The Wiener-Itô Chaos Decomposition and Multiple Wiener integrals

Xi Geng

Abstract

In these notes, we give a self-contained introduction to the Wiener-Itô chaos decomposition theorem, which is among the most fundamental tools in stochastic analysis. In its classical form, the theorem asserts that the Hilbert space of square integrable functionals on the Wiener space (the path space $\mathcal{W}$ equipped with the law μ of Brownian motion) admits a decomposition into orthogonal components:

$$L^2(\mathcal{W},\mu)=\bigoplus_{n=0}^{\infty}\mathcal{H}_n,$$

where each component $\mathcal{H}_n$ is generated by the so-called Hermite polynomial functionals of degree n. In addition, as discovered by K. Itô in his renowned paper [1], $\mathcal{H}_n$ coincides with the space of multiple Wiener integrals of order n.

We adopt an elementary approach to construct the classical Hermite polynomials by using Gram-Schmidt orthogonalisation. This also yields the decomposition theorem in the case of the one dimensional Gaussian measure. Our approach to the general case follows the main line of D. Nualart [4] by extracting the essential structure of the Wiener space and working in the framework of Gaussian probability spaces.

Contents

1 Motivation

We know from the classical theory of Fourier series that the Hilbert space $L^2([-\pi, \pi], dx)$ (dx is the Lebesgue measure) has an orthonormal basis (ONB) given by the functions $\{\frac{1}{\sqrt{2\pi}} e^{inx} : n \in \mathbb{Z}\}$. As a result, every $f \in L^2([-\pi, \pi], dx)$ admits an expansion

$$f(x) = \sum_{n \in \mathbb{Z}} \frac{c_n}{\sqrt{2\pi}} e^{inx} \tag{1.1}$$

which converges in the sense of L^2. Another way of looking at this property is to realise that the function e^{inx} is a 2π-periodic eigenfunction of the Laplace operator $\Delta = \frac{d^2}{dx^2}$ with eigenvalue $-n^2$:

$$\frac{d^2}{dx^2} e^{inx} = -n^2 e^{inx}.$$

As a result, when viewing Δ as a differential operator

$$\Delta : \mathcal{C}^2_{2\pi}([-\pi, \pi]) \subseteq L^2([-\pi, \pi], dx) \to L^2([-\pi, \pi], dx)$$

($\mathcal{C}^2_{2\pi}([-\pi, \pi])$ is the space of twice continuously differentiable functions with period 2π), the space $L^2([-\pi, \pi], dx)$ admits an orthogonal decomposition into eigenspaces of Δ :

$$L^2([-\pi, \pi], dx) = \bigoplus_{n \geqslant 0} E_{-n^2}, \tag{1.2}$$

where $E_{-n^2} = \text{Span}\{e^{inx}, e^{-inx}\}$ is the eigenspace associated with the eigenvalue $-n^2$. This is often known as the L^2-*spectral decomposition theorem* for the Laplace operator (on the circle).

To see where the circle comes from, the main observation is that periodic functions on $\mathbb{R}$ can be equivalently viewed as functions defined on the circle. As a result, the Laplacian is equivalently viewed as the Laplacian on the circle. The compactness of the circle is critical to expect the decomposition (1.2), which can further be generalised to the Laplace operator on any compact Riemannian manifolds. The situation becomes drastically different if we remove compactness, for instance if we move to the space $L^2(\mathbb{R}^1, dx)$. In this case, there are no integrable eigenfunctions of Δ. Moreover, smooth eigenfunctions are parametrised continuously: e^{itx} is an eigenfunction for each $t \in \mathbb{R}$. This also accounts for the replacement of Fourier series theory by Fourier transform theory in order to expect a proper inversion theorem.

On the other hand, if we replace the Lebesgue measure dx by the Gaussian measure

$$\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

many nice things will occur. Firstly, we do have an L^2-decomposition theorem

$$L^2(\mathbb{R}^1, \gamma) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n, \tag{1.3}$$

where the $\mathcal{H}_n$'s are one dimensional subspaces generated by the so-called *Hermite polynomials*. Moreover, (1.3) can indeed be viewed as the spectral decomposition of some differential operator. To describe this, we take a second look at the Lebesgue measure case. The relation between Δ and dx is revealed through the introduction of a Brownian motion: the Laplacian is the generator of Brownian motion, and dx is the invariant measure of Brownian motion in the sense that

$$\int_{\mathbb{R}} P_t f(x) dx = \int_{\mathbb{R}} f(x) dx$$

for a rich class of test functions f. In a similar way, it is reasonable to expect that in the decomposition (1.3), the $\mathcal{H}_n$'s are the eigenspaces of the generator of a Markov process whose invariant measure is the Gaussian measure γ. Indeed, such a Markov process is the so-called *Ornstein-Uhlenbeck process* defined by the SDE

$$dX_t = -X_t + \sqrt{2} dB_t,$$

which can be solved explicitly as $X_t = \sqrt{2} \int_0^t e^{-(t-s)} dB_s$. The generator of X_t is the differential operator given by

$$\mathcal{L} = \frac{d^2}{dx^2} - x\frac{d}{dx},$$

and the Gaussian measure γ is the invariant measure of X_t. It turns out that (1.3) is indeed a decomposition of $L^2(\mathbb{R}^1, \gamma)$ into orthogonal eigenspaces of $\mathcal{L}$ (cf. Section 2.5 below).

Unlike the Lebesgue measure which is sensitive to dimension, an important feature of the Gaussian measure is that many of its analytic properties (isoperimetric inequalities, concentration property, log-Sobolev inequalities etc.) are dimension free and extends naturally to infinite dimensions (e.g. on the path space equipped with the law of Brownian motion). In particular, the aforementioned decomposition theorem and the related spectral interpretation have natural extensions to general Gaussian probability spaces. In addition, in the case of Brownian motion, the homogeneous component $\mathcal{H}_n$ has an elegant connection with multiple Wiener integrals, which was discovered by K. Itô in his renowned paper [1]. This general decomposition theorem, known as the *Wiener-Itô chaos decomposition*,, plays a fundamental role in modern stochastic analysis. It has several important applications such as in the differential calculus on Wiener space and Stein's method for Gaussian approximations. We refer the reader to [3, 4] for a discussion on these applications.

The aim of these notes is to give a self-contained introduction to the Wiener-Itô chaos decomposition theorem and its connection with multiple Wiener integrals. In Section 2, we construct the Hermite polynomials from the elementary viewpoint of orthogonalisation and establish the decomposition theorem in $\mathbb{R}^1$. In Section 3, we discuss the Cameron-Martin structure of the Wiener space that is critical for generalising the theorem to the Brownian motion case. This leads us to the general framework of Gaussian probability spaces. In Section 4, we establish the decomposition theorem on arbitrary irreducible Gaussian probability spaces which applies to general continuous Gaussian processes. In Section 5, we discuss its connection with multiple Wiener integrals.

2 The one dimensional case and Hermite polynomials

As a toy model, we first look at the one dimensional case. Nonetheless, this part contains most of the essential ideas and structures for the more abstract development.

Let γ be the standard Gaussian measure on $\mathbb{R}^1$, namely

$$\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

We often use

$$w(x) \triangleq \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

to denote the Gaussian density. The inner product over $L^2(\mathbb{R}^1, \gamma)$ is denoted as

$$\langle f, g \rangle \triangleq \int_{\mathbb{R}^1} f(x) g(x) \gamma(dx),$$

and the L^2-norm is simply denoted as $\|\cdot\|$. Our aim in this section is to understand the structure of $L^2(\mathbb{R}^1, \gamma)$ by decomposing it into certain orthogonal subspaces. Essentially the same kind of decomposition will be obtained on the Wiener space.

2.1 Construction of an orthonormal basis of $L^2(\mathbb{R}^1, \gamma)$

It is an important observation that all polynomials are square integrable with respect to γ, due to the rapid decay of the kernel $w(x)$. This leads to a natural way of constructing an orthonormal basis (ONB) of $L^2(\mathbb{R}^1, \gamma)$. Indeed, we can apply the Gram-Schmidt orthogonalisation procedure to the linearly independent family

$$\{1, x, x^2, x^3, \cdots\}$$

of functions to obtain an orthonormal system

$$\{\bar{H}_0(x), \bar{H}_1(x), \bar{H}_2(x), \cdots\},$$

both of which generate the space of polynomials. Since polynomials are rich enough to approximate continuous functions, it is not surprising that this system has to be complete so that it becomes an ONB.

To be more precise, we recall from linear algebra that the Gram-Schmidt orthogonalisation procedure starts with $\bar{H}_0(x) \triangleq 1$ and inductively we have

$$\bar{H}_n(x) \triangleq \alpha_n \cdot \big(x^n - \sum_{k=0}^{n-1} \langle x^n, \bar{H}_k(x)\rangle \bar{H}_k(x)\big), \quad n \geqslant 1, \tag{2.1}$$

where $\alpha_n > 0$ is a normalising constant so that $\|\bar{H}_n\| = 1$. As a general property of orthogonalisation, we know that

$$\mathrm{Span}\{\bar{H}_0(x), \bar{H}_1(x), \cdots, \bar{H}_n(x)\} = \mathrm{Span}\{1, x, \cdots, x^n\} = \mathcal{P}_n \tag{2.2}$$

for each $n \geqslant 0$, where $\mathcal{P}_n$ denotes the space of polynomials of degree n. As a result, we only need the following lemma to conclude that $\{\bar{H}_0, \bar{H}_1, \bar{H}_2, \cdots\}$ is an ONB of $L^2(\mathbb{R}^1, \gamma)$.

Lemma 2.1. *The space $\mathcal{P}$ of polynomials is dense in $L^2(\mathbb{R}^1, \gamma)$.*

Proof. Let $f \in L^2(\mathbb{R}^1, \gamma)$ be an element that is orthogonal to all polynomials. We claim that

$$\int_{\mathbb{R}^1} f(x) e^{itx} \gamma(dx) = 0 \quad \forall t \in \mathbb{R}. \tag{2.3}$$

If this is true, the Fourier transform of the signed measure

$$\nu(A) \triangleq \int_A f(x)\gamma(dx), \quad A \in \mathcal{B}(\mathbb{R}^1)$$

is identically zero, which then implies that $\nu = 0$. As a result, $f = 0$ γ-a.s.

To show the claim (2.3), we argue in a slightly more general way. Let $c \in \mathbb{C}$ be fixed. For each n, define

$$S_n(x) \triangleq \sum_{k=0}^{n} \frac{c^k x^k}{k!}.$$

Then

$$\|e^{cx} - S_n(x)\| \leqslant \sum_{k=n+1}^{\infty} \frac{|c|^k}{k!} \cdot \|x^k\| = \sum_{k=n+1}^{\infty} \frac{|c|^k}{k!} \sqrt{\int_{\mathbb{R}^1} x^{2k} \gamma(dx)}.$$

A use of the $2k$-moment formula for the Gaussian measure γ reveals that the order of the general term in the above series is $\frac{C^k}{\sqrt{k}k^{k/2}}$. As a result, we know that

$$S_n(x) \to e^{cx} \quad \text{in } L^2(\mathbb{R}^1, \gamma)$$

as $n \to \infty$. Since $f \perp \mathcal{P}$, we conclude that $f \perp e^{cx}$ (for any $c \in \mathbb{C}$). In particular, the claim (2.3) holds. ∎

Remark 2.1. It is almost never an issue when one extends the consideration to complex-valued functions, as one can always consider the real and imaginary parts separately. An alternative argument without using Fourier transform of signed measures is the following. By using the Stone-Weierstrass theorem, the claim (2.3) implies $f \perp \varphi$ for any continuous periodic function φ. This further implies that $f \perp \psi$ for any bounded continuous function ψ (by choosing a periodic function φ that equals ψ on an arbitrarily large interval $[-M, M]$). A standard measure-theoretic argument then shows that the latter property is sufficient to conclude $f = 0$ γ-a.s.

Theorem 2.1. *The family* $\{\bar{H}_0, \bar{H}_1, \bar{H}_2, \cdots\}$ *is an ONB of* $L^2(\mathbb{R}^1, \gamma)$.

Proof. We already know that the family is orthonormal and generates the space of polynomials. According to Lemma 2.1, we know that the family is also complete. $\square$

Remark 2.2. The reason we use the notation $\bar{H}_n$ is to save the more common notation H_n for the standard Hermite polynomials.

2.2 Rodrigues' formula for $\bar{H}_n$

Our next effort is to figure out the shape of each $\bar{H}_n$. This is essential for deeper considerations as well as generalisations to the Wiener space / more general Gaussian probability spaces.

In the first place, it is clear from the equation (2.1) that $\bar{H}_n$ is a polynomial of degree n with leading monomial $\alpha_n x^n$. In addition, $\bar{H}_n$ is orthogonal to $\mathcal{P}_{n-1}$ since $\mathcal{P}_{n-1}$ is spanned by $\bar{H}_0, \cdots, \bar{H}_{n-1}$. Before proceeding further, we make two useful observations whose proofs are straight forward.

Fact 1. For any $n \geqslant 1$, we have

$$w(x)^{-1} \cdot \frac{d}{dx}\big(x^{n-1}w(x)\big) = -x^n + \psi_n(x), \quad \psi_n \in \mathcal{P}_{n-2}, \tag{2.4}$$

where we recall that $w(x) \triangleq \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the Gaussian density and $\mathcal{P}_{-1} \triangleq \{0\}$.

Fact 2. For any $n \geqslant 0$ and any polynomial $q \in \mathcal{P}$, we have

$$\int_{\mathbb{R}} \bar{H}_n'(x)q(x)\gamma(dx) = -\int_{\mathbb{R}} \bar{H}_n(x) \cdot w(x)^{-1}\frac{d}{dx}\big(q(x)w(x)\big)\gamma(dx).$$

More concisely,

$$\langle \bar{H}_n', q\rangle = -\langle \bar{H}_n, w^{-1} \cdot (qw)'\rangle \quad \forall q \in \mathcal{P}. \tag{2.5}$$

This is a direct consequence of integration by parts.

The above simple facts allow us to compute the leading coefficient α_n of $\bar{H}_n$ easily.

Lemma 2.2. *For each* $n \geqslant 1$, *we have*

$$\bar{H}_n' = \frac{\alpha_{n-1}}{\alpha_n}\bar{H}_{n-1}.$$

By using the above relation recursively, we have $\alpha_n = \frac{1}{\sqrt{n!}}$. *In particular,* $\bar{H}_n' = \sqrt{n}\bar{H}_{n-1}$.

Proof. Since $\bar{H}_n' \in \mathcal{P}_{n-1}$, we can write

$$\bar{H}_n' = \sum_{k=0}^{n-1} c_k \bar{H}_k.$$

By orthogonality, the coefficients c_k are given by

$$c_k = \langle \bar{H}_n', \bar{H}_k \rangle = -\langle \bar{H}_n, w^{-1} \cdot (\bar{H}_k w) \rangle, \quad 0 \leqslant k \leqslant n-1,$$

where the second identity follows from (2.5) with $q = \bar{H}_k$. For $k \leqslant n-2$, from (2.4) we know that $w^{-1} \cdot (\bar{H}_k w)'$ is a polynomial of degree $k+1 \leqslant n-1$. As a result, $c_k = 0$ for such k's. For $k = n-1$, again from (2.4) we see that $c_{n-1} = \alpha_{n-1} \langle \bar{H}_n, x^n \rangle$. On the other hand, if we take the inner product with $\bar{H}_n$ on both sides of (2.1), we have $\alpha_n \langle \bar{H}_n, x^n \rangle = 1$. Therefore, $c_{n-1} = \frac{\alpha_{n-1}}{\alpha_n}$. The first assertion thus follows.

For the second assertion, by further differentiation and using the first part recursively, we have

$$\bar{H}_n^{(n)} = \frac{\alpha_{n-1}}{\alpha_n} \cdot \frac{\alpha_{n-2}}{\alpha_{n-1}} \cdot \frac{\alpha_{n-3}}{\alpha_{n-2}} \cdot \cdots \cdot \frac{\alpha_0}{\alpha_1} \bar{H}_0 = \frac{1}{\alpha_n} \quad (\bar{H}_0 = 1).$$

Since the leading term of $\bar{H}_n$ is $\alpha_n x^n$, we also know that $\bar{H}_n^{(n)} = n! \alpha_n$. Therefore,

$$\frac{1}{\alpha_n} = n! \alpha_n \implies \alpha_n = \frac{1}{\sqrt{n!}}.$$

The final assertion of the lemma requires no comment. □

In order to derive an explicit formula for the function $\bar{H}_n$, we need to look deeper into the relation between $\bar{H}_{n-1}$ and $\bar{H}_n$. From (2.4) we know that $w^{-1} \cdot (\bar{H}_{n-1} w)' \in \mathcal{P}_n$, and thus we can write

$$w^{-1} \cdot (\bar{H}_{n-1} w)' = \sum_{k=0}^{n} c_k \bar{H}_k,$$

where

$$c_k = \langle w^{-1} \cdot (\bar{H}_{n-1} w)', \bar{H}_k \rangle = -\langle \bar{H}_k', \bar{H}_{n-1} \rangle, \quad 0 \leqslant k \leqslant n.$$

It is clear that $c_k = 0$ when $k \leqslant n-1$. In addition, from Lemma 2.2 we have

$$c_n = -\langle \bar{H}_n, \bar{H}_{n-1} \rangle = -\sqrt{n} \langle \bar{H}_{n-1}, \bar{H}_{n-1} \rangle = -\sqrt{n}.$$

Consequently,

$$w^{-1} \cdot (w \bar{H}_{n-1})' = -\sqrt{n} \bar{H}_n. \tag{2.6}$$

This relation leads us to the following so-called *Rodrigues' formula* for $\bar{H}_n$.

Theorem 2.2. *For each $n \geqslant 0$, we have*

$$\bar{H}_n(x) = \frac{(-1)^n}{\sqrt{n!}} \cdot e^{x^2/2} \cdot \frac{d^n}{dx^n} \left(e^{-x^2/2} \right).$$

Proof. By using the relation (2.6) recursively, we have

$$\begin{aligned}
\bar{H}_n(x) &= -\frac{1}{\sqrt{n}} \cdot w(x)^{-1} \cdot \frac{d}{dx}\big(w(x)\bar{H}_{n-1}\big) = -\frac{1}{\sqrt{n}} \cdot e^{x^2/2} \cdot \frac{d}{dx}\big(e^{-x^2/2}\bar{H}_{n-1}\big) \\
&= -\frac{1}{\sqrt{n}} \cdot e^{x^2/2} \cdot \frac{d}{dx}\Big(e^{-x^2/2} \cdot \Big(-\frac{1}{\sqrt{n-1}} e^{x^2/2}\frac{d}{dx}\big(e^{-x^2/2}\bar{H}_{n-2}\big)\Big)\Big) \\
&= \frac{(-1)^2}{\sqrt{n(n-1)}} e^{x^2/2} \cdot \frac{d^2}{dx^2}\big(e^{-x^2/2}\bar{H}_{n-2}\big) \\
&= \frac{(-1)^2}{\sqrt{n(n-1)}} e^{x^2/2} \cdot \frac{d^2}{dx^2}\Big(e^{-x^2/2}\Big(-\frac{1}{\sqrt{n-2}} e^{x^2/2}\frac{d}{dx}\big(e^{-x^2/2}\bar{H}_{n-3}\big)\Big)\Big) \\
&= \frac{(-1)^3}{\sqrt{n(n-1)(n-2)}} e^{x^2/2}\frac{d^3}{dx^3}\big(e^{-x^2/2}\bar{H}_{n-3}\big) = \cdots = \frac{(-1)^n}{\sqrt{n!}} e^{x^2/2}\frac{d^n}{dx^n}\big(e^{-x^2/2}\big).
\end{aligned}$$

This gives the desired formula. □

2.3 The generating function and basic properties of $\bar{H}_n$

Rodrigues' formula implicitly suggests that the $\bar{H}_n$'s may arise as the coefficients of the Taylor expansion of certain function (a generating function). To elaborate this idea, let us define

$$F(x,t) \triangleq e^{xt-t^2/2}, \quad (x,t) \in \mathbb{R}^2.$$

By writing

$$F(x,t) = e^{x^2/2-(x-t)^2/2}$$

and using Rodrigues' formula, we see that

$$\frac{\partial^n F(x,t)}{\partial t^n}\bigg|_{t=0} = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}\big(e^{-x^2/2}\big) = \sqrt{n!}\bar{H}_n(x).$$

A standard application of Taylor's expansion gives the following useful fact.

Proposition 2.1. *$F(x,t)$ is the generating function of the $\bar{H}_n$'s in the sense that*

$$F(x,t) = \sum_{n=0}^{\infty} \frac{\bar{H}_n(x)}{\sqrt{n!}} t^n. \tag{2.7}$$

In other words, $\bar{H}_n(x)/\sqrt{n!}$ is the n-th coefficient in the Taylor expansion of $F(x,t)$ with respect to t.

Using the viewpoint of generating function, we can easily summarise the essential properties of $\bar{H}_n$.

Proposition 2.2. *The functions $\{\bar{H}_n : n \geqslant 0\}$ satisfy the following three relations:*

(i) $\bar{H}_n(-x) = (-1)^n\bar{H}_n(x)$, namely $\bar{H}_n$ is an odd function if n is odd and it is an even function if n is even;
(ii) $\bar{H}_n' = \sqrt{n}\bar{H}_{n-1}$;
(iii) $\sqrt{n}\bar{H}_n(x) = x\bar{H}_{n-1}(x) - \sqrt{n-1}\bar{H}_{n-2}(x)$.

Proof. (i) This follows from the observation that

$$F(-x,t) = F(x,-t)$$

and the expansion 2.7.

(ii) This is a restatement of Lemma 2.2 but let us derive it by using the generating function. Observe that $F(x,t)$ satisfies the following PDE:

$$\frac{\partial F(x,t)}{\partial x} = tF(x,t).$$

As a result, we have

$$\sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} \bar{H}_n(x)' = \sum_{n=0}^{\infty} \frac{t^{n+1}}{\sqrt{n!}} \bar{H}_n(x).$$

The relation follows by comparing the coefficients of t^n on both sides.

(iii) This relation can be obtained in a similar way as in (ii) by observing that $F(x,t)$ satisfies another PDE:

$$\frac{\partial F(x,t)}{\partial t} = (x-t)F(x,t).$$

□

2.4 Hermite polynomials

In probability theory, we often work with the normalisation

$$H_n(x) \triangleq \frac{\bar{H}_n(x)}{\sqrt{n!}}, \quad n \geqslant 0.$$

Definition 2.1. The polynomial H_n is called the n-th *Hermite polynomial* over $\mathbb{R}^1$.

We summarise the essential properties of H_n in the following result. They are direct translations of what we have obtained previously.

(i) *Rodrigues' formula*:

$$H_n(x) = \frac{(-1)^n}{n!} e^{x^2/2} \frac{d^n}{dx^n} \left(e^{-x^2/2}\right);$$

(ii) *Parity*:

$$H_n(-x) = (-1)^n H_n(x);$$

(iii) *1st recursive relation*:

$$H_n' = H_{n-1}; \tag{2.8}$$

(iv) *2nd recursive relation*:

$$(n+1)H_{n+1}(x) = xH_n(x) - H_{n-1}(x). \tag{2.9}$$

Remark 2.3. The first few Hermite polynomials are given by

$$H_0(x) = 1,\ H_1(x) = x,\ H_2(x) = \frac{x^2-1}{2},\ H_3(x) = \frac{x^3-3x}{6} \text{ etc.}$$

The following generalised orthogonality property is useful when extending the discussion to general Gaussian probability spaces.

Lemma 2.3. *Let (X,Y) be a jointly Gaussian random vector such that*

$$\mathbb{E}[X]=\mathbb{E}[Y]=0,\ V[X]=V[Y]=1.$$

Then

$$\mathbb{E}[H_m(X)H_n(Y)]=\begin{cases}\frac{1}{n!}\mathbb{E}[XY]^n, & m=n,\\ 0, & \textit{otherwise.}\end{cases}$$

Proof. The joint moment generating function of (X,Y) is given by

$$\mathbb{E}[e^{sX+tY}]=e^{\frac{1}{2}(s^2+t^2+2st\mathbb{E}[XY])}.$$

Equivalently, we have

$$\begin{aligned}e^{st\mathbb{E}[XY]}&=\mathbb{E}\big[e^{sX-\frac{1}{2}s^2}\cdot e^{tY-\frac{1}{2}t^2}\big]=\mathbb{E}[F(X,s)F(Y,t)]\\&=\mathbb{E}\big[\big(\sum_{m=0}^{\infty}H_m(X)s^m\big)\cdot\big(\sum_{n=0}^{\infty}H_n(Y)t^n\big)\big]\\&=\sum_{m,n=0}^{\infty}\mathbb{E}[H_m(X)H_n(Y)]s^mt^n.\end{aligned}$$

The result follows by expanding the left hand side into an (s,t)-series and comparing the coefficients of s^mt^n on both sides. □

Theorem 2.1 can be restated in the following form. Its extension to general Gaussian probability spaces (including Brownian motion as a basic example) is the goal of later sections.

Theorem 2.3. *The Hilbert space $L^2(\mathbb{R}^1,\gamma)$ admits the following decomposition:*

$$L^2(\mathbb{R}^1,\gamma)=\bigoplus_{n=0}^{\infty}\mathcal{H}_n, \tag{2.10}$$

where $\mathcal{H}_n\triangleq\mathrm{Span}\{H_n(x)\}$ and $\mathcal{H}_m\perp\mathcal{H}_n$ for all $m\neq n$.

2.5 The spectral perspective

Let us examine Theorem (2.3) from the viewpoint of spectral decomposition. We want to identify a differential operator $\mathcal{L}$ such that the $\mathcal{H}_n$'s in the decomposition (2.10) are the eigenspaces of $\mathcal{L}$. As mentioned in the introduction, an essential point is that the Gaussian measure γ should be the invariant measure of a Markov process X_t whose generator is $\mathcal{L}$, in the sense that

$$\int_{\mathbb{R}}P_tf(x)\gamma(dx)=\int_{\mathbb{R}}f(x)\gamma(dx)\quad\forall f \tag{2.11}$$

where P_t is the transition semigroup of X_t generated by $\mathcal{L}$. In particular,

$$\int_{\mathbb{R}} \mathcal{L}f(x)\gamma(dx) = 0 \quad \forall f.$$

To derive the shape of $\mathcal{L}$, A natural idea is to look for $\mathcal{L}$ as a perturbation of the Laplacian by a first order differential operator. Namely,

$$\mathcal{L} = \frac{d^2}{dx^2} + b(x)\frac{d}{dx},$$

with some function $b(x)$ to be determined. By the definition of the generator, we have

$$\int_{\mathbb{R}} \mathcal{L}f(x)\gamma(dx) = \int_{\mathbb{R}} \big(f''(x) + b(x)f'(x)\big)\gamma(dx) \quad \forall f.$$

Assuming all functions have at most polynomial growth at infinity, a simple integration by parts shows that

$$\int_{\mathbb{R}} (x + b(x))f'(x)\gamma(dx) = 0 \quad \forall f.$$

As a result, we have $b(x) = -x$ and thus

$$\mathcal{L} = \frac{d^2}{dx^2} - x\frac{d}{dx}.$$

From the perspective of stochastic calculus, the associated Markov process with generator $\mathcal{L}$ is given by the SDE:

$$dX_t = -X_t dt + \sqrt{2}dB_t, \ X_0 \sim N(0,1). \tag{2.12}$$

By solving it explicitly, one finds that

$$X_t = e^{-t}X_0 + \sqrt{2}\int_0^t e^{-(t-s)}dB_s.$$

In particular, X_t is a stationary Gaussian process and it is known as the *Ornstein-Uhlenbeck process*. Explicit calculation shows that the invariant measure of X_t is the Gaussian measure γ in the sense of (2.11).

On the other hand, from the recursive relations (2.8) and (2.9), it is plain to check that

$$H_n''(x) - xH_n'(x) = -nH_n(x).$$

In particular, $\mathcal{H}_n$ is the eigenspace of $\mathcal{L}$ associated with the eigenvalue $-n$ ($n \geqslant 0$). Moreover, when viewed as an unbounded linear operator

$$\mathcal{L} : \mathcal{C}_b^2(\mathbb{R}^1) \subseteq L^2(\mathbb{R}^1, \gamma) \to L^2(\mathbb{R}^1, \gamma),$$

it can be shown that $\{-n : n \geqslant 0\}$ are the only possible eigenvalues of $\mathcal{L}$. As a consequence, Theorem 2.3 provides an L^2-spectral decomposition of the differential operator $\mathcal{L}$ into orthogonal eigenspaces.

3 Gaussian probability spaces

As we have mentioned, an important feature of Gaussian measures is that many of their properties are dimension free and extends to infinite dimensions naturally. The most canonical example is the Wiener measure (the law of Brownian motion) on the so-called Wiener space. Our next goal is to generalise Theorem 2.1 to the Wiener space. However, there are several specific structures on the Wiener space that are irrelevant to the development. Restricting ourselves to the special setting of Brownian motion will conceal the essence of the theorem. Therefore, we first spend some time extracting the essential structure of the Wiener space on which the theorem is based.

3.1 The Wiener space and its Cameron-Martin subspace

Let $\mathcal{W}$ be the space of continuous functions $w : [0,1] \to \mathbb{R}^1$ with $w_0 = 0$. $\mathcal{W}$ is a Banach space under the supremum norm. We equip $\mathcal{W}$ with the Borel σ-algebra $\mathcal{B}(X)$ (the σ-algebra generated by open subsets of $\mathcal{W}$). Let μ be probability measure on $\mathcal{B}(X)$ defined by the law of Brownian motion over $[0,1]$. The probability space $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$ is known as the *Wiener space* over $[0,1]$, and the probability measure μ is called the *Wiener measure*.

The classical Wiener-Itô decomposition theorem asserts that the Hilbert space $L^2(\mathcal{W}, \mu)$ admits the following orthogonal decomposition

$$L^2(\mathcal{W}, \mu) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n, \tag{3.1}$$

where each $\mathcal{H}_n$ is a closed subspace generated by certain Hermite polynomial functionals on $\mathcal{W}$ of degree n. The essential ingredient governing this decomposition is an intrinsic Hilbert structure embedded the Wiener space $\mathcal{W}$, known as the *Cameron-Martin subspace*, which we now describe.

Observe that the canonical process

$$W_t(w) \triangleq w_t, \quad w \in \mathcal{W}$$

is a standard Brownian motion over $[0,1]$ under μ. We define $\mathcal{H}_1$ be the closure (in $L^2(\mathcal{W}, \mu)$) of the linear subspace spanned by $\{W_t : 0 \leqslant t \leqslant 1\}$. Since Gaussian distributions are closed weak convergence, it is clear that all elements in $\mathcal{H}_1$ are Gaussian random variables on $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$. The subspace $\mathcal{H}_1$ will be the $n = 1$ component in the decomposition (3.1). Given an element $Z \in \mathcal{H}_1$, we can construct an associated path by

$$h_t \triangleq \mathbb{E}^\mu[Z W_t], \quad 0 \leqslant t \leqslant 1. \tag{3.2}$$

It is easy to see that $h \in \mathcal{W}$. Let H be the space of paths $h \in \mathcal{W}$ that arise in this way.

Definition 3.1. We define an inner product structure on H by

$$\langle h_1, h_2 \rangle_H \triangleq \mathbb{E}^\mu[Z_1 Z_2],$$

where h_i is associated with Z_i ($i = 1, 2$). The Hilbert space $(H, \langle h_1, h_2 \rangle_H)$ is known as the *Cameron-Martin subspace* of the Wiener space $\mathcal{W}$.

Remark 3.1. There are two technical points to check: the inner product is well defined, and H is indeed a Hilbert space under this inner product. It can also be seen that H is separable. We leave these tedious details as an exercise.

From the definition of H, we know that $(H, \langle \cdot, \cdot \rangle_H)$ and $(\mathcal{H}_1, \langle \cdot, \cdot \rangle_{L^2(W,\mu)})$ are isometrically isomorphic. The following result describes the explicit shape of H. We use $\dot{h}$ to denote the time derivative of a function $h : [0, 1] \to \mathbb{R}^1$.

Theorem 3.1. *The Cameron-Martin subspace is equivalently given by*

$$H = \{h \in \mathcal{W} : h \text{ is absolutely continuous and } \dot{h} \in L^2([0, 1], dt)\}, \tag{3.3}$$

where the inner product structure is

$$\langle h_1, h_2 \rangle_H = \int_0^1 \dot{h}_1(t)\dot{h}_2(t)dt. \tag{3.4}$$

Proof. Let $\tilde{H}$ be the Hilbert space defined by the right hand side of (3.3) whose inner product $\langle \cdot, \cdot \rangle_{\tilde{H}}$ is defined by the right hand side of (3.4). To put the definition in another way, every element $h \in \tilde{H}$ is given by

$$h_t = \int_0^t \varphi_s ds, \quad 0 \leqslant t \leqslant 1, \tag{3.5}$$

where $\varphi \in L^2([0, 1], dt)$. Correspondingly, the inner product $\langle \cdot, \cdot \rangle_{\tilde{H}}$ is induced from the one on $L^2([0, 1], dt)$. We wish to show that

$$H = \tilde{H}, \ \langle \cdot, \cdot \rangle_H = \langle \cdot, \cdot \rangle_{\tilde{H}}.$$

The main idea is to verify this on a suitable dense subspace.

Let H' be the subspace of H consisting of those h's defined by (3.2) with

$$Z \in \mathcal{H}_1' \triangleq \operatorname{Span}\{W_u : 0 \leqslant u \leqslant 1\} \subseteq \mathcal{H}_1.$$

It is clear that H is the closure of H'. In parallel, let $\tilde{H}'$ be the subspace of H' consisting of those h's defined by (3.5) with

$$\varphi \in E \triangleq \operatorname{Span}\{\mathbf{1}_{[0,u]} : u \in [0, 1]\} \subseteq L^2([0, 1], dt).$$

We also observe that $\tilde{H}$ is the closure of $\tilde{H}'$. As a result, it is sufficient to show that $H' = \tilde{H}'$ and the two inner products are identical on H'. But this is immediate since the two spaces $\mathcal{H}_1'$ and E are in one-to-one correspondence through

$$W_u \longleftrightarrow \mathbf{1}_{[0,u]}, \quad u \in [0, 1].$$

The random variable $Z \triangleq W_u$ gives rise to

$$h_t = \mathbb{E}[ZW_t] = t \wedge u, \quad 0 \leqslant t \leqslant 1,$$

while the L^2-function $\varphi \triangleq \mathbf{1}_{[0,u]}$ gives rise to the same path

$$h_t = \int_0^t \mathbf{1}_{[0,u]}(s)ds = t \wedge u, \quad 0 \leqslant t \leqslant 1.$$

The two inner products are identical since

$$\mathbb{E}[W_u^2] = u = \|\mathbf{1}_{[0,u]}\|^2_{L^2([0,1],dt)}.$$

$\square$

Remark 3.2. Many authors directly define the Cameron-Martin subspace by (3.3). We take a different viewpoint which is more robust and applies to arbitrary continuous Gaussian processes.

Remark 3.3. The Cameron-Martin subspace plays a fundamental role in the analysis of Brownian motion and Wiener functionals such as solutions to SDEs. Its significance lies in the renowned *Cameron-Martin transformation theorem*, which asserts that the Wiener measure is quasi-invariance along directions in H (namely, the measure μ^h induced by the translation $w \mapsto w + h$ along any given direction $h \in H$ is absolutely continuous with respect to μ), while it is singular along directions in H^c (namely, μ^h and μ are singular to each other for any $h \in H^c$). This transformation theorem suggests that a proper notion of differential calculus on the Wiener space needs to respect the Cameron-Martin structure H (i.e. only differentiation along H-directions is meaningful under the Wiener measure μ). The development of such a theory as well as its applications is one of the main themes in stochastic analysis (the Malliavin calculus).

A natural question is, given an absolutely continuous path $h : [0,1] \to \mathbb{R}^1$ with $\dot{h} \in L^2([0,1],dt)$, what is the corresponding $Z \in \mathcal{H}_1$ that satisfies (3.2)? To answer this question, we first introduce the notion of Wiener integrals. For any indicator function $\varphi = \mathbf{1}_{[u,v]} \in L^2([0,1],dt)$, we set

$$\mathcal{I}(\varphi) \triangleq \int_0^1 \varphi_t dW_t \triangleq W_v - W_u.$$

By linear extension, this defines a mapping $\mathcal{I}$ from the subspace of step functions (linear combinations of the $\mathbf{1}_{[u,v]}$'s) into $L^2(\mathcal{W},\mu)$. This mapping is an isometry in the sense that

$$\langle \mathcal{I}(\varphi), \mathcal{I}(\psi)\rangle_{L^2(\mathcal{W},\mu)} = \langle \varphi, \psi\rangle_{L^2([0,1],dt)} \quad \forall \text{ step functions } \varphi, \psi, \tag{3.6}$$

which can be checked by plain calculation. As a result, $\mathcal{I}$ extends to an isometric embedding

$$\mathcal{I} : L^2([0,1],dt) \to L^2(\mathcal{W},\mu).$$

From the definition of $\mathcal{H}_1$, we further see that the image of $\mathcal{I}$ is precisely $\mathcal{H}_1$. In other words, $\mathcal{I}$ defines an isometric isomorphism between $L^2([0,1],dt)$ and $\mathcal{H}_1$. As a result,

$$\mathcal{I}(\varphi) \sim N(0, \|\varphi\|^2_{L^2([0,1],dt)}) \quad \forall \varphi \in L^2([0,1],dt).$$

In fact, more than this is true: the family

$$\{\mathcal{I}(\varphi) : \varphi \in L^2([0,1],dt)\}$$

is a Gaussian system with mean zero and covariance structure given by (3.6).

Definition 3.2. For any $\varphi \in L^2([0,1], dt)$, the random variable $\mathcal{I}(\varphi)$ on $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$ is known as the *Wiener integral* of φ. Symbolically we write

$$\mathcal{I}(\varphi) = \int_0^1 \varphi_t dW_t \text{ or more canonically } \mathcal{I}(\varphi)(w) = \int_0^1 \varphi_t dw_t, \ w \in \mathcal{W}.$$

Remark 3.4. Wiener integrals are special cases of Itô's stochastic integrals in stochastic calculus.

Using the notion of Wiener integrals, a direct corollary of Theorem 3.1 is the following.

Proposition 3.1. *For any $h \in H$, the corresponding Gaussian random variable $Z \in \mathcal{H}_1$ in the relation (3.2) is given by $Z = \mathcal{I}(\dot{h})$.*

Proof. Essentially we need to check that

$$h_t = \mathbb{E}[\mathcal{I}(\dot{h})W_t] \quad \forall h \in H \text{ and } t \in [0,1].$$

Given fixed t, the Cameron-Martin path $h^t_\cdot \in H$ associated with $Z \triangleq W_t$ is

$$h^t_s = s \wedge t, \ 0 \leqslant s \leqslant 1.$$

In addition, the Wiener integral of $\dot{h}^t$ is exactly W_t. According to isometry property (3.6), for any $h \in H$ we have

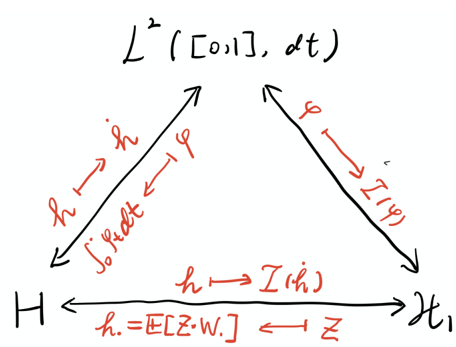
$$\mathbb{E}[\mathcal{I}(\dot{h})W_t] = \mathbb{E}[\mathcal{I}(\dot{h})\mathcal{I}(\dot{h}^t)] = \langle \dot{h}, \dot{h}^t \rangle_{L^2([0,1],dt)} = \langle \dot{h}, \mathbf{1}_{[0,t]} \rangle_{L^2([0,1],dt)} = h_t.$$

The desired relation then follows. $\square$

The relations among the three isometrically isomorphic Hilbert spaces

$$L^2([0,1], dt), \ H, \ \mathcal{H}_1$$

are summarised in the following diagram.



The essential structure from the Wiener space

By abuse of notation, we write

$$W : H \to \mathcal{H}_1,\ W(h) \triangleq \mathcal{I}(\dot{h}) = \int_0^1 \dot{h}_t dW_t$$

to denote the mapping at the bottom line of the previous diagram.

To summarise the essential structure obtained so far, we have a mean zero Gaussian family

$$\{W(h) : h \in H\}$$

indexed by a separable Hilbert space H, whose covariance structure is given by

$$\mathbb{E}[W(h_1)W(h_2)] = \langle h_1, h_2 \rangle_H.$$

As we will see, this is the only structure needed for establishing the aforementioned decomposition theorem (3.1).

3.2 Irreducible Gaussian probability spaces

Having the previous essential structure in mind, we now introduce the following definition (cf. Malliavin [2]). Let H be a separable Hilbert space that is given and fixed.

Definition 3.3. An *irreducible Gaussian probability space* is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which is defined a Gaussian family $\{W(h) : h \in H\}$ such that

(i) $\mathcal{F} = \sigma(W(h) : h \in H)$;
(ii) For any $h_1, h_2 \in H$, we have

$$\mathbb{E}[W(h_1)W(h_2)] = \langle h_1, h_2 \rangle_H. \tag{3.7}$$

The family $\{W(h) : h \in H\}$ is called an *isonormal Gaussian family* with respect to H.

The Wiener space is a special example of an irreducible Gaussian probability space. More generally, given any Gaussian process $\{X_t : 0 \leqslant t \leqslant 1\}$ with continuous sample paths, there is an associated irreducible Gaussian space in which the underlying Hilbert space is the corresponding Cameron-Martin subspace. The construction, in particular of this Cameron-Martin subspace (cf. 3.1), follows the same line as in the Brownian motion case.

The following property is parallel to the case of the Wiener space. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{W(h) : h \in H\})$ be an irreducible Gaussian probability space.

Lemma 3.1. *The mapping*

$$W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P}),\ h \mapsto W(h) \tag{3.8}$$

is an linear isometric embedding.

Proof. The only thing that needs to be checked is linearity:

$$W(ch_1 + h_2) = cW(h_1) + W(h_2), \quad \forall h_1, h_2 \in H \text{ and } c \in \mathbb{R}. \tag{3.9}$$

To this end, by using the relation (3.7) we easily find

$$\mathbb{E}[(W(ch_1 + h_2) - cW(h_1) - W(h_2))^2] = 0.$$

Therefore, the equation (3.9) holds in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. □

4 The Wiener-Itô decomposition theorem on irreducible Gaussian probability spaces

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{W(h) : h \in H\})$ be a given fixed irreducible Gaussian probability space. Recall that H_n $(n \geqslant 0)$ is the n-th Hermite polynomial over $\mathbb{R}^1$. We define $\mathcal{H}_0 \triangleq \mathbb{R}$, and for each $n \geqslant 1$ we define $\mathcal{H}_n$ to be the L^2-closure of the subspace

$$\text{Span}\{H_n(W(h)) : h \in H,\ \|h\|_H = 1\}. \tag{4.1}$$

Note that $\mathcal{H}_1 = \{W(h) : h \in H\}$. For $n \geqslant 2$, the condition $\|h\|_H = 1$ in (4.1) is needed.

Definition 4.1. The closed subspace $\mathcal{H}_n$ is called the n-th *Wiener chaos* over the given Gaussian probability space. Elements in $\mathcal{H}_n$ are often called *homogeneous Wiener polynomials* of degree n.

The main result of this section is the following decomposition theorem which generalises the one dimensional case in Section 2.

Theorem 4.1 (The Wiener-Itô chaos decomposition theorem)**.** *The space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ admits the following decomposition:*

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n,$$

where the Wiener chaoses $\mathcal{H}_n$ are orthogonal to each other: $\mathcal{H}_m \perp \mathcal{H}_n$ for any $m \neq n$.

Proof. The orthogonality of the $\mathcal{H}_n$'s is a direct consequence of Lemma 2.3. To establish the decomposition, it remains to show that:

$$F \in L^2(\Omega, \mathcal{F}, \mathbb{P}),\ F \perp \mathcal{H}_n\ \forall n \implies F = 0.$$

Let F be such an element. Since $\mathcal{F}$ is generated by the family $\{W(h) : h \in H\}$, to show $F = 0$ it is enough to show that the signed measure

$$\nu(\Gamma) = \mathbb{E}[F\mathbf{1}_{\{(W(h_1),\cdots,W(h_r))\in\Gamma\}}], \quad \Gamma \in \mathcal{B}(\mathbb{R}^r)$$

is zero for any given r and $h_1, \cdots, h_r \in H$. To this end, note that the Fourier transform of ν is given by

$$\begin{aligned}
\hat{\nu}(t_1, \cdots, t_r) &= \int_{\mathbb{R}^r} e^{i(t_1x_1+\cdots t_rx_r)} d\nu = \mathbb{E}[Fe^{i(t_1W(h_1)+\cdots+t_rW(h_r))}] \\
&= \mathbb{E}[Fe^{iW(h)}] \quad (h \triangleq t_1h_1 + \cdots + t_rh_r) \\
&= \sum_{n=0}^{\infty} \frac{i^n\|h\|_H^n}{n!}\mathbb{E}[FW(\bar{h})^n] \quad (\bar{h} \triangleq h/\|h\|_H).
\end{aligned}$$

Since $W(\bar{h})^n \in \text{Span}\{1, H_1(W(\bar{h})), \cdots, H_n(W(\bar{h}))\}$ (cf. (2.2)), by the assumption on F we conclude that $\hat{\nu} = 0$, and thus $\nu = 0$. Since r and $h_1, \cdots, h_r$ are arbitrary, it follows that $F = 0$. □

Similar to the case of $\mathbb{R}^1$, it can be show that the sum of the first n Wiener chaoses gives the space of "polynomial functionals" of degree n. To make this precise, let $\mathcal{P}_n^0$ be the linear subspace spanned by elements of the form $p(W(h_1), \cdots, W(h_r))$, where $r \geqslant 1$, $h_1, \cdots, h_r \in H$ and p is a polynomial in r variables whose degree is at most n. Let $\mathcal{P}_n$ be the L^2-closure of $\mathcal{P}_n^0$ and also define $\mathcal{C}_n \triangleq \oplus_{k=0}^n \mathcal{H}_k$.

Proposition 4.1. *For each $n \geqslant 0$ we have $\mathcal{C}_n = \mathcal{P}_n$.*

Proof. It is clear that $\mathcal{C}_n \subseteq \mathcal{P}_n$, since

$$H_k(W(h)) \in \mathrm{Span}(1, W(h), \cdots, W(h)^k) \subseteq \mathcal{P}_n^0, \quad \forall\, k \leqslant n.$$

For the other direction, according to Theorem 4.1, it is enough to show that $\mathcal{P}_n^0$ is perpendicular to all those $\mathcal{H}_m$'s with $m > n$. More specifically, let $p(W(h_1), \cdots, W(h_r)) \in \mathcal{P}_n^0$ and $h \in H$ with $\|h\|_H = 1$. We wish to see that

$$p(W(h_1), \cdots, W(h_r)) \perp H_m(W(h)) \tag{4.2}$$

where $m > n$. To this end, let $\{h, e_1, \cdots, e_s\}$ be the orthonormal system obtained from the family $\{h, h_1, \cdots, h_r\}$ by applying the Gram-Schmidt orthogonalisation procedure. By the linearity of W, we can write

$$p(W(h_1), \cdots, W(h_r)) = q(W(h), W(e_1), \cdots, W(e_s)).$$

where q is a polynomial of degree at most n. Each monomial on the right hand side has the form

$$W(h)^a W(e_1)^{a_1} \cdots W(e_s)^{a_s}, \quad a \leqslant n.$$

Since $\{W(h), W(e_1), \cdots, W(e_s)\}$ are independent and $W(h)^a \perp H_m(W(h))$ $(a \leqslant n < m)$, we see that

$$\begin{aligned} &\mathbb{E}[W(h)^a W(e_1)^{a_1} \cdots W(e_s)^{a_s} H_m(W(h))] \\ &= \mathbb{E}[W(h)^a H_m(W(h))] \cdot \mathbb{E}[W(e_1)^{a_1}] \cdots \mathbb{E}[W(e_s)^{a_s}] = 0. \end{aligned}$$

The property (4.2) thus follows. $\square$

Our next task is to identify an ONB for each $\mathcal{H}_n$ (and thus for $L^2(\Omega, \mathcal{F}, \mathbb{P})$). Let us begin by fixing an ONB $\{e_1, e_2, \cdots\}$ of H which exists due to separability. We define Λ to be the set of all sequences $\mathbf{a} = (a_1, a_2, \cdots)$ in which $a_i \in \mathbb{N}$ and there are at most finitely many non-zero components. For each $\mathbf{a} = (a_1, a_2, \cdots) \in \Lambda$, we set

$$|\mathbf{a}| \triangleq \sum_{i=1}^\infty a_i, \;\; \mathbf{a}! \triangleq \prod_{i=1}^\infty a_i!,$$

and define

$$\Phi_{\mathbf{a}} \triangleq \sqrt{\mathbf{a}!} \prod_{i=1}^\infty H_{a_i}(W(e_i)).$$

We also set

$$\Lambda_n \triangleq \{\mathbf{a} \in \Lambda : |\mathbf{a}| = n\}.$$

Theorem 4.2. *For each $n \geqslant 1$, $\{\Phi_{\mathbf{a}} : \mathbf{a} \in \Lambda_n\}$ is an ONB of $\mathcal{H}_n$. As a consequence, $\{\Phi_{\mathbf{a}} : \mathbf{a} \in \Lambda\}$ is an ONB of $L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

To prove the theorem, we need the following technical lemma.

Lemma 4.1. *Suppose that $h_n^i \to h^i$ in H as $n \to \infty$ ($1 \leqslant i \leqslant r$). Let p be a polynomial in r variables. Then for any $\alpha \geqslant 1$, we have*

$$p(W(h_n^1), \cdots, W(h_n^r)) \to p(W(h^1), \cdots, W(h^r)) \quad \text{in } L^\alpha(\Omega, \mathcal{F}, \mathbb{P})$$

as $n \to \infty$.

Proof. We can express

$$\begin{aligned} & p(W(h_n^1), \cdots, W(h_n^r)) - p(W(h^1), \cdots, W(h^r)) \\ & = \sum_{j=1}^r \Big(\int_0^1 \partial_{x_j} p\big((1-t)\xi_n + t\xi\big) dt\Big) \cdot W(h_n^j - h^j), \end{aligned}$$

where

$$\xi_n \triangleq (W(h_n^1), \cdots, W(h_n^r)), \ \xi \triangleq (W(h^1), \cdots, W(h^r)).$$

Since $\{h_n^i : n \geqslant 1\}$ are bounded in H, it is not hard to see that

$$\Big\| \int_0^1 \partial_{x_j} p\big((1-t)\xi_n + t\xi\big) dt \Big\|_{L^{2\alpha}} \leqslant C \quad \forall n \geqslant 1 \text{ and } 1 \leqslant j \leqslant r,$$

where C is a constant depending on the polynomial p and $\sup_{n,j} \|h_n^j\|_H$. By using Hölder's inequality, we have

$$\begin{aligned} & \|p(W(h_n^1), \cdots, W(h_n^r)) - p(W(h^1), \cdots, W(h^r))\|_{L^\alpha} \\ & \leqslant \sum_{j=1}^r \Big\| \int_0^1 \partial_{x_j} p\big((1-t)\xi_n + t\xi\big) dt \Big\|_{L^{2\alpha}} \cdot \|W(h_n^j - h^j)\|_{L^{2\alpha}} \\ & \leqslant C' \cdot \sup_{1 \leqslant j \leqslant r} \|W(h_n^j - h^j)\|_{L^{2\alpha}} \leqslant C'' \sup_{1 \leqslant j \leqslant r} \|h_n^j - h^j\|_H, \end{aligned}$$

which converges to zero as $n \to \infty$ by the assumption. □

Proof of Theorem 4.2. According to Lemma 2.3, we have

$$\mathbb{E}[\Phi_{\mathbf{a}} \Phi_{\mathbf{b}}] = \sqrt{\mathbf{a}!\mathbf{b}!} \prod_{i=1}^\infty \mathbb{E}[H_{a_i}(W(e_i)) H_{b_i}(W(e_i))] = \begin{cases} 1, & \text{if } \mathbf{a} = \mathbf{b}, \\ 0, & \text{if } \mathbf{a} \neq \mathbf{b}. \end{cases}$$

This shows that the system $\{\Phi_{\mathbf{a}} : \mathbf{a} \in \Lambda\}$ is orthonormal. In addition, if we define $\mathcal{L}_n$ to be the L^2-closure of $\text{Span}\{\Phi_{\mathbf{a}} : \mathbf{a} \in \Lambda_n\}$, then $\mathcal{L}_m \perp \mathcal{L}_n$ whenever $m \neq n$. It remains to show that $\mathcal{L}_n = \mathcal{H}_n$.

For this purpose, we first show that

$$\mathcal{P}_n = \mathcal{L}^{(n)} \triangleq \bigoplus_{k=0}^n \mathcal{L}_k. \quad (\mathcal{L}_0 \triangleq \mathbb{R})$$

The fact that $\mathcal{L}^{(n)} \subseteq \mathcal{P}_n$ is obvious. For the other direction, let $p(W(h_1), \cdots, W(h_r)) \in \mathcal{P}_n^0$. For each $1 \leqslant j \leqslant r$, we can find

$$h_n^j \to h^j \quad \text{in } H$$

where h_n^j is a linear combination of the basis elements $e_1, e_2, \cdots$. According to Lemma 4.1, we know that

$$p(W(h_n^1), \cdots, W(h_n^r)) \to p(W(h^1), \cdots, W(h^r)) \quad \text{in } L^2(\Omega, \mathcal{F}, \mathbb{P}).$$

On other other hand, since

$$h_n^j \in \mathrm{Span}\{e_1, e_2, \cdots\},$$

we see that $p(W(h_n^1), \cdots, W(h_n^r))$ is a linear combination of monomials of the form

$$W(e_{i_1})^{a_1} \cdots W(e_{i_s})^{a_s}$$

where $a_1 + \cdots + a_s \leqslant n$. Since

$$W(e_{i_l})^{a_l} \in \mathrm{Span}\{1, H_1(W(e_{i_l})), \cdots, H_{a_l}(W(e_{i_l}))\},$$

we see that $W(e_{i_1})^{a_1} \cdots W(e_{i_s})^{a_s} \in \mathcal{L}^{(n)}$. Consequently, $\mathcal{P}_n \subseteq \mathcal{L}^{(n)}$.

Combining with Proposition 4.1, we have shown that $\mathcal{L}^{(n)} = \mathcal{P}_n = \mathcal{C}_n$. To prove $\mathcal{L}_n = \mathcal{H}_n$, let $X \in \mathcal{L}_n$. Since $X \in \mathcal{C}_n$, we can write

$$X = Y + Z, \quad Y \in \mathcal{C}_{n-1}, Z \in \mathcal{H}_n.$$

By taking inner product with Y, we have

$$\langle X, Y \rangle_{L^2} = \langle Y, Y \rangle_{L^2} + \langle Z, Y \rangle_{L^2} = \langle Y, Y \rangle_{L^2}.$$

On the other hand, since $Y \in \mathcal{C}_{n-1} = \mathcal{L}^{(n-1)}$ and $X \in \mathcal{L}_n$, we know that $\langle X, Y \rangle_{L^2} = 0$. It follows that $\langle Y, Y \rangle_{L^2} = 0$ (i.e. $Y = 0$) and thus $X = Z \in \mathcal{H}_n$. This shows $\mathcal{L}_n \subseteq \mathcal{H}_n$. The other direction is obtained in a similar way. Therefore, we conclude that $\mathcal{L}_n = \mathcal{H}_n$, which also completes the proof.

□

Example 4.1. Consider $\Omega = \mathbb{R}^d$ equipped with the standard d-dimensional Gaussian measure $\mathbb{P}$ on the Borel σ-algebra $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$. In this case, we have $H = \mathbb{R}^d$ and the isonormal Gaussian family is defined by

$$W(h) : \mathbb{R}^d \to \mathbb{R}, \; W(h)(x) = \langle h, x \rangle_{\mathbb{R}^d}$$

for each $h \in H$. Let $\{e_1, \cdots, e_d\}$ be the canonical ONB of H. The family

$$\Big\{ \frac{1}{\sqrt{a_1! \cdot \cdots \cdot a_d!}} H_{a_1}(W(e_1)) \cdots H_{a_d}(W(e_d)) : a_1 + \cdots + a_d = n \Big\}$$

is an ONB of $\mathcal{H}_n$.

Remark 4.1. The spectral interpretation of the Wiener-Itô decomposition is still valid in the infinite dimensional case. If we consider the Wiener space $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$, the decomposition (3.1) is precisely the spectral decomposition for the generator $\mathcal{L}$ of a $\mathcal{W}$-valued Markov process (the Ornstein-Uhlenbeck process) whose invariant measure is μ. For each $n \geqslant 0$, the space $\mathcal{H}_n$ is the eigenspace of $\mathcal{L}$ associated with the eigenvalue $-n$. We refer the reader to [5] for a deeper discussion along this line.

5 Multiple Wiener integrals and their connection with Wiener chaoses

In the case of the Wiener space, the n-th Wiener chaos has an elegant connection with multiple Wiener integrals. Recall that the Cameron-Martin subspace H is isometrically isomorphic to $L^2([0,1], dt)$ and hence can be identified with the latter. This L^2-nature of H is the key structure that supports the connection with multiple Wiener integrals.

To filter out other irrelevant structures, we consider a general irreducible Gaussian probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{W(h) : h \in H\})$ where the Hilbert space H is given by an L^2-space. We need the following definition to make precise our assumption on H.

Definition 5.1. Let $(T, \mathcal{B}, \mu)$ be a measure space. A subset $B \in \mathcal{F}$ is called an *atom* of μ if $\mu(B) > 0$ and

$$A \in \mathcal{B},\ A \subseteq B \implies \mu(A) = 0 \text{ or } \mu(A) = \mu(B).$$

A measure space $(T, \mathcal{B}, \mu)$ is said to be *atomless* if there are no atoms of μ.

From now on, unless otherwise stated we always assume that $H = L^2(T, \mathcal{B}, \mu)$, where $(T, \mathcal{B}, \mu)$ is a σ-finite atomless measure space and H is separable. The atomless assumption plays a critical role in this part (cf. Section 5.1.4 below). In the case of the Wiener space, we have $T = [0,1]$, $\mathcal{B} = \mathcal{B}([0,1])$ and $\mu = dt$. The situation of σ-finite measures is relevant when we consider $T = [0, \infty)$. It can be shown that the Lebesgue measure is atomless.

5.1 Construction of multiple Wiener integrals

We first construct the multiple Wiener integrals. More specifically, for each $n \geqslant 1$ we wish to define

$$\int_{T^n} f(t_1, \cdots, t_n) dW_{t_1} \cdots dW_{t_n} \tag{5.1}$$

where $f \in L^2(T^n, \mathcal{B}^n, \mu^n)$ and $(T^n, \mathcal{B}^n, \mu^n)$ is the n-th product space of $(T, \mathcal{B}, \mu)$. The underlying idea of the construction is similar to the case of the Wiener integral: we first write down a natural definition of (5.1) for a class of "elementary" functions f, and then rely on a suitable isometry property to pass to the limit. Here a crucial effort is to identify what this class of "elementary" functions should be.

We begin with some basic definitions. A function $f(t_1, \cdots, t_n)$ is said to be *symmetric* if

$$f(t_{\sigma(1)}, \cdots, t_{\sigma(n)}) = f(t_1, \cdots, t_n) \quad \forall \sigma \in \mathcal{S}_n,$$

where $\mathcal{S}_n$ denotes the set of permutations of order n. Given an arbitrary function $f(t_1, \cdots, t_n)$, its *symmetrisation* is defined by

$$\tilde{f}(t_1, \cdots, t_n) \triangleq \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} f(t_{\sigma(1)}, \cdots, t_{\sigma(n)}). \tag{5.2}$$

It is clear that $\tilde{f}$ is symmetric, and a function f is symmetric if and only if $f = \tilde{f}$. We use $L^2_S(T^n, \mathcal{B}^n, \mu^n)$ to denote the subspace of symmetric functions in $L^2(T^n, \mathcal{B}^n, \mu^n)$. Mathematically, $L^2_S(T^n, \mathcal{B}^n, \mu^n)$ is the image of the symmetrisation operator defined by (5.2) on $L^2(T^n, \mathcal{B}^n, \mu^n)$.

Next, we set

$$\mathcal{B}_0 \triangleq \{A \in \mathcal{B} : \mu(A) < \infty\}.$$

For each $A \in \mathcal{B}_0$, we write $W(A) \triangleq W(\mathbf{1}_A)$ (note that $\mathbf{1}_A \in H$). It is obvious that $W(A) \sim N(0, \mu(A))$. Given $n \geqslant 1$, we define $\mathcal{E}_n$ to be the class of functions $f \in L^2(T^n, \mathcal{B}^n, \mu^n)$ having the form

$$f(t_1, \cdots, t_n) = \sum_{i_1, \cdots, i_n = 1}^{m} a_{i_1, \cdots, i_n} \mathbf{1}_{A_{i_1} \times \cdots \times A_{i_n}}, \tag{5.3}$$

where $\{A_1, \cdots, A_m\}$ is any given collection of disjoint subsets in $\mathcal{B}_0$, and the coefficient $a_{i_1, \cdots, i_n} = 0$ *if there are repeated indices* in $(i_1, \cdots, i_n)$. For each $f \in \mathcal{E}_n$ given by the form (5.3), we define its multiple Wiener integral to be

$$I_n(f) \triangleq \sum_{i_1, \cdots, i_n = 1}^{m} a_{i_1, \cdots, i_n} W(A_{i_1}) \cdot \cdots \cdot W(A_{i_n}). \tag{5.4}$$

The following result summarises the construction of the multiple Wiener integral and its essential properties.

Theorem 5.1 (Multiple Wiener integrals). *For each $n \geqslant 1$, the mapping $I_n : \mathcal{E}_n \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is well defined and extends to a unique bounded linear operator*

$$I_n : L^2(T^n, \mathcal{B}^n, \mu) \to L^2(\Omega, \mathcal{F}, \mathbb{P}).$$

In addition, the following properties hold true:

(i) For any $f \in L^2(T^n, \mathcal{B}^n, \mu^n)$, we have

$$I_n(f) = I_n(\tilde{f}),$$

where $\tilde{f}$ is the symmetrisation of f defined by (5.2).
(ii) For any $f \in L^2(T^p, \mathcal{B}^p, \mu^p)$ and $g \in L^2(T^q, \mathcal{B}^q, \mu^q)$, we have

$$\mathbb{E}[I_p(f) I_q(g)] = \begin{cases} p! \langle \tilde{f}, \tilde{g} \rangle_{L^2}, & p = q, \\ 0, & \text{otherwise}. \end{cases} \tag{5.5}$$

In particular, if we introduce an inner product on $L^2_S(T^n, \mathcal{B}^n, \mu^n)$ by

$$\langle f, g \rangle_{L^2_S} \triangleq n! \langle f, g \rangle_{L^2}, \quad f, g \in L^2_S(T^n, \mathcal{B}^n, \mu^n),$$

then under this inner product I_n is an isometric embedding from $L^2_S(T^n, \mathcal{B}^n, \mu^n)$ into $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Remark 5.1. Symbolically we also use

$$\int_{T^n} f(t_1, \cdots, t_n) dW_{t_1} \cdots dW_{t_n}$$

to denote the multiple Wiener integral $I_n(f)$. Note that I_1 is just the mapping W defined by (3.8) (the Wiener integral in the case of the Wiener space).

The essential point in the proof of this theorem is to establish the isometry property (5.5) for I_n on the space $\mathcal{E}_n$. It is then a consequence of the denseness of $\mathcal{E}_n$ that I_n admits a unique extension to $L^2(T^n, \mathcal{B}^n, \mu^n)$ that preserves the same property. The rest of this subsection is devoted to the proof of Theorem 5.1. Some of the steps are quite technical and less inspiring. Nonetheless, it is important to see how the off-diagonal assumption in the definition $\mathcal{E}_n$ (i.e. $a_{i_1,\cdots,i_n} = 0$ if $(i_1, \cdots, i_n)$ has repeated indices) plays a basic role when deriving the isometry property (5.5), while the subspace $\mathcal{E}_n$ is still rich enough to generate $L^2(T^n, \mathcal{B}^n, \mu^n)$. The proof of the theorem can be skipped if the reader is convinced by these points.

We break down the proof in the several major steps.

5.1.1 Step one: the mapping $I_n : \mathcal{E}_n \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is well defined

Lemma 5.1. *The definition of $I_n(f)$ given by (5.4) does not depend on the particular representation of f. In particular, $I_n : \mathcal{E}_n \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a well defined linear operator.*

Proof. Suppose that f admits two representations:

$$f = \sum_{i_1,\cdots,i_n=1}^{k} a_{i_1,\cdots,i_n} \mathbf{1}_{A_{i_1} \times \cdots \times A_{i_n}} = \sum_{j_1,\cdots,j_n=1}^{l} b_{j_1,\cdots,j_n} \mathbf{1}_{B_{j_1} \times \cdots \times B_{j_n}}, \tag{5.6}$$

where $\{A_1, \cdots, A_k\}$ and $\{B_1, \cdots, B_l\}$ are both disjoint families in $\mathcal{B}_0$. We may assume without loss of generality that none of the A_i's are equal to $\emptyset$, and each A_i does appear on some term in the first summation with $a_{i_1,\cdots,i_n} \neq 0$. The assumption applies to the B_j's. Under this assumption, we claim that

$$\cup_{i=1}^{k} A_i = \cup_{j=1}^{l} B_j. \tag{5.7}$$

Indeed, given $1 \leqslant i \leqslant k$, suppose that A_i appears within a term

$$a_{i_1,\cdots,i,\cdots,i_n} \mathbf{1}_{A_{i_1} \times \cdots \times A_i \times \cdots \times A_{i_n}}, \quad a_{i_1,\cdots,i,\cdots,i_n} \neq 0.$$

Note that the indicator sets appearing in the representation of f are disjoint. As a result, if we pick an arbitrary

$$\mathbf{t} = (t_1, \cdots, t, \cdots, t_n) \in A_{i_1} \times \cdots \times A_i \times \cdots \times A_{i_n},$$

then $f(\mathbf{t}) \neq 0$. Using the second representation in (5.6), this implies that $\mathbf{t} \in B_{j_1} \times \cdots \times B_{j_n}$ for some $(j_1, \cdots, j_n)$. In particular, $t \in B_j$ for some j, and thus $A_i \subseteq \cup_{j=1}^{l} B_j$. The other direction is similar.

As a consequence of (5.7), we can now write

$$\begin{aligned} f &= \sum_{i_1,\cdots,i_n=1}^{k} \sum_{j_1,\cdots,j_n=1}^{l} a_{i_1,\cdots,i_n} \mathbf{1}_{C_{i_1 j_1} \times \cdots \times C_{i_n j_n}} \\ &= \sum_{i_1,\cdots,i_n=1}^{k} \sum_{j_1,\cdots,j_n=1}^{l} b_{j_1,\cdots,j_n} \mathbf{1}_{C_{i_1 j_1} \times \cdots \times C_{i_n j_n}}, \end{aligned} \tag{5.8}$$

where $C_{i_rj_r} \triangleq A_{i_r} \cap B_{j_r}$. In particular, whenever $(i_1, \cdots, i_n)$ and $(j_1, \cdots, j_n)$ are such that $C_{i_1j_1}, \cdots, C_{i_nj_n}$ are all non-empty, we must have

$$a_{i_1,\cdots,i_n} = b_{j_1,\cdots,j_n}.$$

For those indices where one of $C_{i_rj_r} = \emptyset$ we can simply drop this term in the expansion of (5.8). As a result, we have

$$\begin{aligned}
\sum_{i_1,\cdots,i_n=1}^{k} a_{i_1,\cdots,i_n} W(A_{i_1}) \cdots W(A_{i_n}) &= \sum_{i_1,\cdots,i_n=1}^{k} \sum_{j_1,\cdots,j_n=1}^{l} a_{i_1,\cdots,i_n} W(C_{i_1j_1}) \cdots W(C_{i_nj_n}) \\
&= \sum_{i_1,\cdots,i_n=1}^{k} \sum_{j_1,\cdots,j_n=1}^{l} b_{j_1,\cdots,j_n} W(C_{i_1j_1}) \cdots W(C_{i_nj_n}) \\
&= \sum_{j_1,\cdots,j_n=1}^{l} b_{j_1,\cdots,j_n} W(B_{j_1}) \cdots W(B_{j_n}),
\end{aligned}$$

which shows that $I_n(f)$ is well defined. It is routine to check that $\mathcal{E}_n$ is a vector space, and the linearity of I_n is also obvious. $\square$

5.1.2 Step two: $I_n(f) = I_n(\tilde{f})$

Lemma 5.2. *For any $f \in \mathcal{E}_n$, we have $I_n(f) = I_n(\tilde{f})$ where $\tilde{f}$ is the symmetrisation of f.*

Proof. It is enough to verity the assertion for

$$f = \mathbf{1}_{A_1 \times \cdots \times A_n},$$

where $A_1, \cdots, A_n \in \mathcal{B}_0$ are disjoint. In this case, we have

$$\tilde{f} = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \mathbf{1}_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}}$$

which is also an element of $\mathcal{E}_n$. By the definition of I_n, we have

$$I_n(\tilde{f}) = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} W(A_{\sigma(1)}) \cdots W(A_{\sigma(n)}) = W(A_1) \cdots W(A_n) = I_n(f).$$

$\square$

5.1.3 Step three: the isometry property of I_n

This part relies on the off-diagonal assumption in the definition of $\mathcal{E}_n$ in a crucial way.

Lemma 5.3. *For any $f \in \mathcal{E}_p$ and $g \in \mathcal{E}_q$, we have*

$$\mathbb{E}[I_p(f)I_q(g)] = \begin{cases} p!\langle \tilde{f}, \tilde{g} \rangle, & p = q, \\ 0, & \textit{otherwise}. \end{cases}$$

Proof. We may assume without loss of generality that f, g are associated with the same disjoint family $\{A_1, \cdots, A_m\}$.

Let us first consider $p \neq q$. By linearity, it suffices to look at the case when

$$f = \mathbf{1}_{A_{i_1}\times\cdots\times A_{i_p}},\ g = \mathbf{1}_{A_{j_1}\times\cdots\times A_{j_q}}.$$

In this case, we have

$$I_p(f) = W(A_{i_1})\cdots W(A_{i_p}),\ I_q(g) = W(A_{j_1})\cdots W(A_{j_q}).$$

Since the indices in $(i_1, \cdots, i_p)$ are all distinct (the same for $(j_1, \cdots, j_q)$) and $p \neq q$, we see that

$$\mathbb{E}[I_p(f)I_q(g)] = 0.$$

Now consider the case when $p = q$, and let

$$f = \sum_{i_1,\cdots,i_p=1}^{m} a_{i_1,\cdots,i_p}\mathbf{1}_{A_{i_1}\times\cdots\times A_{i_p}},\ g = \sum_{i_1,\cdots,i_p=1}^{m} b_{i_1,\cdots,i_p}\mathbf{1}_{A_{i_1}\times\cdots\times A_{i_p}}.$$

By definition, we have

$$\begin{aligned} I_p(f) &= \sum_{i_1,\cdots,i_p=1}^{m} a_{i_1,\cdots,i_p} W(A_{i_1})\cdots W(A_{i_p}) \\ &= \sum_{1\leqslant i_1<\cdots<i_p\leqslant m} \big(\sum_{\sigma\in\mathcal{S}_p} a_{i_{\sigma(1)},\cdots,i_{\sigma(p)}}\big) W(A_{i_1})\cdots W(A_{i_p}) \\ &= \sum_{1\leqslant i_1<\cdots<i_p\leqslant m} \tilde{a}_{i_1,\cdots,i_p} W(A_{i_1})\cdots W(A_{i_p}), \end{aligned}$$

where

$$\tilde{a}_{i_1,\cdots,i_p} \triangleq \sum_{\sigma\in\mathcal{S}_p} a_{i_{\sigma(1)},\cdots,i_{\sigma(p)}}.$$

Similar equation holds for $I_p(g)$. It follows that

$$\begin{aligned} &\mathbb{E}[I_p(f)I_p(g)] \\ &= \mathbb{E}\big[\big(\sum_{1\leqslant i_1<\cdots<i_p\leqslant m} \tilde{a}_{i_1,\cdots,i_p} W(A_{i_1})\cdots W(A_{i_p})\big)\big(\sum_{1\leqslant j_1<\cdots<j_p\leqslant m} \tilde{b}_{j_1,\cdots,j_p} W(A_{j_1})\cdots W(A_{j_p})\big)\big] \\ &= \sum_{1\leqslant i_1<\cdots<i_p\leqslant m} \tilde{a}_{i_1,\cdots,i_p}\tilde{b}_{i_1,\cdots,i_p}\mathbb{E}[W(A_{i_1})^2]\cdots\mathbb{E}[W(A_{i_p})^2] \\ &= \sum_{1\leqslant i_1<\cdots<i_p\leqslant m} \tilde{a}_{i_1,\cdots,i_p}\tilde{b}_{i_1,\cdots,i_p}\mu(A_{i_1})\cdots\mu(A_{i_p}). \end{aligned}$$

On the other hand, we have

$$\tilde{f} = \sum_{i_1,\cdots,i_p=1}^{m} a_{i_1,\cdots,i_p}\big(\frac{1}{p!}\sum_{\sigma\in\mathcal{S}_p}\mathbf{1}_{A_{i_{\sigma(1)}}\times\cdots\times A_{i_{\sigma(p)}}}\big).$$

Note the the expression inside the above bracket is independent of the ordering of $i_1, \cdots, i_p$. Therefore, we can further write

$$\tilde{f} = \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \tilde{a}_{i_1, \cdots, i_p} \big(\frac{1}{p!} \sum_{\sigma \in \mathcal{S}_p} \mathbf{1}_{A_{i_{\sigma(1)}} \times \cdots \times A_{i_{\sigma(p)}}} \big).$$

Similar equation holds for $\tilde{g}$. It follows that

$$\begin{aligned} \langle \tilde{f}, \tilde{g} \rangle_{L^2} &= \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \tilde{a}_{i_1, \cdots, i_p} \tilde{b}_{i_1, \cdots, i_p} \cdot \frac{1}{(p!)^2} \sum_{\sigma \in \mathcal{S}_p} \mu(A_{i_1}) \cdots \mu(A_{i_p}) \\ &= \frac{1}{p!} \sum_{1 \leqslant i_1 < \cdots < i_p \leqslant m} \tilde{a}_{i_1, \cdots, i_p} \tilde{b}_{i_1, \cdots, i_p} \mu(A_{i_1}) \cdots \mu(A_{i_p}) \\ &= \mathbb{E}[I_p(f) I_p(g)]. \end{aligned}$$

This gives the desired isometry property. □

Observe that, for any $f \in \mathcal{E}_n$ we have

$$\|\tilde{f}(t_1, \cdots, t_n)\|_{L^2} \leqslant \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \|f(t_{\sigma(1)}, \cdots, t_{\sigma(n)})\|_{L^2} = \|f\|_{L^2}.$$

As an immediate consequence of Lemma 5.3, we see that the linear operator $I_n : \mathcal{E}_n \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is continuous:

$$\|I_n(f)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \sqrt{n!} \cdot \|\tilde{f}\|_{L^2(T^n, \mathcal{B}^n, \mu^n)} \leqslant \sqrt{n!} \cdot \|f\|_{L^2(T^n, \mathcal{B}^n, \mu^n)}.$$

5.1.4 Step four: $\mathcal{E}_n$ is dense in $L^2(T^n, \mathcal{B}^n, \mu^n)$

In order to expect a unique extension of I_n to the space $L^2(T^n, \mathcal{B}^n, \mu^n)$, the last missing piece is the fact that $\mathcal{E}_n$ is dense in $L^2(T^n, \mathcal{B}^n, \mu^n)$. The non-diagonal assumption in $\mathcal{E}_n$ creates extra technical challenge to establish this fact. We develop the proof via the following route of approximations:

$$\mathcal{E}_n \rightsquigarrow \mathbf{1}_{A_1 \times \cdots \times A_n} \text{ (each } A_i \in \mathcal{B}_0) \rightsquigarrow \mathbf{1}_A \ (\mu^n(A) < \infty) \rightsquigarrow f \in L^2(T^n, \mathcal{B}^n, \mu^n).$$

Recall that $\mathcal{B}_0$ is the collections of measurable sets in $\mathcal{B}$ with finite μ-measure.

Lemma 5.4. *The subspace* $\mathrm{Span}\{\mathbf{1}_A : A \in \mathcal{B}_0\}$ *is dense in* $L^2(T, \mathcal{B}, \mu)$.

Proof. Since μ is σ-finite, we can find a sequence $T_n \in \mathcal{B}_0$ $(n \geqslant 1)$ such that $T_n \uparrow T$. Let $f \in L^2(T, \mathcal{B}, \mu)$. By monotone convergence, we have

$$f \mathbf{1}_{\{|f| \leqslant M\} \cap T_n} \to f \quad \text{in } L^2$$

as $M, n \to \infty$. As a result, given $\varepsilon > 0$, there exists M and n such that

$$\|f^M \mathbf{1}_{T_n} - f\|_{L^2} < \varepsilon,$$

where $f^M \triangleq f\mathbf{1}_{\{|f|\leqslant M\}}$. Next, since f^M is uniformly bounded, from measure theory we can find a simple function φ (i.e. linear combination of indicator functions) such that

$$\|f^M - \varphi\|_\infty < \frac{\varepsilon}{\mu(T_n)^{1/2}}.$$

Therefore, we have

$$\int_T |\varphi - f^M|^2 \mathbf{1}_{T_n} d\mu \leqslant \varepsilon^2 \text{ or } \|\varphi \mathbf{1}_{T_n} - f^M \mathbf{1}_{T_n}\|_{L^2} < \varepsilon.$$

The result then follows as

$$\varphi \mathbf{1}_{T_n} \in \text{Span}\{\mathbf{1}_A : A \in \mathcal{B}_0\}.$$

□

As a direct corollary, for each $n \geqslant 1$, the subspace

$$\text{Span}\{\mathbf{1}_E : E \in \mathcal{B}^n, \mu^n(E) < \infty\}$$

is dense in $L^2(T^n, \mathcal{B}^n, \mu^n)$. The next point of approximation is contained in the lemma below.

Lemma 5.5. *Let $E \in \mathcal{B}^n$ be such that $\mu^n(E) < \infty$. Then $\mathbf{1}_E$ can be approximated in the L^2-sense by linear combinations of indicator functions of the form $\mathbf{1}_{A_1 \times \cdots \times A_n}$ where $A_1, \cdots, A_n \in \mathcal{B}_0$.*

Proof. Recall from the construction of product measures that

$$\mu^n(E) = \inf\Big\{\sum_{m=1}^{\infty} \mu(A_{1,m}) \cdots \mu(A_{n,m}) : E \subseteq \cup_{m=1}^{\infty} A_{1,m} \times \cdots \times A_{n,m}\Big\}.$$

Since $\mu^n(E) < \infty$, given $\varepsilon > 0$, we find a covering $\cup_{m=1}^{\infty} A_{1,m} \times \cdots \times A_{n,m}$ of E such that

$$\Big|\sum_{m=1}^{\infty} \mu^n(A_{1,m} \times \cdots \times A_{n,m}) - \mu^n(E)\Big| < \varepsilon^2.$$

In particular, $\cup_{m=1}^{\infty} A_{1,m} \times \cdots \times A_{n,m}$ has finite μ^n-measure and

$$\big\|\mathbf{1}_{\cup_{m=1}^{\infty} A_{1,m} \times \cdots \times A_{n,m}} - \mathbf{1}_E\big\|_{L^2} < \varepsilon.$$

When N is large enough, we have

$$\big\|\mathbf{1}_{\cup_{m=1}^{N} A_{1,m} \times \cdots \times A_{n,m}} - \mathbf{1}_E\big\|_{L^2} < \varepsilon. \tag{5.9}$$

By intersecting with T_l (for large l) if necessary (where $T_l \uparrow T$, $\mu(T_l) < \infty$), we may further assume that

$$A_{i,m} \in \mathcal{B}_0 \quad \forall i = 1, \cdots, n, \ m \geqslant 1,$$

and (5.9) remains true. The result then follows by observing that

$$\mathbf{1}_{\cup_{m=1}^{N} A_{1,m} \times \cdots \times A_{n,m}} \in \text{Span}\big\{\mathbf{1}_{B_1 \times \cdots \times B_n} : B_1, \cdots, B_n \in \mathcal{B}_0\big\}.$$

□

As the final point of approximation, it remains to prove the following lemma . The atomless asumption on μ plays a critical role here (the key property is Proposition A.1 in the appendix).

Lemma 5.6. *Let $A_1, \cdots, A_n \in \mathcal{B}_0$. Then $\mathbf{1}_{A_1\times\cdots\times A_n}$ can be approximated by elements in $\mathcal{E}_n$ in the sense of L^2.*

Proof. We first identify a disjoint $\mathcal{A}$ family of subsets so that each A_i is the union of some members in $\mathcal{A}$. To this end, let

$$\Xi \triangleq \{\omega = (\omega_1, \cdots, \omega_n) : \omega_i = \pm 1 \ \forall i \text{ and at least one of } \omega_i = 1\}.$$

For each $\omega \in \Xi$, we define

$$A^\omega \triangleq A_1^{\omega_1} \cap \cdots \cap A_n^{\omega_n},$$

where $A_i^{-1} \triangleq A_i^c$. It is clear that $\mathcal{A} \triangleq \{A^\omega : \omega \in \Xi\}$ is a disjoint family of subsets in $\mathcal{B}_0$, and

$$A_i = \cup_{\omega\in\Xi,\omega_i=1} A^\omega \tag{5.10}$$

for each $1 \leqslant i \leqslant n$.

Next, let $\varepsilon > 0$ be given. According to Proposition A.1, for each $\omega \in \Xi$ we can find a partition of A^ω in which every member has μ-measure less than ε. Since the A^ω's are disjoint for different ω's, this leads to an entire family $\{B_1, \cdots, B_m\}$ of disjoint subsets such that $\mu(B_i) < \varepsilon$ for all i and each A_i is the union of some members from this family. As a result, we can formally express

$$\begin{aligned} \mathbf{1}_{A_1\times\cdots\times A_n} &= \sum_{i_1,\cdots,i_n=1}^{m} \omega_{i_1,\cdots,i_n} \mathbf{1}_{B_{i_1}\times\cdots\times B_{i_n}} \quad (\omega_{i_1,\cdots,i_n} = 0 \text{ or } 1) \\ &= \sum_{(i_1,\cdots,i_n)\in\mathcal{I}} \omega_{i_1,\cdots,i_n} \mathbf{1}_{B_{i_1}\times\cdots\times B_{i_n}} + \sum_{(i_1,\cdots,i_n)\in\mathcal{J}} \omega_{i_1,\cdots,i_n} \mathbf{1}_{B_{i_1}\times\cdots\times B_{i_n}}, \end{aligned} \tag{5.11}$$

where $\mathcal{I}$ consists of those n-tuples $(i_1, \cdots, i_n)$ in which there are no repeated indices, and $\mathcal{J} \triangleq \mathcal{I}^c$. Note that the first summation on the right hand side is an element in $\mathcal{E}_n$.

Let us estimate the second summation in (5.11). Recall that $\mathcal{J}$ consists of those n-tuples $(i_1, \cdots, i_n)$ with repeated indices. The main observation is

$$\sum_{(i_1,\cdots,i_n)\in\mathcal{J}} \mu(B_{i_1}) \cdots \mu(B_{i_n}) \leqslant \binom{n}{2} \cdot \big(\sum_{i=1}^{m} \mu(B_i)^2\big) \cdot \big(\sum_{i=1}^{m} \mu(B_i)\big)^{n-2}, \tag{5.12}$$

for the obvious combinatorial reason that each term on the left and side appears at least once in the full expansion of the right hand side, while the latter involves repeated counting. The right hand side of (5.12) can further be bounded above by

$$\varepsilon \cdot \binom{n}{2} \cdot \big(\sum_{i=1}^{m} \mu(B_i)\big)^{n-1} = \varepsilon \cdot \binom{n}{2} \cdot \mu\big(\cup_{i=1}^m B_i\big)^{n-1} = \varepsilon \cdot \binom{n}{2} \cdot \mu\big(\cup_{i=1}^n A_i\big)^{n-1},$$

The last identity follows from the fact that

$$\cup_{i=1}^m B_i = \cup_{i=1}^n A_i,$$

which is clear from the construction of the B_i's. As a result,

$$\sum_{(i_1,\cdots,i_n)\in\mathcal{J}} \omega_{i_1,\cdots,i_n}\mu(B_{i_1})\cdots\mu(B_{i_n}) \leqslant \varepsilon\cdot\binom{n}{2}\cdot\mu\big(\cup_{i=1}^n A_i\big)^{n-1}.$$

Since n and the A_i's are given fixed and ε is arbitrary, from (5.11) we conclude that $\mathbf{1}_{A_1\times\cdots\times A_n}$ can be approximated by elements in $\mathcal{E}_n$ in the sense of L^2. □

To summarise all the previous steps, we have now shown that $\mathcal{E}_n$ is dense in $L^2(T^n,\mathcal{B}^n,\mu^n)$ and $I_n:\mathcal{E}_n\to L^2(\Omega,\mathcal{F},\mathbb{P})$ is a bounded linear operator. Therefore, it admits a unique bounded linear extension to $L^2(T^n,\mathcal{B}^n,\mu^n)$ satisfying the same isometry property (5.3). In particular, if we define a new inner product on $L_S^2(T^n,\mathcal{B}^n,\mu^n)$ by

$$\langle\cdot,\cdot\rangle_{L_S^2(T^n,\mathcal{B}^n,\mu^n)} \triangleq n!\langle\cdot,\cdot\rangle_{L^2(T^n,\mathcal{B}^n,\mu^n)}, \tag{5.13}$$

then the restriction of I_n on to $L_S^2(T^n,\mathcal{B}^n,\mu^n)$ is an isometric embedding into $L^2(\Omega,\mathcal{F},\mathbb{P})$.

Example 5.1. Consider the Wiener space over $[0,1]$ in which we have $H=L^2([0,1],dt)$. Given a function $f\in L^2([0,1]^n,(dt)^n)$, we define the *iterated Itô integral* of f by

$$\begin{aligned}
J_n(f) &\triangleq \int_{0<t_1<\cdots<t_n<1} f(t_1,\cdots,t_n)dW_{t_1}\cdots dW_{t_n} \\
&\triangleq \int_0^1\Big(\int_0^{t_n}\cdots\Big(\int_0^{t_3}\Big(\int_0^{t_2} f(t_1,\cdots,t_n)dW_{t_1}\Big)dW_{t_2}\Big)\cdots dW_{t_{n-1}}\Big)dW_{t_n},
\end{aligned} \tag{5.14}$$

where the each of above integrals is understood in the sense of Itô. We claim that for any symmetric $f\in L_S^2([0,1]^n,(dt)^n)$,

$$I_n(f)=n!J_n(f). \tag{5.15}$$

To sketch the proof of this fact, we first consider the case when f is the symmetrisation of

$$\mathbf{1}_{[s_1',s_1'']\times\cdots\times[s_n',s_n'']}$$

where $s_i''\leqslant s_{i+1}'$ for each i. In this case,

$$I_n(f)=W([s_1',s_1''])\cdots W([s_n',s_n''])=\big(W_{s_n''}-W_{s_n'}\big)\cdots\big(W_{s_1''}-W_{s_1'}\big).$$

On the other hand, since

$$f=\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\mathbf{1}_{[s'_{\sigma(1)},s''_{\sigma(1)}]\times\cdots\times[s'_{\sigma(n)},s''_{\sigma(n)}]},$$

Due to the special ordering in (5.14), after applying J_n the only surviving term is the one corresponding to $\sigma=\mathrm{id}$. As a result, we have

$$J_n(f)=\frac{1}{n!}J_n(\mathbf{1}_{[s_1',s_1'']\times\cdots\times[s_n',s_n'']})=\frac{1}{n!}\big(W_{s_n''}-W_{s_n'}\big)\cdots\big(W_{s_1''}-W_{s_1'}\big)=\frac{1}{n!}I_n(f).$$

For the general case, one needs to show (in a similar spirit to Section 5.1.4) that the linear subspace generated by those f's of the above form is dense in $L_S^2([0,1]^n,(dt)^n)$, and the iterated Itô integral J_n also possesses a similar isometry property. This allows us to pass to the limit to obtain (5.15). We let the reader to think about the technical details.

5.2 The connection between multiple wiener integrals and Wiener chaoses

Finally, we establish the important result that I_n defines an isometric isomorphism between $L_S^2(T^n, \mathcal{B}^n, \mu^n)$ (equipped with the inner product (5.13)) and the n-th Wiener chaos $\mathcal{H}_n$. The core of its proof relies on a product formula for multiple Wiener integrals which we discuss in what follows. This product formula is also of independent interest.

5.2.1 The product structure for multiple Wiener integrals

We show that the product $I_p(f) \cdot I_q(g)$ of two multiple Wiener integrals can be expressed as a *linear combination* of multiple Wiener integrals of degrees up to $p+q$.

To this end, we first introduce a few notation. We again work in the setting of Section 5.1. In particular, $H = L^2(T, \mathcal{B}, \mu)$ where μ is a σ-finite atomless measure. For simplicity, we denote $L^2(T^n, \mathcal{B}^n, \mu^n)$ (respectively, $L_S^2(T^n, \mathcal{B}^n, \mu^n)$) by $L^2(T^n)$ (respectively, $L_S^2(T^n)$).

Definition 5.2. Let $f \in L^2(T^p)$ and $g \in L^2(T^q)$. The *tensor product* of f and g is a function of $p+q$ variables defined by

$$f \otimes g(t_1, \cdots, t_p, t_{p+1}, \cdots, t_{p+q}) \triangleq f(t_1, \cdots, t_p) \cdot g(t_{p+1}, \cdots, t_{p+q}).$$

For $1 \leqslant r \leqslant p \wedge q \triangleq \min\{p, q\}$, the *$r$-th contraction* of f and g is a function of $p+q-2r$ variables defined by

$$\begin{aligned} & f \otimes_r g(t_1, \cdots, t_{p-r}, t_{p+1}, \cdots, t_{p+q-r}) \\ & \triangleq \int_{T^r} f(t_1, \cdots, t_{p-r}, \mathbf{s}) g(t_{p+1}, \cdots, t_{p+q-r}, \mathbf{s}) \mu^r(d\mathbf{s}). \end{aligned}$$

For convenience we also set $f \otimes_0 g \triangleq f \otimes g$. We use $f \tilde{\otimes} g$ and $f \tilde{\otimes}_r g$ to denote their symmetrisations respectively.

We leave it as an exercise to show that

$$\|f \otimes g\|_{L^2(T^{p+q})} = \|f\|_{L^2(T^p)} \cdot \|g\|_{L^2(T^q)}$$

and

$$\|f \otimes_r g\|_{L^2(T^{p+q-2r})} \leqslant \|f\|_{L^2(T^p)} \cdot \|g\|_{L^2(T^q)}.$$

In particular, both of $f \otimes g$ and $f \otimes_r g$ are square integrable.

The following result is an important step towards the general product formula.

Proposition 5.1. *Let $f \in L_S^2(T^p)$ and $g \in L^2(T)$. Then*

$$I_p(f) \cdot I_1(g) = I_{p+1}(f \otimes g) + p I_{p-1}(f \otimes_1 g). \tag{5.16}$$

Proof. By linearity, we may assume that $f = \tilde{\mathbf{1}}_{C_1 \times \cdots \times C_p}$ and $g = \mathbf{1}_D$ where $C_1, \cdots, C_p \in \mathcal{B}_0$ are disjoint and $D \in \mathcal{B}_0$. By writing

$$C_i = (D \cap C_i) \cup (D^c \cap C_i),$$

$$D = \cup_{i=1}^p (D \cap C_i) \cup \big(D \cap (C_1 \cup \cdots \cup C_p)^c\big),$$

and using linearity again, we may further assume without loss of generality that

$$f = \tilde{\mathbf{1}}_{A_1 \times \cdots \times A_p},\ g = \mathbf{1}_{A_0} \text{ or } \mathbf{1}_{A_1},$$

where $A_0, A_1, \cdots, A_p \in \mathcal{B}_0$ are disjoint.

Case 1: $g = \mathbf{1}_{A_0}$.

We have

$$I_p(f) \cdot I_1(g) = W(A_0)W(A_1)\cdots W(A_p).$$

On the other hand, we also have

$$I_{p+1}(f \otimes g) = I_{p+1}\big(f \tilde{\otimes} g\big) = I_{p+1}\big(\tilde{\mathbf{1}}_{A_0 \times A_1 \times \cdots \times A_p}\big) = W(A_0)W(A_1)\cdots W(A_p),$$

and

$$f \otimes_1 g(t_1, \cdots, t_{p-1}) = \int_T f(t_1, \cdots, t_{p-1}, s) g(s) \mu(ds) = 0,$$

since A_0 is disjoint from $\{A_1, \cdots, A_p\}$. The equation (5.16) thus follows.

Case 2: $g = \mathbf{1}_{A_1}$.

Given $\varepsilon > 0$, let $B_1, \cdots, B_m$ be a partition of A_1 such that $\mu(B_i) < \varepsilon$ for each i (cf. Proposition A.1). Then we can write

$$\begin{aligned} I_p(f)I_1(g) &= W(A_1)^2 W(A_2)\cdots W(A_p) = \big(\sum_{i=1}^m W(B_i)\big)^2 W(A_2)\cdots W(A_p) \\ &= \sum_{i \neq j} W(B_i)W(B_j)W(A_2)\cdots W(A_p) + \sum_{i=1}^m \big(W(B_i)^2 - \mu(B_i)\big) W(A_2)\cdots W(A_p) \\ &\quad + \mu(A_1)W(A_2)\cdots W(A_p). \end{aligned}$$

We now analyse the three terms on the right hand side separately.

The first term is a multiple Wiener integral given by $I_{p+1}(h_\varepsilon)$ where

$$h_\varepsilon \triangleq \sum_{i \neq j} \mathbf{1}_{B_i \times B_j \times A_2 \times \cdots \times A_p}.$$

We claim that $\tilde{h}_\varepsilon \to f \tilde{\otimes} g$ in $L^2(T^{p+1})$ as $\varepsilon \to 0$. Indeed, this follows from

$$\begin{aligned} &\big\|\tilde{h}_\varepsilon - f \tilde{\otimes} g\big\|_{L^2(T^{p+1})}^2 \\ &= \big\|\tilde{h}_\varepsilon - \tilde{\mathbf{1}}_{A_1 \times A_1 \times A_2 \times \cdots \times A_p}\big\|_{L^2(T^{p+1})}^2 \leqslant \big\|h_\varepsilon - \mathbf{1}_{A_1 \times A_1 \times A_2 \times \cdots \times A_p}\big\|_{L^2(T^{p+1})}^2 \\ &= \sum_{i=1}^m \mu(B_i)^2 \mu(A_2)\cdots\mu(A_p) \leqslant \varepsilon \mu(A_1)\cdots\mu(A_p). \end{aligned}$$

By the continuity of I_{p+1}, we have

$$I_{p+1}(h_\varepsilon) = I_{p+1}(\tilde{h}_\varepsilon) \to I_{p+1}(f \tilde{\otimes} g) = I_{p+1}(f \otimes g) \quad \text{in } L^2(\Omega, \mathcal{F}, \mathbb{P})$$

as $\varepsilon \to 0$.

We claim that the second term

$$R_\varepsilon \triangleq \sum_{i=1}^{m} \big(W(B_i)^2 - \mu(B_i)\big) W(A_2)\cdots W(A_p)$$

converges to 0 in L^2 as $\varepsilon \to 0$. This follows from

$$\begin{aligned}\mathbb{E}[R_\varepsilon^2] &= \Big(\sum_{i=1}^{m} \mathbb{E}\Big[\big(W(B_i)^2 - \mu(B_i)\big)^2\Big]\Big)\mu(A_2)\cdots\mu(A_p) \\ &= 2\Big(\sum_{i=1}^{m}\mu(B_i)^2\Big)\mu(A_2)\cdots\mu(A_p) \leqslant 2\varepsilon\mu(A_1)\cdots\mu(A_p).\end{aligned}$$

Finally, we claim that the last term $\mu(A_1)W(A_2)\cdots W(A_p)$ equals $pI_{p-1}(f\otimes_1 g)$. To this end, we introduce the following notation for convenience. We put a line on a set of variables in a function to denote the corresponding permuted sum. For instance

$$f(\overline{t_1, t_2, \cdots, t_p}, t_{p+1}, \cdots, t_q) \triangleq \sum_{\sigma\in\mathcal{S}_p} f(t_{\sigma_1}, \cdots, t_{\sigma_p}, t_{p+1}, \cdots, t_q). \tag{5.17}$$

Using this notation, we have

$$\begin{aligned} f\otimes_1 g(t_1,\cdots,t_{p-1}) &= \int_T \tilde{\mathbf{1}}_{A_1\times\cdots\times A_p}(t_1,\cdots,t_{p-1},s)\mathbf{1}_{A_1}(s)\mu(ds) \\ &= \frac{1}{p!}\int_T \mathbf{1}_{A_1\times\cdots\times A_p}(\overline{t_1,\cdots,t_{p-1},s})\mathbf{1}_{A_1}(s)\mu(ds) \\ &= \frac{1}{p!}\int_T \mathbf{1}_{A_1\times\cdots\times A_p}(s,\overline{t_1,\cdots,t_{p-1}})\mathbf{1}_{A_1}(s)\mu(ds) \\ &= \frac{\mu(A_1)}{p}\cdot\frac{1}{(p-1)!}\mathbf{1}_{A_2\times\cdots\times A_p}(\overline{t_1,\cdots,t_{p-1}}) \\ &= \frac{\mu(A_1)}{p}\tilde{\mathbf{1}}_{A_2\times\cdots\times A_p}(t_1,\cdots,t_{p-1}). \end{aligned}$$

It follows that

$$I_{p-1}(f\otimes_1 g) = \frac{\mu(A_1)}{p}\cdot W(A_2)\cdots W(A_p)$$

which gives the desired claim.

Putting the three terms together, we have

$$I_p(f)I_1(g) = I_{p+1}(h_\varepsilon) + R_\varepsilon + I_{p-1}(f\otimes_1 g),$$

and the equation (5.16) follows by letting $\varepsilon \to 0$. □

The extension of Proposition 5.1 to the general case relies on the following rather technical lemma. Its proof is not inspiring and the reader is encouraged to take the lemma as granted.

Lemma 5.7. *Given $p \geqslant q \geqslant r$, let $f \in L^2_S(T^p)$ and $g \triangleq g_1 \tilde{\otimes} g_2$, where $g_1 \triangleq \tilde{\mathbf{1}}_{A_1\times\cdots\times A_{q-1}}$, $g_2 \triangleq \mathbf{1}_{A_q}$ and the sets $A_1, \cdots, A_q \in \mathcal{B}_0$ are disjoint. Then we have*

$$f\tilde{\otimes}_r g = \frac{r(p+q-2r+1)}{q(p-r+1)}\big(f\tilde{\otimes}_{r-1}g_1\big)\otimes_1 g_2 + \frac{q-r}{q}\big(f\tilde{\otimes}_r g_1\big)\tilde{\otimes} g_2. \tag{5.18}$$

Proof. Recall that

$$\begin{aligned}
&f\otimes_r g(t_1,\cdots,t_{p-r},t_{p+1},\cdots,t_{p+q-r})\\
&\triangleq \int_{T^r} f(t_1,\cdots,t_{p-r},\mathbf{s})g(t_{p+1},\cdots,t_{p+q-r},\mathbf{s})\mu^r(d\mathbf{s}).
\end{aligned}$$

We express

$$\begin{aligned}
g(t_{p+1},\cdots,t_{p+q-r},\mathbf{s}) &= \tilde{\mathbf{1}}_{A_1\times\cdots\times A_q}(t_{p+1},\cdots,t_{p+q-r},\mathbf{s})\\
&= \frac{1}{q!}\sum_{\sigma\in\mathcal{S}_q}\mathbf{1}_{A_1\times\cdots\times A_q}(\sigma(t_{p+1},\cdots,t_{p+q-r},\mathbf{s}))\\
&= \frac{1}{q!}\Big(\sum_{\sigma\in\mathcal{S}'_q}+\sum_{\sigma\in\mathcal{S}''_q}\Big)\mathbf{1}_{A_1\times\cdots\times A_q}(\sigma(t_{p+1},\cdots,t_{p+q-r},\mathbf{s})),
\end{aligned} \tag{5.19}$$

where $\sigma(\cdots)$ means permuting the variables inside the bracket by σ. In the above summation, $\mathcal{S}'_q$ consists of those permutations σ under which the last position of

$$\sigma(t_{p+1},\cdots,t_{p+q-r},\mathbf{s}) \tag{5.20}$$

is not an s-variable, and $\mathcal{S}''_q$ consists of those permutations under which the last position of (5.20) is an s-variable. Using the decomposition (5.19), we can formally write $f\otimes_r g$ into two parts:

$$f\otimes_r g = I + J.$$

Let us first look at I. We use the notation introduced in (5.17) and also use $(\cdots,\hat{t},\cdots)$ to denote the tuple obtained by removing the variable t. From the definition of I we have

$$\begin{aligned}
I &= \frac{1}{q!}\sum_{\sigma\in\mathcal{S}'_q}\int_{T^r} f(t_1,\cdots,t_{p-r},\mathbf{s})\mathbf{1}_{A_1\times\cdots\times A_q}(\sigma(t_{p+1},\cdots,t_{p+q-r},\mathbf{s}))\mu^r(d\mathbf{s})\\
&= \frac{1}{q!}\sum_{j=1}^{q-r}\int_{T^r} f(t_1,\cdots,t_{p-r},\mathbf{s})\mathbf{1}_{A_1\times\cdots\times A_{q-1}}(\overline{t_{p+1},\cdots,\hat{t}_{p+j},\cdots t_{p+q-r},\mathbf{s}})g_2(t_{p+j})\mu^r(d\mathbf{s})\\
&= \frac{1}{q}\sum_{j=1}^{q-r}\Big(\int_{T^r} f(t_1,\cdots,t_{p-r},\mathbf{s})\tilde{\mathbf{1}}_{A_1\times\cdots\times A_{q-1}}(t_{p+1},\cdots,\hat{t}_{p+j},\cdots t_{p+q-r},\mathbf{s})\mu^r(d\mathbf{s})\Big)g_2(t_{p+j})\\
&= \frac{1}{q}\sum_{j=1}^{q-r} f\otimes_r g_1(t_1,\cdots,t_{p-r},t_{p+1},\cdots,\hat{t}_{p+j},\cdots,t_{p+q-r})\cdot g_2(t_{p+j}).
\end{aligned}$$

After symmetrisation, we obtain

$$\tilde{I} = \frac{q-r}{q}(f\tilde{\otimes}_r g_1)\tilde{\otimes} g_2(t_1,\cdots,t_{p-r},t_{p+1},\cdots t_{p+q-r}). \tag{5.21}$$

Next, we look at the term J. By definition,

$$
\begin{aligned}
J &= \frac{1}{q!} \sum_{\sigma \in \mathcal{S}_q''} \int_{T^r} f(t_1, \cdots, t_{p-r}, \mathbf{s}) \mathbf{1}_{A_1 \times \cdots \times A_q}(\sigma(t_{p+1}, \cdots, t_{p+q-r}, \mathbf{s})) \mu^r(d\mathbf{s}) \\
&= \frac{1}{q!} \sum_{j=1}^{r} \int_{T^r} f(t_1, \cdots, t_{p-r}, \mathbf{s}) \mathbf{1}_{A_1 \times \cdots \times A_{q-1}}(\overline{t_{p+1}, \cdots, t_{p+q-r}, \mathbf{s} \backslash \{s_j\}}) g_2(s_j) \mu^r(d\mathbf{s}) \\
&= \frac{1}{q} \sum_{j=1}^{r} \int_{T} \Big(\int_{T^{r-1}} f(t_1, \cdots, t_{p-r}, s_j, \mathbf{s} \backslash \{s_j\}) \\
&\quad \cdot g_1(t_{p+1}, \cdots, t_{p+q-r}, \mathbf{s} \backslash \{s_j\}) \mu^{r-1}(d(\mathbf{s} \backslash \{s_j\})) \Big) g_2(s_j) \mu(ds_j) \quad \text{(by the symmetry of } f\text{)} \\
&= \frac{r}{q} \int_T f \otimes_{r-1} g_1(s, t_1, \cdots, t_{p-r}, t_{p+1}, \cdots, t_{p+q-r}) g_2(s) \mu(ds).
\end{aligned}
$$

To compare this with $(f \tilde{\otimes}_{r-1} g_1) \otimes_1 g_2$, we compute:

$$
\begin{aligned}
&(f \tilde{\otimes}_{r-1} g_1) \otimes_1 g_2(t_1, \cdots, t_{p-r}, t_{p+1}, \cdots, t_{p+q-r}) \\
&= \frac{1}{(p+q-2r+1)!} \int_T f \otimes_{r-1} g_1(\overline{t_1, \cdots, t_{p-r}, s, t_{p+1}, \cdots, t_{p+q-r}}) g_2(s) \mu(ds).
\end{aligned}
$$

The crucial observation is that, if the s-variable is shuffled to any of the last $q-r$ positions the integrand is zero, due to the definitions of g_1 and g_2. In addition, the location of s at any of the first $p-r+1$ positions results in the same value for the integral, due to the symmetry of f. As a result, we have

$$
\begin{aligned}
&(f \tilde{\otimes}_{r-1} g_1) \otimes_1 g_2(t_1, \cdots, t_{p-r}, t_{p+1}, \cdots, t_{p+q-r}) \\
&= \frac{p-r+1}{(p+q-2r+1)!} \int_T f \otimes_{r-1} g_1(s, \overline{t_1, \cdots, t_{p-r}, t_{p+1}, \cdots, t_{p+q-r}}) g_2(s) \mu(ds) \\
&= \frac{p-r+1}{(p+q-2r+1)} \cdot \frac{q}{r} \cdot \tilde{J}. \qquad (5.22)
\end{aligned}
$$

The desired equation (5.18) follows from (5.21) and (5.22). $\square$

We are now able to establish the general product formula for multiple Wiener integrals.

Theorem 5.2. *Let $f \in L_S^2(T^p)$ and $g \in L_S^2(T^q)$. Then*

$$
I_p(f) I_q(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(f \otimes_r g).
$$

Proof. We use induction on q and always assume that $p \geqslant q$. The case when $q=1$ is contained in Proposition 5.1. Suppose that the claim is true for $q-1$ and now let $g \in L_S^2(T^q)$. By linearity, we may assume that

$$
g = g_1 \tilde{\otimes} g_2, \ g_1 \triangleq \tilde{\mathbf{1}}_{A_1 \times \cdots \times A_{q-1}}, \ g_2 \triangleq \tilde{\mathbf{1}}_{A_q},
$$

where $A_1, \cdots, A_q \in \mathcal{B}_0$ are disjoint. In particular, $g_1 \otimes_1 g_2 = 0$ and by Proposition 5.1 we have

$$I_q(g) = I_{q-1}(g_1) \cdot I_1(g_2) + (p-1) I_{p-2}(g_1 \otimes_1 g_2) = I_{q-1}(g_1) \cdot I_1(g_2).$$

It follows that

$$\begin{aligned}
I_p(f) I_q(g) &= I_p(f) I_{q-1}(g_1) I_1(g_2) \\
&= \sum_{r=0}^{q-1} r! \binom{p}{r} \binom{q-1}{r} I_{p+q-1-2r}(f \tilde{\otimes}_r g_1) I_1(g_2) \quad \text{(induction hypothesis)} \\
&= \sum_{r=0}^{q-1} r! \binom{p}{r} \binom{q-1}{r} \big(I_{p+q-2r}((f \tilde{\otimes}_r g_1) \tilde{\otimes} g_2) \\
&\quad + (p+q-1-2r) I_{p+q-2-2r}((f \tilde{\otimes}_r g_1) \otimes_1 g_2) \quad \text{(Proposition 5.1)} \\
&= \sum_{r=0}^{q-1} r! \binom{p}{r} \binom{q-1}{r} I_{p+q-2r}((f \tilde{\otimes}_r g_1) \tilde{\otimes} g_2) \\
&\quad + \sum_{r=1}^{q} (r-1)! \binom{p}{r-1} \binom{q-1}{r-1} (p+q-2r+1) I_{p+q-2r}((f \tilde{\otimes}_{r-1} g_1) \otimes_1 g_2) \\
&= \sum_{r=0}^{q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(f \tilde{\otimes}_r g). \quad \text{(Lemma 5.7)}
\end{aligned}$$

This concludes the induction step. $\square$

5.2.2 The isometry between $L_S^2(T^n, \mathcal{B}^n, \mu^n)$ and $\mathcal{H}_n$

Using the product formula for multiple Wiener integrals (in fact we only need Proposition 5.1), we can now establish the following isometry property. Recall that the space $L_S^2(T^n)$ is equipped with the inner product (5.13).

Theorem 5.3. *For each $n \geqslant 1$, the multiple Wiener integral operator I_n is an isometric isomorphism between $L_S^2(T^n, \mathcal{B}^n, \mu^n)$ and $\mathcal{H}_n$.*

Proof. The key observation is that

$$n! H_n(W(h)) = \int_{T^n} h(t_1) \cdots h(t_n) dW_{t_1} \cdots dW_{t_n} \tag{5.23}$$

for any $h \in H = L^2(T)$ with $\|h\|_H = 1$. The case when $n = 1$ is obvious. For the inductive step, suppose that (5.23) is true up to degree n. We write

$$h^{\otimes n}(t_1, \cdots, t_n) = h(t_1) \cdots h(t_n).$$

According to the recursive relation (2.9) for Hermite polynomials and Proposition 5.1, we have

$$\begin{aligned}(n+1)H_{n+1}(W(h)) &= W(h)H_n(W(h)) - H_{n-1}(W(h))\\ &= \frac{1}{n!}I_1(h)I_n(h^{\otimes n}) - \frac{1}{(n-1)!}I_{n-1}(h^{\otimes(n-1)})\\ &= \frac{1}{n!}\big(I_{n+1}(h^{\otimes(n+1)}) + nI_{n-1}(h^{\otimes n}\otimes_1 h)\big) - \frac{1}{(n-1)!}I_{n-1}(h^{\otimes(n-1)}).\end{aligned}$$

Note that $h^{\otimes n}\otimes_1 h = h^{\otimes(n-1)}$ since $\|h\|_H = 1$. Therefore, we arrive at

$$(n+1)H_{n+1}(W(h)) = \frac{1}{n!}I_{n+1}(h^{\otimes(n+1)}),$$

which completes the inductive step.

By using a density argument (based on the definition of $\mathcal{H}_n$ and the isometry property (5.5) of I_n), the relation (5.23) already implies that every element in $\mathcal{H}_n$ is the multiple Wiener integral of some function $f \in L_S^2(T^n)$. It remains show that the image of I_n is contained in $\mathcal{H}_n$. Due to Theorem 4.1, it is enough to see that $I_n(f) \perp \mathcal{H}_m$ whenever $f \in L_S^2(T^n)$ and $m \neq n$. But this follows from the fact that

$$I_n(f) \perp \frac{1}{m!}I_m(h^{\otimes m}) = H_m(W(h)), \quad \forall h \in H \text{ with } \|h\|_H = 1 \text{ and } \forall m \neq n.$$

□

Example 5.2. In the Brownian motion case, we have $T = [0,1]$ and $\mu = dt$. By taking $h = 1 \in L^2([0,1], dt)$ in the formula (5.23), we find

$$\begin{aligned} n!H_n(W_1) &= \int_{[0,1]^n} dW_{t_1}\cdots dW_{t_n}\\ &= n!\int_0^1\Big(\int_0^{t_n}\cdots\Big(\int_0^{t_2} dW_{t_1}\Big)\cdots dW_{t_{n-1}}\Big)dW_{t_n}.\end{aligned}$$

Appendix A Atomless measures

We collect a few basic facts about atomless measures that are used when proving the denseness of $\mathcal{E}_n$ as well as the product formula (5.16).

Let $(T, \mathcal{B}, \mu)$ be an atomless measure space (cf. Definition 5.1). Note that for each $B \in \mathcal{B}$, the restriction of μ on $B \cap \mathcal{B}$ is also atomless. For this reason and the fact that we always work with sets of finite μ-measure, we may restrict ourselves to finite atomless measure spaces. An essential property of the atomless property is that μ takes values continuously. This is the content of the following result.

Theorem A.1. *Suppose that $c \triangleq \mu(T) < \infty$. Then for any $t \in [0, c]$, there exists $A \in \mathcal{B}$ such that $\mu(A) = t$.*

Proof. The main idea is to use Zorn's lemma. Define

$$\Gamma \triangleq \left\{ S : D \to \mathcal{B} \mid D \subseteq [0,c], S \text{ increasing}, \ \mu(S(t)) = t \ \forall t \in D \right\}.$$

Here "S increasing" means

$$t_1 < t_2 \text{ in } D \implies S(t_1) \subseteq S(t_2).$$

We define a partial order on Γ by

$$S \preceq S' \iff D \subseteq D', \ S'|_D = S.$$

It can be checked that every chain in Λ has an upper bound. According to Zorn's lemma, Γ has a maximal element which is denoted as $S_m : D_m \to \mathcal{B}$.

We now show that $D_m = [0,c]$. To see this, first note that $c \in D_m$, for otherwise we may add the point c into D_m and assign $S_m(c) \triangleq T$, contradicting the maximality of S_m. Suppose on the contrary that $s \notin D_m$ for some $s < c$. Let

$$u^* \triangleq \sup\{u : u \in D_m, \ u < s\}, \ v^* \triangleq \inf\{v : v \in D_m, \ v > s\}.$$

We claim that $u^* = v^*$. Indeed, pick two sequences $u_k, v_k \in D_m$ such that $u_k \uparrow u^*$ and $v_k \downarrow v^*$. Set

$$A_k \triangleq S_m(u_k), \ B_k \triangleq S_m(v_k), \ A \triangleq \cup_k A_k, \ B \triangleq \cap_k B_k.$$

It follows that

$$\mu(A) = \lim_{k\to\infty} \mu(A_k) = \lim_{k\to\infty} u_k = u^*$$

and similarly $\mu(B) = v^*$. If $u^* < v^*$, then $\mu(B \backslash A) = v^* - u^* > 0$. By the definition of the atomless property, there exists $C \in \mathcal{B}$ such that $A \subseteq C \subseteq B$ and $u^* < \mu(C) < v^*$. We set $s^* \triangleq \mu(C)$, add the point s^* into D_m, and assign $S_m(s^*) \triangleq C$. This contradicts the maximality of S_m. As a result, we have $u^* = v^* = s$. Now by adding s into D_m and assigning $S_m(s) \triangleq \cup_k A_k$, we have $\mu(S_m(s)) = s$ and again a contradiction with the maximality of S_m. Consequently, $D_m = [0,c]$ and the assertion of the theorem follows from this. □

As an important consequence of Theorem A.1, we have the following property which is used in the main text.

Proposition A.1. *Let $(T, \mathcal{B}, \mu)$ be a finite atomless measure space. Then for any $\varepsilon > 0$, there exists a partition $T = \cup_{i=1}^m B_i$, such that $\mu(B_i) < \varepsilon$ for each i.*

Proof. Fix a number $0 < c < \varepsilon$. We first find $B_1 \in \mathcal{B}$ such that $\mu(B_1) = c$. The existence of B_1 is guaranteed by Theorem A.1. Then we consider $\mu|_{T\backslash B_1}$ and find $B_2 \in \mathcal{B}$ such that $\mu(B_2) = c$. We continue this procedure inductively, which stops after finitely many steps (say n) since $\mu(T) < \infty$ and $c > 0$. The family

$$B_1, \cdots, B_n, (B_1 \cup \cdots \cup B_n)^c$$

gives the desired partition. □

References

[1] K. Itô, Multiple Wiener integral, *J. Math. Soc. Japan* 3 (1951) 157–169.

[2] P. Malliavin, *Stochastic analysis*, Springer-Verlag, 1997.

[3] I. Nourdin and G. Peccati, *Normal approximations with Malliavin calculus*, Cambridge University Press, 2012.

[4] D. Nualart, *The Malliavin calculus and related topics*, 2nd Edition, Springer, 2006.

[5] I. Shigekawa, *Stochastic analysis*, Translations of Mathematical Monographs, AMS, 1998.