Brownian Motion

Peter Rudzis

June 7th, 2017

Contents

1 Introduction **1**

2 Preliminaries **3**
2.1 Probability Theory: Basic Definitions 3
2.2 The Probabilistic Way of Thought 4
2.3 Independence 5
2.4 Characteristic functions 7
2.5 Stochastic Processes 9

3 Gaussian Distribution **12**

4 Wiener Process **17**

5 Construction of the Wiener process **20**
5.1 Haar functions 21
5.2 The Main Construction 22

6 Markov Property **26**
6.1 Conditional Expectation 26
6.2 Stopping Times 29
6.3 The Wiener Process as a random variable taking values in $C([0,\infty),\mathbb{R}^n)$ . . 30
6.4 The Weak Markov Property 31
6.5 The Strong Markov Property 33

7 Appendices **35**

1 Introduction

Attempts by persons such as Robert Brown to understand the motion of pollen grains suspended in a fluid (1828) and Bachelier to understand the stock market (1900) provided

impetus for the development of a mathematical theory of Brownian motion during the twentieth century [3, p. 301]. Today, the theory of Brownian motion has assumed a central role in modern probability theory, and it has deep connections in other fields of mathematics as well. Brownian motion is not only an applied tool, but also a concept of enormous theoretical importance.

The purpose of this paper is to give a mathematical exposition on Brownian motion, emphasizing the concept's theoretical underpinnings and basic properties. We spend §2 reviewing the concepts from probability theory needed for the definition and construction of Brownian motion. Since Brownian motion is a Gaussian process, it is useful to develop some general facts about Gaussian distribution in §3. In §4 we define the Wiener process, which is the mathematical model for Brownian motion, and prove some of its some elementary properties. (Note that while the term "Wiener process" refers to a particular *model* for the *concept* of Brownian motion, following tradition, we will sometimes be sloppy and refer to both by the same name. When we refer to Brownian motion, we always have the Wiener process in mind.) A construction for the Wiener process is provided in §5 . Finally §6 introduces conditional expectation and stopping times, and proves two remarkable homogeneity conditions satisfied by Brownian motion, known as the weak and strong Markov properties.

This paper does *not* cover the many applications of Brownian motion in mathematics and various scientific fields. However, in writing this paper, we had applications to problems in differential equations in mind. In particular, the Markov property of Brownian motion can be utilized to obtain solutions to the Dirichlet problem on a very general class of domains (see Bass [1, Chapter 2]). As such, this paper would provide most of the background needed for this direction of study. Certainly other applications, found in for example [1] or [5], would become accessible as well.

The most essential prerequisite for understanding this paper is a knowledge of measure theory. Familiarity with (measure-theoretic) probability and the basics of Fourier analysis is also important. While the results which we need from probability theory are presented in §2, this section is intended as a review rather than a thorough introduction. We also rely on results from Fourier analysis for several of our proofs, and these results are presented in the Appendices.

This paper draws upon a number of different sources. Most of the basic material on probability theory can be found in Durrett [3], and material on measure theory can be found in Folland [4] and Rudin [7]. The construction we give for the Wiener process is presented by Bass [1] and Roger and Williams [6] but is originally due to Ciesielski [2]. Our proofs of the Markov properties more or less follow those of Bass [1], with some modifications.

Acknowledgements

I would like to express my sincere thanks to Jim Morrow, my thesis supervisor, from whom I first learned about about the connection between Brownian motion and the Dirichlet problem. This paper is in a large part a product of his excellent advice, insight, and encouragement. I also would like to thank my parents and siblings, my professors at University of Washington, and everyone who has contributed to my education thus far.

2 Preliminaries

2.1 Probability Theory: Basic Definitions

Let Ω is a set, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a measure on $\mathcal{F}$. If $P(\Omega) = 1$, we call the measure space $(\Omega, \mathcal{F}, P)$ a **probability space**. Measurable functions on Ω are called **random variables**, and are most commonly denoted by the letters X, Y, or Z. The codomain of a random variable X will usually be $\mathbb{R}$ or $\mathbb{R}^n$; however, it could be any measurable space. Later in this paper, for example, we will study random variables taking values in $C([0,\infty), \mathbb{R}^n)$, the space of continuous functions from $[0,\infty)$ into $\mathbb{R}^n$ (equipped with a certain σ-algebra to be defined later). If $(M, \mathcal{M})$ is a measurable space, every random variable $X : \Omega \to M$ has an associated σ-algebra $\sigma(X) \subset \mathcal{F}$, defined by $\sigma(X) = \{X^{-1}A : A \in \mathcal{M}\}$. It is called the **$\sigma$-algebra generated by** X.

Probabilistic terminology differs from the standard terminology in analysis in a number of cases. For example, if $A \in \mathcal{F}$, the function 1_A, defined by $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$, is called the **indicator function** for A. In other fields of analysis, this function would be called the *characteristic function* for A and would be denoted by χ_A. A certain property $T(\omega)$, depending on the elements $\omega \in \Omega$, is said to hold **almost surely** (abbreviated a.s.) if T holds true except on a P-null set, in contrast to the usual term *almost everywhere* used in other fields of analysis. For this paper, we will use the conventional probabilistic terms and notation when working with a probability space, and use the usual analytical terms and notation when working with another measure space (for example $\mathbb{R}^n$ with Lebesgue measure). This should not cause any serious confusion.

If a random variable X takes values in $\mathbb{R}^n$, the integral of X with respect to P is called the **expected value** of X and is denoted by $E[X]$. Integration over a set $A \in \mathcal{F}$ is denoted $E_A[X]$. However, if we happen to be working with multiple probability measures, to avoid ambiguity we may instead use the usual integration notation $\int X dP$. Of course, the definition of expected value only applies if $X \geq 0$ or $E|X| < \infty$. If $E|X| < \infty$, we write $X \in L^1(P)$. More generally, if $E|X|^p < \infty$ for any for $0 < p < \infty$, we write $X \in L^p(P)$.

Among the primary objects of interest in probability theory are the **distributions** of random variables. (This term should not to be confused with a related but distinct concept of the same name, occurring in other fields of analysis and physics.) Given a random variable X, taking values in a measurable space $(M, \mathcal{M})$ (e.g. $M = \mathbb{R}^n$ and $\mathcal{M} = \mathcal{B}^n$, the Borel σ-algebra) the **law, or distribution**, of X is the probability measure P_X on $\mathcal{M}$ defined by

$$P_X(B) = P(X \in B) \qquad \text{for all } B \in \mathcal{M}. \tag{1}$$

If $(M, \mathcal{M}) = (\mathbb{R}^n, \mathcal{B}^n)$, sometimes the distribution of X is given by a **density function**, that is a nonnegative function ρ on $\mathbb{R}^n$ such that $P_X(B) = \int_B \rho dm$, where m denotes Lebesgue measure. If X and Y are random variables, the notation $X =^d Y$ means that $P_X = P_Y$.

If $X_1, ..., X_m$ are random variables taking values in measurable spaces $(M_1, \mathcal{M}_1), ..., (M_k, \mathcal{M}_k)$ respectively, we can regard the vector $[X_1, ..., X_k]^T$ as a random variable in the product space $\prod_{j=1}^k M_j$, equipped with the product σ-algebra $\bigotimes_{j=1}^k \mathcal{M}_j$ (see Folland [4, §1.2]). The **joint distribution** of $X_1, ..., X_k$ is then defined to be the distribution of $[X_1, ..., X_k]^T$ in $\bigotimes_{j=1}^k M_j$. For example, if all of the X_j's take values in $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$, then the joint dis-

tribution of the X_j's would be the distribution of $[X_1, .., X_k]^T$ in $\mathbb{R}^k$ with the Borel σ-algebra $\mathcal{B}^k$.

Remark 2.1. Because the measure P is finite, a probability space is often easier to work with than more general measure spaces. For example, we have the especially simple situation that $L^q(P) \subset L^p(P)$ whenever $q \geq p$. To prove this, suppose $X \in L^q(P)$, let $A = \{\omega \in \Omega : |X| \leq 1\}$, and let $B = \{\omega \in \Omega : |X| > 1\}$. Then $E_A|X|^p \leq P(A) < \infty$, while $|X|^p \leq |X|^q$ on B so $E_B|X|^p \leq E_B|X|^q < \infty$. Therefore, $E|X|^p = E_A|X|^p + E_B|X|^p < \infty$.

Remark 2.2. In addition, verifying dominated convergence theorem is often easier in the probabilistic setting. Specifically, suppose that X_k is a sequence of random variables such that $X_k \to X$ a.s, and suppose we want to show that $\lim_{k\to\infty} E[X_k] = E[X]$. The dominated convergence theorem always applies when the sequence is bounded, say $|X_k| < b$ for all k, because we may take the constant random variable b as our dominating function. While in other settings in analysis using the dominated convergence theorem may require a careful argument, proofs in this paper will often just use the phrase "by dominated convergence" with no further explanation, when it is obvious we are working with a bounded sequence of random variables.

Before moving on, let us prove the following useful fact

Proposition 2.3 (Borel-Cantelli Lemma)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and suppose $\{A_n\}_{n=1}^{\infty}$ is a sequence of sets in $\mathcal{F}$ such that $\sum P(A_n) < \infty$. Then $P(\cap_{n=1}^{\infty} \cup_{j=n}^{\infty} A_n) = 0$.*

Proof. Observe that $P(\cap_{j=1}^{\infty} \cup_{n=j}^{\infty} A_n) = \lim_{j\to\infty} P(\cup_{n=j}^{\infty} A_j)$. If $\sum P(A_n) < \infty$, then

$$\lim_{j\to\infty} P(\cup_{n=j}^{\infty} A_j) \leq \lim_{j\to\infty} \sum_{n=j}^{\infty} P(A_n) \to 0 \qquad \text{as } j \to \infty. \tag{2}$$

Hence, the result follows. □

2.2 The Probabilistic Way of Thought

A notable characteristic of probability theory is that, in contrast to other fields of analysis, the spaces on which functions (i.e. random variables) are defined, play a somewhat peripheral role. If we want to understand some "space" K (e.g. a subset of $\mathbb{R}^n$, a graph, a set continuous functions, a collection of real world data, etc.), the probabilistic approach to the problem is to study random variables $X : \Omega \to K$, taking advantage of their formal properties. We draw conclusions about K itself by examining the joint and individual distributions of the random variables, which weight objects in K in various ways.

Within this (somewhat over-simplified) paradigm, essentially any probability space will do, provided that the random variables have the same formal properties, the same distributions, and the same joint distributions. This idea is made precise by the notion of an **extension** of a probability space. Given a probability space $(\Omega, \mathcal{F}, P)$, a probability space $(\Omega', \mathcal{F}', P')$ is said to extend the original probability space if there exists a measurable mapping $\varphi : \Omega' \to \Omega$, such that the mapping $\varphi^{-1} : \mathcal{F} \to \mathcal{F}'$ preserves probabilities,

i.e. $P(A) = P'(\varphi^{-1}(A))$ for all $A \in \mathcal{F}$. If $(M, \mathcal{M})$ is a measurable space, let us temporarily introduce the notation $V_\Omega(M, \mathcal{M})$ to denote the set of all random variables from Ω into $(M, \mathcal{M})$. The next proposition says that probability space extensions are characterized by the fact that they preserve distributions. Hence, from an abstract point of view, probability theory can regarded as the study of properties of probability spaces which are preserved under extensions.

Proposition 2.4. *A probability space $(\Omega', \mathcal{F}', P')$ extends a probability space $(\Omega, \mathcal{F}, P)$ if and only if, for each measurable space $(M, \mathcal{M})$, there exists a map $\Phi : V_\Omega(M, \mathcal{M}) \to V_{\Omega'}(M, \mathcal{M})$ such that, for all $X \in V_\Omega(M, \mathcal{M})$, $X =^d \Phi(X)$.*

Remark 2.5. Of course, the map Φ also preserves joint distributions, since the joint distribution of a collection of random variables is the distribution of a particular random variable.

Proof. If $(\Omega', \mathcal{F}', P')$ extends $(\Omega, \mathcal{F}, P)$ via the mapping $\varphi : \Omega' \to \Omega$, and $(M, \mathcal{M})$ is a measurable, space define $\Phi : V_\Omega(M, \mathcal{M}) \to V_{\Omega'}(M, \mathcal{M})$ by $\Phi(X) = X \circ \varphi$. Then for any $A \in \mathcal{M}$, $P(X \in A) = P(X^{-1}A) = P'(\varphi^{-1}X^{-1}A) = P'(X \circ \varphi \in A)$. Hence $X =^d \Phi(X)$.

Conversely, suppose such a map Φ exists for all measurable space $(M, \mathcal{M})$. Take $(M, \mathcal{M}) = (\Omega, \mathcal{F})$, let I be the identity random variable on $(M, \mathcal{M})$, and let $\varphi = \Phi(I)$. Then φ is a measurable map from Ω' into Ω. Moreover, if $A \in \mathcal{F}$, then $P(A) = P(I \in A) = P'(\Phi(I) \in A) = P'(\varphi^{-1}(A))$. Hence φ^{-1} preserves probabilities. Hence $(\Omega', \mathcal{F}', P')$ extends $(\Omega, \mathcal{F}, P)$ via the mapping φ. □

2.3 Independence

One concept more or less unique to probability theory is that of *independence*. This concept comes in a number of different guises, which we enumerate below.

- If $(\Omega, \mathcal{F}, P)$ is a probability space, two sets $F_1, F_2 \in \mathcal{F}$ are said to be **independent** if $P(F_1 \cap F_2) = P(F_1)P(F_2)$.
- Two sub-σ-algebras $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$ are said to be independent if, for every $F_1 \in \mathcal{F}_1$ and every $F_2 \in \mathcal{F}_2$, F_1 and F_2 are independent.
- Two random variables X_1 and X_2 are said to be independent if the σ-algebras which they generate are independent.
- A finite collection of sets $F_1, F_2, ..., F_n$ in $\mathcal{F}$ is said to be independent if, whenever $I \subset \{1, 2, ..., n\}$, $P(\cap_{i \in I} F_i) = \prod_{i \in I} P(F_i)$.
- Independence of a finite collection of σ-algebras $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ and independence of a finite collection of random variables $X_1, X_2, ...X_n$ are defined in obvious analogy to the previous cases.
- Finally, an arbitrary collection $\{F_\alpha\}_{\alpha \in A}$ of sets in $\mathcal{F}$ is said to be independent if every finite subcollection is independent, and similarly for arbitrary collections of σ-algebras and of random variables.

Independent random variables have excellent formal properties. For example, expectation distributes over products of random variables.

Proposition 2.6. *Suppose $X_1, ..., X_n$ is a collection of independent random variables in $L^1(P)$. Then*

$$E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n]. \tag{3}$$

Proof. First, suppose that $X_j = 1_{A_j}$, $1 \leq j \leq n$, where each $A_j \in \mathcal{F}$. Note that each $A_j \in \sigma(X_j)$ and $X_1 X_2 \cdots X_n = 1_{A_1 \cap \cdots \cap A_n}$. By independence

$$E[X_1 \cdots X_n] = P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n) = E[X_1] \cdots E[X_n]. \tag{4}$$

Since the expected value operator is linear, it then follows that (3) holds whenever $X_1, ..., X_n$ are simple functions. If we take $X_1, ..., X_n$ to be arbitrary elements of $L^1(P)$, since the simple functions are dense in $L^1(P)$, for each $1 \leq j \leq n$ there exists a sequence of simple functions $\{X_{j,k}\}_{k=1}^n$ such that $X_{j,k} \to X_j$ in L^1. In fact, we may assume that $|X_{j,1}| \leq |X_{j,2}| \leq \cdots \leq |X_j|$ (see Folland [4, Theorem 2.10]). The equality (3) then follows by dominated convergence, where $|X_j|$ is our dominating function. □

One reason why the concept of extensions for probability spaces is powerful is that it allows us to easily obtain large collections of independent random variables. To show how this may be accomplished, first we need to describe how the idea of product measure is extended to arbitrary products of measure spaces. We will not present the full construction here. For a detailed argument, see Williams [9] To sketch the key ideas of the construction, let $\{(\Omega_\alpha, \mathcal{F}_\alpha, P_\alpha) : \alpha \in A\}$ be an arbitrary collection of probability spaces, and let $\Omega = \prod_{\alpha \in A} \Omega_\alpha$. For each $\alpha \in A$, let $\pi_\alpha : \Omega \to \Omega_\alpha$ be the projection map, and let $\mathcal{F}$ be the σ-algebra generated by the projection maps π_α (i.e. $\mathcal{F}$ is the smallest σ-algebra such that each π_α is measurable). Let $\mathcal{A}_0$ be the collection of all finite intersections $\cap_{j=1}^m \pi_{\alpha_j}^{-1}(B_j)$ where $\alpha_1, ..., \alpha_m \in A$ and each $B_j \in \mathcal{F}_{\alpha_j}$. Let $\mathcal{A}$ be the set of all finite disjoint unions of sets in $\mathcal{A}_0$. Observe that the σ-algebra generated by the sets in $\mathcal{A}$ is precisely $\mathcal{F}$. Moreover, one can prove that $\mathcal{A}$ is an algebra, i.e. is closed under finite intersections and the taking of complements. Now each element $R \in \mathcal{A}$ takes the form

$$R = \bigcup_{i=1}^n \bigcap_{j=1}^{m_i} \pi_{\alpha_{i,j}}^{-1}(S_{i,j}), \tag{5}$$

where for each i, $\alpha_{i,1}, ..., \alpha_{i,m_i}$ are distinct elements of A, each each $S_{i,j} \in \mathcal{F}_{\alpha_{i,j}}$. Here we are thinking of R is a union of n disjoint rectangles, with sides $S_{i,j}$. We define a the set function P_0 on $\mathcal{A}$ by

$$P_0(R) = \sum_{i=1}^n \prod_{j=1}^{m_i} P_{\alpha_{i,j}}(S_{i,j}), \tag{6}$$

which is just the sum of the "areas" of each rectangle. Note that in particular $P_0(\Omega) = 1$. Of course, the representation (5) of elements of $\mathcal{A}$ is in general not unique, so one has to check that P_0 is well-defined. Having established this, one then proves

Lemma 2.7. *P_0 is a premeasure on the algebra $\mathcal{A}$.*

We remark that even though as a product space Ω may have infinitely many factors, the definition of P_0 involves only finite products, and consequently the proof of the proposition above follows essentially verbatim the typical argument that is given in the construction of product measures on finite product spaces (see e.g. Folland [4, §2.5]). Using Carathéodory's extension theorem (see [4, Theorems 1.11 and 1.14]), one then can prove

Proposition 2.8. *There exists a unique probability measure P on $\mathcal{F}$ such that $P|_{\mathcal{A}} = P_0$.*

In particular, the measure P is the outer measure on $\mathcal{F}$ induced by P_0. In this construction, the σ-algebra $\mathcal{F}$ generated by the projection maps π_α is called the **product σ-algebra** on Ω, and the measure P is called the **product measure** on Ω. The concept of product measure allows us to prove the following result, which will play an essential role in the construction of Brownian motion.

Proposition 2.9. *Suppose $\{X_\alpha\}_{\alpha\in A}$ is any (finite or infinite) collection of random variables on the probability space $(\Omega, \mathcal{F}, P)$. Then there exists an extension $(\Omega', \mathcal{F}', P')$ and an independent collection of random variables $\{Y_\alpha\}_{\alpha\in A}$ on Ω' such that $Y_\alpha =^d X_\alpha$ for all $\alpha \in A$.*

Proof. Let $\Omega' = \prod_{\alpha\in A} \Omega$ and let $\mathcal{F}'$ be the corresponding product σ-algebra. For each $\alpha \in A$, let $\pi_\alpha : \Omega' \to \Omega$ be the projection onto α'th factor of Ω'. As above, we define P' to be the corresponding product measure on $\mathcal{F}'$. Hence, for each $\alpha \in A$ and $F \in \mathcal{F}$, $P'(\pi_\alpha^{-1}A) = P(A)$, and if $F_1, ..., F_n \in \mathcal{M}$, and $\alpha_1, ..., \alpha_n$ are distinct elements of A, then

$$P'(\cap_{j=1}^n \pi_{\alpha_j}^{-1} F_j) = \prod_{j=1}^n P(F_j). \tag{7}$$

Moreover, it is clear that the probability space $(\Omega', \mathcal{F}', P')$ is an extension of $(\Omega, \mathcal{F}, P)$, because any one of the maps π_α is a measurable map such that π_α^{-1} preserves probabilities.

Let $(M, \mathcal{M})$ be the measurable space which is the common codomain of the X_α's. For each $\alpha \in A$, define $Y_\alpha = X_\alpha \circ \pi_\alpha$. Given distinct elements $\alpha_1, ..., \alpha_n$ in A and sets $B_1, ..., B_n \in \mathcal{M}$, we have

$$\begin{aligned} P'(Y_{\alpha_j} \in B_j, \text{ for } 1 \le j \le n) &= P'(\cap_{j=1}^n \pi_{\alpha_j}^{-1} X_{\alpha_j}^{-1} B_j) \\ &= \prod_{j=1}^n P'(\pi_{\alpha_j}^{-1} X_{\alpha_j}^{-1} B_j) = \prod_{j=1}^n P'(Y_{\alpha_j} \in B_j), \end{aligned} \tag{8}$$

here using equation (7). This completes the proof. □

2.4 Characteristic functions

Characteristic functions provide us with an efficient method to check when two random variables X and Y taking values in $\mathbb{R}^n$ have the same distribution. To describe these functions, we begin by defining the **Fourier transform of a finite measure**. If μ is a finite measure (think probability measure), the Fourier transform of μ is simply the function

$$\hat{\mu}(\xi) = \int e^{-2\pi i x\cdot\xi} d\mu(x) \qquad (\xi \in \mathbb{R}^n). \tag{9}$$

(This is of course in direct analogy to the definition of the usual Fourier transform of a function f on $\mathbb{R}^n$, denoted by $\hat{f}$. See Appendix B.)

If X is a random variable taking values in $\mathbb{R}^n$, let P_X denote the distribution of X. The **characteristic function** of X is the function $\varphi = \varphi_X$ defined to be the Fourier transform of the measure P_X, i.e.

$$\varphi_X = \widehat{P_X}. \tag{10}$$

A simple but useful observation is that, for any random variable X,

$$\varphi_X(\xi) = \int e^{-2\pi i \xi \cdot x} dP_X(x) = E[e^{-2\pi i \xi \cdot X}]. \tag{11}$$

Further, in the special case where the distribution of X is given by a density function ρ, then

$$\varphi_X(\xi) = \int e^{-2\pi i x \cdot \xi} dP_X(x) = \int e^{-2\pi i x \cdot \xi} \rho(x) dx = \hat{\rho}(\xi). \tag{12}$$

Characteristic functions are useful because of the following result, which we will prove using the Fourier inversion theorem (see Theorem 7.6),

Theorem 2.10. *Any finite measure μ on $\mathcal{B}^n$ is uniquely determined by its Fourier transform $\hat{\mu}$.*

To prove this fact, first we need a lemma.

Lemma 2.11. *Suppose μ is a finite measure on $\mathcal{B}^n$, and suppose $f \in L^1(\mathbb{R}^n)$. Then*

$$\int \hat{f}(x) d\mu(x) = \int f(\xi) \hat{\mu}(\xi) d\xi. \tag{13}$$

Proof. This is simply a matter of applying Fubini's Theorem. We have

$$\begin{aligned} \int \hat{f}(x) d\mu(x) &= \iint e^{-2\pi i x \cdot \xi} f(\xi) d\xi d\mu(x) \\ &= \int f(\xi) \int e^{-2\pi i x \cdot \xi} d\mu(x) d\xi = \int f(\xi) \hat{\mu}(\xi) d\xi, \end{aligned} \tag{14}$$

as required. □

Proof of Theorem 2.10. Suppose μ_1 and μ_2 are two finite Borel measures such that $\hat{\mu_1} = \hat{\mu_2}$. Since the closed rectangles in $\mathbb{R}^n$ generate the Borel sets, to prove that $\mu_1 = \mu_2$, it is enough to prove that $\mu_1(R) = \mu_2(R)$ for any closed rectangle $R \subset \mathbb{R}^n$. Let $f = \chi_R$, and note that $f \in L^1$. Further, $|f^\vee(x)| \leq \int |f(x)| dx = m(R) < \infty$, where m is Lebesgue measure on $\mathcal{B}^n$), so $f^\vee$ is bounded. Then, by Corollary 7.9, there exists a sequence of smooth L^1 functions $\{f_j\}$ such that $f_j \to f$ a.e., and each $f_j^\vee \in L^1$. By the inversion theorem (Theorem 7.6), $\widehat{f_j^\vee} = f_j$. Thus by Lemma 2.11 and equality of $\hat{\mu_1}$ and $\hat{\mu_2}$,

$$\int f_j(x) d\mu_1(x) = \int f_j^\vee(\xi) \hat{\mu_1}(\xi) d\xi = \int f_j^\vee(\xi) \hat{\mu_2}(\xi) d\xi = \int f_j(x) d\mu_2(x). \tag{15}$$

By dominated convergence, this equality implies

$$\mu_1(R) = \lim_{j \to \infty} \int f_j(x) d\mu_1(x) = \lim_{j \to \infty} \int f_j(x) d\mu_2(x) = \mu_2(R). \tag{16}$$

This completes the proof. □

The uniqueness of the Fourier transform of finite measures clearly implies

Corollary 2.12. *Suppose X and Y are two random variable in $\mathbb{R}^n$. Then $\varphi_X = \varphi_Y$ iff $X =^d Y$.*

The situation where we need to check that two random variables have equal distributions will come up quite often in this paper. Corollary 2.12 will therefore be an indispensable result.

2.5 Stochastic Processes

Stochastic process is a broad term which may refer to anything we think of as a random quantity evolving over time. As the main stochastic process of interest in this paper is Brownian motion (see §4.1), we will use this subsection to present some basic results about stochastic processes which can be easily proved in general, and we will provide further refinements later in the paper.

In its most basic definition, a stochastic process is parametrized family of random variables $\{X_t\}_{t\in T}$ taking values in a measurable space $(M, \mathcal{M})$, where $T \subset [0, \infty)$. For this paper, we will almost always take $T = [0, \infty)$ and $(M, \mathcal{M}) = (\mathbb{R}^n, \mathcal{B}^n)$. Note a stochastic process $\{X_t\}$ is associated with two collections of functions. First, for each fixed $t \in T$, there is the random variable

$$X_t : \Omega \to M, \quad \omega \mapsto X_t(\omega), \tag{17}$$

and second, for each fixed $\omega \in \Omega$, there is the function

$$X(\omega) : T \to M, \quad t \mapsto X_t(\omega). \tag{18}$$

The later function is called the **sample path** (or sometimes just path) of X corresponding to ω.

According to the basic definition of a stochastic process given above, $\{X_t\}$ is simply an indexed collection of random variables which may not be related to each other in any meaningful way. In order to put greater emphasis on the sample paths, it is useful to specialize definition of a stochastic process as follows. Let M^T denote the set functions $f : T \to M$. For each $t \in T$, let $\pi_t : M^T \to M$ be the projection map onto the t'th factor, i.e. $\pi_t(f) = f(t)$ for all $f \in M$. We define $\mathcal{M}^T$ to be the product σ-algebra generated by the projection maps π_t. In other words, $\mathcal{M}^T$ is the smallest σ-algebra such that all the maps π_t are measurable. We now make the following official definition.

Definition 2.13 (Stochastic process)**.** *A* stochastic process *is a random variable X from Ω into the space $(M^T, \mathcal{M}^T)$.*

Having established this definition, we will fix T and M for the discussion that follows. The usual notation π_t for projection maps is not quite adequate for our needs, so we introduce following notation.

Notation 2.14.

- If $t \in T$, as before $\pi_t : M^T \to M$ denotes the projection map onto the t'th factor, that is $\pi_t(f) = f(t)$ for $f \in M^T$.

- If $U \subset T$, we define $\pi_U : M^T \to M^U$ by $\pi_U(f) = f|_U$ for $f \in M^T$.
- If $t \in U$, we define $\pi_t^U : M^U \to M$ by $\pi_t^U = f(t)$ for $f \in M^U$.
- Finally, if $V \subset U$, we define $\pi_V^U : M^U \to M^V$ by $\pi_V^U(f) = f|_V$ for $f \in M^U$.

For the discussion that follows, for any $U \subset T$, we will assume that the product space M^U is equipped with the product σ-algebra $\mathcal{M}^U$, generated by the projection maps π_t^U. Hence each projection map π_t^U is measurable. Moreover, we have

Proposition 2.15. *Suppose $V \subset U \subset T$. Then π_V^U is measurable.*

Proof. Observe that, for any $t \in U$, $\pi_t^U = \pi_t^V \circ \pi_V^U$. Now $\mathcal{M}^V$ is generated by sets of form $A = (\pi_t^V)^{-1}B$, where $B \in \mathcal{M}$. Observe that $(\pi_V^U)^{-1}A = (\pi_t^V \circ \pi_V^U)^{-1}B = (\pi_t^U)^{-1}B \in \mathcal{M}^U$, because $\mathcal{M}^U$ is generated by sets of this form. Therefore π_V^U is measurable. □

Define a **cylinder** of $\mathcal{M}^T$ to be a set of form $C = \pi_S^{-1}E$, where $S \subset T$ is *finite* and $E \in \mathcal{M}^S$. Note that a cylinder $C \in \mathcal{M}^T$ because π_S is measurable by Proposition 2.15. Let $\mathcal{C}^T \subset \mathcal{M}^T$ denote the set of all cylinders of $\mathcal{M}^T$. When working with the σ-algebra $\mathcal{M}^T$, the following fact is sometimes useful.

Proposition 2.16. *The collection $\mathcal{C}^T$ is an algebra. Moreover, $\mathcal{C}^T$ is a generating set for $\mathcal{M}^T$*

Proof. The latter statement is clear, because $\mathcal{C}^T$ is a subset of $\mathcal{M}^T$ and contains the collection of sets $\{\pi_t^{-1}B : t \in T \text{ and } B \in \mathcal{M}\}$ which generates $\mathcal{M}^T$. As for the rest of the claim, let $S \subset T$ be finite, and let $E \in \mathcal{M}^S$. Then

$$(\pi_S^{-1}E)^c = \pi_S^{-1}(E^c) \in \mathcal{C}^T. \tag{19}$$

Hence $\mathcal{C}^T$ is closed under the taking of complements. Moreover, if $S' \subset T$ is another finite set and $F \in \mathcal{M}^{S'}$, observe that

$$\pi_S^{-1}E \cap \pi_{S'}^{-1}F = \pi_{S\cup S'}^{-1}G, \tag{20}$$

where $G = \{f \in M^{S\cup S'} : f|_S \in E \text{ and } f|_{S'} \in F\} = (\pi_S^{S\cup S'})^{-1}E \cap (\pi_{S'}^{S\cup S'})^{-1}F$. Since the maps $\pi_S^{S\cup S'}$ and $\pi_{S'}^{S\cup S'}$ are measurable, it follows that $G \in \mathcal{M}^{S\cup S'}$. Therefore, $\pi_S^{-1}E \cap \pi_{S'}^{-1}F \in \mathcal{M}^{S\cup S'}$. We conclude that $\mathcal{C}^T$ is closed under finite intersections. This completes the proof. □

Note that with definition (2.13), it makes sense to speak of the **law, or distribution, of a stochastic process**. If X is a stochastic process, this is simply the measure P_X on $\mathcal{M}^T$ defined by

$$P_X(E) = P(X \in E) \qquad (E \in \mathcal{M}^T). \tag{21}$$

In practice, rather than working directly with the law, or distribution of a stochastic process, it is often easier to work with the finite dimensional distributions, which are defined as follows. Since the maps π_S are measurable, the function

$$\pi_S \circ X \tag{22}$$

is a random variable taking values in the measurable space $(M^S, \mathcal{M}^S)$. Let us denote the law of such a random variable by μ_S. We then define the collection

$$\{\mu_S : S \subset T \text{ is finite}\} \tag{23}$$

to be the **finite-dimensional distributions** of X. Explicitly, if say $S = \{s_1, ..., s_m\} \subset T$, and $A = A_1 \times \cdots \times A_n \subset M^S$, then

$$\mu_S(A) = P(X_{s_1} \in A_1, ..., X_{s_n} \in A_n). \tag{24}$$

This is simply the joint distribution of the random variables $X_{s_1}, ..., X_{s_n}$.

To understand the distribution of a stochastic process, it turns out that the finite-dimensional distributions tell us everything.

Proposition 2.17. *The finite-dimensional distributions of a stochastic process uniquely determine its law.*

Proof. Let X be a stochastic process. By the Carathéodory extension theorem and the fact that the cylinders $\mathcal{C}^T$ form an algebra which generates $\mathcal{M}^T$ (see Folland [4, Theorems 1.11 and 1.14]), the measure P_X is uniquely determined by its values on sets of form $\pi_S^{-1}E$, where $S \subset T$ is finite and $E \in \mathcal{M}^S$. But

$$P_X(\pi_S^{-1}E) = P(X \in \pi_S^{-1}E) = P(\pi_S \circ X \in E) = \mu_S(E). \tag{25}$$

Thus P_X is uniquely determined by the finite-dimensional distributions μ_S. □

Finally, we restrict our attention to the case of a stochastic process $X = \{X_t\}_{t \in T}$, where $T = [0, \infty)$ and $M = \mathbb{R}^n$ with the Borel σ-algebra. A rather important case of this situation is the following.

Definition 2.18 (Continuous stochastic process)**.** *Suppose that $X = \{X_t\}_{t \geq 0}$ is a stochastic process in $\mathbb{R}^n$.*

i. We say that X is continuous *if the paths of X are continuous, i.e. $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega$.*

ii. We say that X is almost surely continuous *if there exists a null set $N \in \mathcal{F}$ such that $t \mapsto X_t(\omega)$ is continuous for all $\omega \in N^c$.*

Note that since stochastic processes are just random variables, it makes sense to speak of two stochastic processes being independent, and we may apply Proposition 2.9 to obtain (possibly large) collections of stochastic processes with given distributions. However, even if we start with continuous stochastic processes, there is no reason that the resulting independent processes should still be continuous. Later in this paper, it will be essential to be able to obtain independent continuous stochastic processes, and with this end in mind we prove the following proposition. The proof is very similar to the proof of Proposition 2.9.

Proposition 2.19. *Suppose $X = \{X_t\}_{t \geq 0}$ and $Y = \{Y_t\}_{t \geq 0}$ are stochastic processes taking values in $\mathbb{R}^n$. Then (after possibly extending the probability space) there exist stochastic processes $X' = \{X'_t\}_{t \geq 0}$ and $Y' = \{Y'_t\}_{t \geq 0}$, such that*

a. *For all* $t \geq 0$, $X_t =^d X'_t$ *and* $Y_t =^d Y'_t$.

b. *For all* $s, t \geq 0$, X'_s *and* Y'_t *are independent.*

c. *Moreover, if* X *and* Y *are almost surely continuous,* X' *and* Y' *may be chosen to be almost surely continuous.*

Proof. If X and Y are initially defined on the probability space $(\Omega, \mathcal{F}, P)$, define a new probability space $(\Omega', \mathcal{F}', P')$ by letting $\Omega' = \Omega \times \Omega$ and letting $\mathcal{F}'$ and P' be the induced product σ-algebra and product measure on Ω'. For $t \geq 0$ and $(\omega_1, \omega_2) \in \Omega' = \Omega \times \Omega$, we define

$$\begin{aligned} X'_t(\omega_1, \omega_2) &= X_t(\omega_1), \text{ and} \\ Y'_t(\omega_1, \omega_2) &= Y_t(\omega_2). \end{aligned} \tag{26}$$

Given a Borel set $B \subset \mathbb{R}^n$, let $E = X_t^{-1}B$. Clearly $X_t'^{-1}B = E \times \Omega$. Therefore, $P'(X'_t \in B) = P'(E \times \Omega) = P(E)P(\Omega) = P(E) = P(X_t \in B)$. Thus $X_t =^d X'_t$. Similarly $Y_t =^d Y'_t$.

Given $s, t \geq 0$, to see that X'_s and Y'_t are independent, let $B, A \subset \mathbb{R}^n$ be a Borel sets, let $E = X_s^{-1}B$ and $F = Y_t^{-1}A$. Then $X_s'^{-1}B = E \times \Omega$ and $Y_t'^{-1}A = \Omega \times F$. Hence $P'(X'_s \in B, Y'_s \in A) = P'(E \times F) = P'(E)P'(F) = P'(X'_t \in B)P(Y'_t \in A)$.

Finally suppose that X has almost surely continuous paths. Then there exists $N \subset \Omega$ such that $P(N) = 0$, and for all $\omega_1 \in N^c$, $X_t(\omega_1)$ is continuous in t. Then it is clear from the definition (26) that $X'_t(\omega_1, \omega_2)$ is continuous for all $(\omega_1, \omega_2) \in N^c \times \Omega$. Since $P'(R \times \Omega) = 1$, this shows that X'_t is almost surely continuous. Similarly, if Y has almost surely continuous paths, then so does Y'. This completes the proof. □

3 Gaussian Distribution

The first ingredient needed for the mathematical description of Brownian motion is Gaussian distribution.

Basic definitions and results

A random variable X has **Gaussian** or **normal** distribution (the names are entirely interchangeable) if its distribution is given by $dP_X(x) = b^{-1}e^{-\pi(x-a)^2/b^2}dx$, where a, b real numbers $b > 0$. However, it is convenient to include in our definition degenerate distribution, where the variance of X is zero. We therefore take the following as our official definition of normal distribution.

Definition 3.1 (normal distribution)**.** *A random variable* X*, taking values in* $\mathbb{R}$*, is said to be* normally distributed*, if either the variance of* X *is zero, or the distribution of* X *is given by the density function*

$$\rho(x) = b^{-1}e^{-\pi(x-a)^2/b^2}, \tag{27}$$

for some $a, b \in \mathbb{R}$ *and* $b > 0$.

Remark 3.2. If the distribution of X is given by (27), then a and b^2 are respectively the mean and variance of X. This may be checked by writing

$$E[X] = \int b^{-1} x e^{-\pi(x-a)^2/b^2} dx \qquad \text{and} \qquad \text{Var}(X) = \int b^{-1}(x-a)^2 e^{-\pi(x-a)^2/b^2} dx, \tag{28}$$

and applying standard integration techniques.

The next result is trivial, but it is important since it will allow us to streamline many arguments to come.

Proposition 3.3. *The distribution of a Gaussian random variable is uniquely determined by its mean and variance.*

Proof. Let X be a Gaussian random variable, let $a = E[X]$ and $b = \text{Var}(X)$. If $b > 0$, the result is immediate from the definition of Gaussian distribution and the previous remark. If $b = 0$, then $X = a$ a.s. In this case, the distribution of X is defined by $P_X(B) = 1$ if $a \in F$ and $P_X(B) = 0$ otherwise, for all Borel sets B. □

This justifies the following notation.

Notation 3.4. *To indicate the distribution of a Gaussian random variable, we write $X \in \mathcal{N}(a, b^2)$, where $a = E[X]$ and $b^2 = Var(X)$.*

Using Proposition 7.8 in Appendix B, we have

Proposition 3.5. *If $X \in \mathcal{N}(a, b^2)$, then the characteristic function of X is*

$$\varphi(\xi) = e^{-\pi b^2 \xi^2} e^{-2\pi i a \xi}, \tag{29}$$

Proof. By (12), if $b > 0$ and ρ is the density function of X, then $\varphi_X = \hat{\rho}$. Using Proposition 7.8, we compute

$$\begin{aligned} \hat{\rho}(\xi) &= b^{-1} \int e^{-2\pi i x \xi} e^{-\pi(x-a)^2/b^2} dx = b^{-1} e^{-2\pi i a \xi} \int e^{-2\pi i y \xi} e^{-\pi y^2/b^2} dy \\ &= e^{-2\pi i a \xi} \int e^{-2\pi i b z \xi} e^{-\pi z^2} dz = e^{-\pi b^2 \xi^2} e^{-2\pi i a \xi}, \end{aligned} \tag{30}$$

which is the desired result. If on the other hand, $b = 0$, then $X = a$ a.s., and so $\varphi_X(\xi) = E[e^{-2\pi i \xi X}] = e^{-2\pi i a \xi}$. □

The next proposition illustrates how Gaussian distribution behaves well under the image of linear maps.

Proposition 3.6. *Suppose $\{X_j\}_{j=1}^m$ is a collection of independent Gaussian random variables, such that each $X_j \in \mathcal{N}(a_j, b_j^2)$, for some real numbers a_j and b_j. Let $Y = \sum_1^m c_j X_j$ be a linear combination of the X_j's, where each $c_j \in \mathbb{R}$. Then $Y \in \mathcal{N}(\sum c_j a_j, \sum c_j^2 b_j^2)$.*

Proof. Note that $E[c_jX_j] = c_jE[X_j] = c_ja_j$ and $\mathrm{Var}(c_jX_j) = c_j^2\mathrm{Var}(X_j) = c_j^2b_j^2$. Hence each $c_jX_j \in \mathcal{N}(c_ja_j, c_j^2b_j^2)$. Therefore, using the previous proposition and Proposition 2.6, we compute

$$\begin{aligned}\varphi_Y(\xi) = E[e^{-2\pi i\xi Y}] = E[\prod_{j=1}^{m} e^{-2\pi i\xi c_jX_j}] = \prod_{j=1}^{m} E[e^{-2\pi i\xi c_jX_j}] \\ = \prod_{j=1}^{m} \varphi_{c_jX}(\xi) = \prod_{j=1}^{m} e^{-\pi c_j^2b_j^2\xi^2}e^{-2\pi ic_ja_j\xi} = e^{-\pi(\sum c_j^2b_j^2)\xi^2}e^{-2\pi i(\sum c_ja_j)\xi}.\end{aligned} \tag{31}$$

But by the previous proposition this is the characteristic function of a random variable in $\mathcal{N}(\sum c_ja_j, \sum c_j^2b_j^2)$. The result follows by uniqueness in Corollary 2.12. □

The following simple estimate will be needed later.

Proposition 3.7. *Suppose $Z \in \mathcal{N}(0,1)$ and $r \geq 1/2\pi$. Then*

$$P(Z \geq r) \leq e^{-\pi r^2}. \tag{32}$$

Proof. Indeed,

$$\begin{aligned}P(Z \geq r) = \int_r^\infty e^{-\pi x^2}dx \leq \int_r^\infty \frac{x}{r}e^{-\pi x^2}dx \\ = (-\frac{1}{2\pi r}e^{-\pi x^2})|_r^\infty \leq e^{-\pi r^2},\end{aligned} \tag{33}$$

as claimed. □

Multidimensional Gaussian distribution

Proposition 3.6 motivates the following definition for Gaussian distribution in $\mathbb{R}^n$.

Definition 3.8 (Multinormal distribution)**.** *A random variable X in $\mathbb{R}^n$ is said to be* normally distributed *if its image $\varphi(X)$ under any linear functional $\phi : \mathbb{R}^n \to \mathbb{R}$ is a Gaussian random variable in $\mathbb{R}$.*

Other names for such a distribution are **(multi)normal distribution** or **Gaussian distribution**. If the components of X are given explicitly, say $X = [X_1, X_2, ..., X_n]^T$, we say that the X_j's are **jointly Gaussian** to mean that X is Gaussian.

Remark 3.9. Gaussian distribution is preserved by linear maps. Indeed, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, and X is a Gaussian random variable in $\mathbb{R}^n$, then TX is a Gaussian random variable in $\mathbb{R}^m$, because if ϕ is a linear functional on $\mathbb{R}^m$, then $\phi \circ T$ is a linear functional on $\mathbb{R}^n$.

If X is a Gaussian random variable in $\mathbb{R}^n$, the analogue for the one-dimensional concept of variance is the **covariance** of X, defined to be the matrix $\mathrm{Cov}(X) = E[XX^T]$. Proposition 3.3 turns out to still be essentially true in higher dimensions.

Proposition 3.10. *The distribution of a Gaussian random variable in $\mathbb{R}^n$ is uniquely determined by its mean and covariance.*

Proof. Suppose X and Y are two Gaussian random variables in $\mathbb{R}^n$ such that $E[X] = E[Y]$ and $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$. By replacing X and Y with $X - a$ and $Y - a$, where $a = E[X] = E[Y]$, we may assume that $E[X] = E[Y] = 0$. By definition of Gaussian distribution in $\mathbb{R}^n$, for any $\xi \in \mathbb{R}^n$, $\xi \cdot X$ and $\xi \cdot Y$ are Gaussian random variables in $\mathbb{R}$. Also $E[\xi \cdot X] = E[\xi \cdot Y] = 0$ since X and Y are mean-zero. Moreover, $\mathrm{Var}(\xi \cdot X) = E[(\xi \cdot X)^2] = E[\xi^T X X^T \xi] = \xi^T \mathrm{Cov}(X) \xi$. Similarly $\mathrm{Var}(\xi \cdot Y) = \xi^T \mathrm{Cov}(Y) \xi$. Hence $\mathrm{Var}(\xi \cdot X) = \mathrm{Var}(\xi \cdot Y)$. Thus by Proposition 3.3 $\xi \cdot X =^d \xi \cdot Y$. Thus by Corollary 2.12 $\varphi_{\xi \cdot X} = \varphi_{\xi \cdot Y}$. Observe that

$$\varphi_{\xi \cdot X}(1) = E[e^{-2\pi i \xi \cdot X}] = \varphi_X(\xi). \tag{34}$$

Similarly, $\varphi_{\xi \cdot Y}(1) = \varphi_Y(\xi)$. Therefore, $\varphi_X(\xi) = \varphi_Y(\xi)$. It then follows by Corollary 2.12 again that $X =^d Y$. □

This proposition justifies introducing the following notation for the distribution of a Gaussian random variable.

Notation 3.11. *To indicate the distribution of a Gaussian random variable in $\mathbb{R}^n$, we write $X \in \mathcal{N}_n(a, S)$, where $a = E[X]$ and $S = Cov(X)$.*

The most fundamental example of an n-dimensional Gaussian random variable is $Z = [Z_1, ..., Z_n]^T$, where the Z_j's are independent $\mathcal{N}(0, 1)$ random variables. The fact that Z is Gaussian follows from Proposition 3.6, which shows that a linear combination of independent Gaussian random variables is Gaussian. We call Z (or any random variable with the same distribution) a **standard normal random variable in** $\mathbb{R}^n$. Observe that, using the notation introduced above, $Z \in \mathcal{N}_n(0, I)$, where I is the $n \times n$ identity matrix. The next proposition shows that any n-dimensional Gaussian random variable is the affine linear image of a standard normal random variable.

Proposition 3.12. *If X is a Gaussian random variable in $\mathbb{R}^n$, then $X =^d BZ + a$, for some $B \in M_n(\mathbb{R})$, $a \in \mathbb{R}^n$, and standard normal random variable Z in $\mathbb{R}^n$. Furthermore, $E[X] = a$ and $Cov(X) = BB^T$.*

First we need a lemma.

Lemma 3.13. *Suppose S is a symmetric, positive semidefinite matrix in $M_n(\mathbb{R})$. Then $S = BB^T$ for some $B \in \mathbb{R}^n$.*

Proof. Since S is symmetric, by the spectral theorem for real matrices, we may write $S = GDG^T$ for some orthogonal matrix G and diagonal matrix D. Since S is positive semidefinite, the diagonal entries of D are nonnegative. Therefore, we may write $D = H^2$, where H is the diagonal matrix whose diagonal entries are the nonnegative square roots of the diagonal entries of D. Thus, if we set $B = GH$, then $S = BB^T$, as required. □

Proof of Proposition 3.12. We may assume $EX = 0$ and prove that $X =^d BZ$, because otherwise we may set $a = EX$ and replace X with $X - a$. Let $S = \mathrm{Var}(X) = E[XX^T]$, and observe that S is a symmetric, positive semidefinite matrix. Indeed, given a nonzero

$z \in \mathbb{R}^n$, $z^T X$ is a Gaussian random variable in $\mathbb{R}$, by definition of Gaussian distribution in $\mathbb{R}^n$. Expected value commutes with linear maps, so $z^T S z = E[z^T X X^T z] = E[(z^T X)^2] \geq 0$. Thus by the lemma we may write $S = BB^T$ for some $B \in M_n(\mathbb{R})$. Let Z be a standard normal random variable in $\mathbb{R}^n$, and observe that $E[BZ] = BE[Z] = 0$ and $\text{Cov}(BZ) = E[BZZ^T B^T] = B\text{Cov}(Z)B^T = BB^T = S$. Thus the mean and covariance of BZ agree with those of X. Thus by 3.10, $X =^d BZ$, as required. □

Corollary 3.14. *Suppose X is a Gaussian random variable in $\mathbb{R}^n$. Let $S = Cov(X)$ and $a = E[X]$. Then*

$$\varphi_X(\xi) = e^{-\pi \xi^T S \xi} e^{-2\pi i \xi \cdot a}. \tag{35}$$

Proof. By the previous proposition, $X =^d BZ + a$, where $a = E[X]$ and $BB^T = S$. If $u \in \mathbb{R}^n$, then $u^T Z = u_1 Z_1 + u_2 Z_2 + \cdots + u_n Z_n$. Hence by Proposition 3.6 $u^T Z \in \mathcal{N}(0, u^T u)$. Then by Proposition 3.5

$$\varphi_{u^T Z}(1) = E e^{-2\pi i u^T Z} = e^{-\pi u^T u}. \tag{36}$$

Substituting $u = B^T \xi$, where $\xi \in \mathbb{R}^n$, yields

$$E e^{-2\pi i \xi^T B Z} = e^{-\pi \xi^T B B^T \xi}. \tag{37}$$

Therefore, the characteristic of function of X is

$$E e^{-2\pi i \xi \cdot X} = E[e^{-2\pi i \xi^T B Z} e^{-2\pi i \xi \cdot a}] = e^{-\pi \xi^T B B^T \xi} e^{-2\pi i \xi \cdot a}, \tag{38}$$

as required. □

We close this section with a result showing that Gaussian random variables behave nicely with respect to limits. This result will be useful to have on hand when we construct the Wiener process in §4.

Proposition 3.15. *Suppose $\{X_k\}_{k=1}^{\infty}$ is a sequence of Gaussian random variables in $\mathbb{R}^n$ such that $X_k \to X$ a.s., and the components of X are in $L^2(P)$. Then X is also a Gaussian random variable.*

Proof. Set $a_k = E[X_k]$ and $a = E[X]$. Set $S_k = Cov(X_k) = E[X_k X_k^T]$ and let $S = Cov(X) = E[XX^T]$. We claim that $a_k \to a$ and $S_k \to S$ as $k \to \infty$. These assertions follow by dominated convergence applied to the components of a_k and S_k. Indeed, by assumption the components of X are in $L^2(P)$ (and hence in $L^1(P)$ by Remark 2.1). The components of each X_k are in $L^2(P)$ since these are Gaussian. Thus, for $1 \leq j \leq n$, $(a_k)_j = E[(X_k)_j] \to E[X_j] = a_j$, by taking $2|X_j|$ as our dominating function. For $1 \leq i, j \leq n$, $(S_k)_{i,j} = E[(X_k)_i (X_k)_j] \to E[X_i X_j] = S_{i,j}$, by taking $2|X_i X_j|$ as our dominating function.

We also observe S is symmetric positive semidefinite. Indeed, S is symmetric because this is true of any covariance matrix. Let v be an eigenvector of S, and let λ be the corresponding eigenvalue. Then $v^T M_k v \geq 0$ for all k, because each M_k is the covariance matrix of a Gaussian random variable and hence is positive semidefinite. Hence $\lambda |v|^2 = v^T M v = \lim_{k \to \infty} v^T M_k v \geq 0$. This shows $\lambda \geq 0$, so S is positive semidefinite.

Next, we compute the characteristic functions. By Corollary 3.14, and the fact that each of the X_k's are normal, we have

$$\varphi_{X_k}(\xi) = E e^{-2\pi i \xi \cdot X_k} = e^{-\pi \xi^T S_k \xi} e^{-2\pi i \xi a_k} \to e^{-\pi \xi^T S \xi} e^{-2\pi i \xi a}, \tag{39}$$

as $k \to \infty$. On the other hand, by dominated convergence,

$$\varphi_{X_k}(\xi) = Ee^{-2\pi i \xi \cdot X_k} \to Ee^{-2\pi i \xi \cdot X} = \varphi_X(\xi), \tag{40}$$

as $k \to \infty$. Therefore, $\varphi_X(\xi) = e^{-\pi \xi^T S \xi} e^{-2\pi i \xi a}$. Since S is symmetric positive definite, by Lemma 3.13, $S = BB^T$ for some $B \in M_n(\mathbb{R})$. Hence φ_X is the characteristic function of the Gaussian random variable $BZ + a$, where $Z \in \mathcal{N}(0, I)$. Thus X is Gaussian, and in particular $X \in \mathcal{N}(a, S)$. □

4 Wiener Process

The Weiner process is the mathematical model for Brownian motion. This section defines the Wiener process, first in $\mathbb{R}$ and then in $\mathbb{R}^n$, and establishes several key invariance properties. Proof of the existence of the Wiener process is deferred to the next section.

Definition 4.1 (Weiner process in $\mathbb{R}$)**.** *Suppose $X = \{X_t\}_{t \geq 0}$ is a real valued stochastic process. We say that X is a* Wiener process *if the following four conditions are satisfied.*

i. $X_0 = 0$ *a.s.*

ii. If $0 \leq t_0 < \cdots < t_m$, then, for $1 \leq j \leq m$, the increments $X_{t_j} - X_{t_{j-1}}$ are independent.

iii. If $0 \leq s < t$, then $X_t - X_s \in \mathcal{N}(0, t - s)$.

iv. The paths of X are almost surely continuous (i.e. there exists a measurable set $A \subset \Omega$ such that $P(A) = 1$ and for all $\omega \in A$, the map $t \mapsto X_t(\omega)$ is continuous).

More generally, if we replace condition (i) with the requirement that $X_0 = x$ a.s. for some $x \in \mathbb{R}$, we say that X is a Wiener process starting from x.

A stochastic process Y is said to be a **Gaussian process** if the finite-dimensional-distributions of Y are Gaussian. Since by condition (iii), for $t \geq 0$, $X_t \in \mathcal{N}(0, t)$, it follows that the Wiener process is an example of a Gaussian process.

Remark 4.2. While strictly speaking the term *Wiener process* refers to the mathematical model for the physical phenomenon of *Brownian motion*, in practice the term Brownian motion is more common and in a mathematical context is understood to refer to the Wiener process. In this paper, we typically use the term Wiener process, but sometimes use the term Brownian motion in more informal discussions.

The next two propositions gives us equivalent characterizations for a Wiener process which are often easier to verify than Definition 4.1.

Proposition 4.3. *Suppose that $Y = \{Y\}_{t \geq 0}$ is a stochastic process with almost surely continuous paths (as in condition (iv) of Definition 4.1). Then Y is a Wiener process starting from x if and only if its finite-dimensional distributions agree with those of a Wiener process starting from x.*

Proof. The direction ($\Rightarrow$) is immediate. Conversely, suppose the finite-dimensional distributions of Y agree with those of a Wiener process X starting from x. Then $P(Y_0 = x) = P(X_0 = x) = 1$, so Y satisfies (i). If $0 \leq t_1 < t_2 < \cdots < t_m$, with $m \geq 2$, then the distribution of $[Y_{t_1}, ..., Y_{t_m}]^T$ in $\mathbb{R}^m$ agrees with that of $[X_{t_1}, ..., X_{t_m}]^T$. Thus, if $T : \mathbb{R}^m \to \mathbb{R}^m$ is the linear map defined by

$$T([x^1, x^2, ..., x^m]^T) = [x^1, x^2 - x^1, ..., x^n - x^{n-1}]^T, \tag{41}$$

then the distribution of $T([Y_{t_1}, ..., Y_{t_m}]^T)$ in $\mathbb{R}^m$ agrees with that of $T([X_{t_1}, ..., X_{t_m}]^T)$. Since X is a Wiener process, it follows that the $Y_{t_j} - Y_{t_{j-1}}$'s are independent and have $\mathcal{N}(t_j - t_{j-1})$ distribution, so (ii) and (iii) are satisfied. Since the paths of Y are almost surely continuous, it follows that Y is a Wiener process. □

Proposition 4.4. *Suppose $Y = \{Y_t\}_{t\geq 0}$ is a Gaussian process with almost surely continuous paths (as in condition (iv) of Definition 4.1). Then Y is a Wiener process if and only if, for all $s, t \in [0, \infty)$, $EY_s = 0$ and $EY_sY_t = \min\{s, t\}$.*

Proof. Suppose Y is a Wiener process, and $s, t \in [0, \infty)$. Assume without loss of generality that $s < t$. Since $Y_s \in \mathcal{N}(0, s)$, $EY_s = 0$, and independence of increments in condition (ii) gives us

$$E[Y_sY_t] = E[Y_s(Y_t - Y_s)] + E[Y_s^2] = E[Y_s]E[Y_t - Y_s] + E[Y_s^2] = s. \tag{42}$$

Conversely, suppose for all $s, t \in [0, \infty)$, $EY_s = 0$ and $EY_sY_t = \min\{s, t\}$. Note that X is a Gaussian process, and as we have already shown $E[X_sX_t] = \min\{s, t\}$. Suppose $t_0, ..., t_q$ are distinct times in $[0, \infty)$. Then $[Y_{t_0}, ..., Y_{t_q}]^T$ and $[X_{t_0}, ..., X_{t_q}]^T$ are Gaussian random variables in $\mathbb{R}^q$, and, for $1 \leq i, j \leq q$, $E[X_{t_i}X_{t_j}] = E[Y_{t_i}Y_{t_j}] = \min\{t_i, t_j\}$. Hence, $[Y_{t_0}, ..., Y_{t_q}]^T$ and $[X_{t_0}, ..., X_{t_q}]^T$ have equal covariance matrices. So, by Proposition 3.10, X and Y have the same finite-dimensional distributions. The result then follows by Proposition 4.4. □

Brownian motion process possesses a number of interesting symmetries. Some basic results on the symmetries of Brownian are enumerated in the following proposition.

Proposition 4.5. *Let X be a Wiener process starting from 0. Then*

a. (Space translation) For all $h \in \mathbb{R}$, $\{X_t + h\}_{t\geq 0}$ is a Wiener process starting from h.

b. (Time translation) For all $a > 0$, $\{X_{t+a} - X_a\}_{t\geq 0}$ is a Wiener process.

c. (Scaling property) For all $c \neq 0$, $\{cX_{t/c^2}\}_{t\geq 0}$ is a Wiener process.

Proof. (a) follows easily from Definition 4.1: We have $X_0 + h = h$ a.s. because X starts from 0. (ii) and (iii) hold because translation does not change the increments $X_t - X_s$, $s, t \in [0, \infty)$. And (iv) holds because translation is a continuous operation.

With regard to (b) and (c), we make the following observation. If $Z = \{Z_t\}_{t\geq 0}$ is a Gaussian process, then so is $\{Z_{f(t)}\}$, whenever f is an injective function from $[0, \infty)$ into $[0, \infty)$. For, if $t_1, ..., t_q$ are distinct points in $[0, \infty)$, then so are $f(t_1), ..., f(t_q)$, and hence the distribution of $(Z_{f(t_1)}, ..., Z_{f(t_q)})$ is Gaussian, since $\{Z_t\}$ is a Gaussian process. Thus, the processes (b) and (c) are Gaussian.

Processes (b) and (c) are almost surely continuous since X is. Moreover, if $0 \leq s < t$, then $E[X_{s+a} - X_s] = 0 - 0 = 0$ and $E[cX_{t/c^2}] = c0 = 0$. Using Proposition 4.4, we also compute

$$E[(X_{t+a} - X_a)(X_{s+a} - X_a)] = (s + a) - 2a + a = s, \tag{43}$$

and

$$E[(cX_{t/c^2})(cX_{s/c^2}) = c^2 \frac{s}{c^2} = s. \tag{44}$$

Thus by Proposition 4.4, processes (b) and (c) are Wiener processes. □

The definition of multidimensional Wiener process is simply the following.

Definition 4.6 (n-dimensional Weiner process)**.** *Suppose $X = \{X_t\}_{t\geq 0}$ is a Wiener process taking values in $\mathbb{R}^n$. We say that X is an* (n-dimensional) Wiener process *if the components are independent, and each component is a Wiener process in $\mathbb{R}$.*

Note that all of the symmetries enumerated in Proposition 2.3 extend in obvious ways to multidimensional Wiener processes. For example, if $\{X_t\}$ is an n-dimensional Wiener process, then so is $\{cX_{t/c^2}\}$, for any $c \neq 0$, where scaler multiplication is defined componentwise. In fact, the n-dimensional Wiener process is more symmetric than first meets the eye. Fundamental to applications to the Dirichlet problem is the following result.

Proposition 4.7 (Rotational symmetry)**.** *Suppose that X is an n-dimensional Wiener process starting from $x \in \mathbb{R}^n$, and $A \in SO(n)$. Then AX a Wiener process starting at Ax.*

Proof. By replacing X with $X - x$, we may assume that $x = 0$. Then $AX_0 = 0$ a.s. since $X_0 = 0$ a.s. Also, the paths of AX are almost surely continuous, since this is true of X. Hence AX satisfies conditions (i) and (iv) of Definition 4.1. In addition, if $0 \leq t_0 < t_1 < \cdots t_m$, then the increments $X_{t_j} - X_{t_{j-1}}$ are independent. Hence the increments $AX_{t_j} - AX_{t_{j-1}}$ are independent, so AX satisfies condition (ii). Finally, suppose that $0 \leq s < t$. Then, for $1 \leq i \leq n$, the component $(X_t)_i - (X_s)_i \in \mathcal{N}(0, t-s)$. Since the components are independent, $X_t - X_s$ is a mean-zero n-dimensional Gaussian random variable, and its covariance matrix is

$$\mathrm{Cov}(X_t - X_s) = E[(X_t - X_s)(X_t - X_s)^T] = (t - s)I.$$

In addition, $AX_t - AX_s$ is an n-dimensional Gaussian random variable. Since the expected value operator commutes with linear maps and A is orthogonal, the covariance matrix of $AX_t - AX_s$ is

$$\begin{aligned}\mathrm{Cov}(AX_t - AX_s) &= E[(AX_t - AX_s)(AX_t - AX_s)^T] \\ &= AE[(X_t - X_s)(X_t - X_s)^T]A^T \\ &= (t - s)AA^T = (t - s)I.\end{aligned}$$

Since the covariance matrices are equal, Proposition 3.10 implies that $X_t - X_s$ and $AX_t - AX_s$ are equally distributed. This proves that AX satisfies condition (iii) of Definition 4.1, so we are done. □

5 Construction of the Wiener process

We have deduced some basic properties of the Wiener process, but so far it is not obvious that such a process should even exist. To emphasize this fact, suppose we modified Definition 4.1 of the Wiener process, by replacing condition (iii) with

(iii)' If $0 \leq s < t$, then $X_t - X_s \in \mathcal{N}(0, (t-s)^p)$,

for some real number $p \neq 1$. If $0 \leq s < t$, then, on the one hand, by (iii)' $E[(X_t - X_s)^2] = (t-s)^p$. On the other hand, using independence of increments in condition (ii) of Definition 4.1, we have

$$\begin{aligned} E[(X_t - X_s)^2] &= E[X_t^2] + E[X_s^2] - 2E[X_t X_s] \\ &= t^p + s^p - 2(E[X_s(X_t - X_s)] + E[X_s^2]) \\ &= t^p + s^p - 2(E[X_s]E[X_t - X_s] + E[X_s^2]) \\ &= t^p + s^p - 2s^p = t^p - s^p. \end{aligned} \tag{45}$$

This gives us a contradiction, because in general $(t-s)^p \neq t^p - s^p$.

The construction of Brownian motion we present is due to Ciesielski [2]. Note that we only need to construct the one-dimensional Wiener process, since the n-dimensional Wiener process is defined in terms of the one-dimensional process. While our construction is one of the easiest ways to obtain the Wiener process, it is still somewhat technical, so before getting our hands dirty, let us sketch our plan.

(1) Our first task is to define a special collection of functions $\{\varphi_{ij} : i \geq 0, \text{ and } 0 \leq j \leq 2^{i-1} - 1\}$, known as **Haar functions**, in the real Hilbert space $L^2[0,1]$. We will prove that this collection of functions is in fact an orthonormal set in the $L^2[0,1]$.

(2) For each φ_{ij}, define $\psi_{ij}(t) = \int_0^t \varphi_{ij}$, where $t \in [0,1]$, and let $\{Y_{ij}\}$ be an collection of independent $\mathcal{N}(0,1)$ random variables. Define $V_0(t) = Y_{00}\psi_{00}$, and for each $i \geq 1$ define

$$V_i(t) = \sum_{j=1}^{2^{i-1}} Y_{ij}\psi_{ij}(t). \tag{46}$$

Our prototype of for Brownian motion will be the process

$$X_t = \sum_{i=0}^{\infty} V_i(t). \tag{47}$$

The difficult part of our argument will be verifying that X_t converges almost surely to a continuous function in t. Once we have proven this result, we will show that X satisfies conditions (i)-(iv) of the Definition 4.1 of the Wiener process. This task will be greatly simplified by the tools of Hilbert space theory applied to the orthonormal set $\{\varphi_{ij}\}$.

(3) The process X_t, obtained in step (2), satisfies all the requirements of Definition 4.1, except that it is only defined for $t \in [0,1]$. However, using *time inversion* (see Proposition 4.5(d)), we will be able extend X_t to the interval $[0,\infty)$ by defining $X_t = tY_{1/t}$ for $t \geq 1$, where Y_s is an independent copy of X_s, for $s \in [0,1]$.

5.1 Haar functions

While Haar functions are conceptually simple, their definition is awkward to write out, and hence we devote a whole subsection to them. [I hope to include some pictures in this section.] We begin with an indexed collection of subintervals of the unit interval $\mathcal{I} = \{I_{ij} : 0 \leq i < \infty, \text{ and } 0 \leq j \leq 2^i - 1\}$, with elements defined as follows. Let $I_{00} = [0,1)$. Let $I_{10} = [0,1/2)$ and $I_{11} = [1/2,1)$. Let $I_{20} = [0,1/4)$, $I_{21} = [1/4,1/2)$, $I_{22} = [1/2,3/4)$, and $I_{23} = [3/4,1)$. Continuing inductively, we let each $I_{i0} = \frac{1}{2}I_{i-1,0} = \{x/2 : x \in I_{i-1,0}\}$ and we let each $I_{ij} = I_{i0} + j2^{-i} = \{x + j2^{-i} : x \in I_{ij}\}$.

A useful decomposition of each interval I_{ij} is obtained as follows. For each i and j, observe that $I_{ij} = I_{i+1,2j} \cup I_{i+1,2j+1}$. Reasoning inductively, we obtain for each integer $k \geq 0$,

$$I_{ij} = I_{i+k,2^k j} \cup I_{i+k,2^k j+1} \cup \cdots \cup I_{i+k,2^k j+2^k-1}. \tag{48}$$

Note that the intervals making up the union in the above expression are disjoint. Further, the intersection of I_{ij} with any other interval in $\mathcal{I}$ of form $I_{i+k,m}$ not appearing in the above expression is empty. We will refer back to decomposition given by (48) a number of times.

For $0 \leq i < \infty$ and $0 \leq j \leq 2^{i-1} - 1$, we define functions $\varphi_{ij} : [0,1] \to \mathbb{R}$ as follows. Let $\varphi_{00} \equiv 1$, and for $i \geq 1$ define

$$\varphi_{ij}(x) = \begin{cases} 2^{(i-1)/2} & x \in I_{i,2j} \\ -2^{(i-1)/2} & x \in I_{i,2j+1} \\ 0 & \text{otherwise.} \end{cases} \tag{49}$$

Hence the support of each function φ_{ij} is $I_{i,2j} \cup I_{i,2j+1} = I_{i-1,j}$. Functions taking this form are called *Haar functions*. The great utility of Haar functions comes from the following result.

Proposition 5.1. *The collection of Haar functions* $\{\varphi_{ij} : 0 \leq i < \infty, \text{ and } 0 \leq j \leq 2^i - 1\}$ *is a complete orthonormal system for the Hilbert space* $L^2[0,1]$.

Proof. First, we prove that the collection $\{\varphi_{ij}\}$ is orthonormal in $L^2[0,1]$. By construction, each interval $I_{i-1,j} = \text{supp}(\varphi_{ij})$ has width $2^{-(i-1)}$, and we have $||\phi_{ij}||_2^2 = \langle \varphi_{ij}, \varphi_{ij} \rangle = \int 2^{i-1}\chi_{I_{i+1,j}} = 1$. Further, if $(i,j) \neq (p,q)$, we claim that $\langle \varphi_{ij}, \varphi_{pq} \rangle = 0$. To see this, note that this is clearly true if $i = p$ because then $\text{supp}(\varphi_{ij}) = I_{i-1,j}$ and $\text{supp}(\varphi_{pq}) = I_{i-1,q}$ are disjoint. If $i \neq p$, assume without loss of generality that $i < p$, and write $p = i + k$, for some positive integer k. Either $\text{supp}(\varphi_{pq})$ and $\text{supp}(\varphi_{ij})$ are disjoint, in which case clearly $\langle \varphi_{ij}, \varphi_{pq} \rangle = 0$; or $\text{supp}(\varphi_{pq})$ and $\text{supp}(\varphi_{ij})$ are not disjoint – that is, $I_{p-1,q} = I_{i-1+k,q}$ and $I_{i-1,j}$ are not disjoint. Then by the decomposition given by equation (48) $I_{p-1,q} \subset I_{i-1,j}$, and further we observe that either $I_{p-1,q}$ is a subset of $I_{i,2j}$ or $I_{p-1,q}$ is a subset of $I_{i,2j+1}$. Hence, either $\varphi_{ij}\varphi_{pq} = 2^{(i-1)/2}\varphi_{pq}$ or $\varphi_{ij}\varphi_{pq} = -2^{(i-1)/2}\varphi_{pq}$. In both cases, $\langle \varphi_{ij}, \varphi_{pq} \rangle = \int_0^1 \varphi_{ij}\varphi_{pq} = 0$. Thus, we have shown that the collection $\{\varphi_{ij}\}$ is orthonormal.

To prove completeness of the collection $\{\varphi_{ij}\}$, we observe that any half-open interval $J = [a, b)$, with endpoints in the dyadic rationals in $[0, 1]$, may be written as

$$J = \bigcup_{j=m}^{n} I_{ij}, \tag{50}$$

for some positive integers i, m, and n, with $0 \leq m \leq n \leq 2^i - 1$. In addition, note that by definition $\varphi_{00} = \chi_{I_{00}}$, and for $i \geq 1$ and $0 \leq j \leq 2^{i-1} - 1$ we have

$$\chi_{I_{i,2j}} = \frac{1}{2}(\chi_{I_{i-1,j}} + 2^{-(i-1)}\varphi_{i-1,j}), \tag{51}$$

and

$$\chi_{I_{i,2j+1}} = \frac{1}{2}(\chi_{I_{i-1,j}} - 2^{-(i-1)}\varphi_{i-1,j}). \tag{52}$$

These two equations just come from the definition of $\varphi_{i-1,j}$ and the fact that $I_{i-1,j} = I_{i,2j} \cup I_{i,2j+1}$. Induction on i then gives us the following result: Every characteristic function $\chi_{I_{ij}}$ may be written as a linear combination of Haar functions. Since $\chi_J = \sum_{j=n}^{m} \chi_{I_{ij}}$, it follows that χ_J may be written as a linear combination of Haar functions. Since the collection of linear combinations of functions of form χ_J, where as above $J \subset [0, 1]$ a half-open interval with end-points in the dyadic rationals, is dense in $L^2[0, 1]$, we conclude that linear combinations of Haar functions are also dense in $L^2[0, 1]$. This shows that the collection $\{\varphi_{ij}\}$ is complete. □

5.2 The Main Construction

Start with the collection of Haar functions $\{\varphi_{ij} : 0 \leq i < \infty, \text{ and } 0 \leq j \leq 2^{i-1}\}$, and, for each φ_{ij}, define the function ψ_{ij} on $[0, 1]$ by

$$\psi_{ij}(t) = \int_0^t \varphi_{ij}(s)ds. \tag{53}$$

We record a useful set of bounds for the functions ψ_{ij}.

Lemma 5.2. *For all i, j, $0 \leq \psi_{ij} \leq 2^{-(i+1)/2}$ and $supp(\varphi_{ij}) = I_{i-1,j}$.*

Proof. This is simply a matter of writing down an explicit expression for ψ_{ij}. By definition, $\varphi_{ij}(x) = 2^{(i-1)/2}$ on $I_{i,2j} = [(2j)2^{-i}, (2j + 1)2^{-i}]$, $\varphi_{ij}(x) = -2^{(i-1)/2}$ on $I_{i,2j+1} = [(2j + 1)2^{-i}, (2j + 2)2^{-i}]$, and $\varphi_{ij}(x) = 0$ otherwise. Hence,

$$\psi_{ij}(t) = \begin{cases} 2^{(i-1)/2}(t - (2j)2^{-i}) & t \in [(2j)2^{-i}, (2j + 1)2^{-i}] \\ 2^{(i-1)/2} - 2^{(i-1)/2}(t - (2j + 1)2^{-i}) & t \in [(2j + 1)2^{-i}, (2j + 2)2^{-i}] \\ 0 & \text{otherwise.} \end{cases} \tag{54}$$

Therefore, $\psi_{ij} \geq 0$, and ψ_{ij} attains its maximum value at $t = (2j+1)2^{-i}$, and this maximum value is $\psi_{ij}((2j + 1)2^{-i}) = 2^{(i-1)/2}2^{-i} = 2^{-(i+1)/2}$. Note that (54) also shows $\text{supp}(\varphi_{ij}) = [(2j)2^{-i}, (2j + 2)2^{-i}] = I_{i-1,j}$. □

Next, let $\{Y_{ij}\}$ be a collection of independent $\mathcal{N}(0,1)$ random variables. (By Proposition 2.9 such a collection exists.) As described in our outline, we define $V_0(t) = Y_{00}\psi_{00}$, and

$$V_i(t) = \sum_{j=1}^{2^{i-1}} Y_{ij}\psi_{ij}(t), \tag{55}$$

for each $i \geq 1$. Our prototype for the Wiener process is

$$X_t = \sum_{i=0}^{\infty} V_i(t). \tag{56}$$

Our next objective is to show that this series indeed converges almost surely to a continuous function in t.

Lemma 5.3. *With probability 1, $\sum_{i=0}^{\infty} V_i(t)$ converges uniformly to a continuous function on $[0,1]$.*

Proof. The main ingredient is the Borel-Cantelli Lemma (Proposition 2.3). Consider the sequence of sets $\{A_i\}_{i=1}^{\infty}$ defined by $A_i = (|V_i(t)| > i^{-2}$ for some $t \in [0,1])$. It is not immediately clear that the sets A_i are measurable. To see that in fact they are, observe that: $|V_i(t)| > i^{-2}$ for some $t \in [0,1]$ if and only if $|V_i(t)| > i^{-2}$ for some $t \in [0,1] \cap \mathbb{Q}$, by continuity of $V_i(t)$ for each fixed ω. Therefore, for each $i \geq 1$, $A_i = \bigcup_{t \in [0,1]\cap\mathbb{Q}} \{\omega : |V_i(t)| > i^{-2}\}$, which is a countable union of measurable sets and hence measurable.

For $i \geq 1$, notice that if $|V_i(t)| > i^{-2}$, then $|Y_{ij}|\psi_{ij}(t) > i^{-2}$ for some $1 \leq j \leq 2^{i-1}$, because $|V_i(t)| \leq \sum_j |Y_{ij}|\psi_{ij}(t)$ and the ψ_{ij}'s have disjoint support by Lemma 5.2. Then $|Y_{ij}|2^{-(i+1)/2} > i^{-2}$ for some $1 \leq j \leq 2^{i-1}$ by the bound obtained in Lemma 5.2. It follows that, for each $i \geq 1$,

$$\begin{aligned} P(A_i) &\leq P(|Y_{ij}|2^{-(i+1)/2} > i^{-2} \text{ for some } 1 \leq j \leq 2^{i-1}) \\ &\leq 2^{i-1} P(|Z| \geq 2^{(i+1)/2} i^{-2}) \\ &\leq 2^{i-1} \exp(-2^i i^{-4}). \end{aligned} \tag{57}$$

Here $Z =^d Y_{ij}$ is an $\mathcal{N}(0,1)$ random variable, and the last line of (57) is obtained from Proposition 3.7. It is easy to see that $\sum 2^{i-1}\exp(-2^i i^{-4})$ converges by, for example, the root test.

Thus $\sum_{i=1}^{\infty} P(A_i) < \infty$, and by the Borel-Cantelli Lemma, $P(\cap_{k=0}^{\infty} \cup_{i=k}^{\infty} A_i) = 0$. In other words, with probability 1, there are only finitely many $i \geq 1$ such that for some $t \in [0,1]$, $|V_i(t)| \geq i^{-2}$. Hence, with probability 1, $\sum_{i=0}^{\infty} V_i(t)$ converges uniformly on $[0,1]$. Since each of the $V_i(t)$'s is continuous in t, uniform convergence implies that $\sum_{i=0}^{\infty} V_i(t)$ is continuous in t. This completes the proof. □

Let $X_t = \sum_{i=0}^{\infty} V_i(t)$. In the next two lemmas, we first show that X_t has finite variance (i.e. $X_t \in L^2(P)$ for each fixed t), and then we obtain more precise result.

Lemma 5.4. *For all $t \in [0,1]$, $Var(X_t^2) < \infty$.*

Proof. To begin, observe that for any integer $N > 0$ and $t \in [0,1]$,

$$\left(\sum_{i=0}^{N} V_i(t)\right)^2 = \sum_{i,k=0}^{N} V_i(t)V_k(t) = \sum_{i,k=0}^{N} \sum_{\substack{0\le j\le 2^{i-1}-1,\\ 0\le l\le 2^{k-1}-1}} Y_{ik}Y_{jl}\psi_{ik}(t)\psi_{jl}(t). \tag{58}$$

Note that $E[Y_{ij}Y_{kl}] = 1$ if $(i,j) = (k,l)$, by standard normal distribution, and otherwise $E[Y_{ij}Y_{kl}] = 0$, by independence. Thus, taking expected values yields

$$E\left(\sum_{i=0}^{N} V_i(t)\right)^2 = \sum_{i=0}^{N}\sum_{j=0}^{2^{i-1}-1} (\psi_{ij}(t))^2 \le \sum_{i=0}^{N} 2^{-(i+1)}, \tag{59}$$

here using Lemma 5.2. Therefore, by Fatou's Lemma, $E[X_t^2] \le \liminf_{N\to\infty} E[(\sum_{i=0}^{N} V_i(t))^2] \le \sum 2^{-(i+1)} < \infty$, as required. □

Lemma 5.5. *For all $s,t \in [0,1]$, $E[X_sX_t] = min(s,t)$.*

This is the one step of our proof that requires tools from Hilbert space theory. The space $L^2[0,1]$ may be equipped with the real Hilbert space inner product

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt, \qquad \text{where } f,g \in L^2[0,1]. \tag{60}$$

Note that, if $\{e_j\}_{j=1}^{\infty}$ is a complete orthonormal system in $L^2[0,1]$, then for any function $f \in L^2[0,1]$

$$\sum_{j=1}^{N}\langle f, e_j\rangle e_j \to f \quad \text{in } L^2 \text{ as } N\to\infty. \tag{61}$$

We will apply this result to the collection of Haar functions $\{\varphi_{ij}\}$, which by Proposition 5.1 is a complete orthonormal system for $L^2[0,1]$.

Proof of Lemma 5.5. Let $s,t \in [0,1]$. First, we claim

$$E[X_sX_t] = \sum_{i,j}\langle 1_{[0,s]}, \varphi_{ij}\rangle\langle 1_{[0,t]}, \varphi_{ij}\rangle. \tag{62}$$

To justify this equality, observe that for any integer $N > 0$

$$\begin{aligned} E\left[\left(\sum_{i=0}^{N} V_i(s)\right)\left(\sum_{k=0}^{N} V_k(t)\right)\right] &= \sum_{i,k=0}^{N} E[V_i(s)V_k(t)] \\ = \sum_{i,k=0}^{N}\sum_{\substack{0\le j\le 2^{i-1}-1,\\ 0\le l\le 2^{k-1}-1}} E[Y_{ij}Y_{kl}\psi_{ij}(s)\psi_{kl}(t)] &= \sum_{i=0}^{N}\sum_{j=0}^{2^{i-1}-1} \psi_{ij}(s)\psi_{ij}(t), \end{aligned} \tag{63}$$

where the last equality follows from the fact that the Y_{ij}'s are independent standard normal random variables, so $E[Y_{ij}Y_{kl}] = 1$ if $(i,j) = (k,l)$ and equals 0 otherwise. For $i \ge 0$

and $0 \leq j \leq 2^{i-1} - 1$, observe that by definition $\psi_{ij}(t) = \langle 1_{[0,t]}, \varphi_{ij} \rangle$. Thus equality (62) will be proved if we can show that $\lim_{N\to\infty} E[(\sum_{i=0}^{N} V_i(s))(\sum_{k=0}^{N} V_k(t))] = E[X_s X_t]$. Note that $\lim_{N\to\infty}(\sum_{i=0}^{N} V_i(s))(\sum_{k=0}^{N} V_k(t)) = X_s X_t$. Thus, if we take $2|X_s X_t|$ as our dominating function (and we can because $X_s, X_t \in L^2(P)$ by Lemma 5.4), this result is implied by dominated convergence.

Next, using bilinearity of the inner product and applying (61) to equation (62), we compute

$$\begin{aligned} E[X_s X_t] &= \left\langle 1_{[0,t]}, \sum_{i,j} \langle 1_{[0,s]}, \varphi_{ij} \rangle \varphi_{ij} \right\rangle \\ &= \langle 1_{[0,t]}, 1_{[0,s]} \rangle = \min(s,t). \end{aligned} \tag{64}$$

This completes the proof. □

For $t \geq 1$, define $Y_t = tX_{1/t}$. Let $X' = \{X_t\}_{t\in[0,1]}$ be an independent copy with a.s. continuous paths of the process X, as in Proposition 2.19. (In other words, X' is a process with a.s. continuous paths such that, for all $s, t \geq 0$, $X'_t =^d X_t$, and X'_s and X_t are independent.) For $t \in [0, \infty)$, define

$$Z_t = \begin{cases} X'_t & t \in [0,1] \\ X'_1 + Y_t - Y_1 & t \in (0,\infty). \end{cases} \tag{65}$$

Our final step is to show

Theorem 5.6. *$Z = \{Z_t\}_{t\geq 0}$ is a Wiener process.*

Proof. By Proposition 4.4, it is enough to prove that (i) the paths of Z are almost surely continuous, (ii) Z is a Gaussian process, and (iii) for all $s, t \in [0, \infty)$, $E[Z_s] = 0$ and $E[Z_s Z_t] = \min(s, t)$.

(i) By Lemma 5.3 and the definitions of X'_t and Y_t, the paths of Z are, with probability 1, continuous at all $t \neq 1$. To see that the paths of Z are almost surely continuous at 1, observe that $\lim_{t\to 1^-} Z_t = \lim_{t\to 1^-} X'_t = X'_1 = Z_1$ a.s., by a.s. continuity of the paths of X'. Also, $\lim_{t\to 1^+} Z_t = \lim_{t\to 1^+}(X'_1 + Y_t - Y_1) = X'_1 = Z_1$ a.s., by a.s. continuity of the paths of Y. Hence Z almost surely continuous at $t = 1$. This shows that the paths of Z are almost surely continuous.

(ii) For each fixed t, Z_t is a linear combination of Gaussian random variables and hence is Gaussian by Proposition 3.6. Given distinct times, $t_0, ..., t_m \in [0, \infty)$, the components of the random vector $[Z_{t_1}, ..., Z_{t_m}]^T$ are Gaussian, and hence $[Z_{t_1}, ..., Z_{t_m}]^T$ is multivariate Gaussian. This shows that the finite-dimensional distributions of Z are Gaussian, so Z is a Gaussian process.

(iii) Suppose $s, t \in [0, \infty)$. We may assume without loss of generality that $s < t$. If $s, t \leq 1$, the result just follows by Lemma 5.5. There are two cases left to check.

Case 1: $s \leq 1$ and $t > 1$. Then $Z_s Z_t = X'_s(X'_1 + Y_t - Y_1) = X'_s X'_1 + tX'_s X_{1/t} - X_1$. Thus, using independence and Lemma 5.5, we have

$$E[Z_s Z_t] = E[X'_s X'_1] + tE[X'_s]E[X_{1/t}] - E[X_1] = E[X'_s X'_1] = s. \tag{66}$$

Case 2: $1 < s < t$. Then $Z_s Z_t = (X_1' + sX_{1/s} - X_1)(X_1' + tX_{1/t} - X_1) = (X_1')^2 + tX_1'X_{1/t} - X_1'X_1 + sX_{1/s}X_1' + stX_{1/s}X_{1/t} - sX_{1/s}X_1 - X_1X_1' - tX_1X_{1/t} + X_1^2$. By independence and Lemma 5.5, $E[X_1X_1'] = E[X_1'X_{1/t}] = E[X_{1/s}X_1'] = 0$. Hence,

$$\begin{aligned} E[Z_sZ_t] &= E[(X_1')^2] + stE[X_{1/s}X_{1/t}] - sE[X_{1/s}X_1] - tE[X_1X_{1/t}] + E[X_1^2] \\ &= 1 + st\frac{1}{t} - s\frac{1}{s} - t\frac{1}{t} + 1 = s. \end{aligned} \tag{67}$$

Hence, in all cases $E[Z_sZ_t] = \min(s,t) = s$, and this completes the proof. □

6 Markov Property

6.1 Conditional Expectation

The Markov property is essentially a statement about the *conditional expectation* of Brownian motion with respect to certain σ-algebras. We first therefore need to define conditional expectation and describe its basic properties.

Definition 6.1 (Conditional expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X an $\mathcal{F}$-measurable random variable, and $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. If $X \in L^1(P)$, the* conditional expectation of X given $\mathcal{F}$*, denoted by $E[X|\mathcal{G}]$, is a $\mathcal{G}$-measurable random variable $Y \in L^1(P)$, which satisfies*

$$\int_G X dP = \int_G Y dP \qquad \text{for all } G \in \mathcal{G}. \tag{68}$$

This definition is of little value if we do not prove.

Proposition 6.2. *If $X \in L^1(P)$ and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-field, then (i) the conditional expectation of X given $\mathcal{G}$ exists, and (ii) it is unique up to almost sure equivalence.*

Remark 6.3. Proposition 6.2 expresses the fact that conditional expectation is defined up to almost sure equivalence. Hence, technically any equality involving $E[X|\mathcal{G}]$ will only hold up to almost sure equivalence. However, the standard convention is to write $Y = E[X|\mathcal{G}]$ with almost sure equivalence understood, rather than to write $Y = E[X|\mathcal{G}]$ a.s., even though the latter is perhaps more correct.

We can establish uniqueness (ii) immediately.

Proof of Proposition 6.2(ii). Suppose Y, Y' are two $\mathcal{G}$-measurable functions in $L^1(P)$ satisfying (68). Let $\epsilon > 0$, and let $A = \{\omega : Y - Y' \geq \epsilon\}$. Then $H \in \mathcal{G}$, and

$$0 = \int_A (X - X) dP = \int_A (Y - Y') dP \geq \epsilon P(A). \tag{69}$$

Hence $P(A) = 0$. Since ϵ is arbitrary, it follows $Y \geq Y'$ a.s. Reversing the roles of Y and Y', we get $Y' \geq Y$ a.s. Hence $Y = Y'$ a.s. □

The existence of conditional expectation is a somewhat deeper matter, which relies on the *Radon-Nikodym theorem* from analysis. If μ and ν are two σ-finite measures on a σ-algebra $\mathcal{F}$, we say that μ is **absolutely continuous** with respect to ν (write $\mu \ll \nu$) if, for any $N \in \mathcal{F}$, $\nu(N) = 0$ implies $\mu(N) = 0$. The special case of Radon-Nikodym theorem which we need is the following result.

Theorem 6.4 (Radon-Nikodym theorem)**.** *Suppose μ and ν are σ-finite measures on $\mathcal{F}$ such that $\mu \ll \nu$. Then there exists a real-valued $\mathcal{F}$-measurable function f such that*

$$\int f d\nu = \mu(F) \qquad \text{for all } F \in \mathcal{F}. \tag{70}$$

For proof of this result, we refer the reader to Durrett [3, Theorem A.4.6], or Rudin [7, Theorem 6.10], or Folland [4, Theorem 3.8]. Using this result, the proof of first part of Proposition 6.2 is fairly simple.

Proof of Proposition 6.2(i). First, let us assume that $X \geq 0$. We define the set function μ on $\mathcal{G}$ by

$$\mu(G) = \int_G X dP \qquad (G \in \mathcal{G}). \tag{71}$$

One easily verifies that μ is a measure on $\mathcal{G}$: $\mu(\emptyset) = 0$ by definition of the integral, and countable additivity follows from dominated convergence. Also, by basic properties of integrals, if $P(G) = 0$ for some $G \in \mathcal{G}$, then $\mu(G) = 0$. Hence μ is abolutely continuous with respect to the measure $P|_{\mathcal{G}}$. Hence, by the Radon-Nikodym theorem, there exists a $\mathcal{G}$-measurable random variable Y such that

$$\int_G X dP = \mu(G) = \int_G Y dP \qquad \text{for all } G \in \mathcal{G}, \tag{72}$$

Hence $Y = E[X|\mathcal{G}]$.

To obtain the result in general, write $X = X_+ - X_-$, where $X_+ = X$ if $X \geq 0$ and $X_+ = 0$ otherwise, and $X_- = |X| - X_+$. Let $Y_+ = E[X_+|\mathcal{G}]$ and $Y_- = E[X_-|\mathcal{G}]$. Let $Y = Y_+ - Y_-$. Then, for any $G \in \mathcal{G}$,

$$\int_G X dP = \int_G X_+ dP - \int_G X_- dP = \int_G Y_+ dP - \int_G Y_- dP = \int_G Y dP. \tag{73}$$

Hence $Y = E[X|\mathcal{G}]$, which is what we needed to show. □

Next we prove some basic properties of conditional expectation. In the statement of the next proposition, we will use the notation $X \in L^1(\mathcal{G})$, where $\mathcal{G}$ is any sub-σ-field of $\mathcal{F}$, to indicate that X is an integrable, $\mathcal{G}$-measurable random variable.

Proposition 6.5. *Suppose $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. Then*

a. *Conditional expectation is linear, i.e. $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$ for all scalers a, b and all $X, Y \in L^1(\mathcal{F})$.*

b. *Suppose $X \in L^1(\mathcal{G})$ and $Y \in L^1(\mathcal{F})$. Then $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$.*

c. *If* $X \in L^1(\mathcal{G})$*, then* $E[X|\mathcal{G}] = X$.

d. *If* $X \in L^1(\mathcal{F})$ *is independent of* $\mathcal{G}$*, then* $E[X|\mathcal{G}] = E[X]$.

e. *Suppose* $\mathcal{H}$ *is a sub-*σ*-algebra of* $\mathcal{G}$*, and* $X \in L^1(\mathcal{F})$*. Then* $E[E[X|\mathcal{G}]|\mathcal{H}] = E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{H}]$.

The fact (e) is sometimes stated informally as "the smallest σ-algebra always wins."

Proof. (a) Let $Z_1 = E[X|\mathcal{G}]$ and $Z_2 = E[Y|\mathcal{G}]$. Then, for any $G \in \mathcal{G}$,

$$\begin{aligned} \int_G (aX + bY)dP &= a\int_G XdP + b\int_G YdP \\ &= a\int_G Z_1 dP + b\int_G Z_2 dP = \int_G (aZ_1 + bZ_2)dP, \end{aligned} \tag{74}$$

as required.

(b) First, suppose $X = 1_A$ for some $A \in \mathcal{G}$, and let $Z = E[XY|\mathcal{G}]$ and $Z' = E[Y|\mathcal{G}]$. Then, for any $G \in \mathcal{G}$

$$\int_G ZdP = \int_{G\cap A} YdP = \int_{G\cap A} Z'dP = \int_G XZ'dP. \tag{75}$$

Since XZ' is $\mathcal{G}$-measurable, this shows that $Z = XZ'$. Hence, by linearity of conditional expectation, $E[XY] = XE[Y]$ whenever X is a $\mathcal{G}$-measurable simple function. Now take X to be an arbitrary random variable in $L^1(\mathcal{G})$. By a basic fact from integration theory, there exists a sequence of $\mathcal{G}$-measurable simple functions X_j such that $X_j \to X$ as $j \to \infty$. Then by dominated convergence, if we let Z and Z' be taken as above, then

$$\int_G ZdP = \int_G XYdP = \lim_{j\to\infty}\int_G X_j YdP = \lim_{j\to\infty}\int_G X_j Z'dP = \int_G XZ'dP, \tag{76}$$

as required.

(c) This is a special case of (b), with $Y \equiv 1$.

(d) For any $G \in \mathcal{G}$, X and 1_G are independent random variables, and hence by Proposition 2.6

$$\int_G XdP = E[X1_G] = E[X]E[1_G] = \int_G E[X]dP, \tag{77}$$

as required.

(e) The fact that $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ follows from the fact that any $\mathcal{G}$-measurable random variable is $\mathcal{H}$-measurable. As for the second inequality, let $H \in \mathcal{H}$, and observe that since $H \in \mathcal{G}$

$$\int_H E[E[X|\mathcal{H}]|\mathcal{G}]dP = \int_H E[X|\mathcal{H}]dP = \int_H XdP, \tag{78}$$

This shows that $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{H}]$. □

6.2 Stopping Times

We will need to know about stopping times in order to describe the strong Markov property. To define stopping times, first we must introduce filtrations, which are also be needed to define the Markov properties.

Definition 6.6 (Filtration)**.** *Let $(\Omega, \mathcal{F}, P)$ is a probability space, a* filtration *is a collection of σ-algebras $\{\mathcal{F}_t\}_{t\geq 0}$ such that*

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F} \qquad \text{whenever } s \leq t. \tag{79}$$

In addition, we say that a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is right-continuous *if, for all $t \geq 0$, $\mathcal{F}_t = \cap_{s\geq t}\mathcal{F}_s$.*

Example 6.7 (Filtration of a stochastic process)**.** A common way in which filtrations arise is as follows. If $Y = \{Y_t\}_{t\geq 0}$ is a stochastic process, one defines $\mathcal{F}_t$ to be the σ-algebra generated by the collection of random variables $\{Y_s : 0 \leq s \leq t\}$ (i.e. $\mathcal{F}_t$ is the smallest σ-algebra for which each Y_s, for $s \leq t$, is measurable). The collection $\{\mathcal{F}_t\}_{t\geq 0}$ clearly satisfies the above definition. Intuitively $\mathcal{F}_t$ represents the package of "potential information" we may extract from X by the time t.

We are now ready to define stopping times. Let $\{\mathcal{F}_t\}_{t\geq 0}$, be a fixed filtration of σ-algebras.

Definition 6.8 (Stopping time)**.** *A random variable $T : \Omega \to [0, \infty)$ is said to be a* stopping time *if, for all $t \geq 0$, $\{\omega : T(\omega) < t\} \in \mathcal{F}_t$.*

Next we prove some elementary properties of stopping times.

Proposition 6.9.

a. T is a stopping time if and only if $\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

b. Any fixed $a \geq 0$ is a stopping time.

c. If T and T' are stopping times, then $T_M = \max(T, T')$ and $T_m = \min(T, T')$ are stopping times.

d. If $\{T_n\}$ is a sequence of stopping times, then $\overline{T} = \sup_n T_n$ is a stopping time, and $\underline{T} = \inf_n T_n$.

e. If S and T are stopping times, then so is $S + T$.

Proof. (a) Observe that $T \leq t$ if and only if $T < q$ for all rational numbers $q > t$. Hence if T is a stopping time, then $\{\omega : T(\omega) \leq t\} = \cap_{q\in(t,\infty)}\{\omega : T(\omega) < q\} \in \mathcal{F}_t$. Conversely, suppose $\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Observe that $T < t$ if and only if $T \leq q$ for some rational $q < t$. Hence $\{\omega : T(\omega) < t\} = \cup_{q\in[0,t)}\{\omega : T(\omega) \leq q\} \in \mathcal{F}_t$. Thus T is a stopping time.

(b) For any $t \geq 0$, $\{\omega : a < t\}$ is equal to $\emptyset$ or Ω, which are both contained in $\mathcal{F}_t$.

(c) Observe that $\{\omega : T_M(\omega) < t\} = \{\omega : T(\omega) < t\} \cap \{\omega : T'(\omega) < t\} \in \mathcal{F}_t$, and $\{\omega : T_m(\omega) < t\} = \{\omega : T(\omega) < t\} \cup \{\omega : T'(\omega) < t\} \in \mathcal{F}_t$

(d) Observe that $\overline{T} \leq t$ if and only if $T_n \leq t$ for all n. Hence $\overline{T} = \cap_n\{\omega : T_n(\omega) \leq t\} \in \mathcal{F}_t$ by (a). Also, $\underline{T} < t$ if and only if $T_n < t$ for some n. Hence $\{\omega : \underline{T} < t\} = \cup_n\{\omega : T_n(\omega) < t\} \in \mathcal{F}_t$.

(e) Observe that

$$\{\omega : S(\omega) + T(\omega) < t\} = \bigcup_{q \in \mathbb{Q} \cap [0,\infty)} \{\omega : q + T(\omega) < t\} \cap \{S(\omega) < q\} \tag{80}$$

Note that $\{\omega : q + T(\omega) < t\} = \{\omega : T(\omega) < t - q\} \in \mathcal{F}_t$ (if $q > t$, then this set is the empty set). Hence $\{\omega : S(\omega) + T(\omega) < t\} \in \mathcal{F}_t$. □

6.3 The Wiener Process as a random variable taking values in $C([0,\infty), \mathbb{R}^n)$

Let us now introduce another point of view from which to consider the Wiener process. Suppose $X = \{X_t\}_{t \geq 0}$ is a Wiener process in $\mathbb{R}^n$ starting from a point x. While before we had avoided assigning any definite meaning to the measure space $(\Omega, \mathcal{F}, P)$ on which X is defined, we will see that a rather nice situation arises when we identify Ω with $C = C([0,\infty), \mathbb{R}^n)$, the set of continuous functions from $[0,\infty)$ into $\mathbb{R}^n$. This identification is accomplished as follows:

For each $t \geq 0$, let $\pi_t : C \to \mathbb{R}^n$ be the usual projection map (i.e. $\pi_t(f) = f(t)$ for all $f \in C$). Let $\mathcal{C}$ be the σ-algebra of subsets of C generated by the projection maps π_t. If X is a Wiener process in $\mathbb{R}^n$ starting from x, define a set function P^x on $\mathcal{C}$ by

$$P^x(A) = P(X \in A) \qquad (A \in \mathcal{C}). \tag{81}$$

To show that the set $\{\omega : X(\omega) \in A\} = X^{-1}A \in \mathcal{F}$, it is enough to prove $X^{-1}A \in \mathcal{F}$ when A is in a generating set for $\mathcal{C}$. The σ-algebra $\mathcal{C}$ is generated by sets of form $\pi_t^{-1}B$, where $t \geq 0$ and $B \in \mathcal{B}^n$. Hence $X^{-1}(\pi_t^{-1}B) = (\pi_t \circ X)^{-1}B = X_t^{-1}B \in \mathcal{F}$, because X_t is a measurable random variable from Ω into $\mathbb{R}^n$. Thus the definition above for P^x makes sense.

The fact that P^x is a measure is easy to see, because first $P^x(\emptyset) = P(\emptyset) = 0$. Second, if $\{A_j\}_{j=1}^n$ is a countable disjoint collection of sets in $\mathcal{C}$, then $P^x(\cup A_j) = P(\cup\{\omega : X \in A_j\}) = \sum_{j=1}^\infty P(X \in A_j)$, here using countable additivity of P and the fact that the sets $\{\omega : X \in A_j\}$ are disjoint.

For each $t \in [0,\infty)$, let $Y_t = \pi_t$ be the projection map on C defined above.

Proposition 6.10. *The process $Y = \{Y_t\}_{t \geq 0}$, defined on the probability space $(C, \mathcal{C}, P^x)$, is a Wiener process in $\mathbb{R}^n$ starting from x.*

Remark 6.11. We can regard Y as a random variable from C into the product space M^T, where $M = \mathbb{R}^n$ and $T = [0,\infty)$, and M^T is equipped with the product σ-algebra $\mathcal{M}^T$, where $\mathcal{M} = \mathcal{B}^n$ is the collection of Borel sets (see §2.5). Indeed, the σ-algebra $\mathcal{M}^T$ is generated by sets of form $\pi_t^{-1}B$ where $B \in \mathcal{M}$ and $t \geq 0$. Hence $\{f \in C : Y(f) \in \pi_t^{-1}B\} = \{f \in C : f(t) \in B\} = \pi_t^{-1}B \in \mathcal{M}^T$, and it follows that Y is measurable. In particular, it is clear that Y satisfies the official Definition 2.13 of a stochastic process.

Proof. The paths of Y are just the elements of C and hence continuous. To see that the finite-dimensional distributions of Y agree with those of X, suppose $\{t_1, ..., t_m\}$ are distinct

elements of $[0,\infty)$, and suppose $B_1, ..., B_m$ are Borel subsets of $\mathbb{R}^n$. Then

$$\begin{aligned} P^x(Y_{t_1} \in B_1, ..., Y_{t_m} \in B_m) &= P^x(\cap_{j=1}^m Y_{t_j}^{-1} B_j) = P(X \in \cap_{j=1}^m Y_{t_j}^{-1} B_j) \\ &= P(Y_{t_j}(X(\omega)) \in B_j \text{ for } j = 1, ..., m) \\ &= P(X_{t_0} \in B_1, ..., X_{t_m} \in B_m). \end{aligned} \tag{82}$$

Thus the finite dimensional distributions of X and Y agree, and the result follows by Proposition 4.3. □

The content of Proposition 6.10 is that to understand the Wiener process, the probability space $(\Omega, \mathcal{F}, P)$, on which the Wiener process X starting from x is defined, may be replaced by a probability space $(C, \mathcal{C}, P^x)$, on which the Wiener process takes an especially simple form as a collection of evaluation maps. The features of the original process X could be quite complicated depending on the measure P, with respect to which it was constructed. Within the new framework, arbitrary features arising from P are hidden away.

This new point of view is so convenient that we will work with it exclusively for the rest of this paper. Rather than breaking with our previous notation for the Wiener process, and using Y, C, and $\mathcal{C}$ from now on, we establish the following

Convention. From now on, we will use Ω to denote the space $C = C([0,\infty), \mathbb{R}^n)$ defined above, and we will use $\mathcal{F}$ to denote the σ-algebra $\mathcal{C}$. The elements $\omega \in \Omega$ are thus continuous functions from $[0,\infty)$ into $\mathbb{R}^n$. A Wiener process starting from $x \in \mathbb{R}^n$ will be understood to be the pair (X, P^x), where P^x is the measure defined above and $X = \{X_t\}_{t \geq 0}$ where each X_t is the evaluation map $\omega \mapsto \omega(t)$ $(\omega \in \Omega)$.

6.4 The Weak Markov Property

Let X be a Wiener process in $\mathbb{R}^n$. We associate with X two filtrations. First, for $t \geq 0$, we define

$$\mathcal{F}_t^0 = \sigma(X_s : s \leq t). \tag{83}$$

that is, the σ-algebra generated by the collection of random variables X_s such that $s \leq t$. Second, for $t \geq 0$, we define the σ-algebras

$$\mathcal{F}_t = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}. \tag{84}$$

Clearly the collections of σ-algebras $\{\mathcal{F}_t^0\}_{t \geq 0}$ and $\{\mathcal{F}_t\}_{t \geq 0}$ are both filtrations, and the latter is a right-continuous filtration. To paraphrase Durrett, the σ-algebra $\mathcal{F}_t^0$ represents the information about X we will know by time t, and $\mathcal{F}_t$ represents the information about X we will know by time t if we are allowed "an infinitesimal peak into the future" [3, p. 307].

The weak Markov property is the following result.

Theorem 6.12 (Weak Markov property)**.** *Suppose f is a bounded, Borel measurable function, and define a process $Y = f(X)$. Then*

$$E^x[Y \circ \tau_s | \mathcal{F}_s] = E^{X_s} Y \qquad P^x\text{-a.s.} \tag{85}$$

To explain the notation, $E^{X_s}Y$ is the random variable $\varphi(X_s)$, where $\varphi(y) = E^y Y$. To prove the theorem, we first prove the result for when f is a complex exponential function. We then extend this result to a larger class of functions using the Fourier inversion theorem. Finally, we obtain the general result using a simple dominated convergence argument.

Lemma 6.13. *If $f(x) = e^{-2\pi i\xi\cdot x}$ for some $\xi \in \mathbb{R}^n$, then the conclusion of Theorem 6.12 holds true.*

Proof. First, we show the result holds if we replace the σ-algebra $\mathcal{F}_s$ with $\mathcal{F}_s^0$. Indeed,

$$\begin{aligned} E^x[Y\circ\tau_s|\mathcal{F}_s^0] &= E^x[e^{-2\pi i\xi\cdot X_{t+s}}|\mathcal{F}_s^0] \\ &= E^x[e^{-2\pi i\xi\cdot X_{t+s}-X_s}e^{-2\pi i\xi\cdot X_s}|\mathcal{F}_s^0] \\ &= E[e^{-2\pi i\xi\cdot(X_{t+s}-X_s)}]e^{-2\pi i\xi\cdot X_s} \\ &= e^{-\pi|\xi|^2 t}e^{-2\pi i\xi\cdot X_s}. \end{aligned} \tag{86}$$

The second to last line above follows from independence of increments together with Proposition 6.5(b) and (d), and the last line follows from Corollary 3.14. On the other hand, by the same corollary, for any $y \in \mathbb{R}^n$,

$$E^y Y = E^y e^{-2\pi i\xi\cdot X_t} = e^{-\pi|\xi|^2 t}e^{-2\pi i\xi\cdot y}. \tag{87}$$

Thus, setting $y = X_s$ and comparing with (86), we see that $E^x[Y\circ\tau_s|\mathcal{F}_s^0] = E^{X_s}Y$.

To obtain the same result for the σ-algebra $\mathcal{F}_s$, we must show that, for any $F \in \mathcal{F}_s$,

$$\int_F Y\circ\tau_s dP^x = \int_F E^{X_s}Y. \tag{88}$$

If $F \in \mathcal{F}_s$, then for all $\epsilon > 0$, $F \in \mathcal{F}_{s+\epsilon}^0$. Hence, by the result we just established, for all $\epsilon > 0$,

$$\int_F Y\circ\tau_{s+\epsilon} dP^x = \int_F E^{X_{s+\epsilon}}[Y]dP^x. \tag{89}$$

Note that $Y\circ\tau_{s+\epsilon} = e^{-2\pi i\xi\cdot X_{s+t+\epsilon}} \to e^{-2\pi i\xi\cdot X_{s+t}}$ pointwise as $\epsilon \to 0$, by continuity. Hence, by dominated convergence, the left side of equation (89) approaches $\int_F Y\circ\tau_s dP^x$ as $\epsilon \to 0$. On the other hand, by equation (87), $E^{X_{s+\epsilon}}[Y] = e^{-\pi|\xi|^2 t}e^{-2\pi i\xi\cdot X_{s+\epsilon}} \to e^{-\pi|\xi|^2 t}e^{-2\pi i\xi\cdot X_s} = E^{X_{s+\epsilon}}[Y]$ pointwise as $\epsilon \to 0$. Hence, by dominated convergence, the right hand side of equation (89) approaches $\int_F E^{X_s}[Y]dP^x$ as $\epsilon \to 0$. Thus the equality (88) is proved. □

Lemma 6.14. *If f is bounded and has compact support, then the conclusion of Theorem 6.12 holds true.*

Proof. Let $K = \text{supp}(f)$. By assumption $|f| \le c$ for some $c > 0$, so $\int |f| \le cm(K) < \infty$. Thus $f \in L^1$. Moreover, taking the inverse Fourier transform of f, we have $|f^\vee| \le \int |f| \le cm(K) < \infty$, so $f^\vee$ is also bounded. Therefore, by Corollary 7.9 there exists a sequence $\{f_j\}_{j=1}^\infty$ of smooth L^1 functions such that $f_j \to f$ a.e., each $f_j^\vee \in L^1$, and $f_j^\vee \to f^\vee$ pointwise. Moreover, by the inversion formula (Theorem 7.6), for all j, $\widehat{(f_j^\vee)} = f_j$ a.e.

Define $Y_j = f_j(X)$. Then $Y_j \to Y$ P^x-a.s. By this we mean that, except for ω in P^x-null set, $Y_j(\omega) \to Y(\omega)$ pointwise as a sequence of continuous functions on $[0,\infty)$. To justify this

statement, let $N \in \mathcal{B}^n$ be the null set where $f_j \not\to f$. Observe that $Y_j(\omega) \not\to Y(\omega)$ if and only if, for some $t \in [0,\infty)$, $X_t(\omega) \in N$. If $t > 0$, since the distribution of X_t is given by a normal density function, it follows that $P^x(X_t \in N) = 0$. By possibly redefining the f_j's on a null set, we may assume that $x \notin N$, in which case we also have $P^x(X_t \in N) = 0$ when $t = 0$.

Given $F \in \mathcal{F}$ and $s \geq 0$, we use the fact that $f_j = \widehat{f_j^\vee}$ and apply Fubini's theorem several times to compute

$$\begin{aligned}
\int_F Y_j \circ \tau_s dP^x &= \int_F f_j(\tau_s X) dP^x = \int_F \int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot \tau_s X} f_j^\vee(\xi) d\xi dP^x \\
&= \int_{\mathbb{R}^n} \left(\int_F e^{-2\pi i \xi \cdot \tau_s X} dP^x \right) f_j^\vee(\xi) d\xi \\
&= \int_{\mathbb{R}^n} \left(\int_F E^{X_s}[e^{-2\pi i \xi \cdot X}] dP^x \right) f_j^\vee(\xi) d\xi \qquad \text{(by Lemma 6.13)} \\
&= \int_F \int_{\mathbb{R}^n} E^{X_s}[e^{-2\pi i \xi \cdot X}] f_j^\vee(\xi) d\xi dP^x \\
&= \int_F E^{X_s} \left[\int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot X} f_j^\vee(\xi) d\xi \right] dP^x \\
&= \int_F E^{X_s} f_j(X) dP^x = \int_F E^{X_s} Y_j dP^x.
\end{aligned} \tag{90}$$

Note that by dominated convergence $\lim E^y Y_j = E^y Y$ for any $y \in \mathbb{R}^n$. Hence $\lim E^{X_s} Y_j = E^{X_s} Y$. Thus again by dominated convergence the first and last quantities appearing in (90) converge to $\int_F Y \circ \tau_s dP^x$ and $\int_F E^{X_s} Y dP^x$ respectively, as $j \to \infty$, and this proves the lemma. □

With these two results on hand, the theorem then follows relatively easily.

Proof of Theorem 6.12. Let f be any bounded measurable function, and for each integer $n \geq 0$, let B_n be the ball of radius n centered at zero. Let $f_n = f\chi_{B_n}$, and let $Y_n = f_n(X)$. Then $f_n \to f$ pointwise, and hence $Y_n \to Y$ pointwise. Given $F \in \mathcal{F}$ and $s \geq 0$, since each f_n is bounded and has compact support, by the previous lemma we have

$$\int_F Y_n \circ \tau_s dP^x = \int_F E^{X_s} Y_n dP^x. \tag{91}$$

Note that by dominated convergence $\lim E^y Y_n = E^y Y$ for any $y \in \mathbb{R}^n$, and in particular $\lim E^{X_s} Y_n = E^{X_s} Y$. Thus again by dominated convergence, the quantities on the right and left in equation (91) converge to $\int_F Y \circ \tau_s dP^x$ and $\int_F E^{X_s} Y dP^x$ respectively, and the theorem is proved. □

6.5 The Strong Markov Property

Let $\{\mathcal{F}_t\}$ be the filtration defined in the previous section. Let T be a fixed stopping time with respect to this filtration. We define

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap (T \leq t) \in \mathcal{F}_t, \text{ for all } t > 0\}. \tag{92}$$

Proposition 6.15. *(i) $\mathcal{F}_T$ is a σ-algebra. (ii) T is $\mathcal{F}_T$-measurable. (iii) X_T is $\mathcal{F}_T$ measurable.*

Proof. (i) Suppose $\{A_j\}_1^\infty \subset \mathcal{F}_T$. Then, for any $t > 0$, $(\cup_1^\infty A_j) \cap (T \leq t) = \cup_1^\infty (A_j \cap (T \leq t)) \in \mathcal{F}_t$, since each $A_j \cap (T \leq t) \in \mathcal{F}_t$. If $B \in \mathcal{F}_T$, then $B \cap (T \leq t) \in \mathcal{F}_t$ and $(T \leq t) \in \mathcal{F}_t$. Hence $(T \leq t) \backslash (B \cap (T \leq t)) = B^c \cap (T \leq t) \in \mathcal{F}_t$. Thus $\mathcal{F}_T$ is closed under countable unions and complementation, so $\mathcal{F}_T$ is a σ-algebra.

(ii) Since the intervals $[0, s]$, $s > 0$, generate the Borel sets in $[0, \infty)$, it is enough to show that, for any $s > 0$, $(T \leq s) \in \mathcal{F}_T$. But this is clearly the case, because, for all $t > 0$, $(T \leq s) \cap (T \leq t) = (T \leq \min(s, t)) \in \mathcal{F}_t$.

(iii) X_T is the composition of functions

$$X_T : (\Omega, \mathcal{F}_T) \xrightarrow{g} ([0, \infty) \times \Omega, \mathcal{B}^{[0,\infty)} \times \mathcal{F}_\infty) \xrightarrow{f} (\mathbb{R}^n, \mathcal{B}^n), \tag{93}$$

where $f(t, \omega) = X_t(\omega)$, $g(\omega) = (T(\omega), \omega)$, and $\mathcal{B}^{[0,\infty)} \times \mathcal{F}_\infty$ is the product σ-algebra. The function f is measurable because it is an $\mathcal{F}_\infty$-measurable function in ω and a continuous function in t. That g is measurable follows from the fact that T is $\mathcal{F}_T$-measurable. □

Some additional notation: We define the maps θ_T and X_T by $\theta_T(\omega)(t) = \omega(T(\omega) + t)$ and $X_T(\omega) = X_{T(\omega)}(\omega)$. We are now ready to state the strong Markov property.

Theorem 6.16 (Strong Markov Property)**.** *Suppose $Y = f(X)$, where f is a bounded, Borel measurable function, and (X, P^x) is a Wiener process. Then*

$$E^x[\theta_T Y | \mathcal{F}_T] = E^{X_T}[Y]. \tag{94}$$

Proof. We must prove that, for any $A \in \mathcal{F}_T$,

$$\int_A f(\theta_T X) dP^x = \int_A E^{X_T}[f(X)] dP^x. \tag{95}$$

Since the continuous functions are dense in the space of Borel-measurable functions on $\mathbb{R}^n$, we may assume that f is continuous. The general result then follows by taking a sequence of continuous functions and applying the dominated convergence theorem.

Define a sequence of stopping times T_n by

$$T_n(\omega) = k/2^n \quad \text{if } T(\omega) \in [(k-1)/2^n, k/2^n) \quad k = 1, 2, 3, ... \tag{96}$$

Then T_n decreases monotonically to T. If $A \in \mathcal{F}_T$, then $A \in \mathcal{F}_{T_n}$ and hence, if we let $A_k = A \cap \{\omega : T_n(\omega) = k/2^n\}$, then $A_k \in \mathcal{F}_{k/2^n}$. By the weak Markov property

$$\begin{aligned} \int_{A_k} f(\theta_{T_n} X) dP^x &= \int_{A_k} f(\theta_{k/2^n} X) dP^x = \int_{A_k} E^{X_{k/2^n}}[f(X)] dP^x \\ &= \int_{A_k} E^{X_{T_n}}[f(X)] dP^x. \end{aligned} \tag{97}$$

Since $\cup_{k=1}^\infty A_k = A$, it follows that

$$\int_A f(\theta_{T_n} X) dP^x = \sum_{k=1}^\infty \int_{A_k} E^{X_{T_n}}[f(X)] dP^x = \int_A E^{X_{T_n}}[f(X)] dP^x. \tag{98}$$

By continuity, for any $t \geq 0$ and $\omega \in \Omega$, $\lim_{n\to\infty} f((\theta_{T_n}X)_t(\omega)) = \lim_{n\to\infty} f(X_{T_n(\omega)+t}(\omega)) = f((\theta_T X)_t(\omega))$. Hence, by dominated convergence, the quantity on the left in (98) approaches $\int_A f(\theta_T X)dP^x$ as $n \to \infty$. As for the quantity on the right, note that the function $\varphi(y) = E^y[f(X)]$ is continuous. This is because $E^y[f(X)] = E^0[f(X+y)]$, and since f is continuous, continuity of φ follows by dominated convergence. Thus, applying dominated convergence to the quantity on the right in (98) yields that $\lim_{n\to\infty} \int_A E^{X_{T_n}}[f(X)]dP^x = \int_A E^{X_T}[f(X)]dP^x$. This completes the proof. □

7 Appendices

Appendix A: Convolutions and Approximate Identities

If f and g are two measurable functions on $\mathbb{R}^n$, the **convolution** of f and g is defined to be the function

$$(f * g)(x) = \int f(x-y)g(y)dy \qquad (x \in \mathbb{R}^n), \tag{99}$$

whenever the above integral is defined.

Convolutions have a number of nice analytical and algebraic properties. These include

Proposition 7.1. *Suppose f, g, h are measurable functions on $\mathbb{R}^n$. Then, whenever the respective convolutions are defined,*

a. $f * g = g * f$.

b. $f * (g * h) = (f * g) * h$.

c. If $f, g \in L^1$, then $f * g \in L^1$, and $||f * g||_1 \leq ||f||_1||g||_1$.

d. Let $f, g \in L^1(\mathbb{R}^n)$. Suppose $g \in C^k(U)$ for some $U \subset \mathbb{R}^n$ and each of the derivatives $\partial^\alpha g$ is bounded for $|\alpha| \leq k$. Then $f * g \in C^k(U)$ and $\partial^\alpha(f * g) = f * \partial^\alpha g$..

Proof. (a) This is just a change variables:

$$f * g(x) = \int f(x-y)g(y)dy = \int f(z)g(x-z)dz = g * f(x) \tag{100}$$

where we have made the substitution $z = x - y$.

(b) By Fubini's Theorem and (a)

$$\begin{aligned} f * (g * h)(x) &= \int f(y) \int g(z-x+y)h(z)dzdy \\ &= \int h(z) \int g(y-x+z)f(y)dydz = (f * g) * h(x). \end{aligned} \tag{101}$$

(c) Using Tonelli's Theorem, we have

$$\iint |f(y)g(x-y)|dydx = \int |f(y)| \int |g(x-y)|dxdy = ||f||_1||g||_1, \tag{102}$$

which is finite by assumption. It then follows that $f * g$ is defined almost everywhere, and

$$||f * g||_1 = \int \left| \int f(y)g(x-y)dy \right| dx \leq \iint |f(y)g(x-y)|dydx = ||f||_1||g||_1, \tag{103}$$

as required.

(d) By boundedness of the partial derivatives, it is clear that $f(y)\partial^\alpha g(x-y)$ is integrable for $|\alpha| \leq k$ and $x \in U$. It follows that for $|\alpha| \leq k$ and $x \in U$ we can differentiate under the integral to obtain

$$\partial^\alpha f * g(x) = \int f(y)\partial^\alpha g(x-y)dy = f * \partial^\alpha g. \tag{104}$$

(For justification of differentiating under the integral, see e.g. Folland [4, Theorem 2.27].) □

Convolutions allow us to approximate a given function by well-behaved functions. Suppose that $\phi \in L^1(\mathbb{R}^n)$, and $\int \phi = 1$. For $t > 0$, define $\phi_t(x) = t^{-n}\phi(t^{-1}x)$. Notice that by changing variables we still have $\int \phi_t = 1$. The collection of functions $\{\phi_t\}_{t>0}$ is known as an **approximate identity**. The reason for this name comes from the following result.

Theorem 7.2. *Suppose $f \in L^1$. Then $f * \phi_t \to f$ in L^1 as $t \to 0$.*

Remark 7.3. One can obtain other forms of convergence, such as uniform convergence or uniform convergence on compact sets, by imposing stronger conditions on f. See for example Theorem 8.14 in Folland [4].

If g is a function on $\mathbb{R}^n$ and $y \in \mathbb{R}^n$, we define $\tau_y g(x) = g(x-y)$. The operator τ_y is sometimes called the translation operator. To prove Theorem 7.2, we will need the following

Fact. If $g \in L^1(\mathbb{R}^n)$, then $\lim_{y\to 0} ||\tau_y g - g||_1 = 0$.

To avoid a somewhat lengthy digression, we refer the reader to Folland [4, Proposition 8.5] for a justification of this fact. The proof of Theorem 7.2 then proceeds as follows.

Proof. We start with the following calculation.

$$\begin{aligned} f * \phi_t(x) - f(x) &= \int [f(x-y) - f(x)]\phi_t(y)dy = \int [f(x-tz) - f(z)]\phi(z)dz \\ &= \int [\tau_{tz}f(x) - f(x)]\phi(z)dz. \end{aligned} \tag{105}$$

Using Fubini's Theorem, this allows us to obtain the following estimate.

$$\begin{aligned} ||f * \phi_t - f||_1 &\leq \iint |\tau_{tz}f(x) - f(x)||\phi(z)|dzdx \\ &= \int \left(\int |\tau_{tz}f(x) - f(x)|dx \right) |\phi(z)|dz \\ &= \int ||\tau_{tz}f - f||_1|\phi(z)|dz. \end{aligned} \tag{106}$$

Note that $||\tau_{tz}f - f||_1 \leq 2||f||_1$. Hence applying dominated convergence to the estimate above shows that $||f * \phi_t - f||_1 \to 0$ as $t \to 0$, as desired. □

Example 7.4. For $x \in \mathbb{R}^n$, define $\phi(x) = e^{-\pi|x|^2}$. We can write

$$e^{-\pi|x|^2} = \prod_{j=1}^{n} e^{-\pi x_j^2}. \tag{107}$$

By a well-known fact from calculus, each $\int e^{-\pi x_j^2} dx_j = 1$. Hence by Fubini's theorem $\int \phi = 1$. Thus by Theorem 7.2,

$$f * \phi_t \to f \text{ in } L^1. \tag{108}$$

But we can say more: Each ϕ_t is a smooth L^1 function, and hence by Propositions 7.1(c) and (d), $f * \phi_t$ is a smooth L^1 function. In fact, by choosing an appropriate sequence $\{t_j\}$, where each $t_j > 0$ and $t_j \downarrow 0$, and defining $f_j = f * \phi_{t_j}$, we obtain a sequence of smooth L^1 functions $\{f_j\}$ such that $f_j \to f$ a.e. (This follows from the fact that any sequence of functions $g_j \to g$ in L^1 has a subsequence which converges to g almost everywhere. See Corollary 2.32 in Folland [4])

Appendix B: The Fourier Transform

Given a function $f \in L^1(\mathbb{R}^n)$, the **Fourier transform** of f is the function

$$\hat{f}(\xi) = \int e^{-2\pi i x\cdot\xi} f(x) dx \qquad (\xi \in \mathbb{R}^n). \tag{109}$$

Typically, the Fourier transform of a function has better smoothness properties than the original function. In particular, by dominated convergence (take $|f|$ as the dominating function), the Fourier transform of the function f in (109) is continuous. Additional mild hypotheses on the function f allow one to differentiate under the integral in (109) multiple times to show that $\hat{f} \in C^k$. Another useful fact is that, under the Fourier transform, convolution becomes multiplication. More precisely, we have

Proposition 7.5. *Let f and g be L^1 functions (whose convolution is well-defined, by Proposition 7.1(c)). Then $\widehat{f * g} = \hat{f}\hat{g}$.*

Proof. The proof is just a matter of applying Fubini's Theorem. For $x \in \mathbb{R}^n$,

$$\begin{aligned}
\widehat{f * g}(x) &= \iint e^{-2\pi i x\cdot\xi} f(\xi - y) g(y) dy d\xi \\
&= \iint e^{-2\pi i x\cdot(\xi-y)} f(\xi - y) e^{-2\pi i x\cdot y} g(y) dy d\xi \\
&= \int \left(\int e^{-2\pi i x\cdot(\xi-y)} f(\xi - y) d\xi \right) e^{-2\pi i x\cdot y} g(y) dy \\
&= \int \hat{f}(x) e^{-2\pi i x\cdot y} g(y) dy = \hat{f}(x)\hat{g}(x),
\end{aligned} \tag{110}$$

as required. □

To recover the original function from its Fourier transform, we introduce *inverse Fourier transform* of a function $g \in L^1$. This is the function

$$g^\vee(x) = \hat{g}(-x) = \int e^{2\pi i x\cdot\xi} g(\xi) d\xi \qquad (x \in \mathbb{R}^n). \tag{111}$$

Observe that, due to the simple relationship above between $g^\vee$ and $\hat{g}$, Proposition 7.5 also holds for the inverse Fourier transform. The inversion formula is the following result.

Theorem 7.6 (Inversion of the Fourier transform)**.** *If $f \in L^1(\mathbb{R}^n)$ and $\hat{f} \in L^1(\mathbb{R}^n)$, let $g = (\hat{f})^\vee$. Then $g, h \in C_0(\mathbb{R}^n)$, and $f = g = h$ a.e.*

Remark 7.7. The same result still holds if we replace $\hat{f}$ by $f^\vee$ and let $g = \widehat{(f^\vee)}$.

For a proof of Theorem 7.6, see for example Theorem 8.26 in Folland [4] or Theorem 9.5 in Rudin [7].

The next result says that the function defined in Example 7.4 is a kind of eigenfunction of the Fourier transform. This result will give us a useful approximate identity. It also provides us with the characteristic function for Gaussian distribution (see Proposition 3.5).

Proposition 7.8. *For $x \in \mathbb{R}^n$, let $\phi(x) = e^{-\pi|x|^2}$. Then $\hat{\phi} = \phi$.*

Proof. The proof has two steps

Step 1. First we show that result holds for dimension $n = 1$. In this case, by definition,

$$\hat{\phi}(\xi) = \int e^{-2\pi i x\xi} e^{-\pi x^2} dx. \tag{112}$$

The derivative of $e^{-2\pi i x\xi} e^{-\pi x^2}$ with respect to ξ is $-2\pi i x e^{-2\pi i x\xi} e^{-\pi x^2}$, which is bounded by $g(x) = 2\pi|x|e^{-\pi x^2}$, which is integrable. Hence we may differentiate under the integral (see Folland [4, Proposition 2.27(b)]) to obtain

$$\begin{aligned}
\hat{\phi}'(\xi) &= \int -2\pi i x e^{-2\pi i x\xi} e^{-\pi x^2} dx \\
&= e^{-2\pi i x\xi} i e^{-\pi x^2} \,|_{-\infty}^{\infty} - \int (i e^{-\pi x^2})(-2\pi i \xi e^{-2\pi i x\xi}) dx \\
&= 0 - 2\pi\xi \int e^{-2\pi i x\xi} e^{-\pi x^2} dx \\
&= -2\pi\xi \hat{\phi}(\xi).
\end{aligned} \tag{113}$$

(The second line above is obtained by integration by parts with $u = e^{-2\pi i x\xi}$ and $dv = -2\pi i x e^{-\pi x^2} dx$.) Thus we have obtained an ODE $\hat{\phi}'(\xi) = -2\pi\hat{\phi}(\xi)$. To solve it, observe that by (113)

$$\frac{d}{d\xi}[\hat{\phi}(\xi) e^{\pi\xi^2}] = \hat{\phi}'(\xi) e^{\pi\xi^2} + 2\pi\xi\hat{\phi}(\xi) e^{\pi\xi^2} = 0. \tag{114}$$

Therefore, $\hat{\phi}(\xi) e^{\pi\xi^2} = \hat{\phi}(0) = \int e^{-\pi x^2} dx = 1$. So indeed $\hat{\phi}(\xi) = e^{-\pi\xi^2}$.

Step 2. Now we generalize dimension $n \geq 1$. For $x \in \mathbb{R}^n$, $|x|^2 = \sum_{j=1}^n x_j^2$. Hence, by Fubini's Theorem,

$$\begin{aligned} \hat{\phi}(\xi) &= \int e^{-2\pi i x\cdot\xi} e^{-\pi|x|^2} dx = \int (\prod_{j=1}^n e^{-2\pi i x_j \xi_j} e^{-\pi x_j^2}) dx_1 dx_2 \cdots dx_j \\ &= \prod_{j=1}^n \int e^{-2\pi i x_j \xi_j} e^{-\pi x_j^2} dx_j = \prod_{j=1}^n e^{-\pi \xi_j^2} = e^{-\pi|\xi|^2}, \end{aligned} \tag{115}$$

here using the result from Step 1. This completes the proof. □

Corollary 7.9. *Suppose $f \in L^1$. Then there exists a sequence $\{f_j\}_{j=1}^\infty$ of smooth L^1 functions such that $f_j \to f$ a.e. and $\hat{f}_j \to \hat{f}$ pointwise. In addition, if $\hat{f}$ is bounded, then the sequence may be chosen so that each $\hat{f}_j \in L^1$.*

Proof. Let ϕ and $\{f_j\}$ be chosen as in Example 7.4. Then as shown in the example, $\{f_j\}$ is a sequence of smooth L^1 functions such that $f_j \to f$ a.e. Moreover, by Proposition 7.5, $\hat{f}_j = \hat{f}\hat{\phi_{t_j}}$. Given $t > 0$, using the previous proposition, we compute

$$\begin{aligned} \hat{\phi}_t(\xi) &= t^{-n} \int e^{-2\pi i x\cdot\xi} e^{-\pi t^{-2}|x|^2} dx = \int e^{-2\pi i y\cdot\xi/t} e^{-\pi|y|^2} dy \\ &= \hat{\phi}(\xi/t) = e^{-\pi|\xi|^2/t^2}. \end{aligned} \tag{116}$$

Thus $\hat{\phi}_t \to 1$ as $t \to 0$, and hence $\hat{f}_j \to \hat{f}$ pointwise as $j \to \infty$. Moreover, $\hat{\phi}_t \in L^1$. So if $\hat{f}$ is bounded, then each $\hat{f}_j \in L^1$. □

Remark 7.10. The result also holds if we replace $\hat{f}_j$ with $f_j^\vee$ and $\hat{f}$ with $f^\vee$.

References

1. R. F. Bass, *Probabilistic Techniques in Analysis*, Probability and its Applications, Springer-Verlag New York, Inc., New York, 1995.

2. Z. Ciesielski, Hölder conditions for realisations of Gaussian processes, *Trans. Amer. Math. Soc.*, **99**(1961) 403-413.

3. R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge University Press, Cambridge, 2010.

4. G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ, 1999.

5. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 3rd ed. Springer-Verlag Berlin, Heidelberg, 1992.

6. L. C. G. Rogers, D. Williams, *Diffusions, Markov Processes and Martingales, Volume 1: Foundations*, 2nd ed. Cambridge University Press, Cambridge, 2000.

7. W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill Book Co., Singapore, 1987.

8. T. Tao, 254A, Notes 3b: Brownian Motion and Dyson Brownian Motion. 18 Jan. 2010. Web 8 June 2017. https://terrytao.wordpress.com/2010/01/18/254a-notes-3b-brownian-motion-and-dyson-brownian-motion/

9. D. Williams, *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.