)| &\le \int_{-\infty}^{\infty} \frac{|p|^{2n+1}}{|p-w|} d\mu \\
&\le \frac{|z|}{\sin\delta} \int_{|p|\le A} |p|^{2n+1} d\mu + \frac{1}{\sin\delta} \int_{|p|\ge A} p^{2n} d\mu \\
&\le \frac{2|z| A^{2n+2}}{\sin\delta} + \frac{1}{\sin\delta} \int_{|p|\ge A} p^{2n} d\mu
\end{aligned}$$

Thus

$$\lim_{z\to 0} |h_{2n}(z)| \le \frac{1}{\sin\delta} \int_{|p|\ge A} p^{2n} d\mu, \quad z \in \Gamma_\delta.$$

Since A is arbitrary and the integral $\int p^{2n} d\mu = c_{2n}$ is convergent we are done. □

Evidently the assymptotic series expansion $g(z) \simeq \sum_{k=0}^{\infty} c_k (-z)^k$ is generally formal, having zero radius of convergence except in the Markov case. However as we will see in the next section, rational functions constructed from this formal series exist which approximate $g(z)$ on the upper half plane $\{\mathrm{Im} z > 0\}$.

Finally we prove two formulas regarding the Hamburger transform of a RT projection. Let $\omega \in S^{n-1}$ be fixed and consider.

$$g(z) = \int_{\infty}^{\infty} \frac{R(\omega, p)}{1+zp} dp$$

By applying the slice theorem, with $F(p) = (1+zp)^{-1}$ we see that the Hamburger transform of a projection can be represented by a similar multivariable integral of f over $\mathbb{R}^n$.

$$\int_{-\infty}^{\infty} \frac{R_f(\omega, p)}{1+zp} dp = \int_{\mathbb{R}^n} \frac{f(x)}{1+z\langle x, \omega\rangle} dx$$

Similarly the GRT slice theorem with $F(p) = (1+zp)^{-1}$ is

$$\int_{-\infty}^{\infty} \frac{GR_f(\omega, p) w(p)}{1+zp} dp = \int_{\mathbb{R}^n} \frac{f(x) w_n(x)}{1+z\langle x, \omega\rangle} dx.$$

2.2 Padé approximants

In this section we recall the necessary definitions and results. Then we will prove some new results to be applied to the Gaussian Radon transform later.

Definition 2.2.1 (Classical definition of Padé approximants)**.** The Padé approximant to a (possibly formal) power series

$$R^{[L/M]}(z) \simeq \sum_{k=0}^{\infty} c_k z^k$$

is a rational function with nummerator (denominator resp.) degree at most L (M resp.), with series equal to $\sum_{k=0}^{N} c_k z^k + O(z^{N+1})$ up to as high an order N as possible.

Let

$$R^{[L/M]}(z) = \frac{P^{[L/M]}(z)}{Q^{[L/M]}(z)} = \frac{a_L z^L + \cdots + a_1 z + a_0}{b_M z^M + \cdots + b_1 z + b_0}.$$

Notice that in general there is a negligable constant common factor between the numerator and denominator, so that with the remaining $L+M+1$ free parameters, we expect an order of accuracy of up to $L+M+1$ constraints, $c_0, c_1, \ldots, c_{L+M}$. Thus we define the $[L/M]$ Padé approximant by the condition,

$$\frac{P^{[L/M]}(z)}{Q^{[L/M]}(z)} = \sum_{k=0}^{L+M} c_k z^k + O\left(z^{L+M+1}\right). \tag{2.2.1}$$

It is helpful to consider the related, and necessary condition

$$P^{[L/M]}(z) = Q^{[L/M]}(z) \left(\sum_{k=0}^{L+M} c_k z^k \right) + O\left(z^{L+M+1}\right). \tag{2.2.2}$$

In the classical theory of Padé approximants (2.2.2) was often taken as a definition. It is always possible to find polynomials of the required degree satisfying this second condition, however they do not necessarily attain the degree of accuracy required by the first. We follow Baker, defining $R^{[L/M]}(z)$ by (2.2.1), provided such a rational function exists.

Definition 2.2.2. The Padé approximant $R^{[L/M]}$ to a (possibly formal) power series is a the unique rational function with nummerator (denominator resp.) degree at most L (M resp.) satisfying the condition (2.2.1). If no such rational function exists we say the Padé approximant does not exist.

Here we note that a sufficient condition for the equivalence of the two definitions, and hence for the existence of the Padé approximant by Baker's definition is that $b_0 = Q^{L/M}(0) \neq 0$.

Equating coeffictients of z^k, $k = 0, 1, \ldots, L+M$ in (2.2.2) gives two linear systems

$$\begin{aligned} a_0 &= b_0 c_0 \\ a_1 &= b_1 c_0 + b_0 c_1 \\ &\vdots \\ a_L &= b_L c_0 + b_{L-1} c_1 + \cdots + b_0 c_L \end{aligned}$$

and,

$$\begin{aligned} 0 &= b_M c_{L-M+1} + b_{M-1} c_{L-M+2} + \cdots + b_0 c_{L+1} \\ 0 &= b_M c_{L-M+2} + b_{M-1} c_{L-M+3} + \cdots + b_0 c_{L+2} \\ &\vdots \\ 0 &= b_M c_L + b_{M-1} c_{L+1} + \cdots + b_0 c_{L+M} \end{aligned}$$

where for convenience we set $c_k = 0$ for $k < 0$. The first systems shows the numerator $P^{[L/M]}$ is determined by the denominator $Q^{[L/M]}$. From the second system we can derive a determinantal formula for $Q^{[L/M]}$. Augmented with the desired definition,

$$Q^{[L/M]}(z) = b_M z^M + b_{M-1} z^{M-1} + \cdots + b_0$$

the system can be written in the form

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ Q^{[L/M]}(z) \end{pmatrix} = \begin{pmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_{L+1} \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ c_L & c_{L+1} & \cdots & c_{L+M} \\ z^M & z^{M-1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} b_M \\ b_{M-1} \\ \vdots \\ b_1 \\ b_0 \end{pmatrix}$$

Solving for b_0 by Cramer's Rule we get,

$$b_0 \begin{vmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_{L+1} \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ c_L & c_{L+1} & \cdots & c_{L+M} \\ z^M & z^{M-1} & \cdots & 1 \end{vmatrix} = \begin{vmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_L & c_{L+1} & \cdots & c_{L+M-1} \end{vmatrix} Q^{[L/M]}(z)$$

The LHS is clearly a polynomial, and nonzero if the RHS minor is nonzero. Thus if the Hankel determinant

$$\begin{vmatrix} c_{L-M+1} & \cdots & c_L \\ \vdots & \ddots & \vdots \\ c_L & \cdots & c_{L+M-1} \end{vmatrix}$$

is nonzero, then up to the afformentioned constant factor,

$$Q^{[L/M]}(z) = \begin{vmatrix} c_{L-M+1} & \cdots & c_{L+1} \\ \vdots & \ddots & \vdots \\ c_L & \cdots & c_{L+M} \\ z^M & \cdots & 1 \end{vmatrix} \tag{2.2.3}$$

By convention we normalize $R^{[L/M]}(z)$ so that $b_0 = Q^{[L/M]}(0) = 1$.
Since the Hankel determinants give a useful form for expressing some conditions of interest we will difine $H(n, m)$ to be the determinant of the $(m+1) \times (m+1)$ Hankel matrix starting with c_n,

$$H(n, m) := \begin{vmatrix} c_n & \cdots & c_{n+m} \\ \vdots & \ddots & \vdots \\ c_{n+m} & \cdots & c_{n+2m} \end{vmatrix}$$

Note in particular that the sufficient condition $Q^{[L/M]}(0) \neq 0$ for the existence of $R^{[L/M]}(z)$ is equivalent to $H(L-M+1, M-1) \neq 0$
We now narrow our focus to Padé approximants to a Humburger series. Let μ be a Borel measure on $\mathbb{R}$ with infinite support and finite moments $c_k = \int p^k \, d\mu$. Recall that the Hamburger transform of μ,

$$g(z) := \int_{\mathbb{R}} \frac{d\mu}{1 + pz}$$

has the assymptotic expansion, in the sense of non-tangential limits,

$$g(z) \simeq \sum_{k=0}^{\infty} c_k(-z^k).$$

We call this formal power series a Hamburger series.

Remark 2.2.3. To simplify the rest of this maybe we redefine?

$$c_k = (-1)^k \int p^k d\mu$$

Lemma 2.2.4. The hamburger moments c_k satisfy the determinental condition $H(2n, m) \neq 0$ for $n, m = 0, 1, \ldots$.

Proof. Consider the quadratic form given by

$$G(\mathbf{x}, \mathbf{y}) := \mathbf{x}^\top \begin{pmatrix} c_{2n} & \cdots & c_{2n+m} \\ \vdots & \ddots & \vdots \\ c_{2n+m} & \cdots & c_{2n+2m} \end{pmatrix} \mathbf{y}$$

If $\mathbf{x} = (x_0, x_1, \ldots, x_m)^\top$ then

$$\begin{aligned} G(\mathbf{x}, \mathbf{x}) &= \sum_{i,j=0}^{m} x_i x_j c_{2n+i+j} \\ &= \sum_{i,j=0}^{m} x_i x_j \int p^{2n+i+j}\, d\mu \\ &= \int \sum_{i,j=0}^{m} x_i x_j p^{2n+i+j}\, d\mu \\ &= \int p^{2n} \left(\sum_{k=0}^{m} x_k p^k \right)^2 d\mu \geq 0. \end{aligned}$$

Thus $G(\mathbf{x}, \mathbf{y})$ is positive semi-definite. Furthermore, equality holds

$$\int p^{2n} \left(\sum_{k=0}^{m} x_k p^k \right)^2 d\mu = 0$$

if and only if μ is supported entirely on the zeros of the polynomial $p^n \sum_{k=0}^{m} x_k p^k$. However by assumption μ has infinite support. So in fact G is strictly positive definite. We conclude, by Sylvester's criterion for example, that assocciated hankel matrix has positive determinant. That is, $D(2n, m) > 0$. $\square$

The previous lemma guarentees the existence of certain Padé approximants, specifically those with $L - M + 1$ even.

Theorem 2.2.5. If J is odd then the Padé approximant $R^{[L/L+J]}(z)$ to the Hamburger series exists in Baker's sense.

With our particular application in mind, we will assume the measure μ is absolutely continuous with respect to the Lebesgue measure and denote it $f(x)dx$ where $f(x)$ is some non-negative Lebesgue measureable function of $\mathbb{R}$. Let $R_M(z)$ be the "offdiagonal" approximant $R^{[M/M+1]}(z)$ to the Hamburger series of μ,

$$R_M(z) = \frac{P_M(z)}{Q_M(z)} \approx \sum_{k=0}^{\infty} c_k (-z)^k \simeq \int_{-\infty}^{\infty} \frac{f(p)}{1 + zp}\, dp$$

where $c_k = \int p^k f(p) dp$. In this section we will prove

Theorem 2.2.6. The off-diagonal Padé approximants $R_M(z)$ to the a determinate Hamburger series $\sum_{k=0}^{\infty} c_k (-z)^k$ exist and converge to $g(z)$ uniformly on compact subsets of the upper half plane $\{\text{Im } z > 0\}$.

An outline of the proof is as follows: We first justify that the approximants exist in Baker's sense. Then it can be shown that limit of any convergent subsequence of R_M must have a representation as the Hamburger transform of some Borel measure

μ which is a solution to our moment problem, and that furthermore such convergent subsequences exist. Thus in order for the sequence R_M to converge to $g(z)$ it will be necessary and sufficient that the moment problem be determinate.

Proposition 2.2.7. If the sequence c_k gives a determinate moment problem then the off-diagonal Padé approximants $R_M(z)$ converge to $g(z)$ locally uniformly.

Proof. The Padé approximant $R_M(z)$ has a convenient representation as an inner product in terms of a finite Jacobi matrix,

$$R_M(z) = \langle \delta_0, (1 + zA_M)^{-1} \delta_0 \rangle.$$

Now since A_M is a real matrix and $\frac{1}{|w|} \geq \frac{1}{|\text{Im}(w)|}$ for any $w \in \mathbb{C}$, we see that

$$|zR_M(z)| = |\langle \delta_0, (z^{-1} + A_M)^{-1} \delta_0 \rangle| \leq \frac{1}{|\text{Im}(1/z)|}$$

and thus

$$|R_M(z)| \leq \frac{1}{|z||\text{Im}(1/z)|} = \frac{|z|}{|\text{Im}(z)|}.$$

Since this bound is independent of R_M, Montel's theorem implies that the off-diagonal Pade approximants form a normal family. It can be shown that the limit of any convergent subsequence has a representation $\int (1 + xz) d\sigma(x)$ where σ is a solution to the Hamburger moment problem. Since the moment problem is determinate, the sequence R_M must converge to $g(z)$ uniformly on compact sets. □

It remains to discuss the determinacy of the Hamburger moment problem. Here we need to add an additional constraint on the measure $d\mu = f(p)dp$, which is that $f(p)$ is L^2 integrable with respect to the Gaussian weight.

Proposition 2.2.8. A function $f(p) \in L^2(\mathbb{R}, e^{-p^2} dp)$ such that the moments

$$c_k = \int_{-\infty}^{\infty} p^k f(p) e^{-p^2} \, dp, \; k = 0, 1 \ldots \tag{2.2.4}$$

are finite, is uniquely determined by those moments.

Proof. It is sufficient to show that if $c_k = 0$ for all $k \geq 0$, then $f \equiv 0$ a.e. The Hermite moments are just linear combinations of 0,

$$\int_{-\infty}^{\infty} H_k(p) f(p) e^{-p^2} \, dp = 0, \; k \geq 0$$

where $H_k(p)$ is the Hermite polynomial of order k. Since the Hermite polynomials are complete in the space $L^2(\mathbb{R}, e^{-p^2} dp)$ then $f \equiv 0$. □

2.3 Convergence results for the Gaussian Radon transform

Now we are in the position to assemble all the previous results into an approximation scheme to be used for reconstruction shapes of unbounded domains.
Take a non-negative Lebesgue measureable function $f(x)$ on $\mathbb{R}^n$ with finite multivariate Gaussian moments

$$c_\alpha^G = \int_{\mathbb{R}^n} f(x) e^{-\|x\|/2} x^\alpha dx$$

We have seen that the multivariate Stieltjes transform of f,

$$g(z) = \int_{\mathbb{R}^n} \frac{f(x) e^{-\|x\|/2}}{1 + \langle x, \omega \rangle z} dx,$$

is equal to the Hamburger transform of $GR_f(\omega, p)$, whose moments in turn are easily computed from the moments of f, for fixed ω. Thus for z in the one dimensional subspace spanned by ω, $g(z)$ can be approximated well by Padé approximants, under the integrability condition $GR_f(\omega, p) \in L^2(\mathbb{R}, e^{-p^2}\ dp)$. On the other hand as an integral on $\mathbb{R}^n$, we can also approximate $g(z)$ by cubature formula,

$$\int_{\mathbb{R}^n} \frac{f(x) e^{-\|x\|^2/2}}{1 + \langle x, \omega \rangle z} dx \approx \sum_{\ell = L} \frac{e^{-\|x_\ell\|^2/2}}{1 + \langle x_\ell, \omega \rangle z} w_\ell f(x_\ell) =: \sum_{\ell = L} C_\ell(z) f(x_\ell).$$

By equating these parallel approximations,

$$\sum_{\ell = L} C_\ell(z) f(x_\ell) \approx R_M(z),$$

with a sufficient quantity of sample points z_j we have arrived at a linear system

$$\sum_{\ell = L} C_\ell(z_j) f(x_\ell) = R_M(z_j), \qquad j = 1, 2, \ldots, J$$

from which we should be able to recover the value of f on our cubature nodes x_ℓ.

2.4 Implementation

While the authors describe forming multivariate "homogeneous" Pade approximants — perhaps first proposed by Cuyt herself in [?Cuyt84] — we opt to bypass this step using univariate approximants instead. Indeed, the only benefit to Cuyt's approximants explicitly cited by the authors, as compared to other multivariate generalizations of the Pade approximants, is that they are "the only satisfying the following powerful slice theorem" [?Cuyt05]. This slice theorem simply states that the homogeneous approximants when restricted to one dimensional subspaces, are equivalent to univariate Pade approximants. Since we only need a finite number of sample points it seems reasonable to compute a series of univariate approximants at various projection angles.
The decision to drop the homogenous approximants has the following potential consequences: On the one hand, the univariate approximants are simpler to compute, with methods built in to many CASs. On the other, while the authors select sample points from a cubic lattice, it would be inefficient (though potentially less so with careful se-

lection of projection angles via, say, Farey sequences) to do so with one dimensional subspaces. Instead we take a collection of equidistant projection angles forming a radial grid of sample points. This seems acceptable, as it is unclear if there is any benefit to the sample points having a particular geometric structure. We should also agknowledge the potential that forming a single homogeneous appoximant is more computationally efficient than many univariate approximants. This requires further investigation, but we note that (anecdotally) in computational tests the formation of approximants is not very intensive compared to the subsequent solving of the high dimensional linear system. Some thought was given to the possibility that sampling from a cubic lattice provides some structure which makes the system easier to solve, but so far we have no hard evidence of this.

A couple of notes on computational details: First we should be explicit about the line of computation. At the outset, the given information is a certain multivariate moment sequence (up to a finite order with some accuracy assumption) which we assume belongs to a region within some fixed bounding box. In evaluating the computational viability of the method we take note of which steps depend on the moments, and which can be precomputed. In order to form univariate approximants we first compute projection moments. While these of course depend on the moment sequence, the multinomial formula used for this computation only depends on the projection angle. Thus with a preditermined set of projection angles this is partially precomputed. Since the cubature formula itself only depends on the choice of quadrature nodes and sample points, it can be precomputed and applied to any moment sequence. The choice of quadrature formula is somewhat arbitrary. Lacking expertise in the area of numerical integration, we follow the authors, opting for the four-point Gauss-Legendre product formula. This formula takes nodes from a nearly cubic lattice, which is implicitly quantized when forming the pixel image. We have not evaluated the error introduced by this quantization but expect it to be minimal and to approach zero with higher pixel resolution.

We don't expect the linear system to have an exact solution, let alone one that whose computation is tractable for high resolutions. We may try Mathematica's LeastSquares method, but Cuyt et al. specifically mention using a truncated singular value decomposition.

For my own benefit, let's review singular value decompositions. An $i \times j$ matrix A of rank r can be written as

$$A = U \Sigma V^\top,$$

where Σ is the diagonal matrix of singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$, and U and V are unitary matrices. The truncated singular value decomposition gives us a reduced rank approximation to A by replacing all but the largest (first) k singular values in Σ with zeros,

$$\tilde{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0)$$

The truncated SVD matrix $\tilde{A} = U \tilde{\Sigma} V^\top$ is optimal in the sense that it is the closest rank-k matrix to A in the Frobenius norm.

The SVD and truncated SVD are used to solve the linear least squares problem as follows: From $U\Sigma V^\top f = p$ we get $f = V\Sigma^{-1}U^\top p$, and similarly $f = \tilde{V}\tilde{\Sigma}^{-1}U^\top p$. Note that the psuedoinverse $\tilde{\Sigma}^{-1}$ is a diagonal matrix made up of the reciprocal singular values.

Mathematica offers a few built in functions for this. LeastSquares packages a number of methods to solve a linear least squares based on the sparsity of the System (our system is dense). SingularValueDecomposition can be restricted to a certain number of singular values. PsuedoInverse can specify a "tolerance" zeroing out singular values less than some proportion of the maximum singular value. While LeastSquares is the simplest option, it does not seem to support truncated solutions. Between SingularValueDecomposition and PsuedoInverse, both can handle truncated problems, but the latter seems — at least superficially — better suited for out application.

On Gaussian shape reconstruction. There remain some minor practical concerns about, for example, the choice of quadrature formula and sample points. An obvious choice would be to continue with the Gauss-Legendre product formula, now applied on a Gaussian-weighted f. Some consideration should be taken as to the viability of other Gaussian quadrature formulas defined particulary for the Gaussian measure.

Now, a fundamental limitation with the proposed method which we have neglected to mention is the practical one: As a computational method we are necessarily restriced to the finite — finite moments, finite order approximations, and most problamatic, bounded domains. In what application would one even be interested in unbounded shape reconstruction, while also having access of Gaussian moments? It would be reasonable to question the use of a computational method in an unbounded context. As with any finite approximation, the unboundedness should be thought of as an idealized limit; perhaps an unbounded approximation is a limit of bounded approximations. But then, the original method applies just as well if all we wanted was a sequence of reconstructions expanding in range. We can only speculate at this point that perhaps this method, being taylored specifically for unbounded reconstruction, could hold some practical advantages over its predecessor; or be satisfied with impracticality.

2.5 An example

$$f : \mathbb{R}^2 \to \mathbb{R}$$

$$\begin{aligned} GR_f(\theta, p) &= \int_{-\infty}^{\infty} f(x(t), y(t))w(t)\,dt \\ &= \int_{-\infty}^{\infty} f(t\sin\theta + p\cos\theta, -t\cos\theta + p\sin\theta)w(t)\,dt \end{aligned}$$

Let $H_\alpha(x) = H_{\alpha_1}(x_1)H_{\alpha_2}(x_2)\cdots H_{\alpha_n}(x_n)$ for $\alpha \in \mathbb{N}_0^n$ and when $n = 2$,

$$H_{i,j}(x, y) = H_i(x)H_j(y)$$

Mathematica table: $\frac{GR_{H_{i,j}}(\theta,p)}{\cos^i\theta\sin^j\theta}$ for $i, j \le 4$

1	p	p^2-1	p^3-3p	p^4-6p^2+3
p	p^2-1	p^3-3p	p^4-6p^2+3	p^5-10p^3+15
p^2-1	p^3-3p	p^4-6p^2+3	p^5-10p^3+15p	$p^6-15p^4+45p^2$
p^3-3p	p^4-6p^2+3	p^5-10p^3+15p	$p^6-15p^4+45p^2-15$	$p^7-21p^5+105p^3$
p^4-6p^2+3	p^5-10p^3+15p	$p^6-15p^4+45p^2-15$	$p^7-21p^5+105p^3-105p$	$p^8-28p^6+210p^4-42$

Conjecture: Based on a Mathematica table (at least up to $i, j \le 4$) it seems that the following formula holds:

$$GR_{H_{i,j}}(\theta, p) = H_{i+j}(p)\cos^i\theta\sin^j\theta$$

We further expect

$$GR_{H_\alpha}(\omega, p) = H_{|\alpha|}(p)\omega^\alpha$$

for $\alpha \in \mathbb{N}_0^n$.
Bogachev (for example): $(H_\alpha)_{\alpha\in\mathbb{N}_0^n}$ is an orthogonal basis in $L^2(\gamma_n)$
Mathematica table: $GR_f(a, p)$ for $f(x, y) = x^i y^i$

1	$p\sin(a)$	
$p\cos(a)$	$(p^2-1)\sin(a)\cos(a)$	$p\cos(a$
$p^2\cos^2(a)+\sin^2(a)$	$p\sin(a)\left((p^2-2)\cos^2(a)+\sin^2(a)\right)$	$\frac{1}{8}(-(p^4-$
$p\cos(a)\left(p^2\cos^2(a)+3\sin^2(a)\right)$	$\frac{1}{4}\sin(2a)\left((p^4-6p^2+3)\cos(2a)+p^4-3\right)$	$\frac{1}{8}p\cos(a)(-8(p^2-3)\cos$

$$f(x) = \begin{cases} 1, & \langle x, \varphi\rangle \le 1/2 \\ 0, & \text{otherwise} \end{cases}$$

Unit ball in $\mathbb{R}^3$? Annulus in $\mathbb{R}^2$? Some assymmetric or off-center examples?

Chapter 3

Discrete Radon transforms

Chapter 3

*

[1] L. C. Petersen, On the relation between the multidimensional moment problem and the one-dimensional moment problem, Math. Scand.,

[2] David Applebaum, Lévy Processes and Stochastic Calculus, Cambridge Studies in Advanced Mathematics, ISBN 0-521-83263-2, Cambridge University Press, 2004

[3] Richard F. Bass, Probabilistic Techniques in Analysis, Probability and its Applications, ISBN 0-387-94387-0, Springer Verlag, 1995

[4] Richard F. Bass, Diffusions and Elliptic Operators, Probability and its Applications, ISBN 0-387-98315-5, Springer Verlag, 1998

[5] Richard F. Bass and Takashi Kumagai, Symmetric Markov Chains on $\mathbb{Z}^d$ with Unbounded Range, Transactions of the American Mathematical Society, Vol. 30, Nr 4, 2041-2075 (2008), AMS, 2008

[6] Zhen-Qing Chen, Zhongmin Qian, Yaozhong Hu and Weian Zheng, Stability and Approximations of Symmetric Diffusion Semigroups and Kernels, Journal of Functional Analysis, 152, 225-280 (1998), Academic Press, 1998

[7] Rama Cont and Peter Tankov, Financial modelling with jump processes, Financial Mathematics Series, ISBN 1-5848-8413-4, Chapman & Hall/CRC, 2004

[8] A. De Masi, P.A. Ferrari, S. Goldstein, W.D. Wick, An Invariance Principle for Reversible Markov Processes, Journal of Statistical Physics, Vol. 55, 787-855 (1989), Springer Verlag, 1989

[9] Joseph L. Doob, Classical Potential Theory and Its Probabilistic Counterpart, Classics in Mathematics ISBN 3-540-41206-9, Springer Verlag, 1984

[10] Richard Durrett, Probability Theory and Examples, Wadsworth & BrooksCole StatisticsProbability Series, ISBN 0-534-13206-5, BrooksCole Publishing Company, 1991

[11] Richard Durrett, Stochastic Calculus, Probability and Stochastics Series, ISBN 0-8493-8071-5, CRC Press, 1996

[12] Lawrence C. Evans, Partial Differential Equations, Graduate Studies in Mathematics, ISBN 0-821-80772-2, American Mathematical Society, 1998

[13] William Feller, An Introduction to Probability Theory and Its Applications Vol. II (2nd Ed.), ISBN 0-471-25709-5, John Wiley & Sons Inc., 1990

[14] G.H. Hardy, J.E. Littlewood and G. Polya, Inequalities (2nd Ed.), ISBN 0-521-05206-8, Cambridge University Press, 1952

[15] Warren P. Johnson, The Curious History of Faà di Bruno's Formula, American Mathematical Monthly, Vol. 109, 217-234, (March 2003), MAA, 2003

[16] Ioannis Karatzas and Steven Schreve, Brownian Motion and Stochastic Calculus (2nd Ed.), ISBN 0-387-97655-8, Springer Verlag, 1988, 1991

[17] Hiroshi Kunita, Stochastic Flows and Stochastic Differential Equations, ISBN 0-521-59925-3, Cambridge University Press, 1990, 1997

[18] Jean-Pierre Lepeltier and Bernard Marchal, Problème des martingales et équations différentielles stochastiques associées à un opérateur inégro-différentiel, Annales de l' Institut Henri Poincaré (B) Vol. 12, 43-103, (1976)

[19] Zhi-Ming Ma and Michael Röckner, Introduction to the Theory of (Non-symmetric) Dirichlet Forms, ISBN 3-540-55848-9, Springer Verlag, 1992

[20] Bernt Øksendal, Stochastic Differential Equations (6th Ed.), ISBN 3-540-04758-1, Springer Verlag, 1985, 2003

[21] L.C.G. Rogers and D. Williams, Diffusions, Markov Processes and Martingales, Vol. II : Itô Calculus (2nd Ed.), ISBN 0-521-77593-0, Cambridge University Press, 2000

[22] Walter Rudin, Functional Analysis, ISBN 0-070-54236-8, McGraw-Hill Science and Engineering, 1991

[23] A.V. Skorokhod, Studies in the Theory of Random Processes, ISBN 9-780-48664240-6, Reading, Mass., Addison-Wesley Pub. Co., 1965

[24] D.W. Stroock and W. Zheng, Markov Chain Approximations to Symmetric Diffusions, Annales de l'Institut Henri Poincaré (B), Vol. 33, 619-649, (1997)